



Unipa Surabaya

Program Studi Statistika

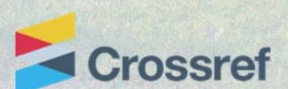
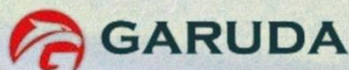
 **Statistika**

ISSN : 2089-0028

E-ISSN : 2654-7511

Jurnal Ilmiah Teori dan Aplikasi Statistika

Volume 15, No 2, DESEMBER 2022



Penanggung Jawab	
Alfisyahrina Hapsery, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Pimpinan Redaksi	
Muhammad Athoillah, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Editor Pelaksana	
Sari Cahyaningtias, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Fenny Fitriani, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Nur Silviyah Rahmi, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Sekar Utami Wijaya S.Stat., M.Si	(Universitas PGRI Adi Buana Surabaya)
Mitra Bestari (Reviewer)	
Dr.rer.pol. Dedy Dwi Prastyo, M.Si	(Institut Teknologi Sepuluh Nopember)
Dr. Drs. Agus Suharsono, M.S	(Institut Teknologi Sepuluh Nopember)
Dr. Bambang Widjanarko Otok	(Institut Teknologi Sepuluh Nopember)
Novri Suhermi, S.Si., M.Si., M.Sc	(Institut Teknologi Sepuluh Nopember)
Shofi Andari, S.Stat., M.Si	(Institut Teknologi Sepuluh Nopember)
Dr. RB Fajriya Hakim, S.Si., M.Si	(Universitas Islam Indonesia)
A'yunin Sofro, S.Si., M.Si., Ph.D.	(Universitas Negeri Surabaya)
Arief Rachman Hakim, S.Si., M.Si	(Universitas Diponegoro)
Dani Al Mahkya, S.Si., M.Si	(Sains Aktuaria Institut Teknologi Sumatera)
Dr. Sri Harini	(Universitas Islam Negeri Maulana Malik Ibrahim)
Dr. Faula Arina, M.Si	(Universitas Sultan Agung Tirtayasa)
Artanti Indrasetimingsih, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Gangga Anuraga, S.Si., M.Si	(Universitas PGRI Adi Buana Surabaya)
Winda Aprianti, S.Si., M.Si	(Politeknik Negeri Tanah Laut)

J Statistika: **Jurnal Ilmiah Teori dan Aplikasi Statistika** merupakan media penerbitan dan pembahasan karya ilmiah dalam bidang ilmu statistika beserta aplikasinya, baik berupa hasil penelitian, bahasan tentang teori, metodologi, komputasi maupun aplikasi statistika dalam bidang lain. Jurnal ini akan terbit dua kali dalam setahun setiap bulan Juli dan Desember.

Redaksi mengundang para pakar dan praktisi dari dalam dan luar lingkungan Universitas PGRI Adi Buana Surabaya untuk menuliskan karya ilmiahnya yang relevan dengan bidang ilmu statistika. Naskah dikirim pada website Jstatistika (<https://jurnal.unipasby.ac.id/index.php/jstatistika>) dengan format A4 margin 4-3-3-3 times new roman 10, spasi 1, disertai biodata penulis. Email redaksi: jstat@unipasby.ac.id. Cp: 085645358346

Pengantar Redaksi

Syukur Alhamdulillah bahwa Jurnal Ilmiah JStatistika, Program Studi Statistika Universitas PGRI Adi Buana Surabaya Volume 15 Nomor 2 dapat terbit bulan Desember 2022 ini. Jurnal Ilmiah JStatistika terbitan kali ini memuat artikel yang membahas tentang **“Pengaruh Konsumsi Rata-Rata, Pengeluaran Pemerintah, dan Ekspor Terhadap PDB Indonesia Tahun 2000-2020 Dengan Pendekatan Error Correction Model (ECM); Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression; Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data Dengan Kelas Tidak Seimbang; Penerapan Metode Clustering SOM dan DBSCAN dalam Mengelompokkan Unmet Need Keluarga Berencana di Nusa Tenggara Barat; Prediksi Profitabilitas Bank Umum Konvensional Pada Masa Pandemi Covid-19; Teknik Oversampling Pada Regresi Logistik Ordinal Dalam Menduga Faktor Yang Memengaruhi Risiko Penyebaran Zona Covid-19 di Kabupaten Garut; Perbandingan Analisis Dekomposisi dan Exponential Smoothing Holt Winters untuk Peramalan Rata-Rata Jumlah KPM PKH di NTB; Faktor – Faktor yang Memengaruhi Permasalahan Stunting di Jawa Barat Menggunakan Regresi Logistik Biner; Pengelompokan Pertumbuhan Ekonomi (PDRB) dan Pengeluaran di Jawa Timur berdasarkan Jumlah UMK serta Faktor-faktor yang Mempengaruhi dengan Model Persamaan Simultan; Penerapan Clustering Time Series Pada Pengelompokan Provinsi di Indonesia (Studi Kasus : Nilai Ekspor Non Migas di Indonesia Tahun 2016-2020); Analisis Partial Least Square Structural Equation Model (PLS-SEM) untuk Pemodelan Penerimaan Sistem Jaringan Informasi Bersama Antar Sekolah (JIBAS)”**.

Jurnal Ilmiah JStatistika menerima artikel ilmiah dari hasil penelitian, laporan / studi kasus, kajian/ tinjauan pustaka, maupun penyegar ilmu statistika, yang berorientasi pada kemuthakhiran ilmu pengetahuan statistika agar dapat menjadi sumber informasi ilmiah yang mampu memberikan kontribusi dalam mengatasi permasalahan yang semakin kompleks. Redaksi mengundang berbagai ilmuwan dari berbagai lembaga pendidikan tinggi maupun penelitian untuk memberikan sumbangan ilmiahnya, baik berupa hasil penelitian maupun kajian ilmiah dalam bidang statistika dan aplikasinya.

Redaksi sangat mengharapkan masukan-masukan dari para pembaca demi semakin meningkatnya kualitas jurnal sebagaimana harapan kita bersama. Semoga artikel ilmiah yang termuat dalam Jurnal Ilmiah JStatistika bermanfaat bagi para akademisi dan profesional yang berkecimpung dalam dunia statistika dan aplikasinya.

Surabaya, Desember 2022

Pimpinan Redaksi

DAFTAR ISI

HALAMAN JUDUL

PENGANTAR REDAKSI

DAFTAR ISI

- ❑ **Pengaruh Konsumsi Rupa, Pengeluaran Pemerintah, dan Ekspor Terhadap PDB Indonesia Tahun 2000-2020 Dengan Pendekatan Error Correction Model (ECM)**
Rezky Yayang Yakhamid
Politeknik Statistika STIS 210 – 218

- ❑ **Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression**
Siva Alfiani¹, Prizka Rismawati Arum²
Universitas Muhammadiyah Semarang 219 - 227

- ❑ **Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data Dengan Kelas Tidak Seimbang**
Putri Septiana Rizky¹, Ristu Haiban Hirzi², Umam Hidayaturrohman³
Universitas Hamzanwadi Selong 228 - 236

- ❑ **Penerapan Metode Clustering SOM dan DBSCAN dalam Mengelompokkan Unmet Need Keluarga Berencana di Nusa Tenggara Barat**
Sausan Nisrina¹, Wiwit Pura Nurmayanti², Basirun³, Kertanah⁴ Muhammad Ghazali⁵
Universitas Hamzanwadi 237 - 244

- ❑ **Prediksi Profitabilitas Bank Umum Konvensional Pada Masa Pandemi Covid-19**
Yustika Haryanti¹, Fauziyah^{2*}
Universitas PGRI Adi Buana Surabaya 245 - 250

- ❑ **Teknik Oversampling Pada Regresi Logistik Ordinal Dalam Menduga Faktor Yang Memengaruhi Risiko Penyebaran Zona Covid-19 di Kabupaten Garut**
Ghina Fauziah¹, Indahwati², Erfiani³, Anwar Fitrianto⁴, Reni Amelia⁵
Institut Pertanian Bogor 251 - 258

- ❑ **Perbandingan Analisis Dekomposisi dan Exponential Smoothing Holt Winters untuk Peramalan Rata-Rata Jumlah KPM PKH di NTB**
M. Hadiyan Amaly¹, Wiwit Pura Nurmayanti², Sausan Nisrina³
Universitas Hamzanwadi 259 - 264

- ❑ **Faktor – Faktor yang Memengaruhi Permasalahan Stunting di Jawa Barat Menggunakan Regresi Logistik Biner**
Silmi Annisa Rizki Manaf¹, Erfiani², Indahwati³, Anwar Fitrianto⁴, Reni Amelia⁵
Institut Pertanian Bogor 265 - 274

- ❑ **Pengelompokan Pertumbuhan Ekonomi (PDRB) dan Pengeluaran di Jawa Timur berdasarkan Jumlah UMK serta Faktor-faktor yang Mempengaruhi dengan Model Persamaan Simultan**
Santi Puteri Rahayu¹, Irene Monica Amanda²
Institut Teknologi Sepuluh Nopember 275 - 285

- ❑ **Penerapan Clustering Time Series Pada Pengelompokan Provinsi di Indonesia (Studi Kasus : Nilai Ekspor Non Migas di Indonesia Tahun 2016-2020)**
Harista Almiatus Soleha¹, Wiwit Pura Nurmayanti², Umam Hidayaturrohman³, Ristu Haiban Hirzi⁴, Ayu Septiani⁵
Universitas Hamzanwadi 286 - 291

- ❑ **Analisis Partial Least Square Structural Equation Model (PLS-SEM) untuk Pemodelan Penerimaan Sistem Jaringan Informasi Bersama Antar Sekolah (JIBAS)**
Jerhi Wahyu Fernanda¹, Vira Luthifiana², M. Khoiril Akhyar³
Institut Agama Islam Negeri (IAIN) Kediri 292 - 297

Pengaruh Konsumsi Rumah, Pengeluaran Pemerintah, dan Ekspor Terhadap PDB Indonesia Tahun 2000-2020 Dengan Pendekatan Error Correction Model (ECM)

Rezky Yayang Yakhamid
Politeknik Statistika STIS
Jalan Otto Iskandardinata, No. 64C, Jakarta 13330
e-mail: 211810559@stis.ac.id

ABSTRAK

Indonesia memiliki produk domestik bruto (PDB) yang cenderung mengalami peningkatan setiap tahunnya. Pertumbuhan PDB suatu negara sangat dipengaruhi oleh pertumbuhan komponen penyusunnya, sehingga pertumbuhan PDB dapat sangat berfluktuasi dan sensitif terhadap kondisi yang dialami sebuah negara. Penelitian ini bertujuan untuk mengetahui pengaruh konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor terhadap PDB Indonesia tahun 2000 hingga 2020. Metode analisis data yang digunakan adalah teknik analisis data deret waktu dengan menggunakan analisis ECM (*Error Correction Model*). Hasil penelitian menunjukkan bahwa konsumsi rumah tangga dan nilai ekspor berpengaruh positif secara signifikan terhadap PDB Indonesia, serta pengeluaran pemerintah berpengaruh positif namun tidak signifikan terhadap PDB Indonesia. Adapun kecepatan penyesuaian model jangka pendek ke jangka panjang sebesar 42,74%.

Kata kunci : Peramalan PDB, model ECM, konsumsi rumah tangga, pengeluaran pemerintah, ekspor

ABSTRACT

Indonesia has Gross Domestic Product (GDP) which tends to increase every year. GDP growth is greatly influenced by growth of the component so that GDP growth can be fluctuate highly and sensitive to country conditions. This study aims to determine the effect of household consumption, government expenditure, dan exports on Indonesia's GDP from 2000 to 2020. Data analysis method used time series data analysis technique using ECM (*Error Correction Model*). The result showed that household consumption and export value had a significantly positive effect on Indonesia's GDP, and government expenditure had a positive but insignificant effect on Indonesia's GDP. The speed of adjustment of short-term to long term model is 42,74%.

Keywords: GDP forecasting, ECM model, private consumption, government expenditure, export

1. PENDAHULUAN

Indonesia adalah negara berkembang yang memiliki pertumbuhan ekonomi yang baik setelah krisis ekonomi tahun 1998. Produk Domestik Bruto (PDB) Indonesia mengalami peningkatan setiap tahunnya. PDB merupakan konsep indikator ekonomi yang paling sering digunakan di negara-negara berkembang di dunia termasuk Indonesia (Mankiw, 2009). Dari data yang dihimpun dari Fred Economic Data, kondisi perekonomian Indonesia setelah krisis moneter 1998 dapat dibilang cukup sehat. PDB Indonesia mencapai 1012,9 triliun rupiah pada triwulan pertama tahun 2000 dan terus berkembang setiap tahunnya hingga mencapai 2786,5 triliun rupiah pada triwulan keempat 2019 sebelum krisis ekonomi 2020 akibat dari adanya pandemi COVID-19 menghantam sebagian besar

negara di dunia termasuk Indonesia (FRED, 2020). Walaupun cenderung tumbuh, PDB sangat sensitif dipengaruhi oleh perubahan komponen-komponen penyusunnya. Oleh karena itu, data pertumbuhan PDB sering kali berfluktuatif tergantung kepada kondisi yang sedang dialami oleh negara tersebut.

Pertumbuhan ekonomi setiap tahunnya sering kali diikuti oleh pertumbuhan komponen-komponen pengeluaran seperti konsumsi rumah tangga, pengeluaran pemerintah, dan nilai ekspor. Konsumsi rumah tangga tercatat 627,06 triliun rupiah pada triwulan pertama tahun 2000 dan naik menjadi 1543,63 triliun rupiah pada triwulan keempat tahun 2019. Pengeluaran pemerintah pertumbuhannya dari 24,74 triliun rupiah pada triwulan pertama tahun 2000 hingga 337,07 triliun rupiah pada triwulan keempat tahun 2020. Untuk

nilai ekspor Indonesia juga demikian mencatatkan pertumbuhan dari 122,79 triliun rupiah pada triwulan pertama tahun 2000 hingga sebesar 748,04 triliun rupiah pada triwulan keempat 2019 (FRED, 2020).

Untuk mengetahui pengaruh masing-masing komponen penyusun PDB terhadap pertumbuhan PDB, sering digunakan metode regresi linear berganda pada beberapa penelitian terdahulu. Padahal, model regresi akan menemui berbagai masalah ketika diterapkan pada data runtun waktu, salah satunya adalah terjadinya kointegrasi antar variabel yang menyebabkan regresi lancung. Salah satu cara yang dapat digunakan untuk mengatasi permasalahan tersebut adalah penggunaan metode *Error Correction Model* (ECM). ECM bertujuan untuk mengidentifikasi hubungan jangka panjang dan jangka pendek yang terjadi karena adanya kointegrasi antar variabel penelitian (Basuki, 2016; Rosadi, 2012).

Berdasarkan latar belakang penelitian, permasalahan yang akan diteliti pada penelitian ini adalah bagaimana pengaruh konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor terhadap PDB Indonesia tahun 2000 hingga 2020. Penelitian ini menggunakan data PDB Indonesia beserta komponen penyusunnya yakni konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor. Oleh karena itu, ruang lingkup penelitian ini adalah Indonesia pada triwulan pertama tahun 2000 hingga triwulan ketiga tahun 2020.

Tujuan dari penelitian ini adalah untuk mengetahui pengaruh konsumsi rumah tangga, pengeluaran rumah tangga, dan ekspor terhadap PDB Indonesia dengan pendekatan metode ECM. Oleh karena itu, hipotesis penelitian yang diajukan pada penelitian ini adalah terdapat adanya pengaruh konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor terhadap PDB Indonesia tahun 2000 sampai dengan 2020 baik secara parsial maupun secara simultan.

2. METODE PENELITIAN

2.1 Tinjauan Referensi

PDB adalah nilai barang dan jasa yang diproduksi di dalam suatu negara baik oleh perusahaan milik warga negara maupun pemerintah di negara tersebut maupun milik negara asing. PDB dapat dihitung dengan cara pengeluaran, cara produk neto, dan cara pendapatan (Sukirno, 1994). Perhitungan PDB dengan cara pengeluaran membagi pengeluaran ke dalam empat (4) jenis komponen yakni konsumsi rumah tangga (C), pembentukan modal/ investasi (I), pengeluaran

pemerintah (G), dan ekspor neto (NX) atau ekspor (X) dikurangi impor (M). Jika ditulis ke dalam persamaan matematis, akan menjadi sebagai berikut (Mankiw, 2009).

$$GDP = C + I + G + NX \quad (1)$$

$$\text{dengan } NX = X - M$$

Konsumsi rumah tangga terdiri atas barang dan jasa yang dibeli oleh rumah tangga (Mankiw, 2009). Konsumsi rumah tangga didefinisikan sebagai nilai perbelanjaan yang dilakukan oleh rumah tangga untuk membeli segala jenis kebutuhannya dalam periode tertentu (Sukirno, 1994).

Investasi terdiri atas barang yang dibeli untuk keperluan masa depan (Mankiw, 2009). Investasi atau pembentukan modal tetap sektor swasta adalah pengeluaran untuk membeli barang modal yang dapat menaikkan produksi barang dan jasa di masa yang akan datang (Sukirno, 1994).

Pengeluaran pemerintah adalah barang dan jasa yang dibeli oleh pemerintahan baik pusat maupun daerah (Mankiw, 2009). Pengeluaran pemerintah terdiri atas konsumsi pemerintah dan investasi pemerintah. Konsumsi pemerintah adalah pembelian barang dan jasa yang akan dikonsumsi seperti gaji, sedangkan investasi pemerintah meliputi pengeluaran untuk membangun sarana dan prasarana (Sukirno, 1994).

Ekspor neto adalah nilai barang dan jasa yang dijual ke negara lain (ekspor) dikurangi nilai barang dan jasa yang negara lain jual kepada negara kita (impor) (Mankiw, 2009). Ekspor suatu negara merupakan barang dan jasa yang dihasilkan di dalam negeri (untuk dijual ke luar negeri), sedangkan impor merupakan produksi negara lain. Oleh karena itu, impor tidak dihitung dalam pendapatan nasional dalam hal ini PDB (Sukirno, 1994).

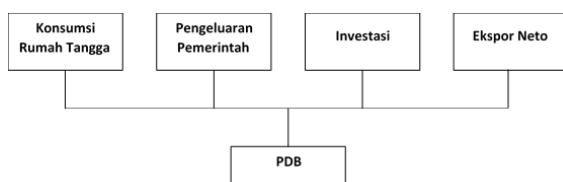
Dari kajian literatur yang telah dilakukan, penulis menemukan terdapat beberapa penelitian yang meneliti PDB di Indonesia dan beberapa hal yang mempengaruhinya. Penelitian-penelitian tersebut sebagian besar menggunakan model *Error Correction Model* (ECM). Berikut merupakan ringkasan dari beberapa penelitian terdahulu tentang PDB Indonesia yang menggunakan teknik analisis deret waktu dengan model ECM.

Pada penelitian Safari dan Fikri yang berjudul “Analisis Pengaruh Ekspor, Pembentukan Modal, dan Pengeluaran Pemerintah Terhadap Pertumbuhan Ekonomi Indonesia” pada tahun 2016, peneliti menggunakan data PDB Indonesia dan penyusunnya tahun 1975-2014. Teknik analisis yang digunakan adalah analisis data deret waktu

dengan model ECM. Model ECM yang didapatkan dari penelitian tersebut adalah $\Delta PDB = 0,012 - 0,038\Delta D_1 + 0,254\Delta EKS + 0,271\Delta GCF - 0,109\Delta PP - 0,684ECT_t + e_t$ dengan PDB adalah produk domestik bruto, D_1 adalah dummy untuk krisis ekonomi, EKS variabel ekspor, GCF adalah penanaman modal, PP adalah pengeluaran pemerintah, ECT_t adalah Error Correction Term, dan e_t adalah residual model. Hasil yang didapatkan antara lain variabel ekspor, pembentukan modal, dan pengeluaran pemerintah secara simultan berpengaruh terhadap PDB baik dalam jangka panjang maupun jangka pendek. Adapun variabel yang tidak berpengaruh adalah variabel *dummy* untuk krisis ekonomi. Adapun kecepatan penyesuaian jangka pendek ke jangka panjang adalah sebesar 68,4%.

Pada penelitian Rafiq yang berjudul “Pengaruh Pengeluaran Konsumsi Rumah Tangga, Investasi dan Pengeluaran Pemerintah Terhadap Pertumbuhan Ekonomi di Indonesia Tahun 2001:T1-2010:T4” pada tahun 2016, peneliti menggunakan data pertumbuhan ekonomi Indonesia dan penyusunnya tahun 1975-2014. Teknik analisis yang digunakan adalah analisis data deret waktu dengan model ECM. Hasil yang didapatkan antara lain variabel konsumsi rumah tangga, penanaman modal dalam negeri, penanaman modal asing, dan pengeluaran pemerintah secara simultan dan parsial berpengaruh positif terhadap pertumbuhan ekonomi baik dalam jangka panjang maupun jangka pendek.

Berdasarkan kajian literatur dan beberapa penelitian-penelitian yang telah dilakukan sebelumnya, maka dapat disimpulkan kerangka pikir yang dapat digambarkan sebagai berikut :



Gambar 1. Kerangka Pikir

2.2 Metode Analisis

Metode analisis yang digunakan pada penelitian ini adalah metode analisis kuantitatif menggunakan analisis data deret waktu. Data yang digunakan merupakan data sekunder yang diperoleh dari situs fred.stlouisfed.org. Adapun observasi yang digunakan adalah tahun 2000 sampai dengan tahun 2020 triwulan ketiga sehingga total keseluruhan terdapat 83 observasi. Variabel yang digunakan dalam data tersebut adalah variabel PDB

triwulanan (PDB) sebagai variabel respons serta konsumsi rumah tangga (KRT), pengeluaran pemerintah (PP), dan nilai ekspor (EKS) sebagai variabel penjelas. Adapun perangkat lunak yang akan digunakan dalam penelitian ini adalah Eviews10.

Teknik analisis data yang digunakan adalah dengan analisis statistika deskriptif dan analisis statistika inferensial berupa analisis data deret waktu. Analisis statistika deskriptif adalah analisis berupa metode-metode yang berkaitan dengan penyajian data sehingga memberikan informasi yang berguna (Walpole, 1982). Analisis deskriptif yang akan digunakan adalah berupa visualisasi data deret waktu ke dalam bentuk diagram garis.

Analisis statistika inferensial mencakup metode yang berkaitan dengan analisis sebagian data untuk sampai pada peramalan atau penarikan kesimpulan keseluruhan data (Walpole, 1982). Analisis statistika inferensial yang digunakan adalah analisis data deret waktu. Data deret waktu adalah salah satu jenis data yang terdiri atas variabel-variabel yang dikumpulkan menurut urutan waktu dalam rentang tertentu untuk kategori tertentu (Rosadi, 2014). Analisis data deret waktu yang digunakan dalam penelitian ini adalah ECM (*Error Correction Model*).

Pemodelan ECM bertujuan untuk mengidentifikasi hubungan jangka panjang dan jangka pendek yang terjadi karena adanya kointegrasi antar variabel penelitian (Basuki, 2016). Keadaan kointegrasi didapatkan apabila Y_t dan X_t keduanya memiliki akar unit (tidak stasioner) dan e_t tidak memiliki akar unit (stasioner). Sedangkan apabila Y_t dan X_t keduanya memiliki akar unit (tidak stasioner) dan e_t memiliki akar unit (tidak stasioner) maka terjadi regresi lancung (*spurious regression*). Pemodelan ECM dapat digunakan apabila antar variabel penelitian terjadi kointegrasi (Rosadi, 2012).

Apabila Y_t dan X_t berkointegrasi, maka sifat hubungan jangka pendek dari variabel respons dan penjelas dapat dinyatakan dalam bentuk ECM (*Error Correction Model*) sebagai berikut :

$$\Delta Y_t = \varphi + \lambda e_{t-1} + \omega \Delta X_t + \varepsilon_t \tag{2}$$

dengan e_{t-1} adalah error dari persamaan regresi antara Y_t dan X_t (persamaan jangka panjang) atau dapat dituliskan $e_{t-1} = Y_{t-1} - (\alpha + \beta X_{t-1})$, dan ε_t adalah komponen error dari pemodelan ECM (Rosadi, 2012). Dalam berbagai literatur, e_{t-1} dapat dituliskan sebagai ECT_t atau *error correction term* pada lag ke-t. Dalam persamaan tersebut terdapat λ yang merupakan koefisien dari error correction

terms (ECT) yang bernilai negatif. λ memiliki interpretasi kecepatan penyesuaian parameter model jangka pendek ke jangka panjang (Enders, 2010).

Dalam mendapatkan persamaan akhir dari ECM, perlu dilakukan beberapa prosedur di antaranya adalah uji stasioneritas, pembentukan model regresi, uji kointegrasi, dan pembentukan model ECM (Basuki, 2012).

Pada model yang stasioner, peramalan masa depan akan dapat dilakukan berdasarkan kondisi di masa lalu. Oleh karena itu, asumsi stasioneritas menjadi penting dalam analisis data deret waktu. Uji stasioneritas dapat dilakukan dengan plot autokorelasi ACF/PACF dan uji akar unit seperti Augmented Dickey Fuller (ADF) test dan Phillip-Peron (PP) test. Pengujian akar unit ADF test dilakukan untuk menguji hipotesis $H_0 : \rho = 0$ (terdapat akar unit) pada persamaan regresi :

$$\Delta Y_t = \alpha + \delta t + \rho Y_{t-1} + \sum_{j=1}^k \phi_j \Delta Y_{t-j} + e_t \quad (3)$$

Hipotesis ditolak apabila nilai p-value pada Eviews10 memiliki nilai kurang dari alfa yang ditetapkan. Apabila hipotesis ditolak, maka variabel yang diuji tersebut dinyatakan tidak memiliki akar unit dan sudah stasioner (Rosadi, 2012). Dalam metode ECM, semua variabel yang akan dimodelkan diharapkan tidak stasioner pada level sehingga terjadi *autokorelasi* antar variabel dan menimbulkan regresi lancung. Selain menguji kestasioneran variabel pada level, diperlukan pula menguji kestasioneran data pada *difference*. Semua variabel yang akan dimodelkan dengan ECM, diharapkan stasioner pada derajat yang sama (Basuki, 2012).

Langkah selanjutnya adalah pembentukan model regresi. Pembentukan model regresi diperlukan untuk membangun persamaan jangka panjang dengan error yang belum terkoreksi. Adapun persamaan regresi yang dimaksud adalah sebagai berikut :

$$Y_t = \alpha + \beta X_t + e_t \quad (4)$$

(Rosadi, 2012)

Setelah melakukan pengujian kestasioneran data pada level dan diketahui data tidak stasioner, langkah selanjutnya adalah melakukan uji kointegrasi. Uji kointegrasi adalah pengujian stasioneritas pada residual (e_t) dari model regresi yang telah dibentuk sebelumnya. Residual model regresi diharapkan stasioner pada tingkat level agar antar variabel X dan Y dapat dikatakan terkointegrasi. Pemodelan ECM dapat dilakukan apabila antar variabel penelitian terkointegrasi (Basuki, 2012).

Pembentukan model ECM dibangun dari ΔY , ΔX , dan residual yang didapatkan dari persamaan jangka panjang (model regresi) yang sudah stasioner di level. Selain variabel penjelas X, model ECM dapat juga disisipkan komponen *lag* variabel respons Y itu sendiri. Secara umum, bentuk model ECM sebagai berikut (Rosadi, 2012).

$$Y_t = \varphi + \delta_1 t + \lambda_1 e_{t-1} + \sum_{i=0}^K \omega_{1,i} \Delta X_{t-i} + \sum_{i=0}^K \omega_{2,i} \Delta Y_{t-i} + \varepsilon_{1,t} \quad (5)$$

Setelah didapatkan model ECM, penelitian perlu dilanjutkan dengan uji asumsi klasik. Pengujian asumsi klasik dilakukan untuk mengetahui terdapat atau tidaknya penyimpangan asumsi klasik pada model yang telah kita dapatkan. Asumsi klasik diperlukan untuk mendapatkan estimator OLS yang BLUE (*Best Linear Unbiased Estimator*). Adapun asumsi klasik yang diharapkan terjadi adalah tidak terjadi multikolinearitas, tidak terjadi heteroskedastisitas, tidak terdapat autokorelasi pada residual, dan residual berdistribusi normal (Rosadi, 2012).

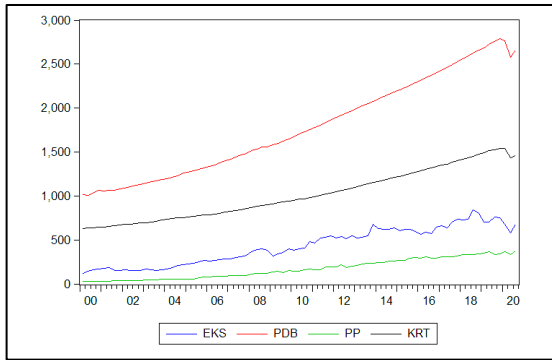
Apabila sampai dengan tahap terakhir terdapat lebih dari satu model, maka diperlukan pemilihan model terbaik. Terdapat beberapa kriteria yang dapat digunakan sebagai dasar pertimbangan dalam pemilihan model terbaik, yakni R-kuadrat (R^2), R-kuadrat terkoreksi (*Adjusted R²*), dan kriteria informasi (Kurniawan, 2016). Adapun kriteria informasi yang terdapat pada output Eviews10 antara lain Akaike Info Criterion (AIC), Schwarz Bayesian Criterion (SBC), dan Hannan-Quinn Criterion (HC). Selain itu, informasi nilai log-likelihood juga dapat digunakan dalam pemilihan model.

3. HASIL DAN PEMBAHASAN

Berikut merupakan hasil dan pembahasan yang didapatkan dari penelitian ini. Bagian ini dibagi menjadi penjelasan mengenai analisis deskriptif dan analisis inferensial dengan data data deret waktu menggunakan *Error Correction Model* (ECM).

3.1 Analisis Deskriptif

Berikut merupakan diagram garis variabel produk domestik bruto (PDB), konsumsi rumah tangga (KRT), pengeluaran pemerintah (PP), dan ekspor (EKS) yang didapatkan dari *output* program Eviews10.



Gambar 2. Diagram Garis Variabel PDB, KRT, PP, dan EKS

Pada diagram garis tersebut terlihat bahwa baik PDB, konsumsi rumah tangga, pengeluaran pemerintah, maupun ekspor memiliki tren yang cenderung naik. Pertumbuhan PDB juga selalu diikuti oleh pertumbuhan konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor. Hal ini mengindikasikan adanya pengaruh konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor terhadap PDB Indonesia. Pada tahun 2008, terlihat variabel ekspor menunjukkan penurunan yang cukup dalam. Hal tersebut dikarenakan adanya krisis ekonomi dunia pada tahun 2008, namun hal tersebut tidak terlalu berdampak pada PDB Indonesia. Sedangkan pada tahun 2020 terlihat terdapat penurunan yang dalam baik pada PDB, konsumsi rumah tangga, dan ekspor. Hal tersebut dikarenakan adanya krisis ekonomi yang mengguncang seluruh negara sebab adanya pandemi COVID-19.

3.2 Analisis Error Correction Model (ECM)

Pembentukan model ECM untuk memodelkan pengaruh konsumsi rumah tangga, pengeluaran pemerintah, dan ekspor dilakukan dalam beberapa tahapan, yakni uji stasioneritas variabel, pembentukan model regresi, uji kointegrasi, pembentukan model ECM, uji asumsi klasik, pemilihan model terbaik, dan interpretasi.

1. Uji Stasioneritas

Pengujian stasioneritas yang pertama adalah untuk menguji apakah model ECM dapat diterapkan atau tidak. Agar model ECM dapat diterapkan, maka hasil pengujian stasioneritas semua variabel yang akan dimodelkan pada level diharapkan tidak stasioner. Adapun variabel yang dilakukan pengujian stasioneritas adalah variabel tanpa transformasi (model 1) dan variabel yang telah ditransformasi menggunakan logaritma natural (model 2).

$H_0 : \delta = 0$ (variabel memiliki akar unit, variabel tidak stasioner)

$H_1 : \delta \neq 0$ (variabel tidak memiliki akar unit, variabel stasioner)

Tingkat signifikansi : $\alpha=5\%$

Statistik uji:

Tabel 1. ADF Unit Root Test (Level)

	Model 1	Model 2
Statistic	0.49327	0.00898
p-value	0.99999	1.00000

Berdasarkan hasil dari statistik uji ADF dan p-value hasil output Eviews10 tersebut, pada taraf uji 5% keseluruhan variabel baik variabel tanpa transformasi maupun variabel yang telah ditransformasi dinyatakan belum cukup bukti untuk dinyatakan stasioner pada level. Oleh karena itu, variabel-variabel tersebut terindikasi memiliki hubungan regresi lancung. Pengujian stasioneritas perlu dilanjutkan untuk mencari ordo differensiasi. Semua variabel yang akan dimodelkan diharapkan stasioner pada derajat yang sama.

Tabel 2. ADF Unite Root Test (Model 1)

Variable	1st Diff	2nd Diff
Fisher Chisq	34.6190*	521.421*
D(EKS)	p = 0.0011*	p = 0.0000*
D(PDB)	p = 0.1362	p = 0.0000*
D(PP)	p = 0.0010*	p = 0.0000*
D(KRT)	p = 0.2070	p = 0.0053*

*signifikan pada level 5%

Berdasarkan hasil dari statistik uji ADF tersebut, pada taraf uji 5% keseluruhan variabel tanpa transformasi (model 1) cukup bukti untuk dinyatakan stasioner pada ordo 2.

Tabel 3. ADF Unite Root Test (Model 2)

Variable	1st Diff
Fisher Chisq	129.155*
D(EKS)	p = 0.0000*
D(PDB)	p = 0.0000*
D(PP)	p = 0.0000*
D(KRT)	p = 0.0000*

*signifikan pada level 5%

Berdasarkan hasil dari statistik uji ADF tersebut, pada taraf uji 5% keseluruhan variabel dengan transformasi (model 2) cukup bukti untuk dinyatakan stasioner pada ordo 1.

2. Pengajuan model

Berdasarkan hasil dari pengujian stasioneritas, terdapat dua persamaan Error Correction Model (ECM) yang diajukan dalam penelitian ini, yakni :

Model 1:

$$\Delta PDB_t = \alpha_0 + \alpha_1 \Delta KRT_t + \alpha_2 \Delta PP_t + \alpha_3 \Delta EKS_t + \alpha_4 ECT_t + \varepsilon_t \quad (6)$$

Model 2:

$$\ln(PDB_t) = \alpha_0 + \alpha_1 \ln(KRT_t) + \alpha_2 \ln(PP_t) + \alpha_3 \ln(EKS_t) + \alpha_4 ECT_t + \varepsilon_t \quad (7)$$

dengan

PDB : Produk Domestik Bruto (PDB);

KRT : Konsumsi Rumah Tangga;

PP : Pengeluaran Pemerintah;

EKS : Nilai Ekspor; dan

$ECT_t = PDB_{t-1} - \overline{PDB}_{t-1}$: Error Correction Term

3. Pembentukan persamaan jangka panjang (regresi linear)

Tabel 4. Koefisien Regresi

Variabel	Model 1	Model 2
C	16.3429	0.7209
KRT	1.5177*	0.8946*
PP	0.6190*	0.0459*
EKS	0.2812*	0.0564*
F-statistics	34344.07*	50221.27*
R-square	0.9992	0.9994
Adj. R-square	0.9992	0.9994
Durbin Watson	0.6977	1.0858

*signifikan pada level 5%

Persamaan jangka panjang yang didapatkan dari output Eviews10 adalah sebagai berikut :

Model 1

$$PDB_t = 16,3429 + 1,5177KRT_t + 0,6190PP_t + 0,2812EKS_t \quad (8)$$

Model 2

$$\ln(PDB_t) = 0,7209 + 0,8946\ln(KRT_t) + 0,0459\ln(PP_t) + 0,0564\ln(EKS_t) \quad (9)$$

Berdasarkan pengujian secara parsial, didapatkan persamaan jangka panjang atau persamaan regresi untuk kedua model yang semua variabel penjelasnya signifikan mempengaruhi variabel respons baik secara simultan maupun parsial. Namun, nilai R-square, adjusted R-square cukup tinggi yakni mendekati 1, nilai statistik uji F

yang sangat signifikan, serta nilai Durbin Watson yang jauh dari 2 mengindikasikan adanya regresi lancung dan autokorelasi pada model regresi tersebut. Oleh karena itu, model harus dikoreksi dengan persamaan jangka pendek ECM.

4. Uji Kointegrasi

Pengujian kointegrasi dilakukan dengan pengujian kestasioneran residual dari persamaan regresi pada level.

Tabel 5. Uji Kointegrasi

	Model 1	Model 2
t-Statistic	-4.1088*	-5.9734*
p-value	0.0016	0.0000

*signifikan pada level 5%

Dari hasil output Eviews10, dengan taraf uji 5%, kedua model cukup bukti untuk menyatakan bahwa residual model regresi yang diujikan tidak memiliki akar unit yang stasioner pada level. Oleh karena itu, antar variabel penjelas dan respons memiliki hubungan kointegrasi baik pada model tanpa transformasi maupun pada model transformasi len. Prosedur dapat dilanjutkan ke pembentukan persamaan ECM untuk mengoreksi model regresi.

5. Pembentukan persamaan jangka pendek ECM

Tabel 6. Koefisien model ECM

Variabel	Model 1	Model 2
C	0.7083	0.0019*
KRT	1.8203*	0.8734*
PP	0.0636	0.0036
EKS	0.0714*	0.0333*
ECT	-0.3063*	-0.4274*
F-statistics	211.1496	56.9546*
R-square	0.9174	0.7473
Adj. R-square	0.9131	0.7342
Durbin Watson	2.1951	1.6716

*signifikan pada level 5%

Dari output Eviews10 tersebut didapatkan dua model ECM sebagai berikut :

Model 1:

$$\Delta PDB_t = 0,7083 + 1,8203\Delta KRT_t + 0,0636\Delta PP_t + 0,0714\Delta EKS_t - 0,3063ECT_t + \varepsilon_t \quad (10)$$

Model 2:

$$\Delta \ln(PDB_t) = 0,0019 + 0,8735\Delta \ln(KRT_t) + 0,0036\Delta \ln(PP_t) + 0,0334\Delta \ln(EKS_t) - 0,4274ECT_t + \varepsilon_t \quad (11)$$

Hasil pengujian parsial dengan taraf uji 5% menyatakan baik model 1 maupun model 2 didapatkan hasil yang signifikan pada variabel penjelas konsumsi rumah tangga (KRT), nilai ekspor (EKS) dan variabel koreksi error (ECT) dalam mempengaruhi perubahan nilai PDB. Adapun variabel penjelas yang tidak signifikan dalam mempengaruhi PDB adalah pengeluaran pemerintah (PP).

Pada model pertama, kecepatan penyesuaian model jangka pendek ke jangka panjang adalah sebesar 30,63%. Sedangkan pada model kedua, kecepatan penyesuaian model jangka pendek ke jangka panjang adalah sebesar 42,74%. Meskipun kedua angka tersebut masih jauh dari 1, akan tetapi dari hasil uji parsial, variabel koreksi ECT signifikan mempengaruhi PDB. Oleh karena itu, variabel ECT tetap dapat diperhitungkan dalam model.

6. Uji asumsi klasik

Setelah mendapatkan model ECM, pengujian asumsi klasik perlu dilakukan untuk mengetahui apakah kedua model dapat dipakai atau tidak. Terdapat 4 (empat) pengujian asumsi klasik yakni pengujian autokorelasi, heteroskedastisitas, multikolinearitas, dan normalitas.

6.1. Uji Autokorelasi

Salah satu asumsi klasik yang harus terpenuhi dalam model regresi adalah tidak terdapatnya autokorelasi. Jika dalam model terdapat autokorelasi maka parameter yang diestimasi akan bias dan model tidak dapat digunakan. Pengujian asumsi autokorelasi pada Eviews10 dapat dilakukan dengan Brusch-Godfrey Serial Correlation LM Test (Basuki, 2016).

H_0 : data tidak terjadi autokorelasi

H_1 : data terjadi autokorelasi

Tingkat signifikansi : $\alpha=5\%$

Tabel 7. Breusch-Godhey LM Test

Statistic	Model 1	Model 2
F-statistic	2.6407	0.3430
Obs*R-squared	5.3959	0.7433

Dari hasil pengujian Breusch-Godhey tersebut, pada taraf uji 5% data belum cukup bukti untuk dinyatakan terjadi autokorelasi. Maka dari itu,

asumsi klasik tidak terdapatnya autokorelasi pada kedua model ECM yang didapatkan telah terpenuhi.

6.2. Uji Heteroskedastisitas

Pengujian asumsi klasik heteroskedastisitas residual bertujuan untuk menguji apakah varians dari residual bersifat konstan atau berubah-ubah. Apabila varians residual berubah-ubah, maka akan mengakibatkan kondisi heteroskedastis yang akan mengakibatkan pelanggaran asumsi klasik (Rosadi, 2012). Pengujian heteroskedastisitas dapat dilakukan dengan Breusch-Pagan-Godhey Test sebagai berikut :

H_0 : data tidak terjadi heteroskedastisitas residual

H_1 : data terjadi heteroskedastisitas residual

Tingkat signifikansi : $\alpha=5\%$

Tabel 8. Breusch-Pagan-Godhey Test

Statistic	Model 1	Model 2
F-statistic	0.8766	1.9012
Obs*R-squared	3.5725	7.3709

Dari hasil pengujian Breusch-Pagan-Godhey tersebut, dengan taraf uji 5% residual kedua model belum cukup bukti untuk dinyatakan terdapat heteroskedastisitas residual. Maka dari itu, asumsi klasik tidak terjadi heteroskedastisitas pada kedua model ECM terpenuhi.

6.3. Uji Multikolinearitas

Dalam model regresi, terdapat asumsi klasik tidak terdapatnya hubungan dependensi antar variabel penjelas. Multikolinearitas adalah kondisi apabila dalam model regresi terjadi hubungan dependensi antar variabel penjelas. Pengujian asumsi klasik multikolinearitas dalam Eviews10 adalah dengan melihat Variance Inflation Factor (VIF). Sebuah variabel penjelas dinyatakan terdapat kolinearitas yang kuat apabila nilai VIF lebih dari 10 (Rosadi, 2012).

Tabel 9. Variance Inflation Factors (VIF)

Variabel	Model 1	Model 2
C	1.5046	0.0000
KRT	0.0067	0.0047
PP	0.0059	0.0001
EKS	0.0006	0.0000
ECT	0.0077	0.0085

Dari hasil pengecekan nilai VIF tersebut, semua variabel penjelas tidak ada yang memiliki $VIF > 10$ pada kedua model. Maka dari itu, asumsi klasik

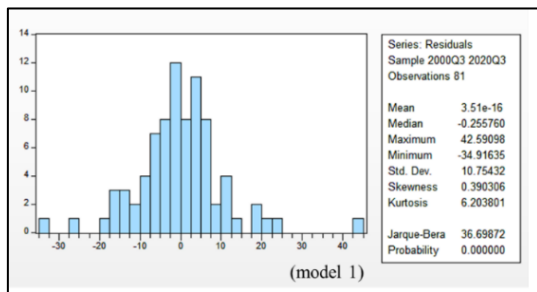
tidak terjadi multikolinearitas variabel penjelas pada kedua model ECM terpenuhi.

kedua model tersebut, maka diperlukan pemilihan model dengan menggunakan beberapa kriteria pemilihan model sebagai berikut :

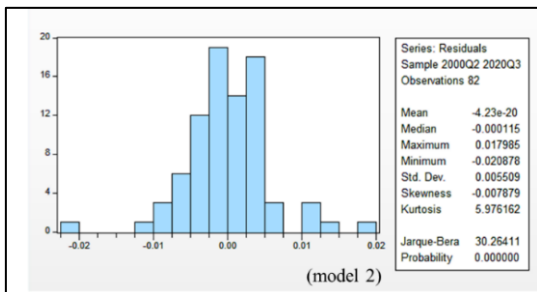
6.4. Uji Normalitas Residual

Asumsi klasik terakhir yang perlu diuji adalah asumsi normalitas residual. Pengujian ini digunakan untuk mengetahui apakah residual dari model mengikuti distribusi normal atau tidak. Pengujian normalitas residual dapat dilakukan dengan uji Jarque-Bera (Basuki, 2016)

H_0 : residual mengikuti distribusi normal
 H_1 : residual tidak mengikuti distribusi normal
 Tingkat signifikansi : $\alpha=5\%$



Gambar 3. Uji Normalitas Jarque-Bera Model 1



Gambar 4. Uji Normalitas Jarque-Bera Model 2

Dari hasil pengujian normalitas dengan statistik uji Jarque-Bera pada taraf uji 5%, residual kedua model tersebut tidak mengikuti distribusi normal. Maka dari itu, asumsi klasik normalitas residual pada kedua model ECM tersebut tidak terpenuhi. Namun, dalam sebagian besar kasus data deret waktu, pelanggaran asumsi normalitas residual dapat diabaikan.

7. Pemilihan model terbaik

Hasil pengujian asumsi klasik menyatakan kedua model ECM tersebut dapat digunakan. Untuk memilih salah satu model yang terbaik di antara

Tabel 10. Pemilihan model terbaik

Kriteria	Model 1	Model 2
R-square	0.9174	0.7473
Adj. R-square	0.9131	0.7342
F statistics	211.149	56.954
Log likelihood	-306.831	310.660
AIC	7.6995	-7.4551
SC	7.8473	-7.3083
HQC	7.7588	-7.3962

Dari kriteria pemilihan model tersebut, untuk memodelkan PDB Indonesia dengan metode ECM lebih direkomendasikan untuk menggunakan model kedua (model dengan transformasi), sebagai berikut:

$$\Delta \ln(PDB_t) = 0,0019 + 0,8735\Delta \ln(KRT_t) + 0,0036\Delta \ln(PP_t) + 0,0334\Delta \ln(EKS_t) - 0,4274ECT_t + \varepsilon_t \tag{12}$$

$$\text{dengan } ECT_t = \ln(PDB_{t-1}) - \ln(\widehat{PDB}_{t-1})$$

Berdasarkan persamaan tersebut, dapat diinterpretasikan antara lain: (1) Variabel konsumsi rumah tangga memiliki pengaruh positif terhadap PDB Indonesia sebesar 89,46% dalam jangka panjang dan sebesar 87,35% dalam jangka pendek; (2) variabel pengeluaran pemerintah berpengaruh positif terhadap PDB Indonesia sebesar 4,59% dalam jangka panjang dan sebesar 0,36% dalam jangka pendek; (3) variabel nilai ekspor memiliki pengaruh positif terhadap PDB Indonesia sebesar 5,64% dalam jangka panjang dan sebesar 3,34% dalam jangka pendek; (4) variabel ECT sebesar -0,4274, artinya fluktuasi keseimbangan jangka pendek akan dikoreksi sebesar 42,74% pada lag pertama dan 57,26% sisanya akan dikoreksi pada lag berikutnya, atau dapat diinterpretasikan sebagai kecepatan penyesuaian model jangka pendek ke model jangka panjang adalah sebesar 42,74%.

4. KESIMPULAN DAN SARAN

Beberapa hal yang dapat disimpulkan berdasarkan model terpilih dan hasil uji signifikansi parsial antara lain : (1) variabel konsumsi rumah tangga, pengeluaran pemerintah, dan nilai ekspor secara simultan berpengaruh signifikan terhadap PDB Indonesia; (2) variabel konsumsi rumah tangga dan nilai ekspor berpengaruh positif dan signifikan terhadap PDB Indonesia dalam jangka panjang maupun jangka pendek; (3) variabel pengeluaran

pemerintah berpengaruh positif dan signifikan terhadap PDB Indonesia dalam jangka panjang namun tidak signifikan mempengaruhi PDB dalam jangka pendek.

Berdasarkan kesimpulan tersebut terdapat beberapa saran dari penelitian ini, antara lain : (1) melihat pengaruh atau kontribusi konsumsi rumah tangga terhadap PDB yang cukup besar, yakni sebesar 87,35%, maka untuk dapat meningkatkan nilai PDB Indonesia secara efektif adalah dengan menaikkan konsumsi rumah tangga masyarakat Indonesia; (2) Meskipun memberikan pengaruh yang cukup kecil, pengeluaran pemerintah dan ekspor tidak dapat diabaikan dalam membantu meningkatkan nilai PDB Indonesia.

Adapun saran untuk penelitian selanjutnya adalah melanjutkan pemodelan ECM yang telah penelitian ini dapatkan agar memenuhi asumsi normalitas dengan melakukan transformasi atau menambah series. Penelitian lanjutan juga dapat dengan menambah variabel-variabel penjelas ekonomi lain seperti investasi, inflasi, dan nilai impor agar model yang dihasilkan dapat lebih menjelaskan variabel respons.

DAFTAR PUSTAKA

- Basuki, Agus T. & Prawoto, Nano. (2016). *Analisis Regresi dalam Penelitian Ekonomi & Bisnis*. Jakarta: Rajawali Pers.
- Enders, Walter. (2010). *Applied Econometric Time Series, Third Edition*. New Jersey: Wiley.
- FRED. (2020). *Fred Economic Data*. <https://fred.stlouisfed.org>
- Kurniawan, Robert & Yuniarto, Budi. (2016). *Analisis Regresi Dasar dan Penerapannya dengan R*. Jakarta : Penerbit Kencana.
- Mankiw, N Gregory. (2009). *Seventh Edition Macroeconomics*. New York: Worth Publishers
- Rafiq, Muhammad. (2016). *Pengaruh Pengeluaran Konsumsi Rumah Tangga, Investasi dan Pengeluaran Pemerintah Terhadap Pertumbuhan Ekonomi di Indonesia Tahun 2001:T1-2010:T4*. [Skripsi] Universitas Lampung. Bandar Lampung.
- Rosadi, Dedi. (2012). *Ekonometrika & Analisis Runtun Waktu Terapan dengan Eviews*. Yogyakarta : Penerbit Andi.

Rosadi, Dedi. (2014). *Aplikasi Runtun Waktu dan Aplikasinya dengan R*. Yogyakarta : UGM Press.

Safari, Menik F. & Fikri, Aula AHS. (2016). *Analisis Pengaruh Ekspor, Pembentukan Modal, dan Pengeluaran Pemerintah Terhadap Pertumbuhan Ekonomi Indonesia*. [Skripsi] Universitas Negeri Yogyakarta. Yogyakarta.

Sukirno, Sadono. (1994). *Makroekonomi : Teori Pengantar, Edisi Ketiga*. Jakarta: Rajagrafindo.

Walpole, Ronald E. (1982). *Pengantar Statistika Edisi ke-3*. Jakarta: Gramedia Pustaka Utama.

Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode *Geographically Weighted Panel Regression*

Siva Alfiani⁽¹⁾, Prizka Rismawati Arum⁽²⁾

Universitas Muhammadiyah Semarang

Jl. Kedungmundu No.18, Kedungmundu Telp. (024) 76740296 Semarang 50273

e-mail: sivaalfiani35@gmail.com dan prizka.rismawatiarum@unimus.ac.id

ABSTRAK

Salah satu tujuan negara adalah meningkatkan pertumbuhan ekonomi. Diperlukan pembangunan ekonomi untuk mewujudkan tujuan tersebut demi mencapai masyarakat yang sejahtera. Salah satu indikator pertumbuhan ekonomi adalah Produk Domestik Regional Bruto (PDRB). Data yang digunakan yaitu data sekunder tentang produk domestik regional bruto, jumlah penduduk miskin, pengeluaran pemerintah, rata-rata lama sekolah, tingkat partisipasi angkatan kerja, fasilitas kesehatan, tingkat pengangguran terbuka, pada tahun 2018 - 2020 di Provinsi Jawa Barat. Ternyata terdapat autokorelasi spasial dalam data tersebut, sehingga pemodelan yang tepat untuk data panel dan terdapat efek spasial dapat dilakukan menggunakan *Geographically Weighted Panel Regression* (GWPR). Dengan menggunakan GWPR diharapkan dapat menghasilkan hasil yang lebih menyeluruh dibandingkan dengan model GWR. Model *Geographically Weighted Panel Regression* yang dihasilkan yaitu model *fixed effect* dengan pembobot *adaptive gaussian kernel* dan *fixed gaussian kernel*. Tujuan dari penelitian ini adalah mengetahui gambaran umum data, mendapatkan model, dan memperoleh model terbaik pertumbuhan ekonomi di Jawa Barat. Hasil akhir dari penelitian ini menunjukkan bahwa model dengan pembobot *Adaptive Gaussian Kernel* lebih baik daripada *Fixed Gaussian Kernel* karena memiliki nilai AIC terkecil dan R^2 terbesar. Nilai AICnya sebesar 2313,117 dan nilai R^2 sebesar 0,7955945.

Kata kunci : *PDRB, Geographically Weighted Panel Regression, Adaptive Gaussian Kernel.*

ABSTRACT

Increasing economic growth is one of the country's goals. To realize this goal, economic development is needed to achieve a prosperous society. Gross Regional Domestic Product (GDP) is one of the indicators of economic growth. The data used are secondary data on gross regional domestic product, number of poor people, government spending, average length of schooling, labor force participation rate, health facilities, open unemployment rate, in 2018 - 2020 in West Java Province. After testing the spatial autocorrelation test, it turns out that there is spatial autocorrelation in the data. Appropriate modeling for panel data and spatial effects can be done using Geographically Weighted Panel Regression (GWPR). Using the GWPR is expected to produce more comprehensive results compared to the GWR model. The resulting Geographically Weighted Panel Regression model is a fixed effect model with adaptive gaussian kernel and fixed gaussian kernel weights. This study aims to find out the general description, get the model, and get the best model of economic growth in West Java. The final results of this study indicate that the Adaptive Gaussian Kernel weighted model is better than the Fixed Gaussian Kernel because it has the smallest AIC value and the largest R^2 . The AIC value is 2313.117 and the R^2 value is 0.7955945.

Keywords : *GDP, Geographically Weighted Panel Regression, Adaptive Gaussian Kernel.*

1. PENDAHULUAN

Salah satu tujuan suatu negara adalah untuk meningkatkan pertumbuhannya (Rahman dan Chamelia, 2015). Dalam upaya mencapai tujuan tersebut dibutuhkan peran serta pemerintah daerah dan masyarakat secara bersama-sama untuk mewujudkan pertumbuhan dan pembangunan

ekonomi. Untuk mewujudkan pertumbuhan ekonomi diperlukan adanya pembangunan ekonomi. Pembangunan merupakan suatu proses perubahan menuju ke arah yang lebih baik untuk mencapai tujuan yakni mewujudkan masyarakat Indonesia yang berkeadilan, berdaya saing, maju, dan sejahtera (Nandita et al. 2019). Pembangunan

ekonomi merupakan suatu proses perubahan kearah yang lebih baik dalam upaya meningkatkan kesejahteraan masyarakat (Parahita et al. 2018).

Salah satu indikator pertumbuhan ekonomi suatu wilayah adalah Produk Domestik Regional Bruto (PDRB). PDRB adalah nilai bersih barang dan jasa akhir yang dihasilkan oleh berbagai kegiatan ekonomi disuatu daerah dalam periode tertentu (Sasana, 2006). Tinggi rendahnya PDRB dipengaruhi oleh beberapa faktor, salah satunya dalam pemerintahan yaitu pengeluaran pemerintah. Pengeluaran pemerintah bagi kepentingan umum dan kesejahteraan umum mempengaruhi meningkat dan turunnya PDRB. Dalam teori Pertumbuhan Ekonomi Neo Klasik menyatakan pertumbuhan ekonomi bergantung padaperkembangan faktor-faktor produksi yaitu : modal, tenaga kerja dan teknologi (Sukirno, 1994). Selain itu terdapat faktor lainnya, yaitu : kemiskinan, pengangguran, dan lama sekolah. Jawa barat merupakan salah satu Provinsi yang memiliki jumlah penduduk yang tertinggi di Indonesia. Berdasarkan data proyeksi penduduk pada tahun 2020 jumlah penduduk di Jawa Barat sejumlah 49565,2 ribu jiwa. Pertumbuhan ekonomi di Jawa Barat dari tahun 2016-2019 terus meningkat hingga pada tahun 2020 mengalami penurunan. Penelitian ini dilakukan pada tahun 2018 hingga 2020 dengan harapan mendapatkan hasil yang lebih baik karena data lebih informatif. Untuk itu perlu dilakukan penelitian dengan melihat variabel-variabel apa saja yang dapat mempengaruhi pertumbuhan ekonomi.

Metode yang dapat digunakan untuk mempelajari pengaruh variabel bebas dan variabel terikat salah satunya metode regresi linier. Estimasi parameter *Ordinary Least Square* (OLS) dapat berlaku global untuk seluruh observasi. Model persamaan global akan memberikan informasi yang akurat untuk wilayah lokal jika tidak ada atau hanya sedikit keragaman antar wilayah lokalnya (Fotheringham, Brunson & Charlton, 2002 dalam Rahayu, 2017). Kondisi wilayah satu dengan yang lainnya berbeda – beda, perbedaan kondisi geografis suatu wilayah memungkinkan adanya heterogenitas spasial. Contohnya di Provinsi Jawa Barat yang memiliki kondisi wilayah geografis yang berbeda-beda antara Kabupaten/Kota satu dengan lainnya, dengan kondisi geografis yang berbeda diharapkan menghasilkan model yang beragam di setiap wilayahnya. Pada kasus keheterogenan berkembang analisis regresi terboboti secara spasial yaitu *Geographically Weighted Regression* (GWR). Model GWR dibangun dari pendekatan titik, yaitu *latitude* dan *longitude* (Rahayu, 2017).

Pemodelan untuk data panel dan terdapat efek spasial dapat dilakukan menggunakan *Geographically Weighted Panel Regression*. Dengan menggunakan GWPR diharapkan dapat menghasilkan hasil yang lebih menyeluruh dibandingkan dengan model GWR. Penelitian Fitri (2019) menggunakan GWPR pada IPM di Jawa Tengah menggunakan model *fixed effect* dan pembobot *adaptive kernel bisquare* dan *adaptive kernel gaussian*, hasil yang didapatkan pembobot *adaptive bisquare kernel* lebih baik karena mendapatkan nilai R^2 terbesar dan AIC terkecil. Penelitian yang dilakukan Surto et al (2020) pada index pembangunan manusia di Kalimantan Barat menggunakan GWPR dengan pembobot *fixed kernel gaussian*, *bisquare*, dan *tricube* dan model *fixed effect* model. Hasil pada penelitian tersebut pembobot *fixed gaussian* lebih baik karena mendapatkan CV paling minimum.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Geographically Weighted Panel Regression*. Data yang digunakan dalam penelitian ini merupakan data kuantitatif, yang merupakan data sekunder tentang produk domestik regional bruto, jumlah penduduk miskin, pengeluaran pemerintah, rata-rata lama sekolah, tingkat partisipasi angkatan kerja, fasilitas kesehatan, tingkat pengangguran terbuka, pada tahun 2018-2020 di Provinsi Jawa Barat yang diperoleh dari website resmi Badan Pusat Statistik yang dirilis pada tahun 2021. Tahapan atau langkah-langkah metode *Geographically Weighted Panel Regression* yaitu :

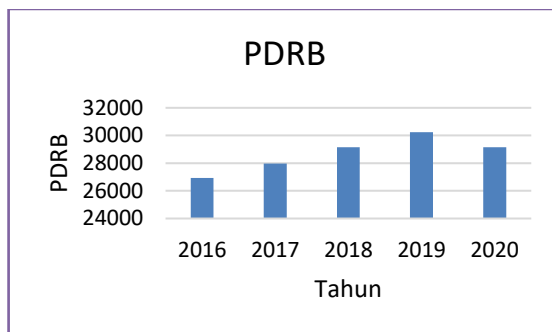
1. Pengumpulan data PDRB dan variabel yang mempengaruhinya.
2. Menentukan pendugaan parameter regresi data panel yaitu : *Common Effect Model* (CEM), *Fixed Effect Model* (FEM), dan *Random Effect Model* (REM).
3. Melakukan uji chow untuk menentukan model CEM dan FEM. Jika nilai p-value < 0,05 maka model yang dipilih adalah FEM.
4. Melakukan uji hausman untuk menentukan model REM dan FEM. Jika nilai p-value < 0,05 maka model yang dipilih adalah FEM.
5. Melakukan uji asumsi terhadap model, yaitu uji normalitas, uji multikolinieritas, uji heteroskedastisitas dan uji autokorelasi.
6. Melakukan uji heterogenitas spasial menggunakan *Breusch - Pagan*.
7. Menghitung jarak *euclidian* antar lokasi.
8. Menentukan *bandwith* optimum berdasarkan kriteria CV minimum.

9. Menghitung masing-masing pembobot *adaptive gaussian kernel* dan *fixed gaussian kernel*.
10. Melakukan estimasi parameter model data panel dengan pembobot *adaptive gaussian kernel* dan *fixed gaussian kernel*.
11. Menghitung nilai R^2 dan AIC.
12. Memilih model terbaik antara pembobot *adaptive gaussian kernel* dan *fixed gaussian kernel*.
13. Menginterpretasikan model.

3. HASIL DAN PEMBAHASAN

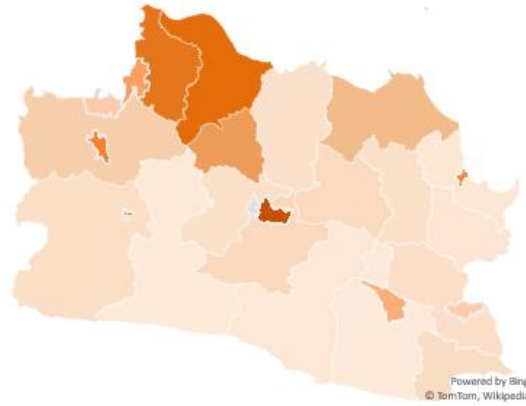
3.1 Gambaran Umum

Jawa Barat merupakan Provinsi yang berada di sebelah barat pulau Jawa. Letak geografis Jawa Barat berada di $5^{\circ}50'-7^{\circ}50'$ lintang selatan dan $104^{\circ}48'-108^{\circ}48'$ bujur timur dengan luas wilayah sebesar 3.710.061,32 hektar. Secara administratif Jawa Barat memiliki 27 Kabupaten/Kota dengan jumlah penduduk terbesar di Indonesia. Berdasarkan data proyeksi penduduk pada tahun 2020 jumlah penduduk di Jawa Barat sejumlah 49.565,2 ribu jiwa. Dengan banyaknya jumlah penduduk dapat mempengaruhi keadaan perekonomian. Tingkat pertumbuhan ekonomi terjadi akibat peningkatan produksi dan jasa. Pertumbuhan ekonomi suatu wilayah dapat dilihat dari PDRBnya.



Gambar 1. Grafik PDRB Jawa Barat Tahun 2016-2020

Gambar 1 memberikan informasi PDRB Provinsi Jawa Barat dari tahun 2016 hingga 2020. PDRB di Jawa Barat terus meningkat dari tahun 2016 hingga tahun 2019 akan tetapi mengalami penurunan pada tahun 2020. Penurunan sebesar 1,104 juta terjadi pada tahun 2020 sehingga PDRB menjadi 29,142 juta. Nilai PDRB tertinggi berada di tahun 2019 sebesar 30,245 juta dan terendah pada tahun 2016 sebesar 26,924 juta.



Gambar 2. Peta Jawa Barat

PDRB disetiap Kabupaten/Kota Jawa Barat berbeda-beda. Pada gambar 2 menunjukkan peta Jawa Barat, semakin pekat warnanya maka semakin tinggi nilai PDRB sebaliknya, semakin cerah warna peta maka semakin rendah nilai PDRBnya. Kota Bandung yang merupakan ibu kota Jawa Barat mempunyai PDRB tertinggi sebesar 769.470.232,29 rupiah. Daerah dengan PDRB terendah yaitu Kabupaten Tasikmalaya sebesar 138.669.613,3 rupiah.

3.2 Model Regresi Data Panel

3.2.1 Uji Normalitas

Uji normalitas bertujuan untuk menguji apakah dalam model regresi, variabel terikat dan variabel bebas memiliki data yang terdistribusi normal atau tidak. Hasil dari uji normalitas adalah sebagai berikut :

- a. Hipotesis
 - H_0 : Residual berdistribusi normal
 - H_1 : Residual tidak berdistribusi normal
- b. Tingkat Signifikansi
 - $\alpha = 0,05$
- c. Statistik Uji
 - P-value = 0,17
- d. Kriteria Penolakan
 - Tolak H_0 pada taraf signifikansi α jika p-value $< 0,05$
- e. Kesimpulan
 - Berdasarkan hasil analisis yang dilakukan menggunakan taraf signifikansi 5% dengan uji shapiro-wilk, didapatkan nilai p-value (0,17) $>$ (0,05) maka keputusan yang diperoleh yakni terima H_0 . Kesimpulannya data berdistribusi normal.

3.2.2 Uji Multikolinieritas

Uji Multikolinieritas bertujuan menguji adanya korelasi antar variabel bebas pada model

regresi. nilai VIF (*Variance Inflation Factor*) digunakan dalam mendeteksi adanya multikolinieritas. Nilai VIF yang kurang dari 10 menunjukkan bahwa dalam data tidak terdapat multikolinieritas. Hasil uji multikolinieritas ditunjukkan sebagai berikut:

- a. Hipotesis
 H_0 : Tidak terdapat multikolinieritas
 H_1 : Terdapat multikolinieritas
- b. Tingkat Signifikansi
 $\alpha = 0,05$
- c. Statistik Uji (nilai VIF pada output)
 Jumlah penduduk = 2,43
 Pengeluaran pemerintah = 4,13
 Rata-rata lama sekolah = 3,14
 TPAK (tingkat partisipasi angkatan kerja)=1,45
 Fasilitas kesehatan = 3,41
 TPT (tingkat pengangguran terbuka) = 1,61
- d. Kriteria Penolakan
 Tolak H_0 pada taraf signifikansi α jika $VIF > 10$
- e. Kesimpulan
 Berdasarkan uji multikolinieritas semua variabel memiliki nilai $VIF < 10$. Kesimpulannya tidak terdapat multikolinieritas di setiap variabelnya. Hasil uji multikolinieritas dapat dilihat dalam tabel berikut :

Tabel 1. Uji Multikolinieritas

Variabel	VIF	Kesimpulan
Jumlah Penduduk Miskin	2,43	Tidak Terdapat Multikolinieritas
Pengeluaran Pemerintah	4,13	Tidak Terdapat Multikolinieritas
Rata-rata Lama Sekolah	3,14	Tidak Terdapat Multikolinieritas
Tingkat Partisipasi Angkatan Kerja	1,45	Tidak Terdapat Multikolinieritas
Fasilitas Kesehatan	3,41	Tidak Terdapat Multikolinieritas
Tingkat Pengangguran Terbuka	1,61	Tidak Terdapat Multikolinieritas

3.2.3 Uji Autokorelasi

Autokorelasi dapat diartikan sebagai adanya korelasi antara observasi satu dengan observasi lain yang berlainan waktu. Konsekuensi dari adanya autokorelasi dalam suatu model regresi adalah varians sampel tidak dapat menggambarkan varians populasinya. Untuk mendiagnosis adanya

autokorelasi dalam suatu model regresi dilakukan melalui pengujian terhadap nilai uji Durbin-Watson.

- a. Hipotesis
 H_0 : Tidak terdapat autokorelasi
 H_1 : Terdapat autokorelasi
- b. Tingkat Signifikansi
 $\alpha = 0,05$
- c. Kriteria Penolakan
 $dw < dL$ = Terjadi Autokorelasi Positif
 $dw > (4 - dL)$ = Terjadi Autokorelasi Negatif
 $dL \leq dw \leq du$ atau $4 - du \leq dw \leq 4 - dL$ = Tidak Ada kesimpulan
 $du < dw < 4 - du$ = Tidak Terjadi Autokorelasi
 Kriteria Penolakan
 Tolak H_0 pada taraf signifikansi α jika p-value $< 0,05$
- d. Statistik Uji
 Sebelum penanganan = 1,6267, p-value = 0,03
 Setelah penanganan = 1,73544, p-value = 0,11
- e. Kesimpulan
 Berdasarkan hasil sebelum penanganan didapatkan hasil $dl(1,4842) \leq 1,6267 \leq du(1,808)$, p-value (0,03) $< 0,05$ keputusan yang diperoleh yaitu menolak H_0 , kesimpulannya terdapat autokorelasi. Setelah dilakukan penanganan yang hasil diperoleh $dl(1,4842) \leq 1,73544 \leq du(1,808)$, p-value (0,11) $> 0,05$ yaitu menerima H_0 , kesimpulannya tidak terdapat autokorelasi.

3.2.4 Uji Heterokedastisitas

Uji heteroskedastisitas digunakan untuk melihat apakah terdapat ketidaksamaan varians dari residual satu ke pengamatan-pengamatan lainnya.

- a. Hipotesis
 H_0 : Tidak terdapat heterogenitas
 H_1 : Terdapat heterogenitas
- b. Tingkat Signifikansi
 $\alpha = 0,05$
- c. Kriteria Penolakan
 Jika p-value $\leq 0,05$ maka tolak H_0 , artinya terdapat heterogenitas.
- d. Statistik Uji
 Nilai p-value pada uji *breusch-pagan*= $2,68 \times 10^{-8}$
- e. Kesimpulan
 Berdasarkan hasil diatas diperoleh nilai p-value ($2,68 \times 10^{-8}$) $< 0,05$, maka keputusannya H_0 ditolak, kesimpulannya data bersifat heterogen. Dengan adanya keragaman antar data tersebut maka dapat dilakukan pemodelan menggunakan GWR-Panel.

3.2.5 Uji Chow dan Uji Hausman

Chow test digunakan untuk memilih antara Model *Common Effect* dan Model *Fixed Effect*. Apabila nilai p-value < 0,05 maka akan dilanjutkan dengan uji hausman. Uji hausman digunakan untuk membandingkan model *Fixed Effect* dengan *Random effect*. Hasil dari uji chow dan uji hausman adalah :

- a. Hipotesis
 - Uji Chow
 - H₀ : Model *Common Effect*
 - H₁ : Model *Fixed Effect*
 - Uji Hausman
 - H₀ : Model *Random Effect*
 - H₁ : Model *Fixed Effect*
- b. Tingkat Signifikansi
 - α = 0,05
- c. Kriteria Penolakan
 - Uji Chow
 - Jika p-value < 0,05 maka tolak H₀, artinya model *fixed effect*.
 - Uji Hausman
 - Jika p-value < 0,05 maka tolak H₀, artinya model *fixed effect*.
- d. Statistik Uji
 - Nilai p-value uji chow (2,20 × 10-16) < 0,05
 - Nilai p-value uji hausman (2,20 × 10-16) < 0,05
- e. Kesimpulan
 - Berdasarkan hasil uji chow nilai p-value (2,20 × 10-16) < 0,05 maka H₀ ditolak jadi model yang didapat *fixed effect* kemudian dilanjutkan dengan uji hausman. Hasil yang didapat yaitu p-value (2,20 × 10-16) < 0,05, maka keputusannya yaitu H₀ ditolak. Kesimpulan yang didapat yaitu model yang diperoleh *fixed effect*.

Berdasarkan dari uji tersebut, model *fixed effect* di Provinsi Jawa Barat yang didapatkan yaitu :

$$\hat{Y}_{it} = 3990241,6 + 2767,2 X_{2,it} - 266474,7 X_{5,it}$$

Nilai intersep di Provinsi Jawa Barat didapatkan dari rata - rata semua intersep disetiap Kabupaten/Kota.

3.3 Model Pertumbuhan Ekonomi Dengan Pembobot Adaptive Gaussian Kernel Dan Fixed Gaussian Kernel Untuk Masing-Masing Kabupaten/Kota.

Langkah awal dari pemodelan adalah pembentukan matriks pembobot yang didasarkan jarak antar lokasi. Fungsi pembobot yang digunakan adalah *Adaptive Gaussian Kernel* dan *Fixed*

gaussian Kernel dimana masing-masing pembobot memerlukan *bandwith* optimum. Penentuan *bandwith* optimum tergantung pada CV.

Tabel 2. Bandwith Optimum Masing-Masing Kabupaten/Kota di Jawa Barat

Kab/Kota	Band width	Kab/Kota	Band width
Bogor	1,73	Karawang	1,33
Sukabumi	1,56	Bekasi	1,64
Cianjur	1,36	Bandung Barat	1,05
Bandung	0,94	Pangandaran	1,98
Garut	1,27	Kota Bogor	1,73
Tasikmalaya	1,59	Kota Sukabumi	1,56
Ciamis	1,71	Kota Bandung	0,94
Kuningan	1,66	Kota Cirebon	1,65
Cirebon	1,65	Kota Bekasi	1,64
Majalengka	1,37	Kota Depok	1,76
Sumedang	8,12	Kota Cimahi	0,96
Indramayu	1,50	Kota Tasikmalaya	1,59
Subang	0,95	Kota Banjar	1,90
Purwakarta	1,13		

Nilai *bandwith* optimum dimasing-masing lokasi berbeda-beda. Nilai *bandwith* optimum tersebut digunakan untuk menentukan mendapat pembobot di masing-masing Kabupaten/Kota.

Setelah dilakukan pemodelan menggunakan regresi data panel selanjutnya dilakukan estimasi parameter yang berpengaruh signifikan terhadap pertumbuhan ekonomi menggunakan metode GWR-Panel dikarenakan adanya kasus heterogenitas spasial atau menunjukkan adanya pengaruh letak geografis terhadap model regresi data panel. Berikut merupakan contoh estimasi parameter Kabupaten/Kota di Jawa Barat :

Tabel 3. Estimasi Parameter dengan Pembobot Adaptive Gaussian Kernel

	Sumedang	Pangandaran	Kota Banjar
Jumlah Penduduk Miskin	-6237,6	2900,5	6332,5

Pengeluaran Pemerintah	2776,8	2600,8	2607,3
Rata-rata Lama Sekolah	-495826,3	391274,2	564386,0
TPAK	23105,6	30285,8	13663,9
Fasilitas Kesehatan	-270595,0	-104449,0	-53739,5
TPT	112965,8	79543,5	70549,0
\hat{y}	-568353,4	-521939,6	-564122,5

Tabel 4. P-value Pembobot *Adaptive Gaussian Kernel*

	Sumedang	Pangandaran	Kota Banjar
Jumlah Penduduk Miskin	0,41	0,79	0,62
Pengeluaran Pemerintah	0,00	0,00	0,00
Rata-rata Lama Sekolah	0,43	0,66	0,59
TPAK	0,48	0,48	0,77
Fasilitas Kesehatan	0,00	0,10	0,47
TPT	0,08	0,27	0,38

Tabel 3 menunjukkan hasil estimasi parameter model dengan pembobot *Adaptive Gaussian Kernel*, dan tabel 4 menunjukkan nilai p.value yang akan digunakan untuk menguji signifikansi dari masing-masing parameter. Berdasarkan tabel 3 dan 4 dapat dimodelkan pertumbuhan ekonomi pada masing-masing Kabupaten/Kota dengan pembobot *Adaptive Gaussian Kernel*. Misalkan model pada Kabupaten Sumedang yaitu :

$$\hat{Y}_{it} = -492,6 + 2776,8 X_{2,it} - 270595,0 X_{5,it}$$

Kabupaten Pangandaran :

$$\hat{Y}_{it} = 305,5 + 2600,8 X_{2,it}$$

Kota Banjar :

$$\hat{Y}_{it} = 432,8 + 2607,3 X_{2,it}$$

Model Kabupaten / Kota yang lainnya dapat ditulis berdasarkan nilai p-value < 0,05.

Tabel 5. Variabel Yang Mempengaruhi Pertumbuhan Ekonomi di Kabupaten/Kota

Jumlah Penduduk Miskin	-	Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar
Pengeluaran Pemerintah	-	Bekasi, Kota Cirebon, Kota Bekasi
Rata-rata Lama Sekolah	-	Bogor, Sukabumi, Cianjur, Bandung, Garut, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya
TPAK (Tingkat Partisipasi Angkatan Kerja)	-	Bekasi, Kota Cirebon, Kota Bekasi
Fasilitas Kesehatan	-	
TPT (Tingkat Pengangguran Terbuka)	-	

Variabel jumlah penduduk tidak mempengaruhi pertumbuhan ekonomi di Kabupaten/Kota Jawa Barat karena nilai p-value yang melebihi 0,05 begitu juga dengan variabel TPAK (Tingkat Partisipasi Angkatan Kerja). Variabel pengeluaran pemerintah mempengaruhi pertumbuhan ekonomi di semua Kabupaten/Kota. Variabel rata-rata lama sekolah dan TPT (Tingkat Pengangguran Terbuka) mempengaruhi pertumbuhan ekonomi di 3 Kabupaten/Kota. Variabel fasilitas kesehatan mempengaruhi pertumbuhan ekonomi di 19 Kabupaten/Kota.

Tabel 6. Estimasi Parameter dengan Pembobot *Fixed Gaussian Kernel*

	Sumedang	Pangandar an	Kota Banjar
Jumlah Penduduk Miskin	-5984,4	-5918,8	-5920,8
Pengeluaran Pemerintah	2768,5	2767,6	2767,7
Rata-rata Lama Sekolah	-500488,6	-500070,6	-500105,3
TPAK	23671,7	23582,9	23584,8
Fasilitas Kesehatan	-266936,5	-266109	-266135,4
TPT	110501,3	109865,1	109883,8
\hat{y}	-563818,3	-455388,1	-602528,2

Tabel 7. P-value Pembobot *Fixed Gaussian Kernel*

	Sumedang	Pangan daran	Kota Banjar
Jumlah Penduduk Miskin	0,50	0,51	0,51
Pengeluaran Pemerintah	0,00	0,00	0,00
Rata-rata Lama Sekolah	0,50	0,50	0,50
TPAK	0,53	0,54	0,54
Fasilitas Kesehatan	0,001	0,001	0,001
TPT	0,15	0,15	0,15

Tabel 6 menunjukkan hasil estimasi parameter model dengan pembobot *Adaptive Gaussian Kernel*, dan tabel 7 menunjukkan nilai p.value yang akan digunakan untuk menguji signifikansi dari masing-masing parameter. Berdasarkan tabel 6 dan 7 dapat dimodelkan pertumbuhan ekonomi pada masing-masing Kabupaten/Kota dengan pembobot *Fixed Gaussian Kernel*.

Misalkan model pada :
Kabupaten Sumedang :

$$\hat{Y}_{it} = -490,3 + 2768,5X_{2i,t} - 266936,5X_{5i,t}$$

Kabupaten Pangandaran :

$$\hat{Y}_{it} = -489,5 + 2767,6X_{2i,t} - 266109X_{5i,t}$$

Kota Banjar :

$$\hat{Y}_{it} = -489,5 + 2767,7X_{2i,t} - 266135,4X_{5i,t}$$

Model Kabupaten / Kota yang lainnya dapat ditulis berdasarkan nilai p-value < 0,05.

Tabel 8. Variabel Yang Mempengaruhi Pertumbuhan Ekonomi di Kabupaten/Kota

Jumlah Penduduk Miskin	-
Pengeluaran Pemerintah	Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar
Rata-rata Lama Sekolah	-
TPAK (Tingkat Partisipasi Angkatan Kerja)	-
Fasilitas Kesehatan	Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar
TPT (Tingkat Pengangguran Terbuka)	-

Variabel jumlah penduduk, rata-rata lama sekolah, TPAK (Tingkat Partisipasi Angkatan Kerja), dan TPT (Tingkat Pengangguran Terbuka) tidak mempengaruhi pertumbuhan ekonomi di Kabupaten/Kota Jawa Barat karena nilai p-value yang melebihi 0,05. Variabel pengeluaran pemerintah dan fasilitas kesehatan mempengaruhi pertumbuhan ekonomi di semua Kabupaten/Kota.

3.4 Model Terbaik Pertumbuhan Ekonomi Menggunakan Geographically Weighted Panel Regression

Model terbaik pertumbuhan ekonomi di Jawa Barat menggunakan metode *Geographically Weighted Panel Regression* adalah menggunakan pembobot *Adaptive Gaussian Kernel*.

Tabel 9. Perbandingan Nilai AIC dan R²

Pembobot	AIC	R2
<i>Adaptive Gaussian Kernel</i>	2313,117	0,7955945
<i>Fixed Gaussian Kernel</i>	2342,484	0,6709164

Dalam pemilihan model terbaik merupakan proses evaluasi dari model untuk mengetahui seberapa besar peluang masing-masing model yang terbentuk sesuai dengan data. Model terbaik untuk pertumbuhan ekonomi di Jawa Barat didapatkan dari nilai AIC terkecil dan nilai koefisien determinasi (R²) terbesar. Pada tabel diatas pembobot *Adaptive Gaussian Kernel* lebih baik karena memiliki AIC terkecil dan nilai koefisien determinasi (R²) terbesar. Nilai AIC didapat sebesar 2313,117 dan nilai R² sebesar 0,7955945. Model terbaik yang didapat, contohnya :

Kota Bandung :

$$\hat{Y}_{it} = -491,18 + 2767,96X_{2,it} - 266338 X_{5,it}$$

Berdasarkan hasil model *Geographically Weighted Panel Regression* yang didapatkan, model dapat diinterpretasikan sebagai berikut :

1. Jika variabel pengeluaran pemerintah mengalami peningkatan 1 satuan maka PDRB di Kota Bandung mengalami kenaikan sebesar 2767,96.
2. Jika variabel fasilitas kesehatan mengalami peningkatan 1 satuan maka PDRB di Kota Bandung mengalami penurunan sebesar 266338.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

1. PDRB tertinggi berada di Kota Bandung yang merupakan ibu kota Jawa Barat sebesar 76947023,29 rupiah. Daerah dengan PDRB terendah yaitu Kabupaten Tasikmalaya sebesar 13866961.3 rupiah.
2. Model pertumbuhan ekonomi di Jawa Barat menggunakan *Geographically Weighted Panel Regression* dengan pembobot *Adaptive Gaussian Kernel* dan *Fixed Gaussian Kernel*

menghasilkan model yang berbeda-beda karena variabel yang mempengaruhinya tergantung ada nilai p-valuenya. Model terbaik pertumbuhan ekonomi di Jawa Barat menggunakan pembobot *Adaptive Gaussian Kernel* karena memiliki nilai AIC terkecil dan R² terbesar. Nilai AIC didapat sebesar 2313,117 dan nilai R² sebesar 0,7955945.

4.2 Saran

Saran untuk penelitian selanjutnya yaitu :

1. Pemodelan *Geographically Weighted Panel Regression* dapat dilakukan menggunakan pembobot lainnya yaitu *adaptive tricube kernel, adaptive bisquare kernel, fixed tricube kernel, dan fixed bisquare kernel* .
2. Variabel bebas dapat ditambah dengan variabel lainnya seperti dana alokasi daerah, modal pemerintah dan investasi.

DAFTAR PUSTAKA

Arum, Prizka Rismawati. 2019. “Analisis Faktor-Faktor Yang Mempengaruhi Jumlah Penduduk Di Kota Semarang Menggunakan Metode Regresi Data Panel.” *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika* 12 (2): 36–41. <https://doi.org/10.36456/jstat.vol12.no2.a2227>

Bruna, Fernando, and Danlin Yu. 2013. “Geographically Weighted Panel Regression.” In *XI Congreso Galego de Estatística e Investigación de Operacións*. Http://Xisgapeio. Udc. Es., 1–23.

Badan Pusat Statistik Provinsi Jawa Barat

F, Kalbuadi A. 2014. “BAB II Tinjauan Pustaka_2010isa.Pdf.” Apriani, no. 1969: 9–66.

Fitri, Erlinda Permata. 2019. “Perbandingan Model Geographically Weighted Panel Regression Dengan Pembobot Adaptive Gaussian Dan Adaptive Bisquare Untuk Indeks Pembangunan Manusia Di Jawa Tengah.”

García Reyes, Luis Enrique. 2017. “Fasilitas Pelayanan Kesehatan.” *Journal of Chemical Information and Modeling* 53 (9): 1689–99.

Lutfiani, Nurul. 2019. Pemodelan Geographically Weighted Regression (GWR) Dengan Fungsi Pembobot Kernel Gaussian Dan Bi-Square. *Unnes Journal of Mathematics*. Vol. 8. <https://doi.org/10.15294/ujm.v8i1.17103>.

Martha, Shantika, Yundari Yundari, Setyo Wira Rizki, and Ray Tamtama. 2021. “Penerapan

- Metode Geographically Weighted Panel Regression (Gwpr) Pada Kasus Kemiskinan Di Indonesia.” BAREKENG: Jurnal Ilmu Matematika Dan Terapan 15 (2) : 241–48.
<https://doi.org/10.30598/barekengvol15iss2pp241-248>.
- Martha, Shantika. 2020. “Fixed Effect Geographically Weighted Panel Regression” 09 (3): 413–22.
- Nandita, Dea Aulia, Lalu Bayu Alamsyah, Enggar Prima Jati, and Edy Widodo. 2019. “Regresi Data Panel Untuk Mengetahui Faktor-Faktor Yang Mempengaruhi PDRB Di Provinsi DIY Tahun 2011-2015.” Indonesian Journal of Applied Statistics 2 (1): 42.
<https://doi.org/10.13057/ijas.v2i1.28950>.
- Nurchayani, I R A. 2017. “Pemilihan Variabel Prediktor Dengan Metode Stepwise Pada Model Geographically Weighted Regression.”
- Parahita, Lantip Lila, Dijan Rahajuni, and Kikin Windhani. 2018. “Analisis Faktor Yang Mempengaruhi Produk Domestik Regional Bruto (PDRB) Sektor Industri Di Provinsi Jawa Barat Tahun 2002-2016.” Journal of Sustainable Competitive Advantage 8 (September): 1–13.
- Pendidikan, Departemen, and Matematika Fpmipa. 2016. “Aplikasi Model Geographically Weighted Regression (Gwr) Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Kasus Gizi Buruk Anak Balita Di Jawa Barat.” Jurnal EurekaMatika 4 (1): 46–63.
- Puteri, Nabila Audia. 2018. "Analisis Pengangguran Terbuka Di Provinsi Jawa Tengah Tahun 2011-2016."
- Prasanti, Tyas Ayu, Triastuti Wuryandari, and Agus Rusgiyono. 2015. “Aplikasi Regresi Data Panel Untuk Pemodelan Tingkat Pengangguran Terbuka Kabupaten/Kota Di Provinsi Jawa Tengah.” None 4 (3): 687–96.
- Rahayu, Nunik Sri. 2017. " Geographically Weighted Panel Regression Untuk Pemodelan Presentase Penduduk Miskin Di Provinsi Jawa Tengah."
- Rahman, Yozi Aulia, and Ayunda Lintang Chamelia. 2015. “Faktor - Faktor Yang Mempengaruhi Pdrb Kabupaten / Kota Jawa Tengah Tahun 2008-2012.” Jejak 8 (1): 88–99.
<https://doi.org/10.15294/jejak.v8i1.3857>.
- Sasana, Hadi. 2006. “Analisis Dampak Desentralisasi Fiskal Terhadap Pertumbuhan Ekonomi Di Kabupaten/Kota Provinsi Jawa Tengah.” Dinamika Pembangunan 3 (2): 145–70.
- Sukirno. 2011. “Teori PDRB” 331 (02): 12–48.
- Udara, Raden. 2013. “Investasi Dan Tenaga Kerja Ekonomi Serta Hubungannya Terhadap ...” Jurnal Ilmu Ekonomi 1 (1): 2–3.

Perbandingan Metode *LightGBM* dan *XGBoost* dalam Menangani Data dengan Kelas Tidak Seimbang

Putri Septiana Rizky⁽¹⁾, Ristu Haiban Hirzi⁽²⁾, Umam Hidayaturrohman⁽³⁾

¹²³Universitas Hamzanwadi Selong

Jl. TGKH. Muhammad Zainuddin Abdul Madjid Pancor, Selong, Lombok Timur, Telp./Fax.
(0376) 22954 NTB 83612

e-mail: putriseptianarizky@gmail.com, ristuastalavista@gmail.com dan
hidayaturrohman4@gmail.com

ABSTRAK

Masalah ketidakseimbangan kelas telah menjadi salah satu tantangan dalam kinerja banyak algoritma klasifikasi. Kelas tidak seimbang merupakan suatu kondisi dimana terdapat dataset yang jumlah kelasnya terdapat perbedaan yang signifikan terhadap masing-masing jumlah kelas. Dalam kumpulan data yang terdiri dari dua kelas, ukuran sampel kategori mayoritas (lebih besar) mendominasi sampel kategori minoritas (lebih kecil) dengan rasio sebesar 1:100, 1:1.000 atau 1:10.000. Dampak ketidakseimbangan ini menyebabkan klasifikasi menjadi buruk dan tidak optimal. Sebagian besar algoritma klasifikasi standar cenderung mengklasifikasikan kelas mayoritas dengan tingkat akurasi tinggi dan kelas minoritas dengan tingkat akurasi rendah, sehingga mengakibatkan terjadinya bias. Dalam banyak aplikasi, lebih penting untuk mengidentifikasi kelas minoritas dari pada kelas mayoritas. Pada penelitian ini diusulkan pendekatan berbasis *ensemble* dengan pengklasifikasi yang digunakan adalah *LightGBM* dan *XGBoost*, kedua metode ini merupakan metode gradien efisien yang beberapa tahun terakhir telah disarankan berdasarkan pohon keputusan sehingga mampu menangani masalah data dengan skala besar. Data yang digunakan diperoleh dari *UCI Repository* dengan 5 data, 3 diantaranya memiliki tingkat ketidakseimbangan tinggi dan sisanya dengan tingkat ketidakseimbangan rendah. Jumlah kelas yang digunakan pada penelitian adalah dua kelas. Hasil penelitian menunjukkan bahwa kinerja metode *XGBoost* dalam akurasi dan sensitivitas lebih baik dibandingkan *LightGBM* di hampir seluruh data. Sedangkan kemampuan dalam menebak kelas minoritas (spesifisitas), metode *LightGBM* lebih baik dibandingkan *XGBoost* dengan nilai keseluruhan rata-rata sebesar 80,41% : 74,64%.

Kata kunci : *LightGBM, XGBoost, Ensemble, Klasifikasi, Kelas Tidak Seimbang*

ABSTRACT

Class imbalance problem had become one of the challenges in the performance of many classification algorithms. Imbalanced class is a condition where there is a dataset in which the number of classes has a significant difference to each number of classes. In a data set consisting of two classes, the sample size of the majority category (larger) dominates the sample of the minority category (smaller) with a ratio of 1:100, 1:1.000 or 1:10.000. The impact of this imbalance causes the classification to be poor and not optimal. Most of the standard classification algorithms tend to classify the majority class with high accuracy and minority class with low accuracy, resulting in bias. In many applications, it is more important to identify the minority class than the majority class-based approach is proposed ensemble with the classifiers used are LightGBM and XGBoost. The data used is obtained from the UCI Repository with 5 data, 3 of which have a high level of imbalance and the rest with a low level of imbalance. The number of classes used in the study were two classes. The results showed that the performance of the XGBoost in accuracy and sensitivity was better than LightGBM in almost all data. While the ability to guess the minority class (specificity), the LightGBM is better than XGBoost with an average overall score of 80.41% :74.64%.

Keywords : *LightGBM, XGBoost, Ensemble, Classification, Imbalanced Class*

1. PENDAHULUAN

Masalah ketidakseimbangan kelas telah menjadi salah satu tantangan dalam kinerja banyak algoritma klasifikasi (Prasetio dan Pratiwi, 2015). *Imbalanced class* (kelas tidak seimbang) merupakan suatu kondisi dimana terdapat dataset yang jumlah kelasnya tidak sama atau terdapat perbedaan yang signifikan terhadap masing-masing jumlah kelas (Astuti dan Lenti, 2021). Sehingga akan memberikan dampak yang tidak baik terhadap klasifikasi (Siringoringo, 2018). Karena suatu algoritma klasifikasi mengumpamakan distribusi kelas pada *dataset* yang relatif seimbang (Pristyanto, 2019). Dalam kumpulan data yang terdiri dari dua kelas, ukuran sampel kategori mayoritas (lebih besar) mendominasi sampel kategori minoritas (lebih kecil) dengan rasio sebesar 1:100, 1:1.000 atau 1:10.000 (Elrahman dan Abraham, 2013).

Dampak ketidakseimbangan ini menyebabkan dampak buruk pada proses *machine learning* seperti klasifikasi maupun klusterisasi, pada akhirnya hasil yang diperoleh buruk dan tidak optimal (Siringoringo, 2018). Sebagian besar algoritma klasifikasi tunggal cenderung mengklasifikasikan kelas mayoritas dengan tingkat akurasi tinggi dan kelas minoritas dengan tingkat akurasi rendah, sehingga mengakibatkan terjadinya bias (Galar dkk, 2011). Dalam banyak aplikasi, lebih penting untuk mengidentifikasi kelas minoritas dari pada kelas mayoritas (Sulistiyowati dan Jajuli, 2020). Misalnya pada masalah medis (Kothan, 2015), masalah klasifikasi teks (Wu dkk, 2014), memprediksi cacat *software* (Saifudin dan Wahono, 2015) dan lain sebagainya.

Terdapat tiga pendekatan untuk menangani kumpulan data dengan kelas tidak seimbang, yaitu level algoritma, level data dan menggabungkan atau memasangkan metode (*Ensemble*) (Yap dkk, 2014). Pada penelitian ini akan berfokus pada penanganan kelas tidak seimbang menggunakan pendekatan *Ensemble learning*. Kelebihan Teknik *Ensemble learning* dibandingkan dengan *Single classifier* adalah mampu menghasilkan banyak pohon keputusan sehingga dalam melakukan prediksi lebih stabil dan efisien daripada *single classifier* (Lessmann dkk, 2015). *Ensemble* adalah algoritma yang menggabungkan beberapa metode klasifikasi tunggal yang bertujuan untuk menghasilkan model klasifikasi yang lebih akurat dan telah berhasil digunakan dalam menangani masalah *machine learning*, seperti pemilihan fitur, estimasi kepercayaan, fitur yang hilang dan kelas data (Zhang dan Ma, 2012). Prinsip kerja metode *Ensemble* menurut Sartono (2015) adalah

menghasilkan banyak pohon klasifikasi dari *dataset* yang digunakan kemudian melakukan pendugaan berdasarkan hasil penggabungan dari masing-masing pohon klasifikasi. Metode ini dirancang dengan tujuan meningkatkan akurasi dari *single classifier* dengan cara melatih beberapa pengklasifikasian yang berbeda (Hakim, 2018). Selanjutnya dari hasil prediksi setiap klasifikasi tersebut digabungkan menjadi prediksi akhir melalui proses yang dinamakan *voting*. Adapun metode berbasis *Ensemble* yang digunakan adalah *LightGBM* dan *XGBoost*, dimana kedua metode ini merupakan metode gradien efisien yang beberapa tahun terakhir telah disarankan berdasarkan pohon keputusan (Al Daoud, 2019:6).

XGBoost merupakan salah satu metode yang berbasis *Gradient Tree Boosting* canggih yang dapat bekerja secara efisien dalam menangani masalah skala besar dengan sumber daya komputasi yang sangat terbatas (Chen dan Guestrin, 2016). Sejak metode ini diperkenalkan, *XGBoost* mampu memenangkan berbagai kompetisi *machine learning* dan menjadi solusi yang kuat, efisien dan berguna untuk memecahkan berbagai masalah klasifikasi. *XGBoost* dikembangkan dengan optimasi 10 kali lebih cepat dibanding metode *Gradient Boosting* lainnya (Chen dan Guestrin, 2016). *XGBoost* memiliki keunggulan dalam kinerja dan kompleksitas waktu serta memori yang terjangkau, telah digunakan pada berbagai bidang penelitian sejak diusulkan mulai dari bidang medis, risiko kredit dan metagenomik (Wang dkk, 2020). Metode *XGBoost* dilaporkan memiliki kemampuan dalam menangani ketidakseimbangan data, hal ini ditunjukkan melalui eksperimen yang telah dilakukan oleh Zhao dkk (2018), menunjukkan bahwa *XGBoost* mampu mengungguli metode lain pada dataset tidak seimbang. Selain itu, *XGBoost* mampu mengerjakan berbagai fungsi seperti klasifikasi, regresi dan rangking (Handayani dkk, 2017).

Metode *LightGBM* dikembangkan oleh *Microsoft* yang memberikan kemampuan yang efisien dari algoritma *Gradient Boosting* (Ke dkk, 2017). *LightGBM* merupakan salah satu metode dalam *Gradient Boosting Decision Tree* (GBDT) baru yang mampu menangani *big data* dan sejumlah fitur besar (Tang dkk, 2020). *LightGBM* merupakan metode yang populer dan secara konsisten mampu untuk memecahkan berbagai masalah klasifikasi. *LightGBM* mempercepat waktu proses 20 kali lebih cepat dari fase pelatihan GBDT konvensional dengan akurasi yang sama (Ke dkk, 2017). Sehingga memberikan hasil model yang lebih efektif. Algoritma *LightGBM* telah banyak diaplikasikan

dalam berbagai klasifikasi dan studi regresi dan mencapai hasil deteksi yang sangat baik yang menunjukkan bahwa *LightGBM* adalah algoritma pengklasifikasi yang efektif (Rufo dkk, 2021).

Untuk mencari metode terbaik dalam menangani masalah ketidakseimbangan kelas, maka pada penelitian ini akan dilakukan pengukuran kinerja pendekatan *Ensemble* dengan pengklasifikasi *LightGBM* dan *XGBoost* dalam menangani permasalahan ketidakseimbangan kelas pada data.

2. METODE PENELITIAN

2.1 LightGBM

LightGBM algoritma yang dirancang oleh *Microsoft Research Asia* menggunakan kerangka *Gradient Boosting Decision Tree* (GBDT) (Ke dkk, 2017). Tujuannya untuk meningkatkan efisiensi komputasi, sehingga masalah prediksi dengan *big data* dapat diselesaikan dengan efisien (Liang dkk, 2020). *LightGBM* memiliki beberapa keunggulan dibandingkan metode GBDT lainnya, yaitu kecepatan pelatihan lebih cepat, efisiensi lebih tinggi, penggunaan memori lebih rendah, tingkat akurasi lebih baik, kemampuan dalam menangani data dengan skala yang besar dan dukungan pembelajaran paralel dan GPU (Rufo dkk, 2021). *LightGBM* adalah kerangka *Gradient Boosting* yang cepat, terdistribusi dan berkinerja tinggi berdasarkan algoritma pohon keputusan yang dapat digunakan untuk peringkat, klasifikasi, regresi dan banyak tugas pembelajaran mesin lainnya (Rufo dkk, 2021).

Diasumsikan bahwa himpunan data mentah dengan contoh $N = \{1, 2, \dots, n\}$ dan model *LightGBM* yang memiliki $T = \{1, 2, \dots, t\}$ pohon dihasilkan. Setelah iterasi t kali, prediksi akhir sama dengan jumlah yang pertama $(1 - t)$ th dan t th. Proses iterasi digambarkan sebagai berikut (Zhang dkk, 2020):

$$y_i^{(t)} = y_i^{(t-1)} + f_i(x_i) \tag{1}$$

Dimana $y_i^{(t)}$ adalah nilai prediksi contoh i th pada iterasi t th. $y_i^{(t-1)}$ menunjukkan model pohon yang dihasilkan sebelumnya dan $f_i(x_i)$ menunjukkan model baru yang dibangun.

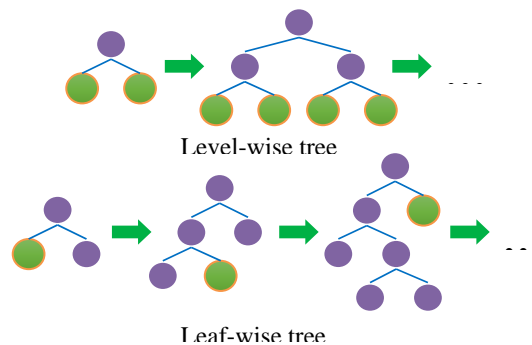
Oleh karena itu, menurut persamaan (1), setiap prediksi baru dihasilkan oleh residu dan prediksi sebelumnya. Proses pelatihan lengkap digambarkan pada persamaan (2), sedangkan istilah regularisasi dapat dihitung menggunakan persamaan (3) yang digunakan untuk mengurangi kompleksitas model dan dapat digunakan untuk meningkatkan kegunaan pada *dataset* lainnya.

$$\begin{cases} y_i^{(0)} = 0 \\ y_i^{(t)} = f_1(x_i) = y_i^{(0)} + f_1(x_i) \\ y_i^{(2)} = f_1(x_i) + f_2(x_i) = y_i^{(1)} + f_2(x_i) \end{cases} \tag{2}$$

$$\begin{aligned} \mathcal{L}^{(t)} &= \sum_{i=1}^n l(y_i, y_i^{(t-1)} + f_t(x_i)) + \sum_{t=1}^T \Omega(f_t) \\ \Omega(f) &= \gamma T + \frac{1}{2} \lambda ||\omega||^2 \end{aligned} \tag{3}$$

Dimana y_i adalah nilai aktual, $y_i^{(t)}$ adalah nilai prediksi. $\sum l$ mewakili jumlah kerugian antara masing-masing kelompok y_i dan $y_i^{(t)}$. dan $\Omega(f_i)$ adalah istilah regularisasi. T adalah jumlah daun, ω adalah bobot daun, λ dan γ adalah koefisien, dengan nilai *default* ditetapkan untuk $\gamma = 0$ dan $\lambda = 1$.

LightGBM memiliki karakteristik yang membedakannya dengan algoritma *tree boosting* lainnya adalah dengan membelah pohon secara memanjang (*leaf-wise tree growth*) dengan yang paling cocok, sedangkan algoritma *tree boosting* lainnya membagi pohon secara mendalam atau sejajar (*level-wise tree growth*) yang ditunjukkan pada gambar 2. Oleh karena itu, ketika tumbuh pada daun yang sama di *LightGBM*, algoritma *leaf-wise* dapat mengurangi lebih banyak kerugian dari pada algoritma *level wise* dan juga menghasilkan akurasi yang jauh lebih baik yang tidak dipenuhi oleh algoritma *boosting* lainnya (Rufo dkk, 2021). Namun, algoritma *leaf-wise* cenderung lebih rentan terhadap *overfitting*.



Gambar 1. Konstruksi *level-wise* dan *leaf-wise*

2.2 XGBoost

XGBoost adalah singkatan dari *Extreme Gradient Boosting*, metode ini pertama kali diperkenalkan oleh Friedman (2001), kemudian

Chen dan Guestrin (2016) mengusulkan metode *XGBoost* sebagai proyek *open-source* yang menerapkan system *machine learning* yang efisien, cepat dan terukur. *XGBoost* merupakan salah satu Teknik *machine learning* untuk mengatasi permasalahan klasifikasi dan regresi berdasarkan Teknik GBDT (Cherif dan Kortebi, 2019). Meskipun *XGBoost* terkenal sebagai model yang kokoh, beberapa studi menyatakan kelemahan metode ini. *XGBoost* merupakan model yang memiliki banyak parameter. Penentuan nilai setiap parameter tergolong sulit sehingga menghabiskan waktu yang tidak sedikit (Siringoringo dkk, 2021:98).

Berbeda dengan GBDT, *XGBoost* memperkenalkan istilah regularisasi dalam fungsi objektif untuk mencegah *overfitting*. Fungsi objektif didefinisikan sebagai berikut:

$$O = \sum_{i=1}^n L(y_i, F(x_i)) + \sum_{k=1}^t R(f_k) + C \quad (4)$$

Dimana $L(y_i, F(x_i))$ adalah fungsi kerugian (*loss function*), $R(f_k)$ menunjukkan istilah regularisasi pada iterasi waktu k dan C adalah istilah konstan yang dapat dihilangkan secara selektif. Istilah regularisasi $R(f_k)$ dinyatakan sebagai:

$$R(f_k) = \alpha H + \frac{1}{2} \eta \sum_{j=1}^H w_j^2 \quad (5)$$

Dimana α mewakili kompleksitas daun, H menunjukkan jumlah daun, η menunjukkan parameter penalti dan w_j^2 adalah hasil *output* dari setiap simpul daun. Secara khusus daun menunjukkan kategori yang diprediksi mengikuti aturan klasifikasi dan simpul daun menunjukkan simpul pohon yang tidak dapat dibagi. Selain itu, alih-alih menggunakan turunan orde pertama di GBDT, serangkaian objektif *taylor* orde kedua diadopsi dalam *XGBoost*. Misalkan kesalahan kuadrat rata-rata digunakan sebagai fungsi kerugian, maka fungsi objektif dapat diturunkan sebagai berikut:

$$O = \sum_{i=1}^n \left[p_i w_{q(x_i)} + \frac{1}{2} (q_i w_{q(x_i)}^2) \right] + \alpha H + \frac{1}{2} \eta \sum_{j=1}^H w_j^2 \quad (6)$$

$q(x_i)$ menunjukkan fungsi yang menetapkan titik data ke daun yang sesuai, p_i dan q_i menunjukkan

turunan pertama dan kedua dari fungsi kerugian masing-masing.

Nilai kerugian akhir dihitung berdasarkan jumlah semua nilai kerugian. Karena sampel sesuai dengan node dalam pohon keputusan, nilai kerugian akhir dapat ditentukan dengan menjumlahkan nilai kerugian node daun. Oleh karena itu, fungsi objektif juga dinyatakan sebagai berikut:

$$O = \sum_{j=1}^T \left[P_j w_j + \frac{1}{2} (Q_j + \eta) w_j^2 \right] + \alpha H \quad (7)$$

Dimana $P_j = \sum_{i \in I_j} p_i$, $Q_j = \sum_{i \in I_j} q_i$, dan I_j menunjukkan semua sampel dalam simpul daun j .

Untuk menyimpulkan, optimalisasi fungsi objektif diubah sebagai kasus memilih minimum fungsi kuadrat. Selain itu, karena pengenalan istilah regularisasi, *XGBoost* memiliki kemampuan yang lebih baik untuk melawan masalah *overfitting*.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Pada penelitian ini menggunakan 5 data yang memiliki karakteristik berbeda antara satu data dengan data yang lainnya. Data diperoleh dari situs *UCI Repository*, kriteria data yang digunakan adalah ketidakseimbangan data, tipe data multivariat dan termasuk data klasifikasi.

Tabel 1. Informasi data

No	Nama Data	Jumlah Data	Jumlah Variabel	Perbandingan kelas
1	<i>Abalone</i>	4.177	9	94:6
2	<i>Car</i>	1.728	7	95:5
3	<i>HTRU-2</i>	17.627	9	91:9
4	<i>Winequality-white</i>	4.659	12	83:17
5	<i>Winequality-red</i>	1521	12	86:14

Berdasarkan tabel 6, dapat dilihat bahwa perbandingan kelas data *Abalone*, *Car* dan *HTRU-2* diatas 90:10 yang menandakan tingkat ketidakseimbangan tinggi, sedangkan tingkat ketidakseimbangan rendah terdapat pada data *winequality-red* dan *winequality-white*. Setiap data yang digunakan pada penelitian ini memiliki karakteristik yang berbeda, yaitu jumlah data, variabel dan perbandingan kelas. Tujuan dari penggunaan kelima data tersebut untuk melihat

kinerja dari *LightGBM* dan *XGBoost* dalam menangani berbagai karakteristik yang berbeda.

3.2 Evaluasi hasil kinerja metode *LightGBM* dan *XGBoost* dalam menangani data dengan kelas tidak seimbang

a Klasifikasi *LightGBM*

Proses klasifikasi pada penelitian ini menggunakan *software R Studio* versi 4.0.2 dengan bantuan *package lightgbm, data.Tabel, caret, dplyr* dan *ROSE*. Algoritma *LightGBM* membutuhkan data dengan tipe matriks, maka peneliti menggunakan fungsi *sparse.model.matrix()*. Keuntungan dari matriks ini adalah dapat mengirimkan variabel dan label serta dapat mengidentifikasi kolom nama yang merupakan label. Serta membagi data menjadi dua bagian yaitu data pelatihan (*training*) sebanyak 75% dan data uji (*testing*) sebanyak 25%.

1) Model *LightGBM*

Pada model *LightGBM* digunakan fungsi *training* yang telah diubah ke bentuk matriks, maka dimasukkan data *train_matrix*. Kemudian menyesuaikan serangkaian model *LightGBM* untuk setiap parameter yang digunakan seperti *max_bin = 5, num_leaves = 35, learning_rate = 0,001, objective = "binary", metric = 'binary_logloss'* dan *nrounds = 1000*.

2) Prediksi *LightGBM*

Hasil pengujian model dianalisis dengan melakukan pencocokan kelas prediksi dengan kelas sebenarnya yang dapat memberikan informasi berupa akurasi, sensitivitas dan spesifisitas dengan menggunakan *confusion matrix*.

Tabel 2. Confusion matrix prediksi *LightGBM*

Nama Data	Akurasi	Sensitivitas	Spesifisitas
<i>Abalone</i>	81.51%	82.14%	71.90%
<i>HTRU-2</i>	94.62%	95.10%	89.76%
<i>Car</i>	95.63%	96.15%	86.36%
<i>Winequality-red</i>	73.46%	71.71%	84.54%
<i>Winequality-white</i>	75.51%	76.62%	69.50%

Berdasarkan tabel 2, dapat dilihat bahwa hasil prediksi diperoleh 2 data dengan nilai akurasi, sensitivitas dan spesifisitas kurang baik yaitu data *Winequality-red* dan *Winequality-white* karena nilai akurasi kurang dari 80% yaitu 73.46% dan 75.51%, nilai sensitivitas kurang dari 80% yaitu 71.71% dan

76.62% serta nilai spesifisitas dari data *Winequality-white* kurang dari 70% yaitu 69.50%. Metode *lightGBM* bekerja dengan optimal pada data *HTRU-2* dan *Car* karena dari kedua data tersebut memiliki nilai akurasi, sensitivitas dan spesifisitas tertinggi dengan nilai diatas 80%.

Untuk melihat seberapa baik atau optimal metode *LightGBM* digunakan untuk melakukan klasifikasi data dengan kelas tidak seimbang, peneliti menggunakan pengukuran nilai AUC (*Area Under the Curve*). Berikut ditampilkan nilai AUC dari setiap *dataset* yang digunakan.

Tabel 3. Nilai AUC metode *LightGBM*

Nama Data	AUC
<i>Abalone</i>	77%
<i>HTRU-2</i>	92,4%
<i>Car</i>	91,3%
<i>Winequality-Red</i>	78,1%
<i>Winequality-White</i>	73,1%

Tabel 3 Menunjukkan nilai AUC yang diperoleh menggunakan metode *LightGBM*. Semakin tinggi nilai AUC yang diperoleh maka kinerja suatu metode dalam melakukan klasifikasi semakin baik. Berdasarkan hasil di atas, kinerja *LightGBM* sangat baik dalam melakukan klasifikasi pada data *HTRU-2* dan *Car* dengan nilai AUC masing-masing sebesar 92.4% dan 91.3%.

b Klasifikasi *XGBoost*

Proses klasifikasi *XGBoost* membutuhkan bantuan *package xgboost, caret, tidyverse* dan *ROSE*. Algoritma *LightGBM* membutuhkan data dengan tipe matriks, maka peneliti menggunakan fungsi *model.matrix()* untuk membuat *dataset* kelas *trained* dan *tested*. Keuntungan dari matriks ini adalah dapat mengirimkan variabel dan label serta dapat mengidentifikasi kolom nama yang merupakan label. Serta membagi data menjadi dua bagian yaitu data pelatihan (*training*) sebanyak 75% dan data uji (*testing*) sebanyak 25%.

1) Model *XGBoost*

Pembuatan model *XGBoost* digunakan data *training* yang telah diubah ke dalam bentuk matriks, maka perlu dimasukkan data *trained*. Parameter yang digunakan pada penelitian ini adalah *nrounds* dan *nthreads*. Parameter *nrounds* memberitahu berapa kali *XGBoost* diulangi, pada penelitian ini menggunakan *nrounds = 1000* dan menggunakan *nthreads = 5*.

2) Prediksi *XGBoost*

Hasil dari pengujian model dianalisis dengan melakukan suatu pencocokan kelas prediksi dengan kelas sebenarnya yang dapat memberikan suatu informasi penting berupa nilai akurasi, sensitivitas dan spesifisitas dengan menggunakan *confusion matrix RStudio*.

Tabel 4. *Confusion matrix* model *XGBoost*

Nama Data	Akurasi	Sensitivitas	Spesifisitas
<i>Abalone</i>	92.54%	94.84%	52.63%
<i>HTRU-2</i>	98%	98.86%	88.49%
<i>Car</i>	97.22%	97.80%	86.36%
<i>Winequality-red</i>	90.03%	91.04%	82.61%
<i>Winequality-white</i>	85.06%	89.78%	63.11%

Nilai akurasi, sensitivitas dan spesifisitas yang diperoleh secara keseluruhan dapat dikatakan baik. Akan tetapi, terdapat 2 data yang memiliki nilai spesifisitas kurang dari 70% yaitu data *Abalone* dan *Winequality-white* dengan nilai spesifisitas masing-masing sebesar 52.63% dan 63.11%. Sedangkan 3 data lainnya yaitu *HTRU-2*, *Car* dan *Winequality-red* memiliki nilai *confusion matrix* baik.

Untuk melihat seberapa optimal metode *XGBoost* digunakan untuk melakukan klasifikasi data dengan kelas tidak seimbang, peneliti menggunakan pengukuran nilai AUC.

Tabel 5. Nilai AUC Metode *XGBoost*

Nama Data	AUC
<i>Abalone</i>	72,2%
<i>HTRU-2</i>	93,7%
<i>Car</i>	92.1%
<i>Winequality-red</i>	86.8%
<i>Winequality-hite</i>	76.4%

Berdasarkan tabel 5 metode *XGBoost* paling optimal/sangat baik digunakan dalam melakukan klasifikasi pada data *HTRU-2*, *Car* dan *Winequality-red* dengan masing-masing nilai AUC sebesar 93.7%, 92.1% dan 76.4%. Sedangkan pada data *Abalone* dan *Winequality-white*, metode *XGBoost* cukup baik digunakan dalam melakukan klasifikasi.

c. Perbandingan Kinerja Klasifikasi Metode *LightGBM* dan *XGBoost*

Setelah diperoleh hasil klasifikasi menggunakan metode *LightGBM* dan *XGBoost*, maka perbandingan nilai akurasi, sensitivitas, spesifisitas dan AUC ditunjukkan pada masing-masing tabel di bawah.

Tabel 6. Perbandingan akurasi metode *LightGBM* dan *XGBoost*

Nama Data	<i>LightGBM</i>	<i>XGBoost</i>
<i>Abalone</i>	81.51%	92.54%
<i>HTRU-2</i>	94.62%	98%
<i>Car</i>	95.63%	97.22%
<i>Winequality-red</i>	73.46%	90.03%
<i>Winequality-white</i>	75.51%	85.06%
Nilai Keseluruhan	84.15	92.57%

Berdasarkan Tabel 6, dapat dilihat bahwa metode *XGBoost* memiliki nilai akurasi lebih tinggi dibandingkan metode *LightGBM*. Nilai akurasi tertinggi untuk *lightGBM* dan pada data *Car* sebesar 95.63% dan untuk *XGBoost* pada data *HTRU-2* sebesar 98%. Selain itu, nilai akurasi terendah menggunakan *LightGBM* terdapat pada data *Winequality-red* dengan akurasi sebesar 73.46%, sedangkan menggunakan *XGBoost* terdapat pada data *Winequality-white* dengan akurasi sebesar 85.06%.

Tabel 7. Perbandingan sensitivitas metode *LightGBM* dan *XGBoost*

Nama Data	<i>LightGBM</i>	<i>XGBoost</i>
<i>Abalone</i>	82.14%	94.84%
<i>HTRU-2</i>	95.10%	98.86%
<i>Car</i>	96.15%	97.80%
<i>Winequality-red</i>	71.71%	91.04%
<i>Winequality-white</i>	76.62%	89.78%
Nilai Keseluruhan	84.34%	94.46%

Kinerja suatu metode dikatakan baik tidak hanya dilihat pada nilai akurasi melainkan mempertimbangkan nilai sensitivitas (ketepatan prediksi kelas positif/mayoritas). Tabel 7 menunjukkan nilai sensitivitas untuk setiap data menggunakan metode *LightGBM* dan *XGBoost*, dimana nilai sensitivitas *XGBoost* sedikit lebih tinggi dibandingkan *LightGBM* pada setiap data yang digunakan. Diperoleh nilai sensitivitas tertinggi terdapat pada data *HTRU-2* menggunakan *XGBoost* dengan nilai sebesar 98.86%. Kemudian untuk nilai sensitivitas terendah terdapat pada data *Winequality-red* menggunakan *LightGBM* dengan nilai sensitivitas sebesar 71.71%.

Tabel 8. Perbandingan spesifisitas metode *LightGBM* dan *XGBoost*

Nama Data	<i>LightGBM</i>	<i>XGBoost</i>
<i>Abalone</i>	71.90%	52.63%
<i>HTRU-2</i>	89.76%	88.49%
<i>Car</i>	86.36%	86.36%
<i>Winequality-red</i>	84.54%	82.61%

<i>Winequality-white</i>	69.50%	63.11%
Nilai Keseluruhan	80.41%	74.64%

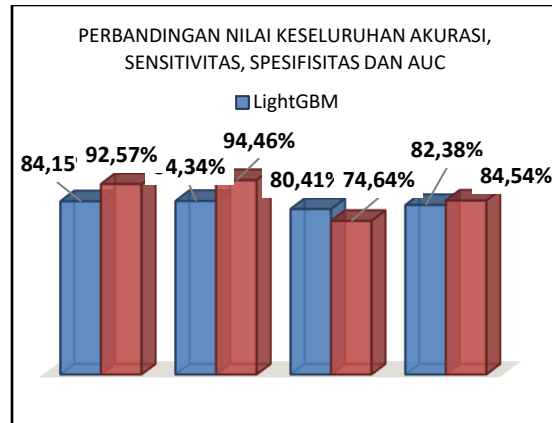
Tabel 8 menunjukkan nilai spesifisitas yang menggambarkan kemampuan metode dalam memprediksi kelas negatif (minoritas). Berdasarkan tabel di atas, terlihat bahwa nilai spesifisitas yang dihasilkan kedua metode dapat dikatakan baik untuk keseluruhan data. Akan tetapi, pada data *Abalone* dan *Winequality-white* nilai spesifisitas rendah. baik menggunakan metode *LightGBM* ataupun *XGBoost*. Nilai spesifisitas terendah terdapat pada data *Abalone* dengan menggunakan metode *XGBoost* sebesar 52.63%. Sedangkan nilai spesifisitas tertinggi terdapat pada data *HTRU-2* menggunakan metode *LightGBM* sebesar 89.76%.

Pada tabel 9 menunjukkan perbandingan nilai AUC dari metode *LightGBM* dan *XGBoost* yang digunakan untuk mengukur seberapa baik/optimal kedua metode tersebut digunakan pada setiap *dataset* yang disajikan. Nilai AUC yang tertinggi diperoleh menggunakan metode *XGBoost* pada data *HTRU-2* sebesar 93.7%, artinya *XGBoost* bekerja dengan optimal dalam melakukan klasifikasi pada data *HTRU-2*. Jika dilihat secara keseluruhan, nilai AUC *XGBoost* sedikit lebih tinggi dibandingkan *LightGBM* pada setiap *dataset* yang digunakan. Maka dengan itu, dapat dikatakan bahwa kinerja dari *XGBoost* lebih optimal dibandingkan *LightGBM*. Akan tetapi, metode *LightGBM* lebih baik/optimal digunakan untuk melakukan klasifikasi pada data *Abalone* dengan nilai AUC sebesar 77%.

Tabel 9. Perbandingan Nilai AUC metode *LightGBM* dan *XGBoost*

Nama Data	<i>LightGBM</i>	<i>XGBoost</i>
<i>Abalone</i>	77%	73.7%
<i>HTRU-2</i>	92.4%	93.7%
<i>Car</i>	91.3%	92.1%
<i>Wine Quality-Red</i>	78.1%	86.8%
<i>Wine Quality-White</i>	73.1%	76.4
Nilai Keseluruhan	82.38%	84.54%

Berikut gambar 2 mengenai tingkat kestabilan *LightGBM* dan *XGBoost* dilihat dari nilai akurasi, sensitivitas, spesifisitas dan AUC.



Gambar 2. Perbandingan Rata-rata keseluruhan nilai dari metode *LightGBM* dan *XGBoost*

Pada gambar 2 menunjukkan hasil perbandingan metode *LightGBM* dan *XGBoost* dengan 5 data dengan kelas tidak seimbang. Secara keseluruhan metode *XGBoost* lebih stabil dari pada metode *LightGBM*. Akan tetapi, nilai keseluruhan rata-rata spesifisitas *LightGBM* lebih tinggi dibandingkan *XGBoost* yaitu 80,41% dan 74,64%. Artinya metode *LightGBM* lebih baik digunakan dalam melakukan prediksi terhadap kelas negatif/minoritas dan tanpa harus mengabaikan kelas mayoritas dan hasil akurasi.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan penjelasan diatas, dapat ditarik kesimpulan dari hasil dan pembahasan yang telah penulis uraikan di atas, diantaranya sebagai berikut:

1. Pengklasifikasian data dengan kelas tidak seimbang menggunakan *LightGBM* dan *XGBoost* memerlukan beberapa parameter pengukuran yang telah ditentukan setelah melakukan beberapa kali percobaan, sehingga diperoleh hasil prediksi klasifikasi yang bagus. Parameter yang digunakan dalam melakukan analisis klasifikasi menggunakan *LightGBM*, yaitu *max_bin* = 5, *num_leaves* = 35, *learning_rate* = 0,001 dan *nrounds* =1000. Sedangkan untuk metode *XGBoost* menggunakan parameter *nrounds* = 1000 dan *nthreads* = 5. Selain itu, membagi data sampel menjadi dua, yaitu *training* 75% dan *testing* 25%.
2. Evaluasi hasil kinerja *LightGBM* dan *XGBoost* dengan menggunakan parameter yang telah ditentukan tersebut menghasilkan nilai akurasi dan sensitivitas terbaik pada data *HTRU-2* menggunakan metode *XGBoost* dengan nilai masing-masing sebesar 98% dan 98,86%.

Sedangkan, untuk nilai spesifisitas terbaik pada data *HTRU-2* menggunakan *LightGBM* dengan nilai sebesar 89,76%. Jika dilihat dari nilai keseluruhan rata-rata, kemampuan metode *LightGBM* lebih baik dalam melakukan prediksi terhadap kelas minoritas dibandingkan metode *XGBoost* dengan nilai 80,41% : 74,64%.

4.2 Saran

Penelitian ini masih dapat dikembangkan dengan membandingkan beberapa metode *Ensemble* lainnya, juga dapat menggunakan data yang lebih spesifik pada saat membandingkan metode serta menggunakan pengukuran kebaikan model yang lainnya. Contohnya perbandingan metode pada jumlah data, jumlah variabel, adanya *outlier* dan lain-lain.

DAFTAR PUSTAKA

- Abd Elrahman, S. M., & Abraham, A. (2013). *A review of class imbalance problem*. Journal of Network and Innovative Computing, 1(2013), 332-340.
- Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system*. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Cherif, I. L., & Kortebi, A. (2019). *On using extreme gradient boosting (XGBoost) machine learning algorithm for home network traffic classification*. In 2019 Wireless Days (WD) (pp. 1-6). IEEE.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). *A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(4), 463-484.
- Hakim, L. (2018). *Metode Ensemble Berbasis Bagging untuk Penanganan Pemodelan Klasifikasi pada Data dengan Kelas Tidak Seimbang*. (Doctoral dissertation). Bogor: Sekolah Pascasarjana. Institut Pertanian Bogor.
- Handayani, A., Jamal, A., & Septiandri, A. A. (2017). *Evaluasi Tiga Jenis Algoritme Berbasis Pembelajaran Mesin untuk Klasifikasi Jenis Tumor Payudara*. Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), 6(4), 394-403.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). *Lightgbm: A highly efficient gradient boosting decision tree*. Advances in neural information processing systems, 30, 3146-3154.
- Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). *Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research*. European Journal of Operational Research, 247(1), 124-136.
- Liang, W., Luo, S., Zhao, G., & Wu, H. (2020). *Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms*. Mathematics, 8(5), 765.
- Pristyanto, Y., Setiawan, NA, & Ardiyanto, I. (2017). *Resampling hybrid untuk menangani kelas yang tidak seimbang pada klasifikasi kinerja siswa di kelas*. Pada tahun 2017 1st International Conference on Informatics and Computational Sciences (ICICoS) (hlm. 207-212). IEEE.
- Rufo, D. D., Debelee, T. G., Ibenthal, A., & Negera, W. G. (2021). *Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM)*. Diagnostics, 11(9), 1714.
- Saifudin, A., & Wahono, R. S. (2015). *Penerapan teknik ensemble untuk menangani ketidakseimbangan kelas pada prediksi cacat software*. IlmuKomputer. com Journal of Software Engineering, 1(1), 28-37.
- Sartono, B. (2015). *Tinjauan terhadap Keunggulan Pohon Klasifikasi Ensemble untuk Memperbaiki Kemampuan Prediksi Pohon Klasifikasi Tunggal*. Journal Biostatistics| Departemen Statistika FMIPA Universitas Padjadjaran, 9(2), 33-38.
- Siringoringo, R. (2018). *Klasifikasi data tidak seimbang menggunakan algoritma SMOTE dan k-nearest neighbor*. Journal Information System Development (ISD), 3(1).
- Siringoringo, R., Perangin-angin, R., & Jamaluddin, J. (2021). *Model Hibrid Genetic-Xgboost dan Principal Component Analysis Pada Segmentasi Dan Peramalan Pasar*. METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi, 5(2), 97-103.

- Sulistiyowati, N., & Jajuli, M. (2020). *Integrasi Naive Bayes Dengan Teknik Sampling SMOTE Untuk Menangani Data Tidak Seimbang*. NUANSA INFORMATIKA, 14 (1), 34-37.
- Tang, M., Zhao, Q., Ding, S. X., Wu, H., Li, L., Long, W., & Huang, B. (2020). *An improved lightGBM algorithm for online fault detection of wind turbine gearboxes*. Energies, 13(4), 807.
- Wang, C., Deng, C., & Wang, S. (2020). *Imbalance-XGBoost: Leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost*. Pattern Recognition Letters, 136, 190-197.
- Wu, Q., Ye, Y., Zhang, H., Ng, M. K., & Ho, S. S. (2014). *ForesTexter: An efficient random forest algorithm for imbalanced text categorization*. Knowledge-Based Systems, 67, 105-116.
- Yap, B. W., Rani, K. A., Rahman, H. A. A., Fong, S., Khairudin, Z., & Abdullah, N. N. (2014). *An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets*. In Proceedings of the first international conference on advanced data and information engineering (DaEng-2013) (pp. 13-22). Springer, Singapore.
- Zhang, C., & Ma, Y. (Eds.). (2012). *Ensemble machine learning: methods and applications*. Springer Science & Business Media.
- Zhang, Y., Zhu, C., & Wang, Q. (2020). *LightGBM-based model for metro passenger volume forecasting*. IET Intelligent Transport Systems, 14(13), 1815-1823.
- Zhao, Z., Peng, H., Lan, C., Zheng, Y., Fang, L., & Li, J. (2018). *Imbalance learning for the prediction of N6-Methylation sites in mRNAs*. BMC genomics, 19(1), 1-10.

Penerapan Metode *Clustering* SOM dan DBSCAN dalam Mengelompokkan *Unmet Need* Keluarga Berencana di Nusa Tenggara Barat

Sausan Nisrina⁽¹⁾, Wiwit Pura Nurmayanti⁽²⁾, Basirun⁽³⁾, Kertanah⁽⁴⁾ Muhammad Ghazali⁽⁵⁾
Universitas Hamzanwadi

Jl. TGKH. Muhammad Zainuddin Abdul Madjid Pancor, Selong, Lombok Timur, NTB 83612
e-mail: sausannisrina96@gmail.com , wiwit.adiwinata3@gmail.com ,
basirun@hamzanwadi.ac.id, kertanah@hamzanwadi.ac.id dan
muhammad.gazali@hamzanwadi.ac.id

ABSTRAK

Perkembangan *big data* terus mengalami kemajuan dengan sangat pesat. Untuk memudahkan mengakses data dan informasi dapat menggunakan metode *clustering*. *Clustering* merupakan proses pengelompokan obyek atas kesamaan karakteristiknya. Metode *clustering* dapat diterapkan pada bidang kesehatan yaitu *unmet need* Keluarga Berencana (KB). *Unmet need* dapat diartikan sebagai tidak terpenuhinya kebutuhan akan alat kontrasepsi pada pasangan usia subur. Metode *clustering* yang digunakan pada penelitian ini yaitu SOM dan DBSCAN. Penelitian bertujuan untuk melihat gambaran *unmet need* di Provinsi Nusa Tenggara Barat yang kedua adalah menguraikan hasil *clustering* dengan metode *Clustering Self Organizing Maps* (SOM) dan terakhir untuk menguraikan hasil *clustering* dengan metode DBSCAN dalam kasus pengelompokan penyebab *unmet need* KB di Provinsi NTB. Berdasarkan hasil analisis di dapatkan bahwa Kabupaten Lombok Timur merupakan kabupaten dengan jumlah *unmet need* KB tertinggi sebanyak 84.72 jiwa, sedangkan terendah di Kota Bima sebesar 4.636 orang. Hasil penelitian dengan metode SOM memiliki 3 *cluster*, 86 kecamatan *cluster* 1, 30 kecamatan *cluster* 2 dan 4 kecamatan *cluster* 3. Untuk metode DBSCAN dengan *eps* 1.20 dan *MinPts* 3, jumlah *cluster* yang terbentuk yaitu 2 *cluster*, *cluster* 1 dengan 97 kecamatan dan *cluster* 2 dengan 4 kecamatan, serta 16 kecamatan yang memiliki data *noise*.

Kata kunci : *Clustering, SOM, DBSCAN, Unmet need, Keluarga Berencana*

ABSTRACT

The development of big data continues to progress very rapidly. To make it easier to access data and information can use the clustering method. Clustering is the process of grouping objects for their similarity of characteristics. The clustering method can be applied to the health sector, namely unmet need Family Planning (KB). Unmet need can be interpreted as the unmet need for contraceptives in couples of childbearing age. The clustering method used in this study is SOM dan DBSCAN. The second study aims to see a picture of unmet need in West Nusa Tenggara Province, the second is to decipher the results of clustering with the Clustering Self Organizing Maps (SOM) method and finally to decipher the clustering results with the DBSCAN method in the case of grouping the causes of unmet need KB in NTB Province. Based on the results of the analysis, it was found that East Lombok Regency is the district with the highest number of unmet needs KB as many as 84.72 people, while the lowest in Bima City is 4,636 people. The results of the study using the SOM method have 3 clusters, 86 cluster 1 sub-districts, 30 cluster 2 sub-districts and 4 cluster 3 sub-districts. For the DBSCAN method with eps 1.20 and MinPts 3, the number of clusters formed is 2 clusters, cluster 1 with 97 sub-districts and cluster 2 with 4 sub-districts, and 16 sub-districts that have noise data.

Keywords : *Clustering, SOM, DBSCAN, Unmet need. Family Planning*

1. PENDAHULUAN

Perkembangan *big data* saat ini terus mengalami kemajuan dengan sangat pesat. Penyebab berkembangnya *big data* dikarenakan

banyaknya kebutuhan manusia akan efisiensi dan kemudahan dalam mengakses data. Untuk memudahkan dalam menyaring data dan informasi dalam bentuk mengelompokkan data, dapat

digunakan salah satu dari metode data mining yaitu *clustering*. *Clustering* merupakan salah satu teknik yang termasuk dalam data mining yang paling umum merujuk pada metodologi partisi elemen dalam kelompok menurut kemiripan karakteristik (Royal, 2022). *Clustering* bertujuan untuk mengelompokkan obyek atas kesamaan yang dimiliki dengan kriteria *cluster* yang terbentuk memiliki homogenitas dan heterogenitas yang tinggi (Halim *et al.*, 2017). Terdapat berbagai jenis metode dalam melakukan analisis *cluster*, mulai dari metode yang sederhana hingga metode yang rumit yaitu dengan menggunakan kecerdasan buatan, seperti jaringan syaraf tiruan. Salah satu metode *clustering* dengan jaringan syaraf tiruan yaitu Kohonen *Self Organizing Maps* (SOM) (Harli *et al.*, 2016).

Self Organizing Map (SOM) diperkenalkan oleh Teuvo Kohonen pada tahun 1982 sebagai salah satu metode yang bersifat *unsupervised learning* atau proses pelatihan yang tidak terawasi (Suwirmayanti, 2020). SOM mampu dalam melakukan pemetaan terhitung cukup valid dan rendah bias, hal ini menyebabkan mampu memetakan kumpulan data berdasarkan ciri-ciri melalui pengaturan diri (*self-organizing*) (Suwirmayanti, 2020). Selain itu, pelopor perkembangan teknik *clustering* berdasarkan kepadatan atau yang biasa dikenal dengan *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN) (Wuryandari *et al.*, 2017). DBSCAN adalah salah satu metode berbasis yang membentuk *cluster* berdasarkan kepadatan yang cukup tinggi ke dalam *cluster*, sementara untuk objek yang tidak masuk ke dalam kelompok manapun dianggap sebagai data pencilon atau *noise* (Risman *et al.*, 2019).

Metode *clustering* dapat diterapkan di hampir semua sektor, salah satunya adalah bidang kesehatan yaitu kebutuhan KB tidak terpenuhi (*unmet need*). *Unmet need* merupakan persentase wanita usia 15-49 tahun yang menikah tetapi tidak menggunakan kontrasepsi, namun tidak ingin hamil dalam waktu dua tahun ke depan (menjarangkan), tidak ingin memiliki anak lagi (membatasi), mengalami kehamilan tidak diinginkan, atau yang sedang dalam masa nifas (Listyaningsih *et al.*, 2016). *Unmet need* dapat diartikan sebagai suatu kebutuhan akan alat kontrasepsi yang tidak terpenuhi Menurut data BKKBN NTB (2020) *Unmet need* di Nusa Tenggara Barat mencapai 12,1% pada tahun 2018 dan meningkat sebesar 2,88 % menjadi 14,98 % di tahun 2019. Ada beberapa faktor penyebab naiknya angka *Unmet need* seperti PUS yang tidak mengetahui tentang KB, biaya mahal, tidak tersedianya fasilitas kesehatan dan sebagainya. BKKBN Provinsi Nusa

Tenggara Barat dalam Renstra tahun 2020 dapat menurunkan persentase *unmet need* dengan target sebesar 14,72 persen pada tahun 2020, diharapkan akan turun menjadi 13,84 persen pada tahun 2024.

Penelitian ini bertujuan untuk untuk melihat gambaran *unmet need* di Provinsi Nusa Tenggara Barat, dan yang kedua adalah menguraikan hasil *clustering* dengan metode *Clustering Self Organizing Maps* (SOM) dan terakhir untuk menguraikan hasil *clustering* dengan metode DBSCAN dalam kasus pengelompokan penyebab *unmet need* KB di Provinsi NTB. Penelitian terkait dengan metode SOM dan DBSCAN telah dilakukan diantaranya oleh (Nastiti *et al.*, 2018) yaitu membandingkan metode SOM dengan *Fuzzy C-Means*, sehingga didapatkan hasil bahwa secara keseluruhan algoritma SOM lebih unggul dalam mengelompokkan kualifikasi data kinerja dosen dibandingkan dengan algoritma *Fuzzy C-Means*. Penelitian lainnya oleh (Qadrini, 2020) membandingkan metode *K-Means* dan DBSCAN untuk mengelompokkan data dasar kompetensi laboratorium ITS Tahun 2017, dengan metode pengelompokan terbaik adalah Metode DBSCAN. Berdasarkan beberapa penelitian sebelumnya, metode SOM dan DBSCAN unggul dibandingkan metode pembandingnya, sehingga peneliti tertarik ingin menganalisis kedua metode tersebut untuk mengelompokkan data *Unmet need* Keluarga Berencana di Provinsi Nusa Tenggara Barat tahun 2021.

2. METODE PENELITIAN

Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif. data yang digunakan yaitu data sekunder yang bersumber dari website resmi Pendataan Keluarga 2021 (Laporan Tabulasi | PK2021 - BKKBN). Variabel yang digunakan pada penelitian yaitu sebagai berikut:

Tabel 1. Keterangan Variabel

Simbol	Keterangan
X1	Tidak mengetahui tentang KB
X2	Tempat pelayanan jauh
X3	Alat/obat KB tidak tersedia
X4	Biaya mahal
X5	Tidak ada petugas pelayanan KB

Teknik analisis yang digunakan dalam penelitian ini adalah analisis deskriptif dan analisis *cluster* dengan menggunakan metode *Self Organizing Map* (SOM) dan *Density-based Spatial Clustering of*

Application with Noise (DBSCAN). Langkah-langkah analisis pada penelitian ini sebagai berikut:

1. Pengumpulan dan *input* data
2. Analisis deskriptif. Analisis deskriptif digunakan untuk memberikan gambaran dari data melalui grafik, tabel, plot dan sebagainya.
3. Standarisasi data. Standarisasi data digunakan jika satuan yang bervariasi atau berbeda. Standarisasi data yang digunakan yaitu *Z-Score*.
4. Menghitung jarak *euclidean* untuk mengukur obyek satu dengan yang lainnya
5. Analisis *clustering* SOM.
 - a. Inisialisasi nilai bobot W_{ij} secara acak.
 - b. Masukkan variabel x untuk setiap vektor *input*, dengan tahapan sebagai berikut:
 - Untuk setiap jarak (j) hitunglah jarak *euclidean* dengan rumus berikut.

$$D(j) = \sum_{i=1}^n (w_{ij} - x_i)^2 \quad (1)$$

Keterangan:
 $D(j)$ = jarak untuk setiap j
 w_{ij} = bobot dari hubungan antara node *input* ke- i terhadap node *output* j
 x_i = node input ke- i
 n = jumlah node pada lapisan input
 - Temukan nilai j sedemikian sehingga $D(j)$ minimum
 - Untuk semua unit j di dalam ketetanggaan j dan untuk semua i dengan rumus:

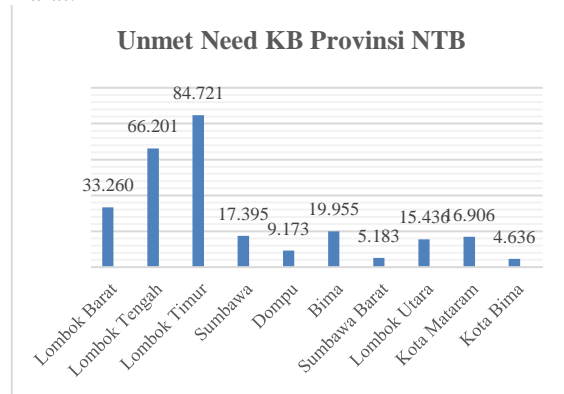
$$W_{ij}(\text{baru}) = w_{ij}(\text{lama}) + \alpha(x_i - w_{ij}(\text{lama})) \quad (2)$$
 - c. *Output* disajikan dalam diagram kipas.
6. Analisis *clustering* DBSCAN.
 - a. Inisialisasi parameter *minimum points* ($minpts$), *epsilon* (eps)
 - b. Tentukan titik awal atau p secara *random*
 - c. Hitung eps atau semua jarak titik yang *density reachable* terhadap p
 - d. Jika titik yang memenuhi eps lebih dari $minpts$ maka titik p adalah *core point* dan *cluster* terbentuk. Jika iya maka *cluster* terbentuk, jika tidak maka dianggap sebagai *noise*.
 - e. Ulangi langkah b-d jika semua titik sudah terproses.

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Data

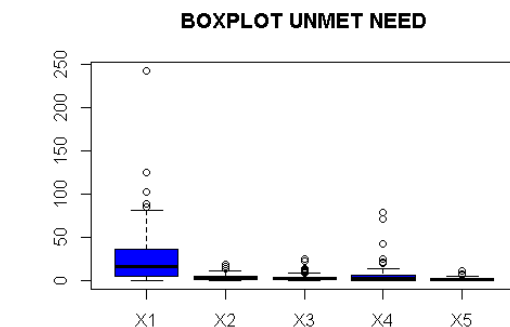
Analisis deskriptif adalah analisis dalam statistik yang digunakan untuk menganalisis data dengan cara memberikan gambaran data yang sebenarnya tanpa bermaksud membuat kesimpulan, biasanya menggunakan visualisasi data berupa tabel grafik, *barplot*, *boxplot* dan lain lain untuk melihat

data *unmet need* KB di Provinsi Nusa Tenggara Barat.



Gambar 1. Jumlah unmet need KB

Berdasarkan Gambar 1, Kabupaten Lombok Timur merupakan kabupaten dengan jumlah *unmet need* KB tertinggi sebanyak 84.721 orang, kemudian kabupaten Lombok Tengah berada di urutan kedua tertinggi dan urutan tertinggi ketiga yaitu kabupaten Lombok Barat, dengan masing-masing angka *unmet need* sebesar 66.201 dan 33.260 orang. Sedangkan jumlah *unmet need* terendah berada di Kota Bima sebesar 4.636 orang. Sebaran data penyebab *unmet need* disajikan dalam Gambar 5 *boxplot* di bawah ini dengan keterangan X1 (tidak tahu tentang KB), X2 (tempat pelayanan jauh), X3 (alat/obat KB tidak tersedia), X4 (biaya mahal), X5 (tidak ada petugas pelayanan KB).



Gambar 2. Boxplot Penyebab Unmet needKB

Output Gambar 2 menjelaskan bahwa dari kelima variabel terdapat outlier, yang berarti setiap data memiliki pencilan atau data yang berada jauh dari rata-rata. Pada variabel X1 memiliki *boxplot* yang berbeda dari *boxplot* variabel lainnya. Hal ini dikarenakan pada X1 jumlah yang tidak mengetahui tentang KB mencapai hingga 2949 jiwa, sedangkan untuk penyebab lainnya seperti tempat pelayanan jauh (X2) dengan jumlah 388 jiwa, alat/obat KB tidak tersedia (X3) 407 jiwa, biaya mahal (X4) 607

jiwa dan tidak ada petugas pelayanan KB (X5) memiliki jumlah 145 jiwa.

3.2. Standarisasi Data

Standarisasi digunakan jika data dalam sebuah penelitian mempunyai satuan yang berbeda. Seperti, satu variabel dengan nilai data puluhan dan variabel yang lain memiliki nilai nol, sehingga memiliki jarak nilai yang terlalu jauh dan akan berpengaruh pada hasil *cluster*. Oleh karena itu, selanjutnya dilakukan proses standarisasi pada data asli dengan menggunakan nilai *Z-Score* dengan rumus sebagai berikut.

$$Z = \frac{(x-\mu)}{\sigma} \tag{3}$$

Keterangan :

Z = *Z-Score*

x = Data yang diamati

μ = Rata-rata populasi

σ = Standar Deviasi

Adapun data hasil standarisasi tersebut dapat dilihat pada Gambar 3 berikut.

No	Kecamatan	X1	X2	X3	X4	X5
1	Gerung	0.8456	-0.0806	-0.5905	0.3495	1.0175
2	Sekotong	2.0134	4.0001	1.3155	1.8168	1.5954
3	Kediri	-0.2905	-0.5907	-0.1140	-0.1089	-0.7162
...
115	Asakota	-0.7008	-0.3357	0.1242	-0.3840	-0.7162
116	Raba	0.0566	-0.5907	0.6007	-0.3840	1.0175
117	Mpunda	-0.6377	-0.8457	2.2686	-0.4757	0.4396

Gambar 3. Hasil Standarisasi Data

3.3. Membentuk Matriks Jarak

Dalam menentukan ukuran jarak yang biasa digunakan dalam analisis *cluster* adalah jarak *euclid*. Jika x_1 dan x_2 merupakan 2 observasi yang berada di jarak *Rn*. Hasil perhitungan untuk melihat jarak kemiripan dari kecamatan satu dengan kecamatan lainnya dapat dilihat pada matriks berikut:

D =

0.0000	4.9128	2.2348	3.6038	2.5575	...	3.3562
4.9128	0.0000	3.5779	6.1221	1.0338	...	2.4762
2.2348	3.5779	0.0000	3.3604	2.9623	...	3.1600
3.6038	6.1221	3.3604	0.0000	5.7445	...	2.4914
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3.3562	2.4762	3.1600	2.4914	1.9161	...	0.0000

Berdasarkan *matrix* tersebut dapat diambil contoh hasil dari jarak *euclidean* antara data 1 kecamatan Gerung dan data 2 kecamatan Sekotong adalah 4.918 sedangkan jarak *euclidean* antara data 1

kecamatan Gerung dengan data 2 kecamatan Kediri adalah 2.2348. Demikian pula untuk interpretasi obyek yang lainnya, semakin kecil jarak antara kedua obyek maka akan semakin mirip karakteristik dari kedua obyek tersebut.

3.4. Self Organizing Map (SOM)

Penelitian ini menggunakan validasi *internal* dengan indeks *silhouette*, *dunn* dan *connectivity* untuk pembentukan jumlah *cluster* pada metode SOM. Penentuan jumlah *cluster* dilihat dari nilai Indeks *dunn* mendekati 1, nilai *silhouette* paling tinggi dan nilai *connectivity* paling kecil. Nilai dari ketiga metode tersebut disajikan dalam Tabel 2 di bawah ini:

Tabel 2. Validasi *Internal Cluster*

	3	4	5
<i>Connectivity</i>	37.7948	59.6591	61.6417
<i>Dunn</i>	0.0488	0.0450	0.4500
<i>Silhouette</i>	0.3808	0.3283	0.2957

Tabel 2 menunjukkan validasi *internal cluster* dengan tiga metode. Berdasarkan hasil validasi *internal cluster* diperoleh nilai *cluster* yang paling baik adalah 3 *cluster* dengan nilai *dunn* dan *silhouette* tertinggi, serta *connectivity* terendah diantara *cluster* lainnya. Berikutnya dilakukan perhitungan algoritma SOM dengan *learning rate* (α) 0.5.

Berikut hasil algoritma perhitungan dengan metode SOM: Inisialisasi berupa bobot (*Wij*) 0,6 yang diperoleh secara acak untuk tiap node. Setelah bobot (*Wij*) diberikan maka jaringan diberikan input (*xi*), dengan ketentuan 5 variabel *X* sebagai variabel input dan 3 *cluster*.

Tabel 3. Bobot Acak

Bobot Acak	C1	C2	C3
X1	0.3	0.6	0.5
X2	0.2	0.2	0.4
X3	0.1	0.3	0.2
X4	0.5	0.5	0.6
X5	0.6	0.4	0.5

Setelah itu lakukan perhitungan jarak vector *d(j)* yang didapat dengan menjumlahkan selisih antara vektor bobot (*Wij*) dengan vektor input (*xi*). Data yang digunakan sebagai percobaan perhitungan menggunakan 2 data yang sudah di standarisasi sebelumnya.

Iterasi 1 data 1

$$D_j = (0.3 - 0.8456)^2 + (0.2 - (-0.0806))^2 + (0.1 - (-0.5906))^2 + (0.5 - 0.3495)^2 + (0.6 - 1.0175)^2 = 1.0503 \text{ (bobot acak 1)}$$

$$D_j = (0.6 - 0.8456)^2 + (0.2 - (-0.0806))^2 + (0.3 - (-0.5906))^2 + (0.5 - 0.3495)^2 + (0.4 - 1.0175)^2 = 1.3362 \text{ (bobot acak 2)}$$

$$D_j = (0.5 - 0.8456)^2 + (0.4 - (-0.0806))^2 + (0.2 - (-0.5905))^2 + (0.6 - 0.3495)^2 + (0.5 - 1.0175)^2 = 1.3060 \text{ (bobot acak 3)}$$

Dari perhitungan data pertama yaitu kecamatan Gerung didapatkan $D_{(j)}$ minimum pada bobot acak 1, maka kecamatan Gerung masuk *cluster* 1. Setelah jarak antara node diketahui maka ditentukan nilai minimum dari perhitungan jarak vektor $D_{(j)}$, maka tahap selanjutnya melakukan update bobot dengan menggunakan rumus persamaan (2). Dari ke tiga vektor jarak maka dipilih vektor jarak dengan nilai minimum yaitu $D_{(j)}$ pada bobot acak 1.

$$W_{1,1(bar)} = 0.3 + 0.5(0.8456 - 0.3) = 0.5728 \text{ (} W_{1,1} \text{)}$$

Tahap ini dilakukan terus-menerus hingga sampai $W_{1,5}$

$$W_{1,2} = 0.2 + 0.5(-0.0806 - 0.2) = 0.0596$$

$$W_{1,3} = 0.1 + 0.5(-0.5905 - 0.1) = -0.2453$$

$$W_{1,4} = 0.5 + 0.5(0.3495 - 0.5) = 0.4248$$

$$W_{1,5} = 0.6 + 0.5(1.0175 - 0.6) = 0.8087$$

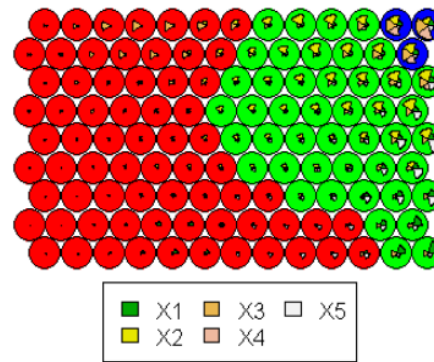
Setelah nilai $W_{ij(bar)}$ didapatkan maka nilai W_{ij} acak diganti

Tabel 4. Update Bobot

Bobot Update	Cluster 1	Cluster 2	Cluster 3
X1	0.5728	0.6	0.5
X2	0.0596	0.2	0.4
X3	-0.2453	0.3	0.2
X4	0.4248	0.5	0.6
X5	0.8087	0.4	0.5

Lakukan langkah yang sama pada data berikutnya untuk melihat bobot vektor dan pembagian *cluster* pada kecamatan lainnya. Selanjutnya dalam proses algoritma SOM didapatkan suatu SOM model dengan menggunakan *software* R akan menghasilkan diagram kipas (*fan*) berikut ini.

Codes plot



Gambar 4. Hasil *cluster* SOM

Berdasarkan Gambar 4 terdapat 117 lingkaran dalam diagram kipas. Jumlah lingkaran menunjukkan jumlah kecamatan/observasi. Masing-masing lingkaran memiliki lima kipas dengan warna berbeda yang menunjukkan terdapat lima variabel. Semakin besar bentuk kipas, maka variabel tersebut mendominasi pada *cluster* tersebut. Begitupun sebaliknya, semakin kecil gambar kipas, maka variabel tersebut tidak mendominasi pada variabel tersebut. Warna pada diagram menunjukkan jumlah *cluster* yang terbentuk yaitu tiga *cluster*. Warna biru untuk *cluster* 1, warna hijau untuk *cluster* 2 dan warna merah untuk *cluster*

Selanjutnya dilakukan perhitungan rata-rata dari masing-masing *cluster* untuk mengetahui kemiripan karakteristik dari masing-masing *cluster* (profilisasi) berdasarkan penyebab *unmet need* Provinsi Nusa Tenggara Barat.

Tabel 5. Profilisasi *Cluster*

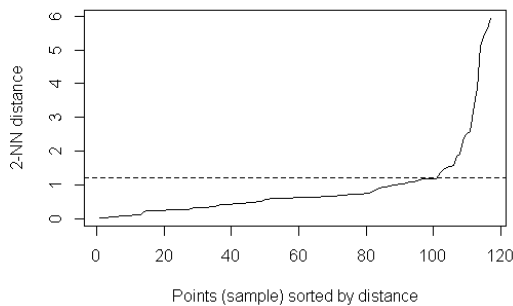
Variabel	Cluster 1	Cluster 2	Cluster 3
X1	29	20	23
X2	4	3	3
X3	4	2	7
X4	6	5	3
X5	1	1	2

Tabel 5 menjelaskan profilisasi atau karakteristik yang mendominasi pada setiap *cluster*. *Cluster* 1 beranggotakan 69 anggota merupakan kecamatan dengan penyebab *unmet need* (kebutuhan KB tidak terpenuhi) paling mendominasi yaitu tidak mengetahui tentang KB, tempat pelayanan jauh dan biaya mahal.. Sedangkan pada *cluster* 2 dengan 45 anggota kecamatan tidak menunjukkan satu pun karakteristik atau penyebab *unmet need* yang

signifikan. Sehingga, dapat dikatakan bahwa pada *cluster 2* memiliki rata-rata penyebab *unmet need* yang sama antar anggota dalam satu *cluster* dan relatif rendah dibandingkan dengan *cluster* yang lainnya. *Cluster 3* memiliki 3 anggota *cluster* dengan penyebab *unmet need* yang mendominasi yaitu alat/obat KB tidak tersedia dan tidak ada petugas pelayanan KB.

3.5. DBSCAN

Nilai *eps* dan *MinPts* yang optimal dipilih berdasarkan garis yang mendekati garis menaik kemudian dipotong secara vertikal pada *plot k* tetangga terdekat. Pada penelitian ini dilakukan pada nilai *k=2*. Nilai tersebutlah yang akan dimasukkan pada algoritme DBSCAN untuk menentukan jumlah *cluster* yang optimal. Berikut ini Gambar 6 dari *plot k-dist* untuk menentukan nilai *epsilon*.



Gambar 5. Plot penentuan *epsilon*

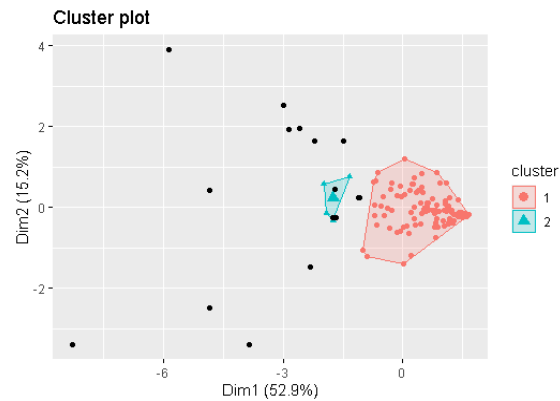
Gambar 5 merupakan plot knn tersebut diketahui bahwa nilai *epsilon* yang optimal berada di rentang titik 0.37 melalui pengamatan *knee of a curve* atau kurva siku merupakan titik dimana kurva terlihat tikungan. Digunakan dalam optimasi jika titik-titik adalah optimal untuk beberapa keputusan, misalnya ketika ada peningkatan dari vertikal ke horizontal. Pemilihan parameter *epsilon* dan *MinPts* yang optimal dilihat dari nilai *silhouette* yang tertinggi dengan memakai beberapa percobaan yang diimplementasikan pada software R. Eksperimen nilai *epsilon* dan *minpts* pada data penyebab *unmet need* di Provinsi Nusa Tenggara Barat ditunjukkan oleh Tabel 6.

Tabel 6. Evaluasi jumlah *cluster* optimal

<i>Epsilon</i>	<i>MinPts</i>	<i>Cluster</i>	<i>Noise</i>	<i>SI</i>
1.20	2	3	14	0.2207
1.20	3	2	16	0.2548
1.25	2	2	14	0.2161

Berdasarkan hasil eksperimen pada Tabel 9 dipilih nilai *epsilon* 1.20 dan *minpts* 3 karena memiliki nilai *silhouette* tertinggi yaitu sebesar 0.2548. Hasil perhitungan menggunakan *Silhouette Index* (SI) berkisar antara nilai -1 hingga 1, dengan -1 merupakan nilai terburuk dan nilai yang mendekati 1 merupakan nilai *cluster* terbaik.

Hasil yang di dapatkan dari proses pengolahan data dan disajikan dalam plot visualisasi sebagai berikut.



Gambar 6. Visualisasi *Cluster*

Gambar 6 terdapat 2 *cluster* yang dihasilkan, *cluster* pertama adalah titik yang berwarna merah dengan 97 kecamatan dan *cluster 2* pada titik yang berwarna hijau dengan 4 kecamatan, sedangkan yang titik hitam merupakan *noise* yang berjumlah 16 kecamatan. Hasil *cluster* pada penelitian ini menunjukkan adanya karakteristik yang berbeda antar *cluster*, analisis identifikasinya ditampilkan pada Tabel 5 berikut.

Tabel 7. Rata-rata *cluster*

	X1	X2	X3	X4	X5
<i>Cluster 1</i>	16	2	2	3	1
<i>Cluster 2</i>	37	11	8	6	2

Tabel 6 menggambarkan karakteristik dari masing-masing variabel. Interpretasi karakteristik masing-masing *cluster* sebagai berikut:

- Penyebab *unmet need* tertinggi berada di *cluster 2* dengan 4 kecamatan memiliki nilai rata-rata semua variabel lebih banyak dibandingkan dibandingkan *cluster 1*. Sehingga dapat dikatakan bahwa pada kecamatan tersebut belum maksimal dalam menangani kasus *unmet need*.
- Cluster* terendah yaitu *cluster 1* dengan 97 kecamatan dengan rata-rata penyebab *unmet need* rendah jika dibandingkan dengan *cluster 2*, sehingga dapat dikatakan bahwa pada

kecamatan tersebut kasus *unmet need* dapat diminimalisir dengan baik.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis Kabupaten dengan angka *unmet need* tertinggi yaitu Kabupaten Lombok Timur dengan jumlah sebanyak 84.721 orang, kemudian kabupaten Lombok Tengah berada di urutan kedua tertinggi dan urutan tertinggi ketiga yaitu kabupaten Lombok Barat, dengan masing-masing angka *unmet need* sebesar 66.201 dan 33.260 orang. Sedangkan jumlah *unmet need* terendah berada di Kota Bima sebesar 4.636 orang. Metode SOM didapatkan *cluster* 1 beranggotakan 86 anggota, *cluster* 2 dengan 30 kecamatan dan *cluster* 3 memiliki 4 kecamatan dengan penyebab *unmet need* yang mendominasi yaitu tempat pelayanan jauh, alat/obat KB tidak tersedia, biaya mahal dan tidak ada petugas KB. Sedangkan untuk metode DBSCAN dengan *eps* 1.20 dan *MinPts* 3. Jumlah *cluster* yang terbentuk yaitu 2 *cluster*, *cluster* 1 dengan 97 kecamatan dan *cluster* 2 dengan 4 kecamatan, serta 16 kecamatan yang memiliki data *noise*.

Penelitian ini menggunakan menggunakan metode SOM dan DBSCAN *Clustering*. Perbedaan metode SOM dan DBSCAN bahwa metode SOM tidak memerlukan parameter dalam analisis pembentukan jumlah *cluster*, sedangkan metode DBSCAN menggunakan parameter *epsilon* dan minimum points. Kedua parameter tersebut akan mempengaruhi hasil *cluster* yang terbentuk. Penelitian ini diharapkan dapat menjadi gambaran bagi pembaca, baik masyarakat maupun lembaga terkait kependudukan maupun kesehatan, dalam meningkatkan akses pelayanan kesehatan yang lebih berkualitas.

DAFTAR PUSTAKA

BKKBN NTB. (2020). Rencana Strategis 2020-2024 Perwakilan BKKBN Provinsi NTB. BKKBN: Nusa Tenggara Barat

Cahyani, A. D., Khotimah, B. K., Rizkillah, R. T., Studi, P., Informatika, T., & Trunojoyo, U. (2014). *Perbandingan Metode SOM (Self Organizing Map) dengan Pembobotan Berbasis RBF (Radial Basis Function) Vol. 7 No. 1 Agustus 2014 ISSN : 1979-8415*. 7(1).

Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer dengan *K-Means Cluster* Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *Computer Engineering, Science and*

System Journal, 3(2), 173.

- Gunawan, W. (2021). Implementasi Algoritma DBSCAN dalam Pengambilan Data Menggunakan Scatterplot. *Techno Xplore : Jurnal Ilmu Komputer Dan Teknologi Informasi*, 6(2), 91–98.
- Halim, N. N., Widodo, E., Statistika, P., Indonesia, U. I., Artikel, I., Bumi, G., Validation, I., Halim, N. N., & Indonesia, U. I. (2017). *Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps*. 1(1), 188–194.
- Harli, E., Fauzi, A., & Kusmanto, T. H. (2016). *Pengelompokan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok*. 2(2), 90–95.
- Kristianto, A., Sedyono, E., & Hartomo, K. D. (2020). Implementation dbscan algorithm to clustering satellite surface temperature data in indonesia. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 6(2), 109–118.
- Listyaningsih, U., Sumini, S., & Satiti, S. (2016). *Unmet need: Konsep Yang Masih Perlu Diperdebatkan*. *Populasi*, 24(1), 72.
- Made, N., Santika, A., Gede, I. K., Putra, D., & Sukarsa, I. M. (2015). *Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan*. 6(3), 185–191.
- Nastiti, R., Suprpto, B., & Gaffar, A. F. O. (2018). Perbandingan Hasil Algoritma Self Organizing Map (SOM) dan Fuzzy C-Means Clustering Untuk Kualifikasi Data Kinerja Dosen. *Prosiding Seminar Nasional Ilmu Komputer Dan Teknologi Informasi*, 3(2), 15–21.
- Qadrini, L. (2020). Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 13(2), 5–11.
- Risman, Syaripuddin, & Suyitno. (2019). Implementasi Metode Dbscan Pada Pengelompokan Kabupaten/Kota Di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat. *Prosiding Seminar Nasional Matematika*, 22–28.
- Royal, S. (2022). *Implementasi Metode K-Means Clustering Tungakan*. 4307(1), 118–124.
- Suwirmayanti, N. L. G. P. (2020). Penerapan Teknik

Clustering Untuk Pengelompokan Konsentrasi Mahasiswa Dengan Metode Self Organizing Map. Jurnal Ilmiah Intech : Information Technology Journal of UMUS, 2(01), 11–20.

Wuryandari, T., Rahmawati, R., Pati, K., Demak, K., & Brebes, K. (2017). *Metode DBSCAN untuk Pengelompokan Kabupaten / Kota Di Provinsi Jawa Tengah.* 5(1), 1–6.

Yordani, R. (2016). *Teori Pertumbuhan Ekonomi Menggunakan Aplikasi Integrasi Self Organizing Map (SOM) dan Sistem Informasi Geografis Visualization Of Clustering Region By Economic Growth Theory Using The Integrasi Self Organizing Map (SOM).* 129–142.

Prediksi Profitabilitas Bank Umum Konvensional Pada Masa Pandemi Covid-19

Yustika Haryanti ⁽¹⁾, Fauziyah ^{*(2)}

Universitas PGRI Adi Buana Surabaya

Jl. Dukuh Menanggal XII, Dukuh Menanggal, Kota Surabaya

e-mail: Yustikaharyanti18@gmail.com, dan Basta.fauziyah@unipasby.ac.id*

ABSTRAK

Tujuan dari penelitian untuk memprediksi profitabilitas bank umum konvensional pada masa pandemi Covid-19, dengan cara mengukur *Return On Assets* (ROA), *Non Performing Loan* (NPL), dan Biaya Operasional dan Pendapatan Operasional (BOPO), karena di masa pandemi sektor perbankan juga merupakan sektor yang terdampak langsung dari pandemi Covid-19. Penelitian ini menggunakan penelitian deskriptif dengan pendekatan kuantitatif. Jenis data menggunakan *time series* yang diperoleh langsung dari laporan kinerja perusahaan perbankan yang sudah diaudit dan diterbitkan di Otoritas Jasa Keuangan (OJK) periode 2020-2021. Hasil penelitian menunjukkan bahwa di masa pandemi Covid-19 prediksi profitabilitas periode 2022 yang diukur menggunakan rasio ROA mengalami peningkatan. Sedangkan prediksi dengan rasio NPL dan BOPO mengalami penurunan.

Kata kunci : *Prediksi, Profitabilitas, Bank Umum Konvensional, Pandemi.*

ABSTRACT

The purpose of this research is to determine the prediction of profitability at conventional commercial banks during Covid-19 pandemic, by measuring Return On Assets (ROA), Non Performing Loans (NPL), and Operational Costs and Operational Income (BOPO), because during the pandemic the banking sector was also one of the sectors that was directly affected by the Covid-19 pandemic. This research based on descriptive research with a quantitative approach. Data was gathered using the documentation method. Types of data used in this research is a time series data obtained indirectly from the performance reports of banking that have been audited and published at the Financial Services Authority (OJK) period 2020-2021. The results showed that during the Covid-19 pandemic, the prediction of profitability for the 2022 period as measured by the ROA ratio has increased. Meanwhile, predictions with the NPL and BOPO have decreased.

Keywords : *Predictions, Profitability, Conventional Commercial Banks, Pandemic.*

1. PENDAHULUAN

Profitabilitas merupakan rasio untuk mengukur kemampuan perusahaan dalam menghasilkan keuntungan dari operasi bisnisnya (Hery, 2017). Dasar dalam mengetahui profitabilitas dapat melalui kinerja keuangan perusahaan perbankan yang dihitung melalui rasio *Return On Assets* (ROA), *Non Performing Loan* (NPL), Biaya Operasional dan Pendapatan Operasional (BOPO) (Kasmir, 2016). Dengan cara tersebut dapat diketahui bagaimana memprediksi profitabilitas perusahaan di masa mendatang.

Situasi yang normal memungkinkan perusahaan dapat memaksimalkan kinerja perusahaan untuk mendapatkan profitabilitas atau laba yang maksimal. Namun, kondisi perekonomian global saat ini sedang dalam keadaan yang

mengkhawatirkan. Pada awalnya kasus pandemi Covid-19 menyebar dengan cepat di seluruh dunia termasuk Indonesia. Dapat dilihat saat terjadinya pandemi Covid-19 pertumbuhan ekonomi Indonesia mengalami kelambatan (Fibrianti, 2021). Organisasi Internasional *World Health Organization* menyatakan bahwa *Coronaviruses* (Cov) dapat menjangkit saluran nafas pada manusia (Nasution et al., 2020). Nama ilmiah virus tersebut yaitu Covid-19. Covid-19 sendiri memiliki gejala klinis batuk kering, disone, demam, dan infeksi paru bilateral filtrat pada pencitraan, gejala ini ditemukan pada masyarakat yang terpapar virus ini.

Perlu diketahui, dampak Covid-19 dalam bidang ekonomi di Indonesia sangat terlihat jelas. Dikatakan seperti itu karena pemerintah menerapkan pembatasan sosial bersekala besar.

Penerapan PSBB (Pembatasan Sosial Bersekala Besar), menghentikan secara menyeluruh kegiatan masyarakat Indonesia. Kegiatan ini dilakukan bertujuan untuk memutus penyebaran virus Covid-19 (Nasution et al., 2020). Sehingga berdampak terhadap sektor perekonomian terkecuali yang bergerak dalam bidang makanan, obat-obatan atau medis dan lembaga keuangan. Lembaga keuangan tidak termasuk dalam pembatasan dikarenakan lembaga ini sebagai penopang sistem keuangan nasional. Namun, penerapan PSBB sangat berdampak besar terhadap kelangsungan pengusaha kecil maupun pengusaha besar, dampak ini terjadi karena pengusaha tidak dapat melakukan kegiatan orasional perusahaan sehingga perusahaan tidak dapat mengelola usahanya. Kondisi ini secara langsung berdampak terhadap perusahaan perbankan.

Sektor perbankan adalah salah satu sektor perekonomian yang terkena dampak Covid-19. Berdasarkan data dari OJK, diprediksi pada akhir 2021 laba bank akan turun antara 30% - 40% dari tahun sebelumnya. Mulai kuartal kedua di tahun 2020, antara bulan April - Juni 2020, laba bank sebelum pajak tercatat turun 19.8% dari tahun sebelumnya (Adam, Safitri & Wahyusi 2018). Penurunan laba ini dikarenakan adanya kredit macet. Data yang diperoleh dari OJK, kredik macet bank mengalami kenaikan 3.22%. Angka tersebut mengalami peningkatan dari bulan Juni sebesar 3.1%. Meskipun dari data tersebut mengalami penurunan keuntungan dan peningkatan kredit macet, akan tetapi data tersebut menunjukkan bahwa likuiditas bank Indonesia cukup memadai dalam menyalurkan kredit untuk mendorong pertumbuhan perekonomian.

Perusahaan perbankan dalam mengantisipasi terjadinya kredit bermasalah dapat melakukan perhitungan dengan rasio NPL. NPL merupakan rasio untuk mengukur kemampuan bank dalam menutupi atau meng-cover risiko terjadinya kegagalan pengembalian kredit oleh debitur (Kasmir, 2016). Dalam mengukur profitabilitas perusahaan dapat menggunakan rasio ROA. Rasio ini dapat digunakan untuk mengetahui sejauh mana kemampuan perusahaan dalam menggunakan seluruh asetnya untuk menghasilkan laba setelah pajak. ROA sebagai indikator kinerja keuangan pada bank, sebagai prediksi kemampuan pengelolaan aset bank dalam mendapatkan laba (Hairunnisa et al., 2021). Sedangkan rasio BOPO untuk mengukur seberapa efisiensi pada bank dalam melakukan kegiatan Operasionalnya. Semakin tinggi biaya Operasional akan berpengaruh terhadap pendapatan profit atau laba.

Salah satu cara untuk mengantisipasi terjadinya penurunan laba atau profitabilitas perusahaan perbankan, dengan melakukan prediksi. Prediksi profitabilitas merupakan profitabilitas masa depan yang disesuaikan dengan industri yang akan diteliti yang kemudian diregresi berdasarkan ukuran profitabilitas (Najib & Rini, 2016). Dari hasil prediksi profitabilitas dapat dilihat bagaimana kemampuan perusahaan mendapatkan laba yang merupakan ukuran dalam kinerja keuangan. Kinerja keuangan tersebut dapat dilakukan dengan mengukur rasio ROA, rasio NPL, dan BOPO. Prediksi tingkat profitabilitas dan laba perusahaan akan memberikan gambaran yang dapat digunakan sebagai strategi keberlangsungan perusahaan (Afkhar & Fauziah, 2021).

Prediksi pada penelitian ini menggunakan *Trend Analysis* dengan bantuan MINITAB. Model yang digunakan yaitu tren Linier, Kuadratik, Eksponensial, dan S-Curve. Model yang dipilih karena data tidak memiliki unsur musiman, data cenderung konstan dan digunakan untuk prediksi jangka Panjang.

2. METODE PENELITIAN

Penelitian ini menggunakan penelitian deskriptif dengan pendekatan kuantitatif. Pendekatan ini menggunakan angka-angka yang didasarkan pada pengumpulan data, penafsiran terhadap data, dan tampilan angka yang dihasilkan. Pendekatan ini terkait dengan variabel peneliti yaitu ROA (X1), NPL (X2), dan BOPO (X3) yang berfokus melakukan prediksi profitabilitas pada masa pandemi Covid-19 dihasilkan bentuk penelitian berupa angka-angka yang bermakna. Penelitian ini hanya menggunakan satu variabel bebas (*Independent variable*). Populasi dan sampel dari penelitian ini yaitu Bank Umum Konvensional di Indonesia yang terdaftar di OJK pada periode 2020-2021. Teknik analisis peneliti menggunakan statistik deskriptif dengan menggunakan Model *Trend Analysis*. *Trend Analysis* digunakan untuk membandingkan data sekarang dengan data masa lalu, *trend analysis* dapat memberikan informasi mengenai rencana jangka panjang dan penilaian terhadap program peningkatan kualitas.

Perhitungan *error* peramalan digunakan pada dua jenis keputusan. Keputusan pertama adalah untuk membandingkan akurasi dan memilih metode peramalan yang paling optimal diantara metode-metode peramalan yang ada. Keputusan kedua adalah untuk mengevaluasi seberapa mendekati kenyataan (keberhasilan atau kegagalan) dan mengukur realibilitas suatu metode peramalan.

Akurasi perhitungan dari kesalahan peramalan di setiap model peramalan dapat dijelaskan dengan membandingkan nilai yang diramal dengan nilai aktual atau nilai yang sedang diamati. Beberapa alat ukur yang dipakai untuk menghitung akurasi peramalan yaitu *mean absolute deviation* (MAD), *mean squared deviation* (MSD), dan *mean absolute percentage error* (MAPE) (Kurniawan, 2018).

2.1. MAPE

MAPE lebih berarti membandingkan MAD karena MAPE menyatakan persentase kesalahan hasil peramalan terhadap permintaan aktual selama periode tertentu yang akan memberikan informasi persentase kesalahan. Secara sistematis, MAPE dirumuskan sebagai berikut :

$$MAPE = \left(\frac{100}{n}\right) \sum At - \frac{Ft}{At}$$

Keterangan :

- At = Permintaan aktual pada periode -t
- Ft = Peramalan permintaan (*Forecast*) pada periode -t
- n = Jumlah periode peramalan yang terlibat

2.2. MAD

MAD merupakan rata-rata kesalahan mutlak selama periode tertentu tanpa memperhatikan apakah hasil peramalan lebih besar atau lebih kecil dibandingkan kenyataannya, secara sistematis MAD dirumuskan sebagai berikut:

$$MAD = \sum \frac{(At-Ft)}{n}$$

Keterangan:

- At = Permintaan aktual pada periode -t
- Ft = Peramalan permintaan (*Forecast*) pada periode -t
- n = Jumlah periode peramalan yang terlibat dalam MAD, kesalahan dengan arah positif atau negatif yang diukur hanya besar kesalahan secara absolut.

2.3. MSD

MSD dihitung dengan menjumlahkan kuadrat semua kesalahan peramalan pada setiap periode dan membaginya dengan jumlah periode peramalan, secara sistematis, MSD dirumuskan sebagai berikut:

$$MSD = \sum \frac{(At-Ft)^2}{n}$$

Keterangan:

- At = Permintaan aktual pada periode -t
- Ft = Peramalan permintaan (*Forecast*) pada periode -t

n = Jumlah periode peramalan yang terlibat

Semakin kecil nilai persentase kesalahan (*percentage error*) dari perhitungan MAD, MSD, dan MAPE maka semakin akurat hasil peramalan tersebut (Kurniawan, 2018).

3. HASIL DAN PEMBAHASAN

3.1. Identifikasi Model Tren Analysis Data

Identifikasi model tren analysis data merupakan langkah pertama dalam memulai prediksi atau ramalan. Peneliti melakukan perbandingan nilai MAPE, MAD, dan MSD pada model *tren Linier, Kuadratik, Eksponensial, dan S-Curve* pada masing-masing profitabilitas. Tabel 1 merupakan hasil perbandingan nilai MAPE, MAD dan MSD untuk ROA, NPL dan BOPO.

Tabel 1. Perbandingan nilai MAPE, MAD, dan MSD untuk ROA

Model Tren	MAPE	MAD	MSD
Linear	15.1869	0.2221	0.068
Kuadratik	13.4942	0.1994	0.053
Eksponensial	15.1215	0.2253	0.0687
S-Curve	-	-	-

Tabel 2. Perbandingan nilai MAPE, MAD, dan MSD untuk NPL

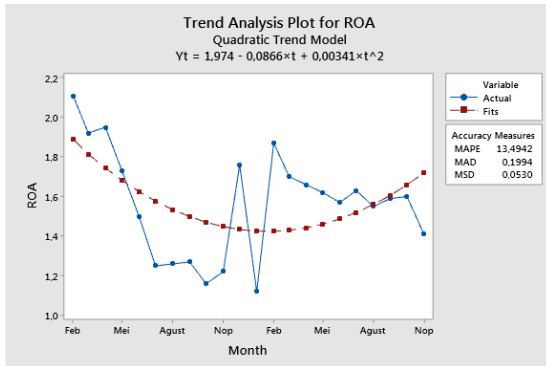
Model Tren	MAPE	MAD	MSD
Linear	1.45370	1.22470	2.32093
Kuadratik	1.26693	1.07802	1.84745
Eksponensial	1.37557	1.16401	2.14163
S-Curve	1.38238	1.18404	2.31815

Tabel 3. Perbandingan nilai MAPE, MAD, dan MSD untuk NPL

Model Tren	MAPE	MAD	MSD
Linear	1.67764	1.50422	4.31474
Kuadratik	1.6322	1.45345	3.3256
Eksponensial	1.67197	1.49977	4.33267
S-Curve	2.35174	2.08117	6.04942

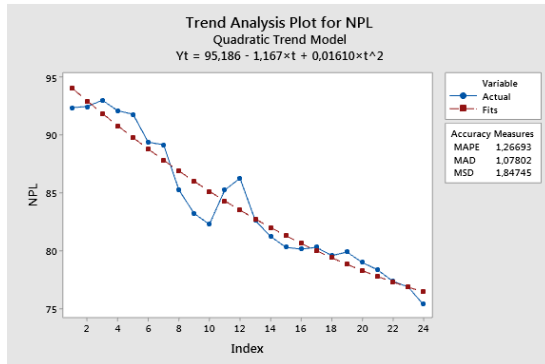
Berdasarkan hasil Identifikasi model *tren analysis*, hasil dari ROA, NPL, dan BOPO diperoleh nilai MAPE, MAD, dan MSD terkecil diantara 4 model tren yaitu model kuadratik. Sehingga dalam penelitian ini *tren analysis* yang digunakan yaitu model *kuadratik*. Gambar 1 merupakan *tren*

analysis untuk ROA, NPL dan BOPO yang menggunakan model *kuadratik*.



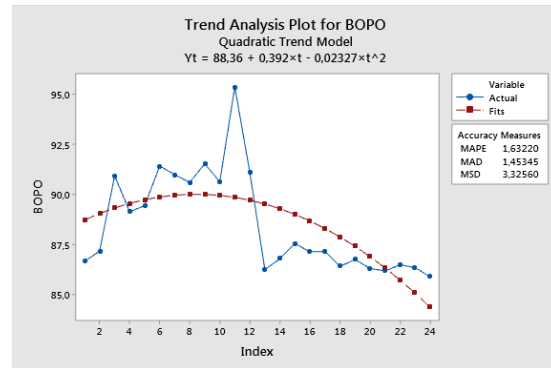
Gambar 1. Tren Analysis Data Kuadratik ROA

Berdasarkan gambar 1 *tren analysis* data *kuadratik* pada ROA membentuk pola parabola menghadap ke atas yang artinya prediksi ROA untuk beberapa periode kedepan akan mengalami kenaikan. Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh (Pratiwi et al., 2017), hasil penelitian menunjukkan bahwa ROA berpengaruh negatif signifikan terhadap prediksi *financial distress*. Artinya, semakin kecil ROA maka semakin besar kemungkinan perusahaan mengalami kebangkrutan.



Gambar 2 Tren Analysis Data Kuadratik NPL

Berdasarkan gambar 2 *tren analysis* data *kuadratik* pada NPL membentuk pola parabola menghadap ke atas. Pola parabola menghadap keatas berarti prediksi NPL untuk periode kedepan mengalami kenaikan. Jika dikaitkan dengan teori signal maka prediksi profitabilitas pada NPL dapat memberi sinyal positif bagi investor dan perusahaan untuk mengantisipasi terjadinya kredit bermasalah sehingga tidak merugikan investor dan perusahaan.



Gambar 3. Tren Analysis Data Kuadratik BOPO

Gambar 3 *tren analysis* data *kuadratik* pada BOPO bergerak naik lalu turun kembali membentuk pola parabola menghadap ke bawah. Pola parabola menghadap ke bawah artinya beberapa periode kedepan prediksi BOPO akan mengalami penurunan. Jika dikaitkan dengan teori signal maka prediksi profitabilitas pada BOPO dapat memberi sinyal terhadap pengelolaan perusahaan sebagai salah satu upaya dalam melakukan taksiran yang tepat antara besaran biaya Operasional perusahaan dengan pendapatan Operasional agar dapat memaksimalkan laba perusahaan.

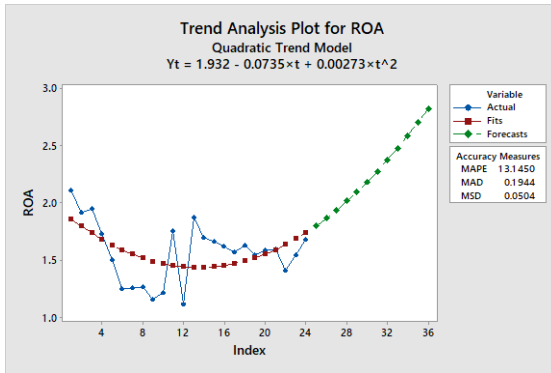
3.2. Prediksi ROA, NPL dan BOPO

Berdasarkan Gambar 1, Gambar 2, dan Gambar 3 didapatkan persamaan *kuadratik* yang dapat digunakan untuk memprediksi ROA, NPL, dan BOPO. Persamaan *kuadratik* disajikan pada tabel 4.

Tabel 4. Persamaan Kuadrat hasil prediksi

No	Profita-bilitas	Persamaan
1	ROA	$Y_t = 1,932 - 0,0735t + 0,00273t^2$
2	NPL	$Y_t = 95,186 - 1,167t + 0,01610t^2$
3	BOPO	$Y_t = 88,36 + 0,392t - 0,02327t^2$

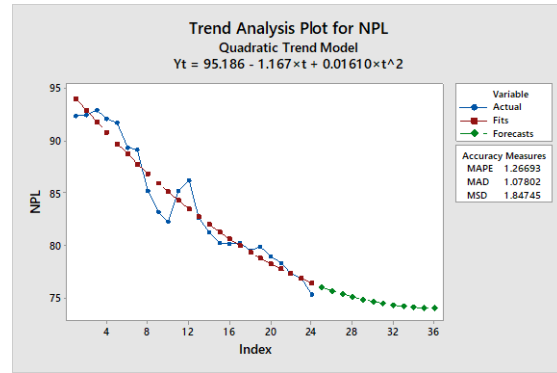
Persamaan kuadrat yang telah didapatkan digunakan untuk melakukan prediksi profitabilitas pada bank umum konvensional untuk bulan Januari-Desember 2022 yang diukur menggunakan ROA, NPL, dan BOPO. Gambar 4 menunjukkan hasil prediksi profitabilitas pada bank umum konvensional dengan bantuan software MINITAB.



Gambar 4. Prediksi ROA

Gambar 4 menunjukkan bahwa prediksi profitabilitas pada bank umum konvensional yang diukur menggunakan ROA didapatkan hasil prediksi periode 2022 yang mengalami peningkatan. Berdasarkan hasil penelitian menunjukkan bahwa prediksi ROA pada bank umum konvensional selama pandemi Covid-19 mengalami peningkatan. ROA merupakan rasio yang digunakan untuk menilai kualitas dan kinerja perusahaan dalam menghasilkan laba bersih dari pemanfaatan aset yang tercantum didalam neraca sehingga kita dapat mengetahui apakah perusahaan dapat menilai sejauh mana kemampuan perusahaan menggunakan asetnya dalam memperoleh laba atau profitabilitas. Semakin tinggi rasio ROA dapat dikatakan bahwasannya bank umum konvensional dapat memperoleh laba atau profitabilitas. Hal ini juga berlaku sebaliknya, apabila rasio ROA mengalami penurunan maka profitabilitas pada bank umum konvensional akan mengalami penurunan.

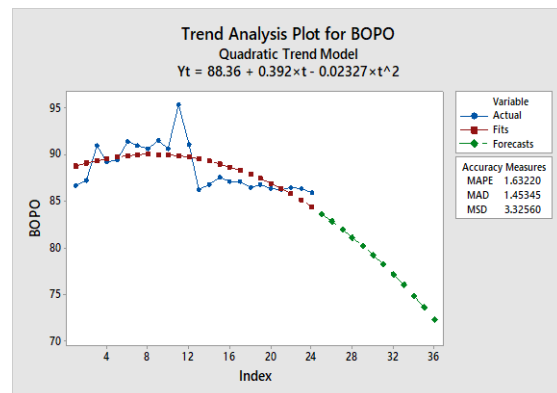
Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh (Pratiwi et al., 2017), hasil penelitian ini menunjukkan bahwa ROA berpengaruh negatif signifikan terhadap prediksi financial distress. Artinya, semakin kecil ROA maka semakin besar kemungkinan perusahaan mengalami financial distress. Jika dikaitkan dengan teori signal maka prediksi profitabilitas pada ROA ini dapat dijadikan sebagai upaya dalam memberi sinyal positif kepada investor dan juga perusahaan berkaitan dengan kinerja perusahaan dalam memperoleh laba atau profitabilitas di masa yang akan datang.



Gambar 5. Prediksi NPL

Gambar 5 menjelaskan bahwa prediksi profitabilitas pada bank umum konvensional yang diukur menggunakan rasio NPL didapatkan hasil prediksi periode 2022 yang mengalami penurunan. Hal tersebut dapat dikatakan bahwa prediksi NPL pada bank umum konvensional selama pandemi Covid-19 mengalami penurunan. NPL merupakan rasio yang digunakan untuk mengukur tingkat kredit bermasalah pada bank. Penurunan rasio NPL ini dapat menunjukkan bahwasannya pendapatan yang seharusnya diterima perusahaan sebagai laba atau profit dapat dimaksimalkan, sehingga berdampak terhadap tingkat profitabilitas bank.

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh (Afkar & Fauziyah, 2021), hasil penelitian ini menunjukkan bahwa *Non Performing Financing (NPF)* mengikuti pola parabola menghadap ke bawah. Pola parabola menghadap ke bawah menunjukkan prediksi NPF untuk beberapa bulan kedepan mengalami penurunan nilai. Jika dikaitkan dengan teori signal maka prediksi profitabilitas pada NPL ini dapat memberi sinyal positif bagi investor dan perusahaan untuk mengantisipasi terjadinya kredit bermasalah sehingga tidak merugikan investor dan perusahaan.



Gambar 6. Prediksi BOPO

Gambar 6 menunjukkan bahwa prediksi profitabilitas pada bank umum konvensional yang diukur menggunakan rasio BOPO didapatkan hasil prediksi periode 2022 yang mengalami penurunan. Hal tersebut dapat dikatakan bahwa prediksi BOPO pada bank umum konvensional selama pandemi Covid-19 mengalami penurunan. BOPO merupakan rasio yang digunakan untuk mengukur efisiensi dan melihat bagaimana kemampuan bank dalam melakukan kegiatan Operasional. Penurunan rasio BOPO ini dapat menunjukkan bahwasannya perusahaan dapat melakukan kegiatan Operasionalnya dengan baik sehingga mendapatkan laba dengan maksimal. Hal ini juga berlaku sebaliknya, apabila rasio BOPO mengalami peningkatan maka akan berakibat terhadap peningkatan biaya Operasional perusahaan dampak tersebut akan berpengaruh terhadap profitabilitas yang akan di dapatkan perusahaan.

Hasil penelitian ini bertolak belakang dengan penelitian yang dilakukan oleh (Afkar & Fauziyah, 2021), hasil penelitian ini menunjukkan bahwa BOPO mengikuti pola parabola menghadap ke atas. Pola parabola menghadap ke atas menunjukkan prediksi BOPO untuk beberapa bulan kedepan mengalami peningkatan. Jika dikaitkan dengan teori signal maka prediksi profitabilitas pada BOPO ini dapat memberi sinyal terhadap pengelolaan profitabilitas perusahaan sebagai salah satu upaya dalam melakukan taksiran yang tepat antara besaran biaya Operasional perusahaan dengan pendapatan Operasional agar dapat memaksimalkan laba perusahaan.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Prediksi perofitabilitas periode 2022 pada bank umum konvensional selama pandemi covid-19 yang diukur menggunakan rasio ROA mengalami peningkatan. Sedangkan prediksi profitabilitas yang diukur menggunakan rasio NPL dan BOPO mengalami penurunan

4.2 Saran

Sebelum memutuskan untuk berinvestasi pada bank umum konvensional yang terdaftar di Otoritas Jasa Keuangan (OJK), sebaiknya terlebih dahulu memperhatikan bagaimana prediksi profitabilitas yang dapat dilihat melalui rasio ROA, NPL, dan BOPO yang dapat menilai bagaimana kinerja perusahaan. Kinerja perusahaan yang stabil atau menjanjikan dapat memberikan keuntungan bagi investor, sehingga investor dapat

mengantisipasi terjadinya kerugian saat berinvestasi.

DAFTAR PUSTAKA

- Afkar, T., & Fauziyah. (2021). Predictions And Trends Profitability For Islamic Commercial Banks In Indonesia During The Covid-19. *International Journal of Economics, Business, and Accounting Research (IJEBAAR)*, 1(1), 188–196.
<https://doi.org/http://dx.doi.org/10.29040/ijeb ar.v5i1.2232>
- Fibrianti, L. (2021). *Kondisi Pertumbuhan Ekonomi saat Pandemi Covid-19 di Indonesia*. Vviva.Co.Id.
<https://www.viva.co.id/vstory/lainnya-vstory/1380397-kondisi-pertumbuhan-ekonomi-saat-pandemi-covid-19-di-indonesia>
- Hairunnisa, Y. I., Mulyantini, S., & Jubaedah, J. (2021). Analisis Profitabilitas Bank Umum Konvensional Sebelum dan Saat Pandemi Covid-19. *Syntax Literate ; Jurnal Ilmiah Indonesia*, 6(10), 5398.
<https://doi.org/10.36418/syntax-literate.v6i10.4355>
- Hery. (2017). *Kajian riset akuntansi : mengulas berbagai hasil penelitian terkini dalam bidang akuntansi dan keuangan* (A. Pramono (ed.)). PT Grasindo.
- Kasmir. (2016). *Analisis Laporan Keuangan* (9 ed). PT.Raja Grafindo Persada.
- Kurniawan, A. (2018). Nalisis Tren Kematian Bayi Dan Anak Di Kota Surabaya. *Jurnal Ilmiah Kesehatan Media Husada*, 6(2), 175–182.
<https://doi.org/10.33475/jikmh.v6i2.36>
- Najib, H., & Rini, R. (2016). Sharia Compliance, Islamic Corporate Governance Dan Fraud Pada Bank Syariah. *Jurnal Akuntansi Dan Keuangan Islam*, 4(2), 131–146.
<https://doi.org/10.35836/jakis.v4i2.23>
- Nasution, D. A. D., Erlina, E., & Muda, I. (2020). Dampak Pandemi COVID-19 terhadap Perekonomian Indonesia. *Jurnal Benefita*, 5(2), 212.
<https://doi.org/10.22216/jbe.v5i2.5313>
- Pratiwi, J., Machmuddah, Z., Akuntansi, P. S., Ekonomi, F., Nuswantoro, U. D., Nakula, J., & Semarang, I. N. (2017). *Diterima : April 2017; Direvisi : Juni 2017; Dipublikasikan : September 2017*. 2(2), 58–67.

Teknik Oversampling Pada Regresi Logistik Ordinal Dalam Menduga Faktor Yang Memengaruhi Risiko Penyebaran Zona Covid-19 di Kabupaten Garut

Ghina Fauziah⁽¹⁾, Indahwati⁽²⁾, Erfiani⁽³⁾, Anwar Fitrianto⁽⁴⁾, Reni Amelia⁽⁵⁾

IPB University

Kampus IPB Baranangsiang, Jl. Raya Pajajaran, Bogor, Jawa Barat-Indonesia

e-mail: ghinafauziah57@gmail.com, indah.stk@gmail.com, erfiani@apps.ipb.ac.id,
anwarstat@gmail.com, dan reniamelia@apps.ipb.ac.id

ABSTRAK

Covid-19 merupakan penyakit yang disebabkan oleh infeksi virus SARS-CoV-2, pertama kali masuk ke Indonesia pada awal tahun 2020. Adanya wabah covid-19 menyebabkan setiap daerah yang ada di Indonesia khususnya kabupaten Garut harus terbagi ke dalam beberapa risiko zona covid-19 sesuai dengan kondisi dari suatu daerah tersebut. Beberapa faktor yang dapat memengaruhi suatu daerah masuk pada risiko zona tertentu dapat ditentukan berdasarkan jumlah kasus positif covid-19, kasus suspek, kasus kontak erat, jumlah desa, dan kepadatan penduduk daerah tersebut. Regresi logistik ordinal merupakan salah satu analisis regresi yang digunakan untuk menganalisa hubungan antara variabel prediktor dan variabel respon, dimana variabel respon tersebut bersifat kategorik dengan skala ordinal. Oleh karena itu, maka digunakan regresi logistik ordinal untuk mengetahui faktor apa saja yang memberikan pengaruh terhadap pembagian risiko zona covid-19 di kabupaten Garut pada bulan Juli tahun 2021. Sebelum melakukan pemodelan regresi logistik ordinal dilakukan terlebih dahulu proses teknik *resampling* dengan metode *oversampling* untuk menangani data yang tidak seimbang pada peubah respon. Berdasarkan pemodelan hasil dari pemodelan serta pengujian secara parsial, didapatkan bahwa peubah bebas yang memiliki pengaruh terhadap risiko zona covid-19 di kabupaten Garut yaitu jumlah desa, kepadatan penduduk, kasus suspek, dan kasus konfirmasi positif dengan nilai akurasi sebesar 85.71%.

Kata kunci : Risiko Zona Covid-19, *Oversampling*, Regresi Logistik Ordinal.

ABSTRACT

Covid-19 is a disease caused by infection with the SARS-CoV-2 virus, which first entered Indonesia in early 2020. The Covid-19 outbreak has caused every region in Indonesia, especially Garut district, to be divided into several risk zones for COVID-19 according to the conditions of the area. Several factors that can affect an area entering a certain risk zone can be determined based on the number of positive cases of COVID-19, suspected cases, close contact cases, the number of villages, and the population density of the area. Ordinal logistic regression is one of the regression analyze the relationship between the predictor variable and the response variable, where the response variable is categorical with an ordinal scale. Therefore, so ordinal logistic regression was used to find out what factors had an influence on the distribution of the risk of the covid-19 zone in Garut district in July 2021. Before doing ordinal logistic regression modeling, the resampling technique process was carried out with the oversampling method to handle the data. unbalanced on the response variabel. Based on the modeling results from modeling and partial testing, it was found that the independent variabels that have an influence on the risk of the covid-19 zone in Garut district are the number of villages, population density, suspected cases, and positive confirmation cases with an accuracy value of 85.71%.

Keywords : Covid-19 Zone Risk, *Oversampling*, Ordinal Logistiks Regression.

1. PENDAHULUAN

Virus corona 2019 atau covid-19 merupakan penyakit menular yang disebabkan oleh virus SARS-CoV-2. Penyakit tersebut telah menyerang

sebagian besar manusia hampir diseluruh negara. Pandemi covid-19 telah dimulai pada tahun 2019 yang terjadi di negara China, diawali dengan menyebarnya virus yang ditemui pada manusia dan

hewan yang menyerang alat vital paru-paru dan memiliki risiko kematian tanpa memandang usia, sehingga semua manusia dapat terkena virus covid-19 tersebut tanpa memandang umur. Awal tahun 2020, *World Health Organization* (WHO) menetapkan bahwa wabah covid-19 telah ditetapkan menjadi masapandemi global dengan penyebaran virus yang sangat cepat dan berpotensi terserang infeksi penyakit covid-19 diseluruh dunia, termasuk Indonesia.

Pada awal tahun 2020 merupakan awal baru bagi seluruh manusia di dunia, tidak terkecuali Indonesia. Pandemi wabah covid-19 telah menyerang hampir ke seluruh wilayah di Indonesia, salah satunya Jawa Barat. Berdasarkan data dari dinas Kesehatan Jawa Barat bahwa dari awal tahun 2020 data konfirmasi positif covid-19 semakin hari semakin meningkat, terutama pada tahun 2021 dimana terjadi lonjakan yang sangat tinggi terhadap konfirmasi positif kasus covid-19. Menurut Pusat Informasi dan Koordinasi Jawa Barat, kabupaten Garut merupakan salah satu dari 15 kabupaten/kota di Jawa Barat dengan kasus konfirmasi positif paling banyak dengan jumlah kasus total di Jawa Barat hingga 28 Mei 2022 yaitu 1.106.499 orang terkonfirmasi positif covid-19. Bulan Juli merupakan salah satu bulan dengan kasus konfirmasi positif covid-19 paling tinggi di kabupaten Garut dibandingkan bulan-bulan pada tahun sebelumnya maupun setelahnya yaitu pada tahun 2021 terjadi sebanyak 4.141 konfirmasi positif covid-19, kemudian pada tahun 2021 terjadi sebanyak 20.633 kasus konfirmasi positif covid-19, dan 2022 hingga 27 Mei sebanyak 6.134 kasus konfirmasi positif covid-19.

Dengan adanya wabah pandemi covid-19, maka dari itu seluruh daerah di Indonesia khususnya kecamatan-kecamatan yang ada di kabupaten Garut menjadi terbagi ke dalam beberapa risiko zona yang salah satunya ditentukan berdasarkan kenaikan atau penurunan jumlah kasus covid-19 di kecamatan tersebut. Tujuan adanya pembagian risiko zona covid-19 untuk mengendalikan kasusnya yang berbasis pada pendataan wilayah masing-masing yang diawasi secara ketat serta membedakan antar wilayah yang memiliki risiko tinggi, risiko sedang, risiko rendah, dan wilayah yang tidak berisiko. Risiko zona ditentukan berdasarkan hasil skoring dan pembobotan dari berbagai indikator kesehatan masyarakat. Maka dari itu, penulis bermaksud melakukan penelitian mengenai pengaruh kasus harian covid-19 meliputi jumlah desa, kepadatan penduduk, suspek, konfirmasi positif, dan kontak erat terhadap risiko zona covid-19 di kabupaten Garut periode bulan Juli tahun 2021.

Dalam menganalisis kasus risiko sebaran zona covid-19 di Kabupaten Garut, digunakan model regresi logistik ordinal. Regresi logistik ordinal digunakan untuk memodelkan hubungan antara peubah respon yang terdiri lebih dari dua kategori dengan satu atau lebih peubah penjelas (Hosmer dan Lemeshow 2000). Adapun alasan mengapa metode yang digunakan adalah regresi logistik ordinal adalah karena kasus penyebaran covid-19 merupakan variabel kategorik (lebih dari 2 kategori) dengan skala ordinal sehingga tepat jika menganalisis sebaran risiko zona covid di kabupaten Garut menggunakan regresi logistik ordinal, berbeda dengan regresi logistik lainnya seperti regresi logistik biner dan regresi logistik multinomial. Selain itu, regresi logistik ordinal dapat menjawab variabel apa saja yang dapat memengaruhi sebaran zona covid-19 di kabupaten Garut. Sebelum melakukan proses pemodelan, dilakukan terlebih dahulu pre-processing data dengan menggunakan teknik resampling dengan tujuan untuk menyeimbangkan distribusi data sehingga mengurangi efek distribusi kelas atau kategori yang tidak seimbang (Jian et al. 2016), adapun teknik resampling yang digunakan adalah random oversampling. Random oversampling digunakan untuk menyeimbangkan data dengan cara menambahkan data secara acak pada kelas minoritas atau paling sedikit (risiko zona covid-19) yaitu pada kelas kategori zona risiko tinggi dan zona risiko rendah.

Berdasarkan penelitian yang telah dilakukan oleh Wara Pramesti dkk mengenai risiko penyebaran covid-19 di Jawa Timur menggunakan regresi logistik ordinal dengan metode *backward elimination*, dari hasil penelitian tersebut dapat disimpulkan bahwa yang memengaruhi risiko penyebaran covid-19 di Jawa Timur adalah jumlah kasus covid-19 yang meninggal dan rata-rata anggota rumah tangga. Dari hasil penelitian tersebut, Pemerintah Jawa Timur dapat membuat kebijakan dengan cara memprioritaskan hal-hal yang dapat menyebabkan penyebaran covid-19 semakin parah.

Tujuan dari penelitian ini : (1) mengetahui gambaran sebaran risiko zona covid-19 di kabupaten Garut; (2) mengaplikasikan metode regresi logistik ordinal untuk mengidentifikasi faktor apa saja yang mempengaruhi sebaran risiko zona covid-19 di kabupaten Garut.

2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai kasus risiko sebaran zona covid-19 kabupaten Garut bulan Juli tahun 2021 yang diperoleh dari Dinas Kesehatan kabupaten Garut.

2.2 Variabel dan Definisi Operasional Variabel

Variabel yang digunakan dalam penelitian ini yaitu risiko penyebaran zona covid-19 sebagai variabel respon berbentuk data ordinal (rendah, sedang, tinggi) yang dinotasikan dengan Y. Variabel bebas terdiri dari jumlah desa (X_1), kepadatan penduduk (X_2), kasus suspek (X_3), kasus konfirmasi positif (X_4), dan kasus kontak erat (X_5).

2.3 Teknik Analisis

Tahapan analisis yang dilakukan dengan rincian sebagai berikut:

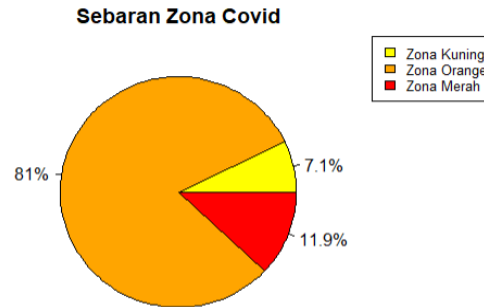
1. Melakukan eksplorasi terhadap data awal
2. Melakukan pengujian korelasi antar variabel bebas.
3. Membagi data menjadi 80% data latih dan 20% data uji.
4. Melakukan proses oversampling.
5. Menyusun model regresi logistik ordinal menggunakan teknik oversampling sebagai berikut:
 - a. Pemodelan regresi logistik ordinal
 - b. Pengujian parameter secara simultan
 - c. Pengujian parameter secara parsial
6. Menyusun model regresi logistik ordinal menggunakan teknik oversampling dengan metode stepwise sebagai berikut:
 - a. Pemodelan regresi logistik ordinal
 - b. Pengujian parameter secara simultan
 - c. Pengujian parameter secara parsial
7. Menentukan akurasi model
8. Menghitung oddsratio

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksplorasi Terhadap Data Awal

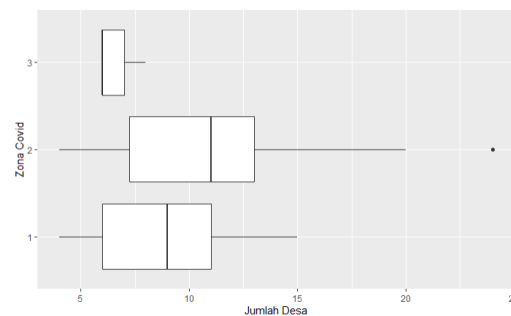
Dari data sampel 42 kecamatan di kabupaten Garut terdapat 7.1% atau 3 kecamatan termasuk pada risiko zona kuning, 81% atau 34 kecamatan termasuk pada risiko zona oranye, dan sisanya sebesar 11.9% atau 5 kecamatan termasuk pada risiko zona merah (Gambar 1). Median jumlah desa yang termasuk pada risiko zona oranye lebih besar dibandingkan median risiko zona lainnya (Gambar 2). Median persentase kepadatan penduduk yang termasuk pada risiko zona oranye lebih besar dibandingkan median risiko zona lainnya (Gambar

3). Median kasus suspek yang termasuk pada risiko zona oranye lebih besar dibandingkan median risiko zona lainnya (Gambar 4). Median kasus konfirmasi positif yang termasuk pada risiko zona oranye lebih besar dibandingkan median risiko zona lainnya (Gambar 5). Median kasus kontak erat yang termasuk pada risiko zona rendah lebih besar dibandingkan median risiko zona lainnya (Gambar 6).



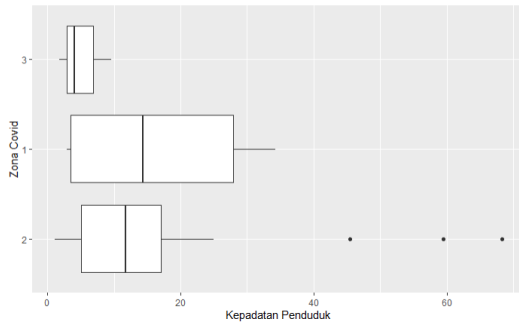
Gambar 1. Risiko sebaran zona covid-19

Pada gambar 1 diatas dapat dilihat bahwa dari 42 kecamatan yang ada di kabupaten Garut terdapat 34 kecamatan (81%) masuk pada zona oranye (tingkat keparahan sedang) terdiri dari kecamatan Wanaraja, Tarogong Kidul, Tarogong Kaler, Talegong, Sukawning, Sukaresmi, Sucinaraja, Singajaya, Selaawi, Samarang, PAsirwangi, Pangatikan, Pamulihan, Pakenjeng, Mekarmukti, Malangbong, Limbangan, Leuwigoong, Leles, Karangpawitan, Kadungora, Garut Kota, Cisurupan, Cilawu, Cikelet, Cikajang, Cihurip, Cigedug, Cibiuk, Cibat, Cibalong, Bungbulang, Bayongbong, dan Banjarwangi. Sedangkan terdapat 3 kecamatan (7.1%) masuk pada zona kuning (zona keparahan rendah) terdiri dari kecamatan Peundeuy, Pameungpeuk, dan Caringin. Kemudian sisanya sebanyak 5 kecamatan (11.9%) masuk pada zona merah (tingkat keparahan tinggi) terdiri dari kecamatan Kersamanah, Karangtengah, Cisompet, Cisewu, dan Banyuresmi.



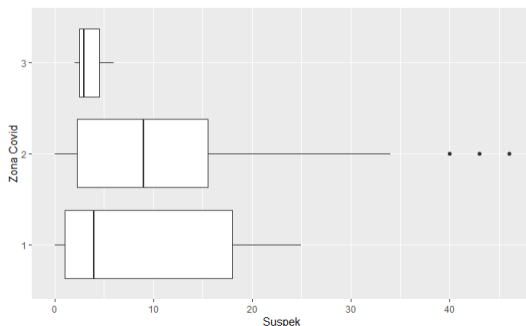
Gambar 2. Jumlah desa menurut risiko sebaran zona covid-19

Berdasarkan gambar 2 diatas, dapat disimpulkan bahwa nilai tengah jumlah desa dari zona kuning, zona oranye, dan zona merah berbeda. Daerah dengan zona oranye memiliki nilai tengah jumlah desa yang paling besar dibandingkan dengan zona lainnya.



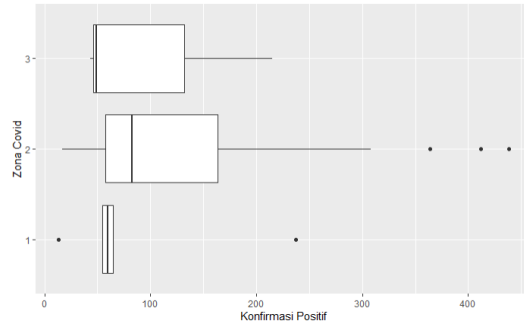
Gambar 3. Persentase laju pertumbuhan penduduk menurut risiko sebaran zona covid-19

Berdasarkan gambar 3 diatas, dapat disimpulkan bahwa nilai tengah kepadatan penduduk dari zona kuning, zona oranye, dan zona merah berbeda. Daerah dengan zona oranye memiliki nilai tengah kepadatan penduduk yang paling besar dibandingkan dengan zona lainnya.



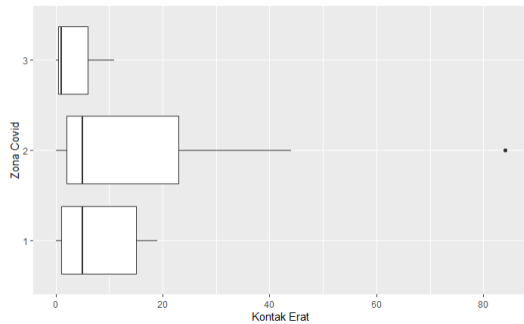
Gambar 4. Kasus suspek menurut risiko sebaran zona covid-19

Berdasarkan gambar 4 diatas, dapat disimpulkan bahwa nilai tengah kasus suspek dari zona kuning, zona oranye, dan zona merah berbeda. Daerah dengan zona oranye memiliki nilai tengah kasus suspek yang paling besar dibandingkan dengan zona lainnya.



Gambar 5. Kasus konfirmasi positif menurut risiko sebaran zona covid-19

Berdasarkan gambar 5 diatas, dapat disimpulkan bahwa nilai tengah kasus konfirmasi positif dari zona kuning, zona oranye, dan zona merah berbeda. Daerah dengan zona oranye memiliki nilai tengah kasus konfirmasi positif yang paling besar dibandingkan dengan zona lainnya.

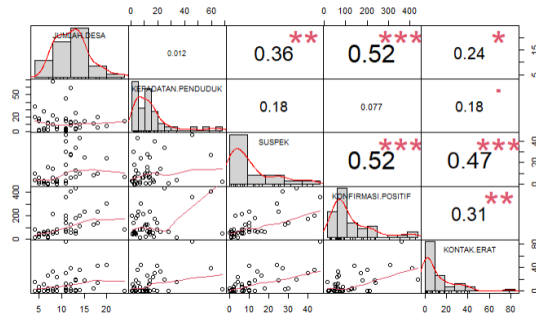


Gambar 6. Kasus kontak erat menurut risiko sebaran zona covid-19

Berdasarkan gambar 6 diatas, dapat disimpulkan bahwa nilai tengah kasus kontak erat dari zona kuning, zona oranye, dan zona merah berbeda. Daerah yang memiliki zona oranye dan zona merah memiliki nilai tengah kasus kontak erat yang paling besar dibandingkan dengan zona lainnya.

3.2 Korelasi Antar Variabel Bebas

Berikut adalah gambar dari korelasi antar variabel bebas.



Gambar 7. Korelasi antar variabel bebas

Berdasarkan gambar 7 diatas ditunjukkan bahwa hasil korelasi antar variabel bebas tidak ada yang lebih besar dari 0.8, maka dapat disimpulkan bahwa tidak terdapat korelasi antar variabel bebas.

3.3 Membagi Data Menjadi 80% Data Latih dan 20% Data Uji

Terdapat 33 kecamatan yang menjadi data latih dan 9 kecamatan menjadi data uji. Dari 33 kecamatan yang menjadi data latih, 3 kecamatan dikategorikan sebagai kecamatan dengan risiko zona tinggi (merah), 28 kecamatan dikategorikan sebagai kecamatan dengan risiko zona sedang (oranye), dan 2 kecamatan dikategorikan sebagai kecamatan dengan risiko zona rendah (kuning).

3.4 Proses Oversampling

Jumlah data pada masing-masing kategori rendah, sedang, dan tinggi tidak sama banyak atau tidak seimbang, sehingga dilakukan resampling menggunakan metode oversampling dengan cara menambah data secara acak pada kelas kategori yang memiliki data paling sedikit atau minoritas, yaitu menambah data secara acak pada risiko zona merah sebanyak 25 data dan pada risiko zona kuning sebanyak 26 data, sehingga setelah dilakukan oversampling maka jumlah data yang dimiliki yaitu 84 data.

3.5 Regresi Logistik Ordinal Menggunakan Teknik Oversampling

1) Pemodelan Regresi Logistik Ordinal

Tabel 1 menyajikan hasil pemodelan regresi logistik ordinal dengan teknik oversampling.

Tabel 1. Model Regresi Logistik Ordinal

Variabel	Koefisien regresi	Standard error
Tinggi Sedang	-4.5116	0.9825
Sedang Rendah	-1.7076	0.7400
Jumlah Desa	-0.2571	0.0892
Kepadatan Penduduk	-0.0878	0.0308

Suspek	-0.1880	0.0455
Konfirmasi Positif	0.0205	0.0043
Kontak Erat	0.0223	0.032

Pada tabel 1 diatas diperoleh dua persamaan regresi logistik ordinal, karena terdapat tiga kategori pada variabel respon sehingga didapatkan model persamaan sebanyak dua model sebagai berikut:

$$\log \text{it}[P(Y \leq 1|X)] = -4.5116 - 0.2571 * \text{JumlahDesa} - 0.0878 * \text{KepadatanPenduduk} - 0.1880 * \text{Suspek} + 0.0205 * \text{KonfirmasiPositif} + 0.0223 * \text{KontakErat} \quad (1)$$

$$\log \text{it}[P(Y \leq 3|X)] = -1.7076 - 0.2571 * \text{JumlahDesa} - 0.0878 * \text{KepadatanPenduduk} - 0.1880 * \text{Suspek} + 0.0205 * \text{KonfirmasiPositif} + 0.0223 * \text{KontakErat} \quad (2)$$

2) Uji Parameter Secara Simultan

Pengujian parameter secara serentak atau simultan digunakan untuk mengetahui apakah variabel bebas berpengaruh signifikan terhadap variabel respon. Hasil uji menggunakan Likelihood Ratio Test dan didapatkan $X^2 = 59.823$ dengan nilai p-value = $1.322e-11$. Karena nilai p-value < 0.05 sehingga dapat disimpulkan bahwa H_0 ditolak, artinya minimal ada satu variabel bebas yang memiliki pengaruh terhadap variabel respon.

3) Uji Parameter Secara Parsial

Tabel 2 menyajikan hasil pengujian parameter secara parsial menggunakan uji Wald.

Tabel 2. Hasil uji parameter secara parsial

Variabel	Nilai wald	p-value
Tinggi Sedang	-4.5116	4.39e-06
Sedang Rendah	-1.7076	2.10e-02
Jumlah Desa	-0.2571	3.97e-03
Kepadatan Penduduk	-0.0878	4.42e-03
Suspek	-0.1880	3.67e-05
Konfirmasi Positif	0.0205	2.27e-06
Kontak Erat	0.0223	4.89e-01

Berdasarkan tabel 2 diatas menunjukkan bahwa variabel jumlah desa, kepadatan penduduk, suspek, dan konfirmasi positif berpengaruh terhadap risiko sebaran zona covid di kabupaten Garut Juli tahun 2021. Hal tersebut dapat dilihat pada nilai p-value yang kurang atau lebih kecil dari $\alpha = 0.05$. Pada variabel kontak erat tidak berpengaruh signifikan terhadap variabel risiko sebaran zona covid-19, hal ini ditunjukkan karena nilai p-value lebih besar dari $\alpha = 0.05$.

3.6 Regresi Logistik Ordinal Menggunakan Teknik Oversampling Disertai Stepwise

1) Pemodelan Regresi Logistik Ordinal

Tabel 3 menyajikan hasil pemodelan regresi logistik ordinal dengan teknik oversampling yang disertai metode stepwise.

Tabel 3. Model Regresi Logistik Ordinal

Variabel	Koefisien regresi	Standard error
Tinggi Sedang	-4.3489	0.9275
Sedang Rendah	-1.5719	0.6975
Jumlah Desa	-0.2472	0.0866
Kepadatan Penduduk	-0.0755	0.0232
Suspek	-0.1759	0.0419
Konfirmasi Positif	0.0203	0.0043

Pada tabel 3 diatas diperoleh dua persamaan regresi logistik ordinal menggunakan teknik oversampling yang disertai dengan metode stepwise, karena terdapat tiga kategori pada variabel respon sehingga didapatkan model persamaan sebanyak dua model sebagai berikut:

$$\log \text{it}[P(Y \leq 1|X)] = -4.3489 - 0.2472 * \text{JumlahDesa} - 0.0755 * \text{KepadatanPenduduk} - 0.1759 * \text{Suspek} + 0.0235 * \text{KonfirmasiPositif} \quad (3)$$

$$\log \text{it}[P(Y \leq 3|X)] = -1.5719 - 0.2472 * \text{JumlahDesa} - 0.0755 * \text{KepadatanPenduduk} - 0.1759 * \text{Suspek} + 0.0235 * \text{KonfirmasiPositif} \quad (4)$$

Hasil estimasi $\alpha_1 = -4.3489$ lebih kecil dari $\alpha_3 = -1.5719$ berarti model ini adalah model kumulatif dengan kemiringan yang sama yaitu model garis regresi yang berdasarkan pada peluang kumulatif kategori respon (Albana 2013). Hal tersebut menunjukkan bahwa peluang kumulatif risiko tinggi lebih kecil daripada peluang kumulatif risiko rendah, berarti setiap kenaikan risiko Covid-19 akan menurunkan peluang logit kumulatif.

2) Uji Parameter Secara Simultan

Uji parameter secara simultan digunakan untuk mengetahui apakah variabel bebas berpengaruh terhadap variabel respon. Hasil uji dengan menggunakan Likelihood Ratio Test dan didapatkan $X^2 = 59.356$ dengan nilai p-value = 3.96e-12. Karena nilai p-value < 0.05 maka dapat disimpulkan bahwa H_0 ditolak, artinya minimal ada satu variabel bebas yang memiliki pengaruh terhadap variabel respon.

3) Uji Parameter Secara Parsial

Tabel 4 menyajikan hasil pengujian parameter secara parsial menggunakan uji Wald.

Tabel 4. Hasil uji parameter secara parsial

Variabel	Nilai wald	p-value
Tinggi Sedang	-4.3489	2.75e-06
Sedang Rendah	-1.5719	2.42e-02
Jumlah Desa	-0.2472	4.31e-03
Kepadatan Penduduk	-0.0755	1.15e-03
Suspek	-0.1759	2.77e-05
Konfirmasi Positif	0.0203	2.41e-06

Berdasarkan tabel 4 diatas menunjukkan bahwa variabel jumlah desa, kepadatan penduduk, suspek, dan konfirmasi positif berpengaruh terhadap risiko sebaran zona covid di kabupaten Garut Juli tahun 2021. Hal ini dapat dilihat pada nilai p-value yang kurang atau lebih kecil dari $\alpha = 0.05$.

3.7 Akurasi Model

Model akurasi digunakan untuk menilai seberapa besar keakurasian model yang dihasilkan. Nilai akurasi didapatkan dari perhitungan *confussion matrix* antara data hasil prediksi dengan data yang sebenarnya pada data uji, kemudian dilakukan pengururan ketepatan antara data prediksi dan data sebenarnya yang dapat menghasilkan tingkat ketepatan dari *confussion matrix* tersebut. Baik model dengan menggunakan teknik oversampling yang disertai stepwise maupun tidak memiliki nilai akurasi yang sama besar yaitu 85,71%. Artinya model regresi logistik ordinal yang didapatkan baik menggunakan metode stepwise ataupun tidak itu baik karena memiliki nilai akurasi yang tinggi.

3.8 Interpretasi Odds Ratio

Interpretasi terhadap model dapat dilihat berdasarkan nilai odds ratio. Nilai odds ratio didapatkan dari nilai koefisien variabel bebas yang di eksponensialkan. Adapun hasil dari nilai odds ratio sebagai berikut:

Tabel 5. Hasil odds ratio

Variabel	Koefisien regresi	Odds ratio
Tinggi Sedang	-4.3489	0.0129
Sedang Rendah	-1.5719	0.2076
Jumlah Desa	-0.2472	0.7809
Kepadatan Penduduk	-0.0755	0.9271
Suspek	-0.1759	0.8386
Konfirmasi Positif	0.0203	1.0205

Interpretasi odds ratio dari output diatas adalah sebagai berikut:

1) Interpretasi Model Risiko Zona Tinggi

- $\text{Exp}(\text{intercept risiko zona tinggi}) = 0.01292 \rightarrow$ tanpa memperhatikan peubah bebasnya, peluang sebaran covid zona tinggi (merah) adalah 0.01292 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya.

- $\text{Exp}(\text{jumlah desa}) = 0.780917 \rightarrow$ Odds kecamatan dengan kategori zona tinggi (merah) akan meningkat sebesar 0.780917 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika jumlah desa naik sebesar satu desa dengan peubah bebas lainnya konstan.

- $\text{Exp}(\text{kepadatan penduduk}) = 0.92719 \rightarrow$ Odds kecamatan dengan kategori zona tinggi (merah) akan meningkat sebesar 0.92719 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika kepadatan penduduk naik sebesar satu desa dengan peubah bebas lainnya konstan.

- $\text{Exp}(\text{suspek}) = 0.83868 \rightarrow$ Odds kecamatan dengan kategori zona tinggi (merah) akan meningkat sebesar 0.83868 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika jumlah kasus suspek naik sebesar satu kasus dengan peubah bebas lainnya konstan.

- $\text{Exp}(\text{konfirmasi positif}) = 1.02058 \rightarrow$ Odds kecamatan dengan kategori zona tinggi (merah) akan meningkat sebesar 1.02058 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika jumlah konfirmasi positif naik sebesar satu kasus dengan peubah bebas lainnya konstan.

2) Interpretasi Model Risiko Zona Rendah

- $\text{Exp}(\text{intercept risiko zona rendah}) = 0.20764 \rightarrow$ tanpa memperhatikan peubah bebasnya, peluang sebaran covid zona rendah (kuning) adalah 0.20764 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya.

- $\text{Exp}(\text{jumlah desa}) = 0.780917 \rightarrow$ Odds kecamatan dengan kategori zona rendah (kuning) akan meningkat sebesar 0.780917 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika jumlah desa naik sebesar satu desa dengan peubah bebas lainnya konstan.

- $\text{Exp}(\text{kepadatan penduduk}) = 0.92719 \rightarrow$ Odds kecamatan dengan kategori zona rendah (kuning) akan meningkat sebesar 0.92719 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika kepadatan penduduk naik sebesar satu desa dengan peubah bebas lainnya konstan.

- $\text{Exp}(\text{suspek}) = 0.83868 \rightarrow$ Odds kecamatan dengan kategori zona rendah (kuning) akan meningkat sebesar 0.83868 kali dibandingkan

peluang untuk masuk pada kategori zona covid lainnya jika jumlah kasus suspek naik sebesar satu kasus dengan peubah bebas lainnya konstan.

- $\text{Exp}(\text{konfirmasi positif}) = 1.02058 \rightarrow$ Odds kecamatan dengan kategori zona rendah (kuning) akan meningkat sebesar 1.02058 kali dibandingkan peluang untuk masuk pada kategori zona covid lainnya jika jumlah konfirmasi positif naik sebesar satu kasus dengan peubah bebas lainnya konstan.

4. KESIMPULAN DAN SARAN

Hasil penelitian dalam mengidentifikasi faktor yang memengaruhi risiko zona covid-19 menggunakan regresi logistik ordinal dengan teknik oversampling diperoleh kesimpulan yaitu:

1. Kecamatan dengan risiko zona sedang memiliki proporsi yang lebih tinggi dibandingkan dengan risiko zona tinggi dan rendah yaitu sebanyak 33 kecamatan termasuk pada risiko zona sedang, 5 kecamatan termasuk pada daerah dengan risiko zona tinggi, dan 3 kecamatan termasuk pada daerah dengan risiko zona rendah.

2. Terdapat dua model regresi logistik ordinal menggunakan teknik oversampling, yaitu model untuk kecamatan dengan risiko zona tinggi dan kecamatan dengan risiko zona rendah. Bentuk umum model dapat dituliskan:

$$\log \text{it}[P(Y \leq 1|X)] = -4.3489 - 0.2472 * \text{JumlahDesa} - 0.0755 * \text{KepadatanPenduduk} - 0.1759 * \text{Suspek} + 0.0235 * \text{KonfirmasiPositif}$$

$$\log \text{it}[P(Y \leq 3|X)] = -1.5719 - 0.2472 * \text{JumlahDesa} - 0.0755 * \text{KepadatanPenduduk} - 0.1759 * \text{Suspek} + 0.0235 * \text{KonfirmasiPositif}$$

Hasil estimasi $\alpha_1 = -4.3489$ lebih kecil dari $\alpha_3 = -1.5719$ berarti model ini adalah model kumulatif dengan kemiringan yang sama yaitu model garis regresi yang berdasarkan pada peluang kumulatif kategori respon (Albana 2013). Hal tersebut menunjukkan bahwa peluang kumulatif risiko tinggi lebih kecil daripada peluang kumulatif risiko rendah, berarti setiap kenaikan risiko Covid-19 akan menurunkan peluang logit kumulatif.

3. Variabel prediktor yang memengaruhi terhadap sebaran zona covid-19 di Kabupaten Garut periode Juli 2021 adalah jumlah desa, kepadatan penduduk, kasus suspek, dan konfirmasi positif dengan nilai akurasi sebesar 85.71%.

DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis*. Inc. New York.
- Ananth, Cande V. (1997). *Regression Models for Ordinal Responses: A Review of Methods and Applications*. *International Journal of Epidemiology*, 2-6.
- Arofah, Irvana. (2018). Analisis Persepsi Biaya Kuliah Perguruan Tinggi dengan Menggunakan Metode Regresi Logistik Ordinal. *Jurnal Statistika dan Aplikasinya*, 1-6.
- B. R. A. Febrilia, S. Rahayu, and B. D. Korida. (2019). "Ordinal Logistic Regression Analysis of Factors Affecting the Length of Student Study". *J. Mat. "MANTIK"*, 1-7.
- Choi, Seung W. (2011). *lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations*. *J Stat Softw*, 2-3.
- Hosmer, D.W. and Lemeshow, S. (2000) *Applied logistik regression*. 2nd Edition. A John Wiley and Sons, Inc. New York.
- Liang, Jiaqi. (2020). *Multinomial and ordinal Logistic regression analyses with multi-categorical variables using*. *Journal of Annals Of Translational Medicine*, 1-7.
- Liu, Xing. (2012). *Ordinal Regression Analysis: Using Generalized Ordinal Logistic Regression Models to Estimate Educational Data*. *Journal Of Modern Applied Statistical Methods*. 2-6.
- Mugirahayu, A. S. (2021). *Penentuan Status Kewaspadaan COVID-19 Pada Suatu Wilayah Menggunakan Metode Fuzzy Inference System (FIS) Mamdani*. *Jurnal Sains dan Edukasi Sains*, 1-2.
- Nugraha, Jaka. (2017). *Pemodelan Data Ordinal, Nominal dan Cacah*. Universitas Islam Indonesia.
- P. Thomas, N.A. Salmon, W. Riana. (2016). *Analisis Regresi Logistik Ordinal*. *Barekeng Jurnal Matematika dan Ilmu Terapan*, 1-6.
- Pramesti, W. (2021). *Regresi Logistik Backward Elimination pada Risiko Penyebaran Covid-19 di Jawa Timur*. *Journal of Statistics and Its Applivation on Teaching and Research*, 1-8.
- S. Imaslihkah, M. Ratna, and V. Ratnasari. (2013). "Analisis Regresi Logistik Ordinal terhadap Faktor-faktor yang mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya". *J. Sains dan Seni Pomits*, 1-6.
- Shen, C., Bar-yam, Y., & New England Complex Systems Institute. (2020). *Color Zone Pandemic Response Version 2*. 2020. Retrieved from <https://necsi.edu/color-zone-pandemic-response-version-2>, diakses tanggal 4 April 2022.

Perbandingan Analisis Dekomposisi dan *Exponential Smoothing Holt Winters* untuk Peramalan Rata-Rata Jumlah KPM PKH di NTB

M. Hadiyan Amaly⁽¹⁾, Wiwit Pura Nurmayanti⁽²⁾, Sausan Nisrina⁽³⁾

Program Studi Statistika, Fakultas MIPA, Universitas Hamzanwadi

Jl. TGKH M. Zainuddin Abdul Madjid No. 132 Pancor (83611) Selong-Lombok Timur-NTB

e-mail: amalyhadiyan17@gmail.com, wiwit.adiwinata3@gmail.com, dan

sausannisrina96@gmail.com

ABSTRAK

Program Keluarga Harapan (PKH) adalah program pemberian bantuan sosial bersyarat kepada keluarga miskin dan rentan terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) dan ditetapkan sebagai Keluarga Penerima Manfaat (KPM). Tujuan penelitian ini untuk meramalkan rata-rata jumlah PKM PKH untuk periode 2021. Hal ini dapat dijadikan dasar bagi instansi pemerintah terkait dalam perencanaan kedepannya. Untuk mendapatkan peramalan terbaik digunakan dua metode yaitu metode dekomposisi dan metode *Exponential Smoothing Holt Winters*. Analisis menggunakan metode dekomposisi meliputi komponen siklis, musiman, *trend*, dan acak kemudian mengalikan nilai dari komponen-komponen tersebut. Sedangkan metode *Exponential Smoothing Holt Winters* merupakan metode yang digunakan untuk meramalkan melalui proses pemulusan (*smoothing*) yang memiliki unsur *trend* dan musiman sekaligus secara bersamaan. Untuk mencapai tujuan penelitian tersebut dilakukan perhitungan kesalahan yang terdiri dari nilai MAPE, MAD, dan MSD. Berdasarkan hasil analisis, diperoleh metode terbaik untuk meramalkan rata-rata jumlah KPM PKH di provinsi NTB adalah *Exponential Smoothing Holt Winters*, karena memiliki nilai *error* lebih kecil dibandingkan dengan metode dekomposisi. Metode *Exponential Smoothing Holt Winters* menggunakan nilai pemulusan $\alpha = 0,3$, $\beta = 0,1$, dan $\gamma = 0,2$ menghasilkan nilai MAPE = 15, MAD = 2761, dan MSD = 1692444.

Kata kunci : Program Keluarga Harapan, Dekomposisi, *Exponential Smoothing Holt Winters*, Peramalan.

ABSTRACT

The Family Hope Program (PKH) is a program to provide conditional social assistance to poor and vulnerable families registered in the Social Welfare Integrated Data (DTKS) and designated as a Beneficiary Family (KPM). The purpose of this study is to forecast the average number of PKH PKM for the 2021 period. This can be used as a basis for relevant government agencies in future planning. To get the best forecasting used a comparison of two methods, namely the decomposition method and the *Exponential Smoothing Holt Winters* method. To achieve the purpose of the study, an error calculation consisted of Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Square Deviation (MSD). Based on the results of data analysis, it is known that the best forecasting method to forecast the average number of KPM PKH in NTB province is the *Exponential Smoothing Holt Winters* method. This method is said to be better because it has a smaller error value compared to the error value of the decomposition method. *Holt Winters' Exponential Smoothing* method uses the smoothing values $\alpha = 0.3$, $\beta = 0.1$, and $\gamma = 0.2$ produces the values MAPE = 15, MAD = 2761, and MSD = 1692444.

Keywords : Family Hope Program, Decomposition, *Exponential Smoothing Holt Winters*, Forecasting.

1. PENDAHULUAN

Permasalahan kependudukan merupakan permasalahan yang dihadapi oleh setiap negara. Berdasarkan Sensus Penduduk 2020 jumlah penduduk Indonesia mencapai 270,20 juta jiwa.

Penduduk Nusa Tenggara Barat berdasarkan Sensus Penduduk 2020 mencapai 5,32 juta jiwa. Jumlah penduduk yang terus meningkat tidak terlepas dari isu-isu kependudukan dan permasalahan sosial, salah satunya adalah kemiskinan. Kemiskinan dapat

dikatakan sebagai permasalahan yang kompleks karena faktor penyebabnya dapat muncul dari internal maupun eksternal. Faktor internal dapat berasal dari diri sendiri, sedangkan untuk faktor eksternal seperti dari lingkungan, pendidikan, keluarga, masyarakat dan lain-lain. Adapun faktor lainnya seperti global yang rendah dan pertumbuhan ekonomi lokal.

Menurut Binti (2017) kemiskinan adalah suatu kondisi terjadi kekurangan atau keterbatasan sesuatu yang umumnya dimiliki seperti sandang, pangan dan papan, hal ini berhubungan erat dengan kualitas hidup. Penyebab kemiskinan diantaranya seperti kelangkaan alat pemenuhan kebutuhan dasar, atau sulitnya akses terhadap pendidikan dan kesehatan. Sebanyak 53,3 juta jiwa atau 20,19% penduduk yang masuk kategori hampir/rentan miskin. Angka ini menggambarkan nominal yang cukup mengkhawatirkan karena jumlah penduduk Indonesia hampir sepertiga masuk kategori miskin dan rentan miskin (Septiadi & Nursan, 2020).

Upaya pemerintah dalam menekan tingginya angka kemiskinan yaitu dengan meningkatkan efektivitas penanggulangan melalui program pengentasan kemiskinan seperti diberlakukannya program berbasis perlindungan sosial salah satunya adalah Program Keluarga Harapan (PKH). Nazaruddin P (2020) menjelaskan Program Keluarga Harapan (PKH) adalah program pemberian bantuan sosial (bansos) bersyarat kepada keluarga miskin dan rentan terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) dan ditetapkan sebagai Keluarga Penerima Manfaat (KPM). Oleh karena itu perlu dilakukan ramalan untuk memperkirakan jumlah penerima PKH guna mendukung upaya pemerintah dalam menurunkan tingkat kemiskinan dan penyaluran dapat tepat sasaran.

Peramalan merupakan salah satu cara untuk memprediksi suatu peristiwa dimasa yang akan datang dengan mempertimbangkan data yang diperoleh pada masa lampau (Yusuf & Anjasari, 2018). Berbagai model peramalan *time series* telah dilakukan oleh peneliti untuk memprediksi sesuatu yang akan terjadi kedepannya, banyak metode yang digunakan dalam pemodelan runtun waktu untuk peramalan diantaranya metode dekomposisi dan metode *Exponential Smoothing Holt Winters*.

Metode dekomposisi merupakan metode yang meliputi komponen siklis, musiman, *trend*, dan acak kemudian mengalikan nilai dari komponen-komponen tersebut. Keunggulan dari metode ini adalah komponen-komponennya dapat dipecah atau didekomposisikan menjadi sub pola yang menunjukkan tiap-tiap komponen deret berkala

terpisah yang mampu meningkatkan ketepatan dalam melakukan peramalan (Yuni et al., 2015).

Sedangkan metode *Exponential Smoothing Holt Winters* digunakan untuk memodelkan data runtun waktu yang memiliki pola musiman, baik mengandung *trend* atau tidak (Jatmiko & Mantik, 2017). Kelebihan dari metode ini sangat baik dalam meramalkan pola data yang memiliki unsur musiman dengan unsur *trend* yang muncul secara bersamaan, serta mudah dalam praktiknya (Safitri et al., 2017).

Penelitian sebelumnya yang dilakukan oleh Yuni (Yuni et al., 2015) yaitu meramalkan jumlah pengunjung perpustakaan Universitas Patimura Ambon menggunakan metode dekomposisi. Kemudian penelitian oleh Safitri (Safitri et al., 2017) yang membandingkan metode *Exponential Smoothing Holt-Winters* dan ARIMA untuk meramalkan jumlah kunjungan wisatawan mancanegara di bandara Ngurah Rai Bali. Berdasarkan latar belakang yang telah dibahas sebelumnya, maka peneliti tertarik untuk membandingkan metode dekomposisi dan metode *Exponential Smoothing Holt Winters* pada data rata-rata jumlah Keluarga Penerima Manfaat (KPM) PKH di NTB.

2. METODE PENELITIAN

2.1 Metode Dekomposisi

Metode dekomposisi adalah mendekomposisikan data deret berkala menjadi pola, yaitu *trend*, siklis dan musiman, dan mengidentifikasi masing-masing komponen tersebut secara terpisah. Hal ini dilakukan untuk meningkatkan kesesuaian peramalan dan membantu pemahaman pola data lebih baik (Yuni et al., 2015). Metode dekomposisi merupakan suatu metode yang menggunakan empat komponen utama dalam meramalkan nilai masa depan. Keempat komponen tersebut antara lain *trend*, musiman, siklus, dan *error* (Makkulau, 2019). Metode ini didasarkan oleh asumsi bahwa data yang digunakan merupakan gabungan dari beberapa komponen, secara sederhana digambarkan sebagai berikut (Kendek, 2014):

$$\begin{aligned} \text{Data} &= \text{Pola} + \text{error} \\ &= f(\text{trend, siklus, musiman}) + \text{error} \end{aligned} \quad (1)$$

Adapun persamaan secara matematis dari pendekatan dekomposisi adalah:

$$X_t = f(T_t, S_t, C_t, I_t) \quad (2)$$

dimana:

X_t = nilai deret berkala (data aktual) pada periode t

T_t = komponen *trend* pada periode t
 S_t = komponen musiman (*seasonal*) pada periode t
 C_t = komponen siklus (*cyclic*) pada periode t
 I_t = komponen kesalahan tidak beraturan (*irregular*) pada periode t
 t = periode (*time*)

2.2 Exponential Smoothing Holt-Winters

Metode *Exponential Smoothing Holt Winters* merupakan metode yang digunakan meramalkan data runtun waktu yang mengalami unsur *trend* dan musiman kecenderungan yang muncul secara sekaligus pada sebuah data deret waktu (Anjasari et al., 2018). Metode *Holt-Winters* didasarkan atas tiga persamaan pemulusan, yaitu persamaan pemulusan keseluruhan, pemulusan komponen *trend* dan pemulusan komponen musiman. Berdasarkan model musiman terdapat dua model dalam metode *Holt-Winters*, yaitu model multiplikatif dan *aditif*. Model musiman multiplikatif digunakan untuk data yang memuat unsur *trend* serta musiman, artinya model ini digunakan jika data menunjukkan adanya kenaikan serta fluktuasi musiman yang semakin meningkat seiring bertambahnya periode waktu pengamatan (Jatmiko & Mantik, 2017). Sama halnya dengan model musiman *aditif* namun model ini digunakan apabila suatu data menunjukkan adanya perkembangan kenaikan serta fluktuasi musiman yang relatif konstan seiring bertambahnya periode waktu pengamatan (Hamidah, 2017). Rumus yang digunakan untuk menghitung hasil peramalan adalah (Jatmiko & Mantik, 2017):

Pemulusan eksponensial:

$$L_t = \alpha(y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (3)$$

Estimasi *trend*:

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (4)$$

Estimasi musiman:

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s} \quad (5)$$

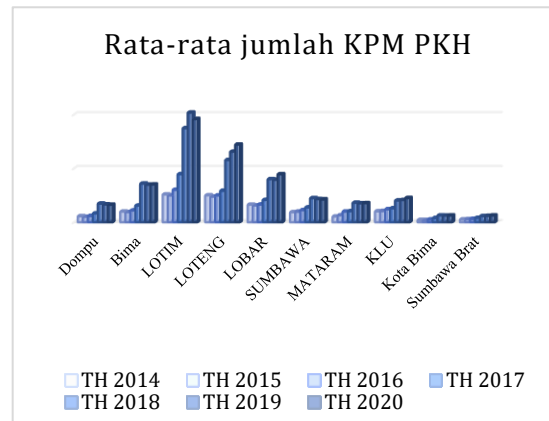
Keterangan :

- L_t = nilai pemulusan eksponensial periode t
- α = konstanta pemulusan data ($0 < \alpha < 1$)
- β = konstanta pemulusan estimasi *trend* ($0 < \beta < 1$)
- μ = konstanta pemulusan untuk estimasi musiman ($0 < \mu < 1$)
- L_s = nilai pemulusan eksponensial periode awal
- y_t = nilai aktual pada periode t
- b_t = estimasi *trend* periode t
- S_t = estimasi musiman periode t
- t = periode ke- t
- s = panjang musim

2.3 Langkah Analisis

Data yang digunakan dalam penelitian ini yaitu data Rekapitulasi Jumlah KPM PKH Provinsi NTB dari tahun 2014 sampai dengan tahun 2020 yang diperoleh dari bidang Perlindungan Jaminan Sosia (LINJAMSOS), Kantor Dinas Sosial Provinsi NTB yang berada di Kota Mataram, Nusa Tenggara Barat. Langkah-langkah dalam menganalisis data menggunakan perbandingan metode dekomposisi dan *exponential smoothing holt winters* sebagai berikut:

1. Melakukan analisis deskriptif atau *explorasi* data rata-rata jumlah KPM PKH provinsi NTB tahun 2014-2020 untuk mengetahui gambaran umum dari data.
2. Melihat plot pola data *time series* apakah sudah membentuk pola musiman atau tidak.
3. Selanjutnya yaitu melakukan analisis peramalan menggunakan perbandingan metode dekomposisi dan *Exponential Smoothing Holt Winters* dengan *software Minitab 18*.
4. Menentukan metode terbaik dengan melihat nilai *error* terkecil pada setiap metode dan hasil peramalan untuk periode kedepan.



Gambar 1. Grafik rata-rata jumlah PKM PKH

3 HASIL DAN PEMBAHASAN

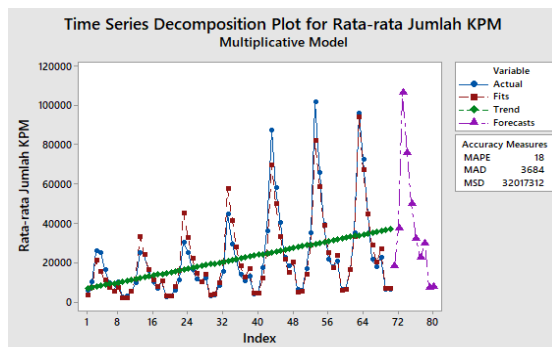
3.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk melihat gambaran karakteristik dari suatu data untuk masing-masing indikator atau variabel, agar data yang ada dapat mudah dipahami oleh pembacanya. Berdasarkan rincian Gambar 1 menunjukkan rata-rata jumlah KPM PKH terendah terdapat pada kabupaten Lombok Utara sebesar 1.958 KPM. Dan rata-rata jumlah KPM PKH tertinggi terdapat pada kabupaten Lombok Timur sebesar 101.704 KPM, hal ini sejalan dengan jumlah penduduk miskin

sebesar 183.840 jiwa. NTB merupakan salah satu provinsi yang termasuk kategori sepuluh besar penduduk miskin tertinggi di Indonesia (Hatuina, 2015). Tahun 2014 Kabupaten Lombok Timur berada pada zona merah (19%) yang artinya masih berada di atas angka kemiskinan nasional (10,96%). Oleh karena itu, kabupaten Lombok Timur perlu mendapat sorotan dalam hal program penanggulangan kemiskinan.

3.2 Dekomposisi

Setelah melakukan analisis deskriptif dengan menampilkan grafik berdasarkan kabupaten serta jumlah KPM PKH dari tahun 2014-2020 selanjutnya melakukan analisis dengan metode Dekomposisi untuk melihat hasil peramalan untuk periode kedepannya. Berikut dibawah ini adalah hasil yang didapatkan.



Gambar 2. Output Dekomposisi

Berdasarkan Gambar 2 menunjukkan *output* hasil peramalan rata-rata jumlah KPM PKH menggunakan metode dekomposisi. Secara deskriptif rata-rata jumlah KPM PKH bersifat musiman, dilihat dari garis warna biru menunjukkan data aktual, garis warna merah menunjukkan nilai estimasi, sedangkan garis warna ungu menunjukkan hasil peramalan. Dengan nilai *error* sebesar MAD = 3684, MAPE = 18 dan MSD = 32017312.

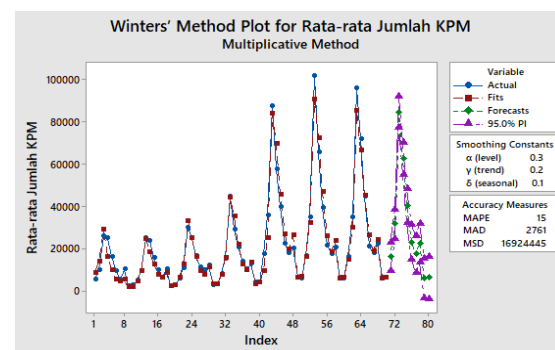
Tabel 1. Hasil Peramalan Metode Dekomposisi

Kabupaten/Kota	TH 2020	Forecast
Dompu	16144	18033
Bima	34959	37366
Lombok Timur	95771	106229
Lombok Tengah	72016	75394
Lombok Barat	44472	49741
Sumbawa	21156	32070
Mataram	17616	22161
Lombok Utara	22361	29676

Kota Bima	6168	7023
Sumbawa Barat	6206	7496

Berdasarkan Tabel 1 terlihat perbandingan data aktual dengan hasil peramalan menggunakan metode dekomposisi pada tahun 2021. Secara deskriptif rata-rata jumlah KPM PKH mengalami peningkatan, hal ini dikarenakan grafik yang berubah atau tidak konstan sehingga berpengaruh terhadap peramalan tahun 2021. Hasil peramalan tertinggi berada di kabupaten Lombok Timur, sedangkan terendah di Kota Bima dengan rata-rata jumlah KPM PKH sebesar 106229 dan 7023.

3.3 Eksponential Smoothing Holt-Winters



Gambar 3. Output Exponential Smoothing Holt-Winters dengan $\alpha = 0,3$, $\beta = 0,2$, dan $\gamma = 0,1$

Berdasarkan Gambar 3 dapat dilihat bahwa terdapat perbedaan pada masing-masing simbol yang terbentuk mengikuti grafik data aktual. Garis dengan simbol warna biru menunjukkan data aktualnya, garis warna merah menunjukkan nilai estimasi, sedangkan garis warna hijau menunjukkan hasil grafik peramalan. Terlihat pada grafik, bahwa hasil peramalan tidak memiliki perbedaan yang terlalu jauh dengan data aktual. Dengan melihat nilai *error* MAD = 2761, MAPE = 15 dan MSD = 1692444. Hal ini dapat dilihat pada perbandingan data aktual dengan hasil peramalan pada tabel berikut:

Tabel 2. Hasil Peramalan Exponential Smoothing Holt-Winters

Kabupaten/Kota	TH 2020	Forecast
Dompu	16144	15897.4
Bima	34959	31452.1
Lombok Timur	95771	84589.6
Lombok Tengah	72016	62381
Lombok Barat	44472	40168.5

Sumbawa	21156	22794.5
Mataram	17616	17334.3
Lombok Utara	22361	22422.1
Kota Bima	6168	5735.9
Sumbawa Barat	6206	6005.3

Berdasarkan Tabel 2 terlihat perbandingan data aktual dengan hasil peramalan dengan menggunakan metode *exponential smoothing holt-winters* pada tahun 2021. Secara deskriptif rata-rata jumlah KPM PKH mengalami penurunan, hal ini dikarenakan grafik yang berubah atau tidak konstan sehingga berpengaruh terhadap peramalan tahun 2021. Hasil peramalan tertinggi berada di kabupaten Lombok Timur, sedangkan terendah di Kota Bima dengan rata-rata jumlah KPM PKH sebesar 84589,6 dan 5735,9.

3.4 Perbandingan Metode Dekomposisi dan Exponential Smoothing Holt-Winters

Tabel 3. Perbandingan Hasil Peramalan

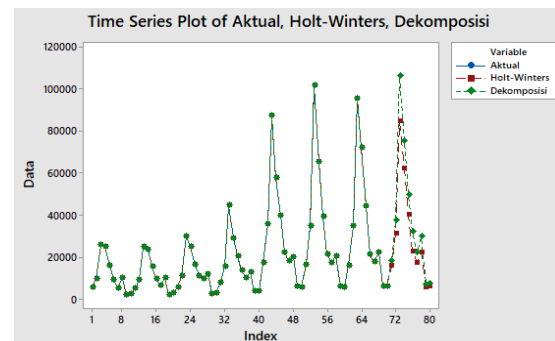
Kabupaten/Kota	Holt-Winters	Dekomposisi
Dompu	15897.4	18033
Bima	31452.1	37366
Lombok Timur	84589.6	106229
Lombok Tengah	62381	75394
Lombok Barat	40168.5	49741
Sumbawa	22794.5	32070
Mataram	17334.3	22161
Lombok Utara	22422.1	29676
Kota Bima	5735.9	7023
Sumbawa Barat	6005.3	7496

Berdasarkan hasil peramalan pada Tabel 3 diketahui bahwa hasil peramalan antara kedua metode tersebut memiliki selisih yang tidak terlalu jauh. Berikut adalah perbandingan nilai MAPE, MAD dan MSD antara metode dekomposisi dan metode *Exponential Smoothing Holt Winters*.

Tabel 4. Penentuan Metode Terbaik

	Holt-Winters	Dekomposisi
MAPE	15	18
MAD	2761	3684
MSD	1692444	32017312

Berdasarkan Tabel 4 dapat dilihat bahwa metode analisis terbaik untuk peramalan data rata-rata jumlah KPM PKH di Nusa Tenggara Barat tahun 2014-2020 adalah menggunakan metode *exponential smoothing holt-winters*. Hal ini dilihat dari nilai MAPE metode *Exponential Smoothing Holt-Winters* lebih kecil dibandingkan dengan nilai MAPE metode. Dekomposisi sehingga metode terbaik yaitu metode *exponential smoothing holt-winters* dengan nilai MAPE sebesar 15. Hasil visualisasi peramalan data rata-rata jumlah KPM PKH sebagai berikut:



Gambar 4. Perbandingan Data Aktual, Holt-Winters dan Dekomposisi

Berdasarkan plot pada Gambar 4 yang menunjukkan perbandingan antara data aktual dan data hasil ramalan dapat diketahui bahwa hasil ramalan sudah hampir mendekati data aslinya. Hal ini dilihat dari plot yang garis berwarna biru, merah, dan hijau yang saling berimpitan

4 KESIMPULAN DAN SARAN

Berdasarkan pembahasan di atas dapat dilihat bahwa metode analisis terbaik untuk peramalan data rata-rata jumlah KPM PKH di Nusa Tenggara Barat tahun 2014-2020 adalah menggunakan metode *eksponential smoothing holt-winters*. Hal ini dilihat dari nilai MAPE metode *eksponential smoothing holt-winters* lebih kecil dibandingkan dengan nilai MAPE metode Dekomposisi sehingga metode terbaik yaitu metode *eksponential smoothing holt-winters* dengan nilai MAPE = 15, MAD = 2761, dan MSD = 1692444. Nilai α , β , dan γ yang paling optimal dalam peramalan rata-rata jumlah KPM PKH menggunakan metode *exponential smoothing holt winters* yaitu $\alpha = 0,3$, $\beta = 0,2$, dan $\gamma = 0,1$.

DAFTAR PUSTAKA

Anjasari, D. H., Listiwikono, E., & Yusuf, F. I. (2018). Perbandingan Metode Double

- Exponential Smoothing Holt Dan Metode Triple Exponential Smoothing Holt-Winters Untuk Peramalan Wisatawan Grand Watu Dodol. *Transformasi: Jurnal Pendidikan Matematika Dan Matematika*, 2(2), 12–25.
- Binti, M. T. (2017). Analisa Pengaruh Pertumbuhan Ekonomi Terhadap Penurunan Tingkat Kemiskinan di Kalimantan Tengah. *AL-KALAM: JURNAL KOMUNIKASI, BISNIS DAN MANAJEMEN*, 3(2), 69-78.
- BPS (2020). *Press Release* Hasil Sensus Penduduk 2020. Dikutip tanggal 13 Mei 2021 melalui website. <https://www.bps.go.id/pressrelease/2021/01/21/1854/hasil-sensus-penduduk-2020.html>
- Hamidah, S. N., Salam, N., & Susanti, D. S. (2017). Teknik Peramalan Menggunakan Metode Pemulusan *Eksponensial Holt-Winters*. *JURNAL MATEMATIKA MURNI DAN TERAPAN EPSILON*, 7(2), 26-33.
- Jatmiko, Y. A., & Mantik, J. M. (2017). Perbandingan Keakuratan Hasil Peramalan Produksi Bawang Merah Metode Holt-Winters dengan Singular Spectrum Analy ... *JURNAL MATEMATIKA "MANTIK,"* 03(01), 13–23.
- Kendek, O. J., Prang, J. D., & Paendong, M. (2014). Prediksi jumlah pengunjung perpustakaan universitas sam ratulangi manado menggunakan metode dekomposisi. *d'CARTESIAN: Jurnal Matematika dan Aplikasi*, 3(1), 73-80.
- Makkulau, M., Raya, R., & Marlinda, S. (2019). Aplikasi Metode Dekomposisi Pada Peramalan Jumlah Kelahiran. In *Seminar Nasional Teknologi Terapan Berbasis Kearifan Lokal* (Vol. 1, No. 1).
- Nazarududin, P. 2020. Pedoman Pelaksanaan Program Keluarga Harapan Diakses pada 29 Februari 2021. <https://pkh.kemensos.go.id>.
- Safitri, T., Dwidayati, N., Safitri, T., Dwidayati, N., & Peramalan, S. P. (2017). PERBANDINGAN PERAMALAN MENGGUNAKAN METODE EXPONENTIAL SMOOTHING HOLT-WINTERS DAN ARIMA. *UNNES Journal of Mathematics*, 6(1), 48–58.
- Septiadi, D., & Nursan, M. (2020). Pengentasan Kemiskinan Indonesia: Analisis Indikator Makroekonomi Dan Kebijakan Pertanian. *Jurnal Hexagro*, 4(1), 1-14.
- Statistik, B. P. (2018). Profil Kemiskinan di Indonesia Maret 2018. *Jakarta (ID): Badan Pusat Statistik*.
- Yuni, S., Talakua, M. W., & Lesnussa, Y. A. (2015). PERAMALAN JUMLAH PENGUNJUNG PERPUSTAKAAN METODE DEKOMPOSISI. *Jurnal Ilmu Matematika Dan Terapan*, 9(1), 41–50.
- Yusuf, F. I., & Anjasari, D. H. (2018). METODE TRIPLE EXPONENTIAL SMOOTHING HOLT-WINTERS UNTUK PERAMALAN JUMLAH WISATAWAN NUSANTARA DI KABUPATEN BANYUWANGI. *Unisda Journal of Mathematics and Computer Science (UJMC)*, 4(2), 1–6.

Faktor – Faktor yang Memengaruhi Permasalahan *Stunting* di Jawa Barat Menggunakan Regresi Logistik Biner

Silmi Annisa Rizki Manaf⁽¹⁾, Erfiani⁽²⁾, Indahwati⁽³⁾, Anwar Fitrianto⁽⁴⁾, Reni Amelia⁽⁵⁾

Institut Pertanian Bogor

Jl. Raya Dramaga Kampus IPB Dramaga Bogor Telp./Fax. (021) 8622642 Jawa Barat 16680

e-mail: silmi_annisa@apps.ipb.ac.id

ABSTRAK

Salah satu bentuk akibat dari kurangnya asupan gizi kronis pada balita adalah *stunting*. *Stunting* merupakan permasalahan kesehatan yang saat ini sedang digencarkan untuk diturunkan angka prevalensinya. Permasalahan kesehatan ini berhubungan erat pada pertumbuhan tinggi badan yang lebih rendah dengan anak seusianya. Berdasarkan data Kemenkes per Agustus 2021, Provinsi Jawa Barat menduduki posisi pertama dengan angka balita *stunting* paling tinggi di Indonesia. Penelitian ini bertujuan untuk mengetahui faktor apa saja yang memengaruhi terjadinya *stunting* pada balita dan memodelkan dengan metode regresi logistik biner untuk wilayah Jawa Barat. Metode ini dapat menunjukkan faktor yang memengaruhi berdasarkan peubah yang signifikan. Regresi logistik biner akan memodelkan hubungan antara satu atau beberapa peubah prediktor dengan peubah respon yang kategorik. Peubah respon didefinisikan sebagai persentase angka balita *stunting* dan dibagi kedalam dua kategori yakni tinggi dan rendah. Pengategorian kelas didasarkan pada nilai median pada persentase angka balita *stunting*. Unit penelitian menuju pada 27 wilayah Kabupaten/Kota di Jawa Barat. Hasil analisis menunjukkan dari 11 peubah prediktor, setelah dilakukan pemodelan terdapat 3 peubah yang berpengaruh signifikan pada taraf nyata 0,10 yakni imunisasi dasar lengkap, tempat pengelolaan makanan yang memenuhi syarat kesehatan, dan penduduk miskin. Model yang terpilih berdasarkan nilai akurasi seimbang terbesar dibandingkan model lainnya yakni dihasilkan nilai akurasi seimbang sebesar 81,59%.

Kata kunci : *Stunting, Jawa Barat, Regresi Logistik Biner.*

ABSTRACT

One of the consequences of chronic lack of nutritional intake in toddlers is stunting. Stunting is a health problem that is currently being intensified to reduce its prevalence. This health problem is closely related to lower height growth with their children age. Based on data from the Ministry of Health as of August 2021, West Java occupies the first position with the highest stunting rate in Indonesia. The purpose of this study is to determine the factors that influence the occurrence of stunting in toddlers and to model it using the binary logistic regression method for the West Java region. This method can show influencing factors based on significant variables. Binary logistic regression will model the relationship between one or more predictor variables and categorical response variables. The response variable is defined as the percentage of stunting and it is categorized into two categories, high and low. Class categorization is based on the median value of the percentage of stunting. The research unit was directed to 27 districts/cities in West Java. The results of the analysis show that from 11 predictor variables, after modeling there are 3 variables that have a significant effect on the 0.10 level of significance, there are complete basic immunization, food management places that meet health requirements, and poor people. The model chosen was based on the highest balanced accuracy value compared to other models, which resulted in a balanced accuracy value of 81.59%.

Keywords : *Stunting, West Java, Binary Logistic Regression.*

1. PENDAHULUAN

Salah satu tujuan utama dari pembangunan berkelanjutan adalah dengan menghilangkan angka kelaparan dan segala bentuk malnutrisi untuk

mencapai ketahanan pangan. Kondisi kekurangan gizi ini berkaitan erat dengan permasalahan yang terjadi saat ini yaitu *stunting*. *Stunting* merupakan suatu permasalahan kurangnya asupan gizi yang

menyakut pada gangguan pertumbuhan atau kondisi gagal tumbuh yakni pada tinggi badan yang lebih rendah dibandingkan dengan anak seusianya (Karyati, 2021).

Permasalahan mengenai *stunting* berkaitan erat dengan kurangnya gizi pada anak usia dini yang dapat berefek pada kondisi pertumbuhannya seperti halnya, *stunting* dapat meningkatkan angka kematian bayi, menurunkan tingkat kecerdasan anak dalam berkembang dan menerima kapasitas belajar, serta dapat menurunkan produktivitas dan daya saing bangsa (Tim Nasional Percepatan Penanggulangan Kemiskinan, 2017).

Menurut Kementerian Sekretariat Negara RI, tahun 2020 prevalensi *stunting* di Indonesia diperkirakan turun menjadi 26,92%. Angka ini turun jika dibandingkan dengan tahun 2019 yaitu 27,65% dan tahun 2018 yaitu 30,80%. Organisasi Kesehatan Dunia memutuskan batas toleransi *stunting* maksimal adalah 20% atau seperlima dari jumlah keseluruhan balita (Urgensi Minimalisasi "Stunting", 2019). Sehingga dengan kata lain, permasalahan mengenai *stunting* di Indonesia dapat dikatakan cukup tinggi dan menjadikan *stunting* menjadi permasalahan yang dihadapi dunia dan dapat berdampak serius pada kualitas sumber daya manusia di masa depan (Karyati, 2021).

Berdasarkan data Kementerian Kesehatan Republik Indonesia per Agustus tahun 2021, Provinsi Jawa Barat menduduki posisi pertama dengan angka balita *stunting* paling tinggi di Indonesia (Data Stunting Nasional, 2021). Setelah di telaah lebih lanjut pada dua tahun sebelumnya untuk wilayah Provinsi Jawa Barat, berdasarkan data yang terlampir, gizi dan angka balita *stunting* masih menjadi tantangan yang harus dibatasi oleh Pemerintah Daerah Provinsi Jawa Barat. Data Dukcapil menyatakan bahwa jumlah balita *stunting* di Jawa Barat ditemukan meningkat sekitar 22% dari tahun 2019 yakni sebesar 226.117 hingga menjadi 276.069 pada tahun 2020 dan diketahui tahun 2021 angka *stunting* di Jawa Barat masih menduduki posisi pertama. Jika dilihat pada tahun 2018, prevalensi angka balita *stunting* sempat turun sebesar 4,89% tetapi kemudian pada dua tahun kemudian tahun 2019 ke 2020 hanya mengalami sedikit penurunan. Namun keadaan buruknya, meningkatnya angka balita *stunting* ini tidak diikuti dengan meningkatnya tenaga kesehatan gizi dan kader posyandu di Jawa Barat (Tim Data Jabar Digital Service, 2022).

Fenomena angka balita *stunting* yang terjadi di Jawa Barat ini menunjukkan bahwa masih ada wilayah di Jawa Barat yang memiliki jumlah kasus

balita *stunting* yang cukup tinggi namun belum didukung secara seimbang oleh tenaga kesehatan gizi yang memadai serta kenaikan yang tinggi dari tahun ke tahun. Hal ini yang menjadi latar belakang perlu dilakukannya penelitian meninjau faktor-faktor apa saja yang dapat memengaruhi terjadinya *stunting* dan peningkatan angka balita *stunting* di Jawa Barat kemudian selanjutnya dapat diberikan penanganan atau pun upaya lebih lanjut oleh pemerintah setempat guna menekan angka *stunting* baik untuk wilayah Jawa Barat itu sendiri maupun di Indonesia.

Roesardhyati dan Kurniawan (2020) dalam penelitiannya melakukan identifikasi faktor yang memengaruhi masalah gangguan pertumbuhan balita pendek menggunakan metode observasi, wawancara, dan studi dokumentasi kepada 120 responden di Posyandu Kecamatan Pagelaran dan Kecamatan Turen. Penelitian tersebut menggunakan uji regresi logistik biner untuk menentukan peubah yang termasuk kedalam faktor yang memengaruhi timbulnya *stunting* pada anak di bawah usia 5 tahun. Hasilnya menunjukkan kejadian *stunting* pada balita dipengaruhi oleh tiga faktor yaitu tingkat pendidikan ibu, pemberian ASI eksklusif, dan berat badan lahir balita. Penelitian lainnya yang dilakukan oleh Mardlatilla dan Ratih (2022) menghasilkan ketepatan klasifikasi sebesar 74% dengan faktor yang memengaruhi yakni mengonsumsi susu formula dan ibu hamil mendapatkan PMT (Pemberian Makanan Tambahan).

Penelitian ini bertujuan untuk mengetahui faktor yang memengaruhi terjadinya *stunting* pada balita dan memodelkan dengan regresi logistik biner di wilayah Jawa Barat. Data persentase angka balita *stunting* terbagi dalam dua kelompok yaitu tinggi dan rendah berdasarkan nilai median pada persentase angka balita *stunting*. Nilai median dipilih menjadi acuan batas pengategorian didasarkan pada nilai median yang kekar terhadap nilai pencilan (*outlier*) dan dianggap sebagai nilai sentral dari suatu kumpulan data. Hasil akhir dari pemodelan yakni peubah yang signifikan yaitu peubah yang diduga menjadi faktor yang memengaruhinya.

2. METODOLOGI PENELITIAN

2.1 Data dan Peubah Penelitian

Data dirujuk dari data sekunder dari suatu publikasi yakni Profil Kesehatan Jawa Barat tahun 2020 yang dipublikasikan langsung oleh Dinas Kesehatan Provinsi Jawa Barat dan Badan Pusat Statistik Provinsi Jawa Barat. Banyaknya observasi dalam penelitian ini mencakup 27 Kabupaten/Kota

yang ada di Jawa Barat.

Tabel 1. Peubah respon Y

Peubah	Keterangan	Kategori	Skala
Y	Persentase Balita <i>Stunting</i>	0. Rendah 1. Tinggi	Nominal

Tabel 2. Peubah prediktor X

Peubah	Keterangan	Skala
X1	Persentase Bayi Baru Lahir Mendapat Inisiasi Menyusui Dini	Rasio
X2	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sumber Sanitasi Layak	Rasio
X3	Persentase Balita yang Diberi ASI Eksklusif	Rasio
X4	Persentase Balita Mendapat Imunisasi Dasar Lengkap	Rasio
X5	Persentase Rumah Tangga yang Memiliki Akses terhadap Layanan Air Minum Layak	Rasio
X6	Rata-rata Lama Sekolah	Rasio
X7	Persentase Tempat Pengelolaan Makanan yang Memenuhi Syarat Kesehatan	Rasio
X8	Persentase Penduduk Miskin	Rasio
X9	Rasio Posyandu terhadap Desa/Kelurahan	Rasio
X10	Rasio Puskesmas terhadap Penduduk	Rasio
X11	Persentase Rumah Tangga Berperilaku Hidup Bersih dan Sehat	Rasio

Tabel 3. Kesimpulan dan kemungkinan tanda koefisien penduga parameter

Peubah	Kesimpulan	Tanda
X1	Semakin tinggi bayi yang baru lahir mendapat inisiasi menyusui dini maka semakin sedikit balita <i>stunting</i> .	Negatif
X2	Semakin tinggi rumah yang memiliki akses sanitasi yang layak maka semakin sedikit balita <i>stunting</i> .	Negatif
X3	Semakin tinggi balita yang diberi ASI eksklusif maka	Negatif

Peubah	Kesimpulan	Tanda
X4	semakin sedikit balita <i>stunting</i> . Semakin tinggi balita yang mendapat imunisasi yang lengkap maka semakin sedikit balita <i>stunting</i> .	Negatif
X5	Semakin tinggi rumah tangga yang memiliki sumber air minum yang layak maka semakin sedikit balita <i>stunting</i> .	Negatif
X6	Semakin panjang jenjang pendidikan seseorang maka angka <i>stunting</i> akan turun.	Negatif
X7	Semakin tinggi tempat pengelolaan yang baik dan memenuhi syarat kesehatan maka angka balita <i>stunting</i> akan turun.	Negatif
X8	Semakin rendah penduduk miskin maka angka balita <i>stunting</i> akan turun.	Positif
X9	Semakin banyaknya posyandu yang ada maka angka balita <i>stunting</i> akan turun.	Negatif
X10	Semakin banyaknya puskesmas yang ada maka angka balita <i>stunting</i> akan turun.	Negatif
X11	Semakin tinggi persentase rumah tangga yang berperilaku hidup bersih dan hidup sehat maka akan menurunkan angka balita <i>stunting</i> .	Negatif

Tanda: Kemungkinan Tanda Koefisien Penduga Parameter

2.2 Prosedur Penelitian

Prosedur penelitian yang dilakukan pada penelitian ini sebagai berikut.

1. Mengategorikan peubah respon Y menjadi dua kategori, Y = 1 (Tinggi) dan Y = 0 (Rendah).
2. Melakukan eksplorasi data dan seleksi peubah berdasarkan pengecekan *information value*, nilai korelasi, multikolinieritas melalui VIF, dan melalui visualisasi data.
3. Memodelkan peubah respon Y dengan peubah prediktor X yang terpilih menggunakan regresi logistik biner.
 - a. Mencari koefisien dugaan parameter dengan regresi logistik biner.

- b. Menyusun model yang diduga.
 - c. Mengevaluasi model menggunakan nilai akurasi, sensitifitas, spesifisitas, dan nilai akurasi seimbang.
 - d. Membuat interpretasi koefisien model regresi logistik biner dari model terbaik dengan nilai odds untuk melihat kecenderungan pengaruh antara peubah prediktor terhadap peubah respon.
4. Melakukan pengujian parameter dengan uji serentak untuk model terbaik yang terpilih dan uji parsial untuk melihat pengaruh dari setiap peubah prediktor terhadap peubah respon.
 5. Melakukan uji kelayakan model menggunakan uji Hosmer-Lemeshow

2.3 Regresi Logistik

Model regresi logistik sama halnya dengan model regresi lainnya yang digunakan dalam dunia statistika. Regresi logistik adalah model yang dapat ditafsirkan untuk menggambarkan hubungan antara peubah respon dengan satu atau sekumpulan peubah prediktor. Hal yang membedakan antara model regresi logistik dengan model regresi linear umumnya terletak di peubah respon yang digunakan. Regresi logistik menggunakan peubah respon biner atau dikotomi (Hosmer, Lemeshow, & Sturdivant, 2013).

Suatu kumpulan dari peubah prediktor p dapat dimodelkan dengan model regresi logistik berganda. Dimana logit dari model logistik berganda dapat dituliskan sebagai berikut.

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (1)$$

Model regresi logistik berganda dapat dituliskan sebagai berikut.

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2)$$

dengan $\pi(x)$ adalah besarnya peluang kelas Y pada observasi, β_0 adalah intersep regresi logistik, β_j adalah koefisien regresi logistik ke- j , x adalah peubah prediktor, dan p adalah banyaknya peubah prediktor.

2.4 Pengujian Parameter

Uji parameter dilakukan untuk meninjau hubungan antara peubah prediktor dengan peubah respon. Pengujian parameter ini dapat dilakukan secara serentak terhadap model atau secara parsial dari masing-masing peubah prediktor.

2.4.1 Uji Serentak dengan Uji G

Uji serentak dilakukan untuk mengetahui peubah prediktor yang berpengaruh nyata secara signifikan pada suatu taraf α terhadap peubah respon secara keseluruhan model.

Hipotesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$$

(Tidak ada peubah X yang memberikan pengaruh terhadap peubah Y)

$$H_1 : \text{Minimal ada satu } \beta_j \neq 0; j = 1, 2, \dots, p$$

(Minimal ada satu peubah X yang memberikan pengaruh terhadap peubah Y)

Statistik Uji:

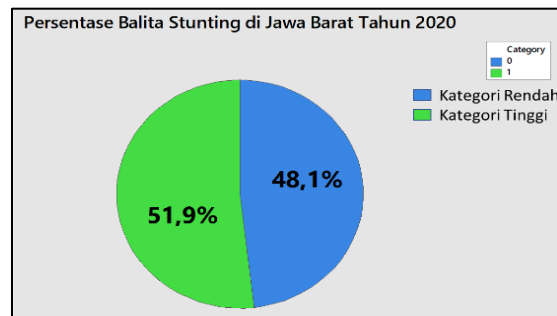
Statistik uji yang digunakan dalam uji ini yaitu uji G

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1 - y_i}} \right] \quad (3)$$

dengan $n_1 = \sum_{i=1}^n y_i$ adalah banyaknya observasi yang terkategori $Y = 1$ dan $n_0 = \sum_{i=1}^n (1 - y_i)$ adalah banyaknya observasi yang terkategori $Y = 0$ (Hosmer, Lemeshow, & Sturdivant, 2013).

Kriteria Penolakan H_0 :

Tolak H_0 jika $G > \chi^2_{ab;\alpha}$ atau nilai- $p < \alpha$



Gambar 1. Karakteristik persentase balita stunting berdasarkan kategori rendah dan tinggi

2.4.2 Uji Parsial dengan Uji Wald

Uji parsial dilakukan untuk mengetahui hubungan antara setiap peubah prediktor dengan peubah respon apakah peubah prediktor tersebut berpengaruh secara nyata (signifikan) terhadap peubah respon.

Hipotesis:

$$H_0 : \beta_j = 0$$

(Peubah X tidak berpengaruh nyata terhadap peubah Y)

$$H_1 : \beta_j \neq 0; j = 1, 2, \dots, p$$

(Peubah X berpengaruh nyata terhadap peubah Y)

Statistik Uji:

Statistik uji yang digunakan dalam uji ini yaitu uji Wald

$$W_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{4}$$

dengan β_j adalah standar error, $se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$ (Hosmer, Lemeshow, & Sturdivant, 2013).

Kriteria Penolakan H₀:

Tolak H₀ jika $G > \chi^2_{db;\alpha}$ atau nilai-p < α

2.4.3 Uji Kelayakan Model

Uji kelayakan model dilakukan untuk melihat seberapa baik model untuk digunakan.

Hipotesis:

H₀ : Model layak digunakan

H₁ : Model tidak layak digunakan

Statistik Uji:

Statistik uji yang digunakan yaitu uji Hosmer-Lemeshow.

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{5}$$

dengan g adalah banyaknya grup, n'_k adalah banyaknya pengamatan dalam grup ke- k , O_k adalah jumlah nilai Y pada grup ke- k , dan $\bar{\pi}_k$ adalah rata-rata dari setiap π untuk grup ke- k (Hosmer, Lemeshow, & Sturdivant, 2013).

Kriteria Penolakan H₀:

Tolak H₀ jika $\hat{C} > \chi^2_{\alpha(g-2)}$ atau nilai-p < α

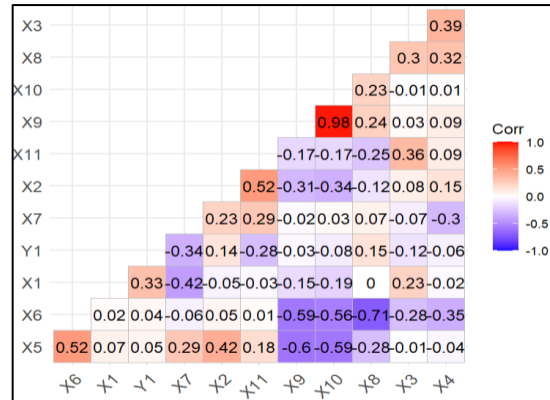
3. HASIL DAN PEMBAHASAN

3.1 Pengategorian Peubah Respon

Tabel 4. Statistik deskriptif peubah respon Y

Peubah	Min	Median	Mean	Maks.
Y	2,54	10,23	9,98	18,97

Pelabelan peubah Y berdasarkan nilai median yaitu 10,23 maka peubah Y dikategorikan 0 jika nilai median dari persentase balita *stunting* bernilai kurang dari 10,23 dan dikategorikan 1 jika nilai median dari persentase balita *stunting* bernilai lebih dari sama dengan 10,23. Total unit observasi yang digunakan terdapat 27 Kabupaten/Kota di Jawa Barat. Dari 27 unit observasi tersebut, sebanyak 14 wilayah memiliki angka balita *stunting* yang terkategori tinggi (Y = 1) yakni sekitar 51,90% dan 13 wilayah lainnya terkategori rendah (Y = 0) yakni 48,10%



Gambar 2. Plot korelasi peubah prediktor X

3.2 Eksplorasi Data dan Seleksi Peubah

3.2.1 Melalui Information Value (IV)

Tabel 5. Seleksi peubah melalui *information value*

Peubah	Information Value
X1	2,559806e-02
X3	2,559806e-02
X2	1,008698e-04
X5	1,008698e-04
X6	1,008698e-04
X7	1,008698e-04
X8	1,008698e-04
X9	1,008698e-04
X10	1,008698e-04
X4	9,695533e-04

Berdasarkan Tabel 5, peubah yang layak digunakan adalah peubah X1 dan X3 karena memiliki IV yang masih dapat digunakan meskipun terkategori tidak berguna. Rentang ukuran nilai untuk IV dijelaskan oleh (Siddiqi, 2006) yakni apabila IV kurang dari 0,02 terkategori tidak berguna (*useless*), 0,02-0,10 lemah, 0,10-0,30 sedang, dan lebih dari 0,30 kuat. Ukuran ini dapat dijadikan salah satu nilai untuk memilih peubah yang relevan (Siddiqi, 2006). Semakin besar nilainya pada suatu peubah mengindikasikan bahwa

peubah tersebut cukup berpengaruh dan dapat membedakan antar kelas kategori (Sartono, Bodro, & Dito, 2020).

3.2.2 Melalui Nilai Korelasi

Nilai korelasi merupakan suatu bentuk ukuran yang dapat menunjukkan antara suatu peubah prediktor yang satu dengan yang lainnya. Semakin kuat hubungan anatar dua peubah maka nilainya akan Semakin besar.

Berdasarkan plot korelasi pada Gambar 2, peubah X9 dan X10 memiliki korelasi yang cukup tinggi yaitu 0,98 dengan kata lain mengindikasikan adanya multikolinieritas. Sehingga penanganannya dapat dilakukan penyeleksian salah satu peubah atau keduanya. Dikatakan adanya multikolinieritas apabila antarpeubah memiliki korelasi yang cukup tinggi. Hal ini menjadi permasalahan serius dan berdampak pada model yang dihasilkan karena adanya multikolinieritas dapat memengaruhi pendugaan nilai parameter (Montgomery, Peck, & Vining, 2012).

3.2.3 Melalui Nilai Variance Inflation Factor

Variance Inflation Factor (VIF) digunakan untuk mengukur tingkat keparahan dari multikolinieritas dalam suatu analisis. Suatu model yang dikatakan baik jika memiliki nilai VIF < 10 artinya tidak ada multikolinieritas antarpeubah. Pada eksplorasi bagian ini, akan dilakukan pengecekan VIF dengan model penuh. Hasil dari pemodelan penuh akan dikeluarkan peubah yang memiliki nilai VIF tertinggi di antara peubah lainnya.

Tabel 6. Nilai VIF dengan model penuh

X	X	X	X	X	X	X	X	X	X9
1	2	3	4	5	6	7	8		
2,	3,	2,	3,	4,	8,	3,	7,	177,	
6	1	5	9	2	2	8	4	7	

Tabel 7. Nilai VIF dengan model penuh (lanjutan)

X10	X11
178,5	3,3

Dalam hal ini, peubah X10 memiliki nilai VIF tertinggi di antara peubah lainnya, sehingga peubah X10 dikeluarkan dari model dan akan dilakukan perhitungan kembali nilai VIF tanpa menggunakan peubah X10.

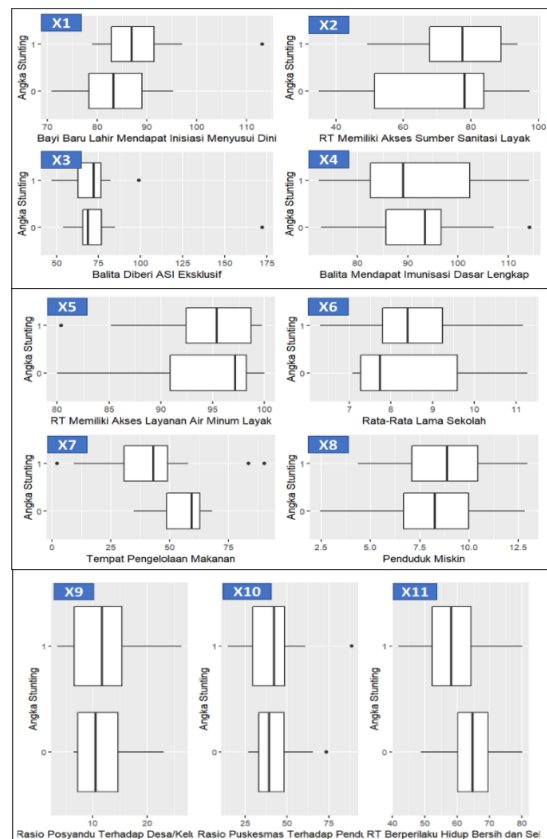
Tabel 8. Nilai VIF dengan model penuh

X	X	X	X	X	X	X	X	X	X
1	2	3	4	5	6	7	8	9	11
2,	2,	1,	3,	4,	4,	5,	6,	3,	2,
5	8	8	2	4	9	3	5	5	6

Berdasarkan hasil eksplorasi dengan nilai VIF, setelah peubah X10 dikeluarkan, model tanpa peubah X10 memiliki nilai VIF kurang dari 10 untuk tiap-tiap peubah prediktor. Sehingga model sudah cukup aman digunakan.

3.2.3 Melalui Eksplorasi Data Hubungan Peubah X_k dan Y

Eksplorasi data dapat digunakan untuk menyeleksi peubah yang akan digunakan. Jika hasil eksplorasi tidak sesuai dengan kondisi yang seharusnya, maka akan dilakukan pemodelan tanpa peubah yang berkaitan. Pada tahap ini akan dilakukan seleksi peubah berdasarkan eksplorasi data hubungan antara peubah prediktor X ke-k dengan peubah respon Y.



Gambar 3. Seleksi peubah X berdasarkan kesesuaian kondisi seharusnya

Tabel 9. Kesesuaian / tidak sesuai dengan kondisi seharusnya

X	X	X	X	X	X	X	X	X	X	X
1	2	3	4	5	6	7	8	9	10	11
T	S	T	S	S	T	S	S	T	T	S

Kode T menandakan bahwa hasil eksplorasi data tidak sesuai dengan kondisi yang seharusnya dan S menandakan hasil eksplorasi data sesuai dengan kondisi yang seharusnya. Kesesuaian hasil eksplorasi data dengan kondisi seharusnya dapat dilihat pada Tabel 3 yang menjelaskan kemungkinan tanda koefisien penduga parameter berdasarkan referensi yang ada. Peubah yang tidak sesuai dengan kondisi seharusnya akan berpengaruh pada model dan simpulan akhir sehingga seleksi peubah melalui metode ini juga diperlukan.

Berdasarkan hasil eksplorasi, peubah yang tidak sesuai dengan kondisi yang seharusnya adalah peubah X1, X3, X6, X9, dan X10. Peubah ini akan memengaruhi hasil analisis, maka selanjutnya akan dilakukan pemodelan tanpa menggunakan peubah yang tidak sesuai dengan kondisi yang seharusnya.

3.3 Penyusunan Model

Dari hasil eksplorasi data untuk melakukan seleksi peubah dengan menggunakan beberapa metode, berikut adalah kesimpulan akhir peubah yang akan digunakan dalam pemodelan.

Tabel 10. Hasil peubah yang akan digunakan

No.	Eksplorasi	Peubah yang digunakan	Peubah tidak digunakan
1.	Korelasi	X1+X2+X3 +X4+X5+X6+X7+X8+X11	X9+X10
2.	Hubungan X_k dengan Y	X2+X4+X5 +X7+X8+x11	X1+X3+X6 +X9+X10
3.	VIF	X1+X2+X3 +X4+X5+X6+X7+X8+X9+X11	X10
4.	Informati on Value	X1+X3	X2+X4+X5 +X6+X7+X8+X9+X10+X11

Tabel 11. Penyusunan model yang terbentuk

Mode l ke-	Peubah yang digunakan	AIC	Peubah signifikan
1.	X1+X2+X3+X4 +X5+X6+X7+X8+X11	40,797	X8
2.	X2+X3+X4+X5 +X7+X8+X11	35,927	X4 X7 X8
3.	X1+X2+X3+X4 +X5+X6+X7+X8+X9+X11	44,534	X1 X2
4.	X1+X3	38,853	X1

Signifikan pada taraf nyata 10%

Pembentukan model tidak hanya dilakukan berdasarkan hasil eksplorasi. Langkah lanjutan akan dilakukan pembentukan model dengan metode *stepwise*.

Tabel 12. Model dengan metode *stepwise*

Model ke-	Hasil <i>stepwise</i>	AIC	Peubah Signifikan
5	X1 + X2 + X11 (Hasil <i>Stepwise Model 3</i>)	35,692	X1 X2 X11
6	X1 (Hasil <i>Stepwise Model 4</i>)	-	-

Signifikan pada taraf nyata 10%

Hasil *stepwise* dari model 1 dan 2 tidak ditampilkan karena hasil *stepwise* menunjukkan peubah yang digunakan sama seperti pada model 2.

Tabel 13. Penyusunan model yang terbentuk

Model ke-	AIC	Peubah yang signifikan	Koefisien penduga parameter
Model 1	40,797	X8	0,9160
Model 2	35,927	X4 X7 X8	-0,1479 -0,1321 0,6786
Model 3	44,534	X1 X2	0,1773 0,0987
Model 4	38,853	X1	0,1097
Model 4	35,692	X1 X2 X11	0,1135 0,0682 -0,1428

Model ke-	AIC	Peubah yang signifikan	Koefisien penduga parameter
Model 6	-	-	-

Signifikan pada taraf nyata 10%

3.4 Evaluasi Performa Model

Evaluasi performa model dilakukan untuk memilih model terbaik. Pada tahap ini, model terbaik dipilih berdasarkan model yang memberikan nilai akurasi seimbang (*balanced accuracy*) yang paling besar di antara model lainnya.

Tabel 14. Evaluasi performa model

Model ke-	A (%)	Sen (%)	Spe (%)	Bal_Ac c (%)
Model 1	81,48	78,57	84,62	81,59
Model 2	81,48	78,57	84,62	81,59
Model 3	77,78	78,57	76,92	77,75
Model 4	55,56	64,29	46,15	55,22
Model 4	62,96	64,29	61,54	62,91
Model 6	55,56	57,14	53,85	55,49

Keterangan:

A: Akurasi ; Sen: Sensitifitas ; Spe: Spesifisitas ; Bal_Acc: Akurasi Seimbang

Model 1 dan model 2 memberikan nilai yang sama untuk setiap nilai akurasi, sensitifitas, spesifisitas, dan akurasi seimbang tetapi pemilihan model terbaik juga meninjau dari kesederhanaan model yang dihasilkan. Model 2 menghasilkan model yang sederhana dan menghasilkan nilai AIC yang lebih kecil jika dibandingkan dengan model 1. Semakin kecil nilai AIC maka metode analisis yang digunakan semakin baik (Revildy, Lestari, & Nalita, 2020). Sehingga model yang terpilih adalah model 2. Model 2 merupakan model tanpa peubah X1, X3, X6, X9, dan X10 yaitu peubah-peubah yang hasil eksplorasi data hubungan antara peubah prediktor X_k dan peubah Y tidak sesuai dengan kondisi yang seharusnya.

Tabel 15. Penduga parameter model terbaik

Peubah	Koefisien	Std. Error	Z-Statistik	Nilai-p
<i>Intercept</i>	1,9495	9,7523	0,200	0,8416
X2	0,0829	0,0510	1,625	0,1041
X4	-0,1479	0,0811	-1,825	0,0680
X5	0,1544	0,1143	1,351	0,1767
X7	-0,1321	0,0607	-2,176	0,0295
X8	0,6786	0,3623	1,873	0,0611
X11	-0,1217	0,0844	-1,442	0,1492

Signifikan pada taraf nyata 10%

Berdasarkan Tabel 15, dapat ditarik simpulan bahwa peubah prediktor yang secara signifikan atau memberikan pengaruh nyata terhadap peubah respon dengan nilai toleransi *alpha* 10%. Hasil dari perdamaan regresi logistik biner didapatkan adalah sebagai berikut:

$$\begin{aligned} \text{Logit } [P(Y = 1)] &= 1,95 + 0,08X2 - 0,15X4 \\ &+ 0,15X5 - 0,13X7 + 0,68X8 \\ &- 0,12X11 \end{aligned}$$

Dari peubah-peubah hasil pemodelan, peubah yang menunjukkan faktor yang memengaruhi angka balita *stunting* adalah peubah-peubah yang signifikan pada taraf nyata 10%. Berikut adalah peubah yang signifikan:

Tabel 16. Peubah yang Signifikan

Peubah	Estimasi koefisien	<i>exp</i> (Estimasi koefisien)
X4	-0,1479	0,8625
X7	-0,1321	0,8762
X8	0,6786	1,9710

Dengan interpretasi koefisien regresi berdasarkan nilai odds adalah sebagai berikut:

➤ $exp(X4) = exp(-0,1479) = 0,8625$

Odds angka balita *stunting* terkategori tinggi akan menurun 0,8625 kali jika persentase balita yang mendapat imunisasi dasar lengkap naik sebesar 1% dengan peubah bebas lainnya konstan.

➤ $exp(X7) = exp(-0,1321) = 0,8762$

Odds angka balita *stunting* terkategori tinggi akan menurun 0,8762 kali jika angka persentase tempat pengelolaan makanan yang telah memenuhi syarat kesehatan naik sebesar 1% dengan peubah bebas lainnya konstan.

➤ $exp(X8) = exp(0,6786) = 1,9710$

Odds angka balita *stunting* terkategori tinggi akan meningkat 1,9710 kali jika persentase penduduk miskin naik sebesar 1% dengan peubah bebas lainnya konstan.

3.5 Pengujian Model

3.5.1 Uji Serentak dengan Uji G

Berikut adalah pengujian model dengan uji serentak menggunakan uji G.

Tabel 17. Hasil uji serentak

G ²	Chi Square tabel	Keputusan
15,4660	10,6440	Tolak H ₀

Karena nilai $G^2 = 15,4660 > Chi-square_{(0,90, 6)} = 10,6440$ artinya tolak H₀. Sehingga sudah cukup bukti untuk menunjukkan bahwa minimal ada satu peubah prediktor X yang memberikan pengaruh terhadap peubah respon Y pada taraf nyata 10%.

3.5.2 Uji Parsial dengan Uji Wald

Berikut adalah hasil pengujian dengan uji parsial menggunakan uji Wald.

Tabel 18. Hasil uji parsial

Peubah	Nilai-p	Keputusan
X2	0,1041	Tak Tolak H ₀
X4	0,0680	Tolak H ₀
X5	0,1767	Tak Tolak H ₀
X7	0,0295	Tolak H ₀
X8	0,0611	Tolak H ₀
X11	0,1492	Tak Tolak H ₀

Nilai-p untuk peubah X4, X7, dan X8 bernilai kurang dari $\alpha = 10\%$, artinya tolak H₀. Sehingga sudah cukup bukti untuk menunjukkan bahwa ketiga peubah tersebut berpengaruh nyata pada taraf nyata 10%.

3.5.3 Uji Kelayakan Model

Berikut adalah hasil pengujian kelayakan model menggunakan uji Hosmer-Lemeshow.

Tabel 19. Hasil uji kelayakan model

Chi-square hitung	Chi-square tabel	Nilai-p	Keputusan
10,1640	13,3615	0,2537	Tidak Tolak H ₀

Karena nilai *Chi-square* hitung bernilai 10,1640 kurang dari *Chi-square* tabel 13,3615 dan nilai-p menunjukkan lebih besar dari $\alpha = 10\%$, artinya tak tolak H₀. Sehingga sudah cukup bukti untuk menunjukkan bahwa model cukup layak digunakan pada taraf nyata 10%.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang telah dilakukan, peubah yang signifikan atau peubah yang berhubungan dengan angka balita *stunting* setelah dilakukan pemodelan untuk menentukan faktor-faktor yang memengaruhi balita *stunting* di Jawa Barat adalah imunisasi dasar lengkap, tempat pengelolaan makanan, dan kondisi penduduk miskin. Model regresi logistik biner merupakan model yang cukup baik digunakan dalam memodelkan dan membandingkan hasil klasifikasi status gizi balita *stunting* secara aktual dengan hasil prediksi sehingga memberikan nilai akurasi, sensitifitas, spesifisitas, dan akurasi seimbang yang cukup baik. Berdasarkan hasil evaluasi model ditunjukkan bahwa hasil akurasi seimbang yang diperoleh model regresi logistik biner sebesar 81,59%.

Saran bagi pemerintah atau pihak terkait setempat dan Dinas Kesehatan Jawa Barat dapat memberikan edukasi kepada masyarakat mengenai *stunting* agar masyarakat menjadi lebih paham dan dapat mencegah terjadinya *stunting* untuk menurunkan angka *stunting*. Berdasarkan hasil eksplorasi statistika deskriptif pada peubah Y yakni persentase angka balita *stunting* di Jawa Barat, wilayah yang terkategori kelas tinggi masih lebih unggul dibandingkan dengan wilayah yang terkategori rendah meskipun tidak jauh untuk perbedaannya tetapi masih terdapat 14 wilayah yang terkategori tinggi sehingga masih perlu dilakukannya suatu upaya untuk menekan angka *stunting* di wilayah-wilayah tersebut.

Saran dari hasil analisis ini untuk penelitian selanjutnya yakni perlu dilakukannya penelitian yang lebih lanjut dengan menambah peubah lain yang lebih kompleks pada penelaitan lanjutan agar dapat menyempurnakan hasil analisis ini agar faktor-faktor yang memengaruhi angka balita *stunting* di Jawa Barat dapat ditekan dan dilakukan penanganan lebih lanjut. Lainnya, dapat menggunakan batas ukuran pengategorian lain yang dapat memperbaiki hasil.

DAFTAR PUSTAKA

- (2021). *Data Stunting Nasional*. Jakarta: Kementerian Kesehatan Republik Indonesia. Retrieved from <https://sigiziterpadu.kemkes.go.id/ppgbm/index.php/Dashboard/>
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression 3rd ed*. New Jersey, United States: John Wiley & Sons Inc.,.
- Karyati, Y. (2021). Pengaruh Jumlah Penduduk Miskin, Laju Pertumbuhan Ekonomi, dan Tingkat Pendidikan terhadap Jumlah Stunting di 10 Wilayah Tertinggi Indonesia Tahun 2010-2019. *Journal Riset Ilmu Ekonomi dan Bisnis*, 101-108.
- Mardlatilla, D. C., & Ratih, I. D. (2022). Analisis Faktor-Faktor yang Berpengaruh terhadap Stunting Menggunakan Regresi Data Logistik Biner (Studi Kasus: Desa Jongbiru Kecamatan Gampengrejo Kabupaten Kediri). *Jurnal Teknik ITS*, 251-257.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis (5th Edition)*. New Jersey: A John Wiley & Sons, Inc., Publication.
- Revildy, W., Lestari, S., & Nalita, Y. (2020). Pemodelan Spatial Error Model (SEM) Angka Prevalensi Balita Pendek (Stunting) di Indonesia Tahun 2018. *Seminar Nasional Official Statistics*, 1224-1231.
- Roesardhyati, R., & Kurniawan, D. (2020). Identifikasi Faktor yang Mempengaruhi Pertumbuhan Balita Pendek (Stunting). *Jurnal Kesehatan MESENCEPHALON*, 142-149.
- Sartono, B., Bodro, D. K., & Dito, G. A. (2020). *Teknik Eksplorasi Data yang Harus dikuasai Data Scientist*. Bogor: PT. Penerbit IPB Press.
- Siddiqi, N. (2006). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey: John Wiley and Sons, Inc.
- Tim Data Jabar Digital Service. (2022, February 25). *Angka Stunting di Jawa Barat Masih 26.1%, Pemerintah Jabar Terus Upayakan dan Pencegahan*. Retrieved from <https://opendata.jabarprov.go.id/id/visualisasi/kasus-balita-stunting-di-jawa-barat-tahun-2019--2020>
- Tim Nasional Percepatan Penanggulangan Kemiskinan. (2017). *100 Kabupaten/Kota Prioritas Untuk Intervensi Anak Kerdil (Stunting)*. Jakarta Pusat: Sekretariat Wakil Presiden Republik Indonesia.
- Urgensi Minimalisasi "Stunting"*. (2019, Februari 26). Retrieved from Kompas: <https://nasional.kompas.com/read/2019/02/26/18055351/urgensi-minimalisasi-stunting#:~:text=WORLD%20Health%20Organization%20%28WHO%29%20menetapkan%20batas%20toleransi%20stunting,berkategori%20sangat%20pendek%20dan%2017%2C1%20persen%20berkategori%20pendek.>

Pengelompokan Pertumbuhan Ekonomi (PDRB) dan Pengeluaran di Jawa Timur Berdasarkan Jumlah UMK serta Faktor-faktor yang Mempengaruhi dengan Model Persamaan Simultan

Santi Puteri Rahayu⁽¹⁾ dan Irene Monica Amanda⁽²⁾
Departemen Statistika, Fakultas Sains dan Analitika Data (FSAD),
Institut Teknologi Sepuluh Nopember (ITS),
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: santi_pr@statistika.its.ac.id dan renemonica@gmail.com

ABSTRAK

Pertumbuhan ekonomi Produk Domestik Regional Bruto (PDRB) diidentifikasi sebagai ukuran kesejahteraan masyarakat. Usaha mikro kecil (UMK) di Indonesia dapat menjadi pendukung dalam pertumbuhan ekonomi karena UMK memiliki karakteristik positif sebagai sektor yang mampu menyediakan lapangan pekerjaan yang besar. Pengeluaran juga dapat mempengaruhi PDRB, karena dapat menunjukkan kesejahteraan masyarakat dalam memenuhi kebutuhan hidupnya. Jawa Timur sebagai salah satu provinsi Indonesia yang memiliki perekonomian yang baik, karena memiliki PDRB terbesar kedua setelah DKI Jakarta dan memiliki UMK yang menjadi pendukung pertumbuhan ekonomi. Adanya hubungan simultan antara PDRB dan pengeluaran yang *over identified* dapat dimodelkan menggunakan metode persamaan simultan 2SLS dan 3SLS. Hasil menunjukkan bahwa estimasi model persamaan simultan lebih baik daripada model persamaan tunggal, berdasarkan kriteria koefisien determinasi maksimum dan kesamaan nilai estimasi. Lebih dari itu, estimasi model persamaan simultan 3SLS ditunjukkan secara empiris bersifat lebih baik dibandingkan model 2SLS, dengan kriteria koefisien determinasi maksimum dan *standard error* minimum. Hasil estimasi model 3SLS menunjukkan bahwa jumlah UMK dan pengeluaran berpengaruh positif terhadap PDRB, tetapi rasio ketergantungan berpengaruh negatif terhadap PDRB. Sementara itu, IPM dan PDRB berpengaruh positif terhadap pengeluaran, tetapi pengangguran berpengaruh negatif terhadap pengeluaran. Hasil konfirmasi pengelompokan estimasi sepuluh daerah PDRB terendah dengan data aktual hanya meliputi empat kabupaten/kota, sedangkan estimasi pengeluaran terdiri dari tiga kabupaten/kota.

Kata kunci : 3SLS, Jawa Timur, PDRB, Pengeluaran, Simultan

ABSTRACT

The economic growth of the Gross Regional Domestic Product (GDP) is identified as a measure of the welfare of the people. Micro and small enterprises (UMK) in Indonesia can be a supporter of economic growth because UMK has positive characteristics as a sector that is able to provide large employment opportunities. Expenditures can also affect GRDP because spending can indicate the welfare of the community in meeting their daily needs. East Java as one of the provinces of Indonesia that has a good economy, because it has the second largest GRDP after DKI Jakarta and has UMK that support economic growth. The existence of a simultaneous relationship between GRDP and expenditures that are over identified can be modeled using the 2SLS and 3SLS simultaneous equation methods. The results show that the estimation of the simultaneous equation model is better than the single equation model, based on the criteria for the maximum coefficient of determination and the similarity of the estimated values. Moreover, the estimation of the 3SLS simultaneous equation model is empirically shown to be better than the 2SLS model, with the criteria of maximum determination coefficient and minimum standard error. The estimation results of the 3SLS model show that the number of UMK and expenditure has a positive effect on GRDP, but the dependency ratio has a negative effect on GRDP. Meanwhile, HDI and GRDP have a positive effect on spending, but unemployment has a negative effect on spending. The results of the confirmation of the grouping of estimates for the ten lowest GRDP regions with actual data only cover for regencies/cities, while the confirmation of the mapping results of the estimated expenditures consists of three regencies/cities.

Keywords : 2SLS, 3SLS, Jawa Timur, PDRB, Expenditure, Simultaneous.

1. PENDAHULUAN

Pertumbuhan ekonomi diidentikkan sebagai ukuran kesejahteraan dan tingkat kehidupan masyarakat di suatu negara atau daerah. Pertumbuhan ekonomi merupakan kenaikan pendapatan nasional secara berarti, pendapatan yang dimaksudkan merupakan pendapatan perkapita (Putong, 2015). Salah satu indikator dari pertumbuhan ekonomi yang sering diamati adalah Produk Domestik Regional Bruto (PDRB), semakin tinggi pertumbuhan ekonomi yang terjadi menandakan bahwa semakin baik kegiatan ekonomi yang diperoleh dari laju pertumbuhan PDRB (Supartoyo dan Sendouw, 2013). Pendapatan yang diterima, baik oleh pemerintah maupun masyarakat, akan digunakan untuk melakukan konsumsi. Konsumsi yang dilakukan masyarakat dapat membantu pertumbuhan ekonomi, konsumsi ini sering disebut sebagai pengeluaran per kapita. Semakin tinggi pengeluaran yang dikeluarkan oleh masing-masing rumah tangga menunjukkan tercukupinya kesejahteraan hidup rumah tangga tersebut karena menunjukkan bahwa masyarakat mampu mencukupi kebutuhan hidupnya. Hubungan timbal balik antara ekonomi keluarga dan ekonomi masyarakat atau nasional terutama terjadi melalui pengeluaran. Pengeluaran atau konsumsi konsumen merupakan bagian terbesar dari permintaan masyarakat, mencapai 60% hingga 80% dari seluruh pendapatan nasional dikeluarkan untuk konsumsi rumah tangga (Putong, 2015).

Provinsi Jawa Timur merupakan salah satu provinsi di Indonesia dengan pertumbuhan ekonomi yang berkembang. Provinsi Jawa Timur juga merupakan provinsi kedua yang memiliki PDRB terbesar, setelah DKI Jakarta, yang pada tahun 2016 Jawa Timur memiliki PDRB sebesar 1405.6 triliun rupiah, dan pada tahun 2018 mencapai 2189.8 triliun rupiah. Nilai PDRB yang tinggi di Provinsi Jawa Timur menunjukkan daya saing Jawa Timur yang komprehensif dengan kinerja tinggi di berbagai aspek ekonomi (Merdikawati dkk., 2014).

Adanya keterkaitan pada variabel-variabel dalam bidang ekonomi tidak dapat diabaikan (Greene, 2012). Sistem persamaan tunggal tidak dapat menggambarkan dependensi antar variabel-variabel ekonomi tersebut sehingga untuk menggambarkan hubungan antar variabel-variabel ekonomi yang bersifat dua arah digunakan sistem persamaan simultan. Penggunaan persamaan simultan dikarenakan estimator model persamaan tunggal *Ordinary Least Square (OLS)* tidak dapat digunakan untuk persamaan simultan sebab *OLS* akan menghasilkan parameter yang bias dan tidak konsisten. Penggunaan metode *Three Stages Least*

Square (3SLS) lebih baik dibandingkan *2SLS* karena memiliki sifat full information, yaitu pada saat menduga parameter persamaan tertentu, metode tersebut sudah mempertimbangkan parameter pada persamaan lainnya (Greene, 2012). Metode *3SLS* juga bersifat lebih konsisten karena merupakan estimasi yang bersifat efisien asimtotik, yaitu hipotesis yang memungkinkan sejumlah variabel eksogen (independen) untuk bertambah bersama dengan ukuran sampel, kondisi pada saling ketergantungan diperlukan untuk konsistensi estimator yang ditetapkan (Greene, 2012).

Penelitian ini bertujuan mengelompokan kabupaten/kota di provinsi Jawa Timur berdasarkan pemodelan hubungan simultan antara PDRB dan pengeluaran berdasarkan jumlah UMK dan faktor-faktor lain yang mempengaruhi, menggunakan metode persamaan simultan *2SLS* dan *3SLS*, dibandingkan dengan metode persamaan tunggal *OLS*. Penelitian relevan sebelumnya (Musaidah & Purba, 2018) dilakukan menggunakan metode *2SLS*, memiliki hasil Kapasitas Fiskal, Pengeluaran Pemerintah Daerah, Penanaman Modal Asing, Penanaman Modal Dalam Negeri, dan Penduduk secara simultan berpengaruh signifikan terhadap Produk Domestik Regional Bruto, serta Produk Domestik Regional Bruto, Pajak Daerah, dan Retribusi Daerah secara simultan berpengaruh signifikan terhadap Kapasitas Fiskal. Penelitian yang dilakukan menggunakan regresi data panel (Lies Maria Hamzah, 2019) dan memiliki hasil tenaga kerja UMKM dan investasi UMKM berpengaruh positif dan signifikan terhadap pendapatan nasional pada sektor UMKM, tetapi jumlah UMKM tidak berpengaruh signifikan pada pendapatan nasional di sektor UMKM. Selanjutnya terdapat penelitian yang dilakukan dengan regresi linier berganda (Syamsuddin, 2013) pertumbuhan penduduk dan rasio kependudukan berpengaruh signifikan dan negatif terhadap pertumbuhan ekonomi, namun angkatan kerja tidak berpengaruh terhadap pertumbuhan ekonomi. Penelitian yang dilakukan dengan menggunakan metode *3SLS* (Rachmawati, 2014) menunjukkan bahwa pertumbuhan ekonomi dipengaruhi inflasi, pengeluaran pemerintah, dan nilai tukar rupiah, sedangkan inflasi dipengaruhi GDP, pengeluaran pemerintah, PMDN, dan suku bunga.

2. METODE PENELITIAN

2.1 Tinjauan Pustaka

A. Matriks Korelasi

Matriks korelasi merupakan matriks dengan elemen-elemennya terdapat hubungan satu sama lain atau merupakan koefisien korelasi dengan nilai

antara -1 hingga 1 dan elemen diagonal bernilai satu. Misalkan \mathbf{X} adalah matriks data dan $\bar{\mathbf{x}}$ adalah matriks rata-rata, serta $\mathbf{X}^T\mathbf{X}$ adalah matriks varians kovarians, terdapat bentuk matriks yaitu pers. (1).

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_j \\ \vdots \\ \bar{x}_p \end{pmatrix} = \frac{1}{n} \begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_p \end{pmatrix} = \frac{1}{n} \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (1)$$

$$\bar{\mathbf{x}} = \frac{1}{n} \mathbf{X}\mathbf{1}$$

Persamaan (1) akan dikalikan dengan vektor $\mathbf{1}^T$ sehingga akan menghasilkan matriks $\bar{\mathbf{x}}\mathbf{1}^T$ yang memiliki bentuk seperti berikut ini.

$$\bar{\mathbf{x}}\mathbf{1}^T = \begin{pmatrix} \bar{x}_1 & \bar{x}_1 & \dots & \bar{x}_1 \\ \bar{x}_2 & \bar{x}_2 & \dots & \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_p & \bar{x}_p & \dots & \bar{x}_p \end{pmatrix} \quad (2)$$

Kemudian kurangkan matriks \mathbf{X} dengan persamaan (2) yang akan dinotasikan sebagai matriks \mathbf{V} , perkalian matriks \mathbf{V} dengan transposenya akan membentuk matriks \mathbf{S} .

$$\mathbf{S} = (\mathbf{X} - \frac{1}{n} \mathbf{X}\mathbf{1}\mathbf{1}^T)(\mathbf{X} - \frac{1}{n} \mathbf{X}\mathbf{1}\mathbf{1}^T)^T$$

$$\mathbf{S} = \mathbf{X}(\mathbf{1} - \frac{1}{n} \mathbf{1}\mathbf{1}^T)\mathbf{X}^T \quad (3)$$

Jika nilai \mathbf{S} diketahui dari persamaan (3), maka \mathbf{S} dapat dihubungkan ke matriks korelasi ρ dengan menghitung matriks $\mathbf{X}^T\mathbf{X}$.

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1j} \\ S_{21} & S_{22} & \dots & S_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \dots & S_{kj} \end{pmatrix} \quad (4)$$

dengan

$$S_{kj} = \sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)$$

Maka menurut Graybill (1983) diketahui bahwa matriks korelasi memiliki bentuk yaitu.

$$\rho = C = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1j} \\ r_{21} & r_{22} & \dots & r_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & r_{kj} \end{pmatrix} \quad (5)$$

dengan

$$r_{kj} = \frac{S_{kj}}{\sqrt{S_{kk}S_{jj}}} \quad k, j = 1, 2, \dots, n$$

Untuk $k = j$ menghasilkan $r = 1$

B. Uji Spesifikasi Hausman

Sistem persamaan simultan memiliki variabel endogen eksplanatori yang memiliki korelasi dengan *error*. Jika tidak terdapat korelasi antara keduanya maka dapat digunakan estimator OLS, tetapi jika terdapat korelasi maka menggunakan estimator lain. Melakukan pengujian korelasi ini menggunakan uji Hausman yang memiliki hipotesis sebagai berikut (Seddighi and Lawler, 2000):

H_0 : Tidak terdapat simultanitas

H_1 : Terdapat simultanitas

Langkah-langkah dalam pengujian Hausman adalah sebagai berikut:

1. Melakukan estimasi persamaan tereduksi dari model persamaan simultan, kemudian melakukan regresi variabel endogen dari setiap persamaan struktural yang ada dengan variabel di sebelah kanan yang terdiri dari variabel *predetermined*. Didapatkan nilai \hat{Y}_i dan sisaan u_i
2. Diketahui bahwa

$$\hat{Y}_i = Y_i + u_i \quad (6)$$

maka substitusikan bentuk ini pada persamaan yang mengandung variabel endogen eksplanatori, kemudian lakukan estimasi menggunakan OLS.

$$y_i = \alpha_i(\hat{Y}_i + u_i) + \beta_i X_i + \varepsilon_i \quad (7)$$

$$y_i = \alpha_i \hat{Y}_i + \alpha_i u_i + \beta_i X_i + \varepsilon_i$$

3. Menggunakan uji individual untuk menguji kesignifikan koefisien regresi dari variabel u_i pada persamaan tersebut. Jika hasil yang didapat menunjukkan bahwa koefisien regresi dari variabel u_i signifikan, maka didapatkan keputusan tolak H_0 dan dapat disimpulkan bahwa terdapat simultanitas. Jika hasil yang didapat menunjukkan bahwa koefisien regresi dari variabel u_i tidak signifikan, maka didapatkan keputusan gagal tolak H_0 dan kesimpulan yang didapat adalah tidak terdapat simultanitas.

C. Sistem Persamaan Simultan

Sistem persamaan simultan merupakan sistem persamaan yang terdiri lebih dari satu variabel tak bebas dan lebih dari satu persamaan yang saling terkait dan dapat dijelaskan secara bersama-sama

(Greene, 2012). Model sistem persamaan simultan dengan bentuk struktural G sebagai peubah endongen dan K sebagai peubah eksogen, secara umum dapat dituliskan dengan rumus seperti berikut ini (Seddighi and Lawler, 2000).

$$\begin{aligned} y_{11}Y_{1t} + y_{12}Y_{2t} + \dots + y_{1G}Y_{Gt} + \beta_{11}x_{1t} \\ + \beta_{12}x_{2t} + \dots + \beta_{1K}x_{Kt} &= \varepsilon_{1t} \\ y_{21}Y_{1t} + y_{22}Y_{2t} + \dots + y_{2G}Y_{Gt} + \beta_{21}x_{1t} \\ + \beta_{22}x_{2t} + \dots + \beta_{2K}x_{Kt} &= \varepsilon_{2t} \\ \dots \\ y_{G1}Y_{1t} + y_{G2}Y_{2t} + \dots + y_{GG}Y_{Gt} + \beta_{G1}x_{1t} \\ + \beta_{G2}x_{2t} + \dots + \beta_{GK}x_{Kt} &= \varepsilon_{Gt} \end{aligned} \tag{8}$$

Secara umum dapat ditulis dalam bentuk yaitu.

$$\Gamma Y_t + B X_t = \varepsilon_t \tag{9}$$

Keterangan :

- Γ : matriks $G \times G$ dari koefisien peubah endogen
- Y_t : vektor $G \times 1$ dari peubah endogen untuk waktu t
- B : matriks $G \times K$ dari koefisien peubah eksogen
- X_t : vektor $K \times 1$ dari peubah eksogen pada waktu t

D. Two Stage Least Square (2SLS)

Metode $2SLS$ digunakan ketika terdapat korelasi antara variabel bebas dengan galat atau *error* (Misno dan Sulistianingsih, 2019). Metode $2SLS$ hampir menyerupai metode ILS dan merupakan metode yang dinamis dalam menciptakan variabel-variabel instrumen dalam menggantikan variabel endogen. Bentuk umum struktural dari metode $2SLS$ adalah seperti berikut ini (Oky, Widiharih, dan Hoyyi, 2012),

$$y_i = b_1y_1 + b_2y_2 + \dots + b_My_M + \gamma_1x_1 + \dots + \gamma_kx_k + \tag{10}$$

Keterangan :

- y_i : variabel endogen ($g=1,2,\dots,M$)
- x_i : variabel eksogen ($h=1,2,\dots,k$)
- b : koefisien dari variabel endogen
- γ : koefisien dari variabel eksogen

Langkah-langkah penyelesaian metode $2SLS$, yang berdasarkan pada metode OLS , dapat dijelaskan dalam 2 langkah sebagai berikut:

1. Menerapkan metode OLS pada persamaan reduksi, yang merupakan persamaan dari setiap variabel endogen dengan semua variabel eksogen. Yang akan memperoleh persamaan

$$\begin{aligned} \hat{Y}_1 &= \pi_{11}x_1 + \pi_{12}x_2 + \dots + \pi_{1k}x_k \\ \hat{Y}_2 &= \pi_{21}x_1 + \pi_{22}x_2 + \dots + \pi_{2k}x_k \\ \dots \\ \hat{Y}_i &= \pi_{i1}x_1 + \pi_{i2}x_2 + \dots + \pi_{ik}x_k \end{aligned} \tag{11}$$

Koefisien dalam bentuk sederhana ($\hat{\pi}$) digunakan untuk mendapatkan sepasang nilai estimasi untuk variabel endogen yaitu $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_M$.

2. Mengganti variabel endogen yang ada di sisi kanan persamaan dengan nilai taksiran $Y_i = \hat{Y}_i + \varepsilon_i$. Pada langkah ini dilakukan substitusi \hat{Y} ke dalam persamaan struktural dan memperoleh bentuk transformasi dari fungsi.

E. Three Stage Least Square (3SLS)

Metode $3SLS$ merupakan metode OLS dalam tiga tahap, tahap pertama adalah menaksir parameter persamaan simultan dari *reduced form* dengan OLS , sehingga akan didapat hasil taksiran persamaan struktural.

$$\begin{aligned} Y_i &= \alpha_{i1}(\hat{Y}_1 + \varepsilon_1) \\ &\quad + \alpha_{i2}(\hat{Y}_2 + \varepsilon_2) + \dots + \alpha_{im}(\hat{Y}_i \\ &\quad + \varepsilon_i) + \beta_{i1}x_1 + \dots + \beta_{ik}x_k + \varepsilon_i \\ Y_i &= \alpha_{i1}\hat{Y}_1 + \alpha_{i1}\varepsilon_1 + \alpha_{i2}\hat{Y}_2 + \alpha_{i2}\varepsilon_2 + \dots + \alpha_{im}\hat{Y}_i \\ &\quad + \alpha_{im}\varepsilon_i + \beta_{i1}x_1 + \dots + \beta_{ik}x_k \\ &\quad + \varepsilon_i \\ Y_i &= \alpha_{i1}\hat{Y}_1 + \alpha_{i2}\hat{Y}_2 + \dots + \alpha_{im}\hat{Y}_i \\ &\quad + \beta_{i1}x_1 + \dots + \beta_{ik}x_k + \varepsilon_i^* \end{aligned} \tag{12}$$

Tahap kedua adalah menaksir kembali hasil taksiran dari tahap pertama menggunakan OLS untuk mendapatkan matriks varians-kovarians variabel *error* yang terdapat dalam persamaan simultan. Tahap ketiga adalah menaksir semua persamaan struktural secara simultan dengan matriks varians-kovarians dengan metode *Generalize Least Square (GLS)* (Greene, 2012).

Matriks kovarians *error* acak pada metode $3SLS$ yang diperoleh

$$E(u_i u_j' | x_1, x_2, \dots, x_M) = \sigma_{ij} I \tag{13}$$

Didapatkan bentuk matriks

$$u_j' | x_1, x_2, \dots, x_M = \begin{pmatrix} E(u_1u_1) & E(u_1u_2) & \dots & E(u_1u_M) \\ E(u_2u_1) & E(u_2u_2) & \dots & E(u_2u_M) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_Mu_1) & E(u_Mu_2) & \dots & E(u_Mu_M) \end{pmatrix} \tag{14}$$

$$u_j' | x_1, x_2, \dots, x_M = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I & \dots & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{22}I & \dots & \sigma_{2M}I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}I & \sigma_{M2}I & \dots & \sigma_{MM}I \end{pmatrix} = \Sigma \otimes I \tag{15}$$

dengan Σ merupakan matriks varians-kovarians, I merupakan matriks identitas. Menggunakan matriks kovarian persamaan (8) dan model instrumental variabel maka didapatkan estimator $3SLS$ yaitu.

$$\tag{16}$$

$$\delta_{3SLS} = Z'(\Sigma \otimes I)Z^{-1}Z'(\Sigma \otimes I)y$$

Dengan Z merupakan semua matriks dari variabel *predetermined*, $\Sigma \otimes I$ merupakan matriks kovarians, dan y adalah matriks variabel endogen.

Berikut adalah langkah-langkah 3SLS.

1. Melakukan estimasi koefisien pada *reduced form* dengan OLS sehingga diperoleh.

$$Y_2 = \Pi_0 + \Pi_1 X_1 + \Pi_2 X_2 + \dots + \Pi_M X_M \quad (17)$$

Oleh karena Y_2 berdasarkan taksiran persamaan *reduced form*, maka variabel ini berlaku sebagai variabel instrumen data asli Y_2 , yaitu variabel baru yang tidak berkorelasi dengan unsur gangguan persamaan, tetapi berkorelasi dengan variabel eksogenen. Terdapat persamaan simultan yaitu

$$Y_2 = \Pi_0 + \Pi_1 x_1 + \Pi_2 x_2 + \dots + \Pi_M x_M + \Omega_j = Y_2 + \Omega_j \quad (18)$$

Sehingga dapat dibentuk matriks Z_j yang terdiri atas nilai variabel endogen dan variabel eksogen persamaan ke-j

$$Z_j = [Y_j \quad X_j]$$

2. Tahap kedua adalah menghitung estimasi δ_j dengan meregresikan Y_j pada Z_j . Hasil δ_j yang didapatkan merupakan hasil estimasi 2SLS untuk masing-masing persamaan dengan estimator δ_{2SLS} , sehingga σ_{ij} dihitung dengan.

$$\sigma_{ij} = \frac{y_i - Z_i \delta_{if} \quad (y_i - Z_i \delta_{if})}{N} \quad (19)$$

3. Tahap ketiga adalah menghitung estimator 3SLS dengan menggunakan estimasi Z dan Σ .

F. Hubungan PDRB dan Pengeluaran dengan Faktor-faktor yang Diduga Berpengaruh

Variabel X_1 (jumlah UMK) diharapkan memiliki hubungan yang positif dengan Y_1 (PDRB) karena jumlah UMK sangat mendominasi dan menyerap tenaga kerja yang sangat banyak sehingga dapat meningkatkan perekonomian di Jawa Timur, potensi perekonomian Jawa Timur juga masih bergantung pada kontribusi UMK (BPS Jatim, 2016). Hubungan antara X_2 (rasio ketergantungan) diharapkan memiliki hubungan yang negatif terhadap Y_1 karena rasio ketergantungan menjadi indikasi yang tidak baik bagi suatu daerah, semakin besar rasio ketergantungan maka semakin besar kapasitas yang harus dimiliki penduduk angkatan kerja dalam menopang penduduk yang bukan angkatan kerja, berdasarkan hal tersebut akan

melemahkan pertumbuhan ekonomi dan akan mempengaruhi PDRB (Pramono dan Suminar, 2019). PDRB dan pengeluaran diharapkan memiliki hubungan yang positif atau searah karena konsumsi yang dilakukan rumah tangga dalam memenuhi kebutuhan hidupnya akan meningkatkan pendapatan daerah dan meningkatkan nilai barang dan jasa yang dikonsumsi..

Persamaan Y_2 (pengeluaran) terdiri dari 2 variabel eksogen, X_3 (IPM) dan X_4 (TPT). Hubungan variabel X_3 dengan Y_2 diharapkan memiliki hubungan yang positif karena IPM menunjukkan terpenuhinya kesejahteraan masyarakat yang tinggal di daerah tertentu dan masyarakat yang dapat memenuhi kebutuhan hidupnya akan melakukan konsumsi yang tinggi, terlebih penduduk akan banyak melakukan konsumsi untuk kebutuhan pangan atau melakukan konsumsi pangan (Wahyu dan Setiawan, 2017). Terdapat variabel X_4 yang diharapkan memiliki hubungan negatif dengan Y_2 karena pengangguran yang semakin meningkat di suatu daerah akan melemahkan perekonomian di daerah tersebut, dan pengangguran menandakan tidak adanya pendapatan yang dapat diterima oleh masyarakat sehingga masyarakat tidak dapat melakukan konsumsi atau pengeluaran untuk kebutuhan hidupnya.

2.2 METODOLOGI

A. Sumber Data

Data 38 kabupaten/kota yang digunakan pada penelitian ini adalah data sekunder yang didapat dari website dari BPS Provinsi Jawa Timur dan publikasi sensus ekonomi tahun 2016, yaitu Potensi Peningkatan Kinerja Usaha Mikro Kecil Provinsi Jawa Timur Analisis Hasil Sensus Ekonomi 2016 (SE2016) Lanjutan. Data yang dipakai adalah PDRB dan pengeluaran per kapita di Provinsi Jawa Timur Tahun 2016 beserta faktor-faktor yang diduga mempengaruhi.

B. Variabel Penelitian

Variabel Spesifikasi model yang dibentuk mengacu pada penelitian-penelitian terdahulu dengan beberapa modifikasi. Model yang dibentuk adalah,

$$Y_{1i} = h_0 + h_1 X_{1i} + h_2 X_{2i} + h_3 Y_{2i} + U_1 \quad (20)$$

Diharapkan $h_1 > 0, h_2 < 0, h_3 > 0$

$$Y_{2i} = g_0 + g_1 X_{3i} + g_2 X_{4i} + g_3 Y_{1i} + U_2 \quad (21)$$

Diharapkan $g_1 > 0, g_2 < 0, g_3 > 0$

Variabel respon dan prediktor yang digunakan dalam penelitian dijelaskan pada Tabel 1.

Tabel 1. Data penelitian

Variabel	Keterangan	Satuan
Variabel Endogen		
y_1	PDRB (Atas Dasar Harga Berlaku 2010)	Triliun Rupiah
y_2	Pengeluaran Per Kapita	Ratusan Ribu Rupiah
Variabel Eksogen		
x_1	Jumlah UMK	Ribuan Unit
x_2	Ratio Ketergantungan	Persen
x_3	Indeks Pembangunan Masyarakat (IPM)	Persen
x_4	Tingkat Pengangguran Terbuka (TPT)	Persen

C. Langkah Penelitian

Langkah-langkah analisis data yang digunakan dalam melakukan penelitian ini sebagai berikut.

1. Mendeskripsikan karakteristik PDRB dan pengeluaran di Provinsi Jawa Timur tahun 2016 dan faktor-faktor yang diduga mempengaruhinya.
2. Membentuk model persamaan simultan dan melakukan identifikasi model
3. Melakukan uji simultanitas *Hausman*
4. Melakukan estimasi model persamaan simultan metode *2SLS* dan *3SLS*
5. Melakukan analisis evaluasi perbandingan antara hasil estimasi metode persamaan simultan *2SLS* dan *3SLS*, dengan kriteria *standard error* minimum dan koefisien determinasi maksimum. Selanjutnya, melakukan analisis evaluasi perbandingan antara hasil estimasi metode persamaan simultan dan metode persamaan tunggal *OLS*, dengan kriteria koefisien determinasi maksimum dan kesamaan nilai estimasi.
6. Melakukan pengujian signifikansi statistik dan asumsi residual pada model yang terpilih.
7. Melakukan interpretasi model dan evaluasi terhadap teori ekonomi yang ada.
8. Melakukan pengelompokan nilai estimasi untuk mengetahui daerah yang memiliki sepuluh nilai estimasi terendah dan memvalidasi dengan data aktual.

3. HASIL DAN PEMBAHASAN

A. Karakteristik PDRB dan Pengeluaran Tahun 2016 dan Faktor-Faktor yang Diduga Berpengaruh.

Analisis statistika deskriptif dilakukan untuk mengetahui karakteristik data. Statistika deskriptif dari PDRB dan pengeluaran di Jawa Timur tahun 2016 beserta faktor-faktor yang diduga berpengaruh tercantum pada Tabel 2.

Tabel 2. Karakteristik PDRB dan pengeluaran serta faktor-faktor yang diduga berpengaruh

Variabel	Mean	Varians	Min	Maks
y_1	451.3832	5607.857	5.3313	451.3832
y_2	8.6412	8.093	5.4804	18.224
x_1	120.2716	5340.215	16501	361079
x_3	44.31579	11.89073	36.35	52.85
x_3	69.79158	28.15054	59.09	80.46
x_4	4.365526	2.744856	1	8.22

Berdasarkan Tabel 2 diketahui nilai PDRB (y_1) paling tinggi daerah Jawa Timur yaitu Rp 451.3832 triliun, sedangkan daerah dengan nilai PDRB terendah adalah sebesar Rp 5.3313 triliun, serta memiliki varians yang cukup besar yaitu 5607.587 yang menandakan penyebaran atau perbedaan data yang cukup besar. Pengeluaran (y_2) tertinggi yaitu sebesar 18.224 ratus ribu rupiah, dan terendah sebesar 5.4804 ratus ribu rupiah. Variansi dari variabel eksogen sangat beragam di setiap daerah, variabel yang memiliki variansi yang besar adalah jumlah UMK (x_1).

B. Hubungan PDRB dan Pengeluaran dengan Faktor-faktor yang Diduga Mempengaruhi

Matriks korelasi dapat digunakan untuk melihat memiliki 2 variabel eksogen yang diduga mempengaruhi, yaitu jumlah UMK dan rasio ketergantungan. Tabel 3 merupakan matriks korelasi persamaan PDRB.

Tabel 3. Matriks korelasi PDRB

	y_1 (PDRB)	y_2 (Pengeluaran)	x_1 (Jumlah UMK)	x_2 (Rasio Ketergantungan)
y_1	1	0.6014	0.6950	-0.5044
y_2	0.6014	1	0.0596	-0.6069
x_1	0.6950	0.0596	1	-0.1420
x_2	-0.5044	-0.6069	-0.1420	1

PDRB dan variabel-variabel yang diduga mempengaruhi memiliki hubungan yang cukup kuat, dengan nilai korelasi yang berada diatas nilai 0,5. Hubungan yang dimiliki y_1 dengan x_1 dan y_2 memiliki hubungan positif atau hubungan searah, namun, terhadap x_2 memiliki hubungan negatif atau berbanding terbalik. Terdapat juga matriks korelasi persamaan pengeluaran dengan variabel eksogen IPM dan TPT pada Tabel 4.

Tabel 4. Matriks korelasi pengeluaran

	y_1 (PDRB)	y_2 (Pengeluaran)	x_3 (IPM)	x_4 (TPT)
y_1	1	0.6014	0.3828	0.4281
y_2	0.6014	1	0.8742	0.4614
x_3	0.3828	0.8742	1	0.5532
x_4	0.4281	0.4614	0.5532	1

Hubungan y_2 dengan x_4 hanya memiliki nilai korelasi sebesar 0,4616 dan berdasarkan teori seharusnya hubungan antara y_2 dengan x_4 memiliki hubungan yang negatif, tetapi pada matriks korelasi memiliki hubungan yang positif. Korelasi antara y_2 dengan x_3 yang memiliki hubungan yang sangat kuat yaitu sebesar 0,874 dan hubungan bersifat positif.

C. Spesifikasi Model dan Uji Hausman Persamaan PDRB dan Pengeluaran

Spesifikasi model yang dibentuk terdiri dari dua persamaan yaitu PDRB dan pengeluaran. Setelah mengetahui bentuk persamaan simultan pada permasalahan maka akan dilakukan identifikasi model untuk mengetahui metode yang digunakan.

Persamaan (20) : $K = 6, k = 3, m = 2$

Identifikasi model : $K - k = 6 - 3 = 3 ; m - 1 = 2 - 1 = 1$

$K - k > m - 1$ maka persamaan *overidentified*

Persamaan (21) : $K = 6, k = 3, m = 2$

Identifikasi model : $K - k = 6 - 3 = 3 ; m - 1 = 2 - 1 = 1$

$K - k > m - 1$ maka persamaan *overidentified*

Sebelum membentuk persamaan simultan, persamaan tersebut akan terlebih dahulu dilakukan uji simultanitas untuk mengetahui ada atau tidaknya hubungan antara variabel endongen dengan variabel gangguan.

Tabel 5. Uji simultanitas (endogenitas) persamaan PDRB dan pengeluaran

Persamaan	Hausman Test	P-value
y_1	2,69795	0,1100
y_2	2,76346	2,76346

Berdasarkan Tabel 5 diketahui bahwa nilai uji simultanitas persamaan y_1 dan y_2 sudah memiliki *p-value*, yaitu 0.11 dan 0.105, yang lebih kecil dibandingkan taraf signifikan sebesar 15% atau 0.15, sehingga diketahui pada kedua persamaan terdapat simultanitas. Kedua persamaan variabel endogen, yaitu y_1 dan y_2 sudah terdapat simultanitas serta memiliki hasil identifikasi yaitu *overidentified* sehingga dapat dilanjutkan untuk menentukan persamaan simultanitas 2SLS dan 3SLS pada kedua persamaan tersebut.

D. Pemodelan Persamaan Simultan PDRB dan Pengeluaran Metode 2SLS dan 3SLS

Persamaan simultan menggunakan 2SLS dilakukan dengan menghitung estimasi dari persamaan tereduksi, kemudian hasil estimasi tersebut akan digunakan untuk membentuk persamaan simultan. Persamaan simultan 3SLS juga dilakukan pada penelitian ini, tahap awal persamaan 3SLS menyerupai persamaan 2SLS, hanya saja pada tahap terakhir atau ketiga akan menggunakan metode SUR. Tabel 6 merupakan hasil estimasi dari PDRB dan pengeluaran dengan metode 2SLS dan 3SLS.

Berdasarkan hasil dari Tabel 6 dapat diketahui bahwa secara keseluruhan, *standar error* atau galat baku dari metode 3SLS memiliki nilai yang lebih rendah dibandingkan hasil dari metode 2SLS. Selain standar error, dapat diketahui juga melalui hasil koefisien determinasi (R^2) untuk masing-masing persamaan dengan menggunakan metode 2SLS dan 3SLS.

Hasil R^2 dari persamaan 2SLS untuk persamaan PDRB adalah sebesar 0.7991, sedangkan pada persamaan pengeluaran adalah sebesar 0.8546. Persamaan 3SLS menghasilkan nilai R^2 untuk persamaan PDRB adalah sebesar 0.7996 dan persamaan pengeluaran memiliki besar R^2 yaitu 0.8548. Koefisien determinasi persamaan simultan 2SLS dan 3SLS tersebut lebih tinggi dari persamaan tunggal OLS, dimana persamaan PDRB adalah sebesar 0.0819, sedangkan pada persamaan pengeluaran adalah sebesar 0.2597. Hasil perbandingan estimasi antara kedua metode persamaan simultan dan metode persamaan tunggal, berdasarkan koefisien determinasi menunjukkan

bahwa kedua model persamaan simultan jauh lebih baik dalam menjelaskan variabilitas kedua variabel endogen, PDRB dan pengeluaran, daripada dengan model persamaan tunggal, demikian pula antara 3SLS terhadap 2SLS.

Selain itu, hasil estimasi kedua metode persamaan simultan, 2SLS dan 3SLS, menunjukkan hasil relatif lebih sama dibandingkan dengan hasil estimasi metode persamaan tunggal OLS. Lebih dari itu, metode estimasi persamaan simultan 3SLS ditunjukkan secara empiris bersifat lebih efisien (varians minimum) dan lebih baik dalam menjelaskan variabilitas kedua variabel endogen dibandingkan metode 2SLS.

Tabel 6. Estimasi parameter persamaan simultan 2SLS dan 3SLS serta persamaan tunggal OLS untuk PDRB & Pengeluaran

Variabel	Koefisien		
	2SLS	3SLS	OLS
Persamaan PDRB			
x_1	0.66	0,67	-5.7
x_2	-3.75	-3,61	232
y_2	1.04	1,05	2612
konstanta	45.66	38.15	-27758
Persamaan Pengeluaran			
x_3	4,48	4.50	0.26
x_4	-1.96	-2.04	-0.70
y_1	0.10	1.00	0.00

Variabel	Standar Error		
	2SLS	3SLS	OLS
Persamaan PDRB			
x_1	0.08	0.08	61.7
x_2	2.16	2.15	1275
y_2	0.29	0.29	1570
konstanta	115.29	115.20	60842
Persamaan Pengeluaran			
x_3	0,41	0.41	0.10
x_4	1.36	1.35	0.31
y_1	0.03	0.03	0.00
konstanta	26.20	6.20	6.03

Secara keseluruhan, berdasarkan hasil *standar error* dan R^2 diketahui bahwa 3SLS memiliki hasil yang lebih baik dibandingkan dengan 2SLS. Analisis selanjutnya adalah menggunakan hasil persamaan simultan dengan 3SLS.

Persamaan PDRB (y_1) yang didapatkan melalui 3SLS yaitu.

$$y_1 = 0,6650209x_1 - 3,606739x_2 + 1,051091y_2 + 38,15429$$

Berdasarkan persamaan yang didapatkan jika jumlah x_1 (usaha mikro kecil atau UMK) bertambah satu satuan atau seribu unit maka secara rata-rata y_1 (PDRB) akan bertambah sebesar 0,6665 triliun rupiah, dengan menganggap variabel independen lain konstan. UMK dianggap sangat berpengaruh terhadap pertumbuhan ekonomi di Indonesia dan memiliki peran penting karena UMK memiliki karakteristik positif sebagai sektor yang mampu menyediakan lapangan pekerjaan yang besar dan mengakomodasi peran masyarakat dan dominan dalam struktur ekonomi. Seperti misalnya, menurut Nugraha dkk. (2019), tahun 2019 terdapat 51.300.000 unit UMKM dan memberikan kontribusi terhadap PDB untuk Rp 2.609 triliun atau mencapai 55.6% PDB dipengaruhi oleh adanya UMKM, sehingga dengan semakin banyaknya UMK yang ada maka akan semakin bertambah pula PDRB. Hasil ini sesuai dengan teori yang ada.

Selanjutnya, terdapat variabel x_2 (ratio ketergantungan) yang diduga mempengaruhi y_1 , hasil yang didapat adalah jika x_2 bertambah satu persen maka secara rata-rata y_1 akan berkurang sebanyak 3.606 triliun rupiah, atau terjadinya perbandingan terbalik. Semakin tinggi ratio ketergantungan di suatu daerah menandakan bahwa beban tersebut semakin menguras daya tahan ekonomi per keluarga di daerah tersebut, pendapatan penduduk yang produktif akan dialokasikan untuk membantu membiayai konsumsi dan kebutuhan masyarakat yang tidak produktif sehingga tidak ada pendapatan yang membantu meningkatkan pertumbuhan ekonomi. Hasil ini sesuai dengan penelitian yang dilakukan oleh Syamsudin (2013), penelitian yang dilakukan untuk mengetahui faktor kependudukan yang berpengaruh terhadap pertumbuhan ekonomi atau PDRB, dan menghasilkan bahwa ratio kependudukan berpengaruh terhadap PDRB dan memiliki hubungan yang berbanding terbalik.

Hubungan variabel y_2 (pengeluaran per kapita) terhadap y_1 didapatkan hasil yang berbanding lurus atau jika y_2 meningkat seratus ribu rupiah maka secara rata-rata pada y_1 juga akan terjadi peningkatan sebesar 1.051 triliun rupiah. Menurut teori Keynes untuk meningkatkan pertumbuhan ekonomi yang diukur melalui pendapatan nasional atau PDRB diperlukan peningkatan dari beberapa faktor ekonomi, salah satunya adalah peningkatan permintaan konsumsi yang dilakukan rumah tangga. Semakin tinggi pendapatan yang diterima maka akan semakin meningkat juga pengeluaran yang

dilakukan, walaupun mengalami peningkatan kecil. Sesuai dengan hasil penelitian yang dilakukan oleh Hakib (2019). bahwa konsumsi atau pengeluaran yang dilakukan rumah tangga di Sulawesi Selatan memiliki pengaruh signifikan dan berpengaruh positif terhadap pertumbuhan ekonomi atau PDRB. Juga terdapat penelitian yang dilakukan oleh Ambarawati (2014), bahwa hubungan antara PDRB dan konsumsi rumah tangga akan memiliki hubungan yang positif dan signifikan.

Selain PDRB terdapat persamaan lain yaitu pengeluaran dengan hasil persamaan yaitu.

$$y_2 = 4,494995x_3 - 2,041822x_4 + 0,1006591y_1 - 223,3324$$

Berdasarkan persamaan yang didapatkan jika jumlah x_3 (IPM) bertambah satu persen maka secara rata-rata y_2 (pengeluaran) juga akan bertambah 4,494 ratus ribu rupiah, hal ini dikarenakan jika IPM bertambah maka akan menunjukkan adanya peningkatan kesejahteraan hidup masyarakatnya, seperti mendapat pendapatan yang cukup, jika keluarga memiliki pendapatan yang cukup maka akan meningkatkan konsumsi per rumah tangga untuk mencukupi kebutuhan hidupnya. Pengeluaran per kapita dan PDRB memiliki hubungan yang berbanding lurus yang menandakan bahwa jika PDRB meningkat maka pengeluaran juga akan meningkat, PDRB berbanding lurus dengan IPM karena IPM menunjukkan Semakin banyak masyarakat yang dapat merasakan pembangunan yang baik karena akan semakin banyak orang yang mampu melakukan konsumsi untuk kehidupan mereka, sehingga IPM juga akan berbanding lurus dengan pengeluaran per kapita.

Selanjutnya terdapat hubungan antara x_4 (tingkat pengangguran terbuka atau TPT) dengan y_2 , dan didapatkan hasil perbandingan terbalik, yaitu jika TPT bertambah satu persen maka secara rata-rata akan adanya pengurangan pengeluaran sebesar 2.041 ratus ribu rupiah. Pengangguran yang semakin meningkat di suatu daerah akan melemahkan perekonomian di daerah tersebut, dan pengangguran menandakan tidak adanya pendapatan yang dapat diterima oleh masyarakat. Pendapatan seseorang akan berbanding lurus dengan kemampuan orang tersebut untuk melakukan konsumsi, dengan tidak adanya pendapatan yang diterima orang yang menganggur maka akan berkurangnya kegiatan konsumsi serta pengeluaran juga berkurang, hasil pada penelitian ini sesuai dengan teori yang ada.

Sama dengan persamaan sebelumnya yaitu PDRB terhadap pengeluaran, pada persamaan pengeluaran terhadap PDRB juga memiliki

hubungan yang positif atau searah karena jika PDRB meningkat maka pengeluaran per kapita juga meningkat. Konsumsi merupakan salah satu komponen utama dari PDRB, karena semakin tinggi pendapatan rumah tangga maka akan semakin tinggi tingkat konsumsinya, dan dengan melakukan konsumsi barang atau jasa akan membantu meningkatkan perkonomian, yaitu PDRB. Konsumsi atau pengeluaran per kapita dalam PDRB mencapai sekitar 60% dari total konsumsi setiap tahunnya (Ambarawati, 2014).

D. Pengujian Signifikansi Parameter PDRB dan Pengeluaran Model 3SLS

Pengujian statistik dapat dilakukan dengan menggunakan uji serentak, uji parsial, dan koefisien determinasi. Pengujian dilakukan pada hasil persamaan simultan 3SLS. Uji serentak untuk mengetahui variabel independen secara serentak dapat berpengaruh terhadap variabel dependen atau tidak. Uji serentak untuk persamaan PDRB dan menghasilkan *p-value* keduanya sebesar 0.000, dengan nilai F masing-masing 136.24 dan 209.4. Nilai *p-value* yang lebih kecil dibandingkan taraf signifikan, yaitu sebesar 0.15. Didapatkan keputusan tolak H_0 sehingga akan didapat kesimpulan bahwa variabel eksogen pada persamaan y_2 secara serentak dapat menjelaskan variabel dependen.

Setelah uji serentak, dilakukan juga uji parsial untuk mengetahui pengaruh masing-masing variabel eksogen terhadap variabel endogennya. Berikut adalah hasil uji z pada persamaan y_1 dengan metode 3SLS yang dapat dilihat pada Tabel 7.

Tabel 7. Uji parsial PDRB 3SLS

Variabel	Koefisien	Standar Error	Nilai z	p-value
Persamaan PDRB				
x_1	0.67	0.08	8.83	0.00
x_2	-3,61	2.15	-1.67	0.09
y_2	1.05	0.29	3.65	0.00
Persamaan Pengeluaran				
x_3	4.50	0.41	10.97	0,00
x_4	-2.05	1.35	-1.51	0,13
y_1	1,00	0.03	3.08	0,00

Berdasarkan uji parsial Tabel 7, pada persamaan y_1 dan y_2 semua variabel tersebut berpengaruh secara parsial terhadap masing-masing persamaan karena memiliki *p-value* yang lebih kecil dibandingkan taraf signifikan, yaitu 0.15.

Koefisien determinasi digunakan untuk mengukur besar pengaruh variabel eksogen terhadap variabel endogen, semakin besar nilainya semakin besar pengaruh variabel eksogen terhadap variabel endogen. Koefisien determinasi juga diperhatikan dalam pengujian signifikansi parameter, koefisien determinasi variabel y_1 memiliki nilai R^2 sebesar 0.7996, yang menandakan bahwa variabel-variabel eksogen pada persamaan y_1 , yaitu x_1 dan x_2 , dapat mempengaruhi y_1 sebesar 79.96%. Sedangkan variabel y_2 memiliki nilai R^2 sebesar 0.8548, nilai tersebut cukup tinggi dan menandakan bahwa variabel-variabel pada persamaan y_2 , yaitu x_3 dan x_4 , berpengaruh terhadap y_2 sebesar 85.48%.

E. Pemeriksaan dan Uji Asumsi Residual Persamaan PDRB dan Pengeluaran

Pemeriksaan multikolinearitas dilakukan pada persamaan PDRB dan pengeluaran yang dapat. Hasil menunjukkan bahwa semua variabel, pada persamaan y_1 dan y_2 memiliki nilai VIF yang lebih kecil dari 10 sehingga tidak terjadi multikolinearitas. Dilakukan uji asumsi residual yang dihasilkan pada pemodelan persamaan simultan antara PDRB dan pengeluaran. Uji distribusi normal dilakukan dengan melihat hasil uji *Kolmogorov-Smirnov*. Berdasarkan pada hasil uji normalitas terhadap residual didapatkan *p-value* sebesar 0.707 dan 0.919, yang lebih besar dibandingkan dengan taraf signifikan, yaitu sebesar 0.15, sehingga didapatkan kesimpulan bahwa residual kedua persamaan sudah berdistribusi normal.

Uji identik digunakan untuk mengetahui residual dari hasil regresi bersifat identik atau tidak. Diketahui bahwa residual dari persamaan y_2 bersifat identik karena memiliki nilai *p-value*, yaitu 0.382, hasil tersebut lebih besar dibandingkan dengan taraf signifikan, yaitu sebesar 0.15, berdasarkan hasil tersebut maka persamaan y_2 simultan memiliki residu yang tidak mengalami heterokedastisitas. Namun, persamaan y_1 memiliki *p-value* yang lebih kecil dibanding taraf signifikan sehingga residual, yaitu sebesar 0.000, mengalami heterokedastisitas, pada penelitian ini diasumsikan residual tidak mengalami heterokedastisitas.

F. Pengelompokan Estimasi PDRB dan Pengeluaran

Setelah mengetahui persamaan yang didapatkan melalui persamaan simultan untuk PDRB dan pengeluaran, maka akan didapatkan nilai estimasi atau nilai perkiraan PDRB dan pengeluaran pada daerah-daerah di Jawa Timur. Perbandingan

antara hasil estimasi dengan nilai aktual PDRB dan dikonfirmasi memiliki hasil yang sama hanya terdapat empat daerah atau hanya sebesar 40% yang sesuai, yaitu Kabupaten Magetan, Kabupaten Trenggalek, Kabupaten Pamekasan, dan Kabupaten Pacitan. Estimasi daerah yang termasuk dalam daerah dengan PDRB terendah yaitu sebesar 60% daerah terendah dapat menjadi informasi atau peringatan awal agar daerah-daerah tersebut lebih diperhatikan, yaitu meliputi Kabupaten Bondowoso, Kabupaten Madiun, Kabupaten Ponorogo, Kabupaten Bangkalan, Kabupaten Ngawi, dan Kabupaten Sampang.

Hasil estimasi persamaan pengeluaran dikonfirmasi hanya memiliki kesesuaian sebesar 30% atau tiga daerah yang sama antara hasil estimasi dengan nilai aktual dan berada pada sepuluh daerah dengan pengeluaran terendah yaitu Kabupaten Sampang, Kabupaten Sumenep, dan Kabupaten Bangkalan. Sama seperti pada PDRB, hasil estimasi pengeluaran daerah yang termasuk dalam sepuluh terendah yaitu sebesar 70% daerah dengan estimasi pengeluaran terendah dapat menjadi informasi atau peringatan awal yang perlu diperhatikan, meliputi Kabupaten Pacitan, Kabupaten Bondowoso, Kabupaten Probolinggo, Kabupaten Lumajang, Kabupaten Situbondo, Kabupaten Jember, dan Kabupaten Pamekasan.

4. KESIMPULAN DAN SARAN

Berdasarkan analisis yang telah dilakukan, metode persamaan simultan yang digunakan pada penelitian ini adalah *3SLS* karena secara empiris menunjukkan hasil yang lebih baik dibandingkan *2SLS* yaitu memiliki *standar error* yang lebih rendah dan nilai R^2 yang lebih tinggi nilainya. Sebagai pendukung juga ditunjukkan bahwa metode estimasi persamaan simultan *2SLS* dan *3SLS* lebih baik daripada estimasi persamaan tunggal *OLS*, dengan koefisien determinasi lebih tinggi dan nilai estimasi yang lebih sama. Berdasarkan hasil persamaan simultan *3SLS* persamaan PDRB, semua hubungan sesuai dengan teori yaitu pada jumlah UMK (x_1) yang tinggi juga akan meningkatkan PDRB, serta ratio ketergantungan (x_2) yang menurun akan meningkatkan PDRB. Pengeluaran (y_2) yang meningkat juga akan meningkatkan PDRB. Persamaan simultan untuk pengeluaran menunjukkan bahwa semakin tinggi IPM (x_3) akan meningkatkan pengeluaran, dan saat pengangguran (x_4) menurun maka pengeluaran akan meningkat. PDRB (y_1) akan semakin meningkatkan pengeluaran juga karena memiliki hubungan positif. Hasil persamaan pengeluaran juga sudah sesuai teori yang ada. Daerah yang memiliki nilai estimasi

PDRB terendah, khususnya yang sesuai dengan data aktual adalah Magetan, Trenggalek, Pamekasan, dan Pacitan. Daerah dengan nilai estimasi pengeluaran yang rendah dan sesuai dengan data aktual pengeluaran adalah Sampang, Sumenep, dan Bangkalan.

Saran yang dapat direkomendasikan untuk penelitian selanjutnya adalah lebih mengkaji beberapa faktor yang berpengaruh terhadap PDRB dan pengeluaran. Bagi pemerintah, sebaiknya mempercepat pengentasan kemiskinan atau melakukan percepatan pertumbuhan ekonomi daerah berpenghasilan rendah, seperti daerah Magetan, Trenggalek, Pamekasan, dan Pacitan, sehingga akan meningkatkan PDRB.

DAFTAR PUSTAKA

- Ambarawati, N. (2014). Analisis Kausalitas Antara Konsumsi Rumah Tangga dengan PDRB Perkapita di Jawa Tengah Periode Tahun 1986-2011. *Naskah Publikasi Fakultas Ekonomi dan Bisnis Universitas Muhammadiyah Surakarta*.
- BPS Jatim (2016), "Potensi Peningkatan Kinerja Usaha Mikro Kecil Provinsi Jawa Timur," *Analisis Hasil SE2016 Lanjutan*.
- Greene, W.H. (2012), *Econometric Analysis 7th Edition*, New Jersey : Prentice Hall.
- Hakib, A. (2019). Pengaruh Konsumsi Rumah Tangga dan Pengeluaran Pemerintah terhadap Pertumbuhan Ekonomi di Sulawesi Selatan Periode 2012-2016. *Jurnal Ekonomi Balance Fakultas Ekonomi dan Bisnis*, 56-71.
- Merdikawati, N. dkk. (2014), *Analisis Daya Saing Provinsi dan Wilayah : Menjaga Momentum Pertumbuhan Indonesia Edisi 2014*, Singapore : World Scientific Publishing.
- Misno dan Sulistianingsih, E. (2019), "Estimasi Model Persamaan Simultan dengan Metode Two Stage Least Square (2SLS)," *Buletin Ilmiah Math, Stat, dan Terapan (Bimaster)*, pp. hal 653-658.
- Nugraha, X., dkk. (2019). *Bunga Rampai Ilmu Hukum Masyarakat Yuris Muda Airlangga*. Yogyakarta: Penerbit Harfeey.
- Oky, R., Widiyah, T., dan Hoyyi, A. (2012), "Pendekatan Sistem Persamaan Simultan dalam Pemodelan Produk Domestik Regional Bruto (PDRB) Provinsi Jawa Tengah Tahun 2000-2010," *Jurnal Gaussian*, pp. hal 199-208.
- Lies Maria Hamzah, D. A. (2019). Pengaruh Perkembangan Usaha Mikro, Kecil, dan Menengah Terhadap Pendapatan Nasional pada Sektor UMKM di Indonesia. *Jurnal Ekonomi Pembangunan*, 215-228.
- Pramono, W.D. dan Suminar, R.E. (2019), *Ekonomi Wilayah untuk Perencanaan Tata Ruang*, Yogyakarta : Penerbit Deepublish.
- Putong, I. (2015), *Ekonomi Mikro : Pengantar Ilmu Ekonomi Makro (Volume 1)*, Jakarta : Buku&Artikel Karya Iskandar Putong.
- Seddighi, H. and Lawler, K. (2000), *Economics : A Practical Approach*, London : Routledge.
- Supartoyo, Y.H. and Sendouw, R.H. (2013), "The Economic Growth and The Regional Characteristics : The Case of Indonesia," *Buletin Ekoonomi Moneter dan Perbankan*,
- Syamsuddin, H. (2013). Analisis Pengaruh Faktor Kependudukan terhadap Pertumbuhan Ekonomi Provinsi Jambi. *Jurnal Paradigma Ekonomika*, vol. 1 no. 7, 73-84.

Penerapan *Clustering Time Series* pada Pengelompokan Provinsi di Indonesia (Studi Kasus : Nilai Ekspor Non Migas di Indonesia Tahun 2016-2020)

Harista Almiatus Soleha⁽¹⁾, Wiwit Pura Nurmawanti⁽²⁾, Umam Hidayaturrohman⁽³⁾, Ristu Haiban Hirzi⁽⁴⁾, Ayu Septiani⁽⁵⁾

Program Studi Statistika, Universitas Hamzanwadi

Jl. TGKH M. Zainuddin Abdul Madjid No. 132 Pancor (83611) Selong-Lombok Timur-NTB

e-mail: harista.almiatus@gmail.com, wiwit.adiwinata3@gmail.com,

hidayaturohman4@gmail.com, ristuastalavista@gmail.com dan septianisuccespg@gmail.com

ABSTRAK

Ekspor non migas merupakan ekspor barang yang bukan berupa minyak dan gas. Tidak semua daerah di Indonesia memiliki potensi yang sama untuk melakukan kegiatan ekspor sehingga setiap daerah memiliki nilai ekspor yang berbeda-beda. Oleh karena itu dilakukan analisis pengelompokan provinsi di Indonesia berdasarkan nilai ekspor non migas tahun 2016 – 2020 menggunakan *cluster time series* dengan metode hierarki *clustering agglomerative* diantaranya *complete linkage* yaitu pengelompokan berdasarkan jarak terbesar antar objek dan *centroid linkage* merupakan nilai tengah observasi pada variabel dalam satu *cluster*, serta menggunakan pengukuran jarak *Euclidean* dan *Dynamic Time Warping* (DTW) dengan tujuan yaitu untuk memperoleh pengukuran jarak terbaik dengan metode yang optimal guna mendapatkan *cluster* yang representatif. Berdasarkan hasil analisis pengelompokan Provinsi di Indonesia terhadap nilai ekspor non migas tahun 2016 - 2020 didapatkan pengukuran jarak kemiripan yang paling baik yaitu jarak DTW dan metode yang optimal yaitu *centroid linkage* berdasarkan pada nilai koefisien *cophenetic* sebesar 0.92 dengan kategori *good cluster* berdasarkan nilai koefisien *silhouette* yaitu 0.60. Sehingga didapatkan tiga *cluster* yaitu nilai ekspor tinggi, sedang dan rendah, diantaranya 24 Provinsi dengan nilai ekspor rendah, 9 Provinsi dengan nilai ekspor sedang, dan 1 Provinsi dengan nilai ekspor tinggi.

Kata kunci : *cluster time series, dynamic time warping, euclid, hierarchical clustering, ekspor.*

ABSTRACT

Non-oil and gas exports are exports of goods that are not in the form of oil and gas. Not all regions in Indonesia have the same potential to carry out export activities so that each region has a different export value. Therefore, an analysis of grouping provinces in Indonesia based on the value of non-oil and gas exports in 2016 - 2020 uses cluster time series with an agglomerative clustering hierarchical method including complete linke, namely grouping based on the largest distance between objects and centroid linkage, which is the median value of observations on variables in one cluster. as well as using Euclidean distance measurement and Dynamic Time Warping (DTW) with the aim of obtaining the best distance measurement with the optimal method to get representative clusters. Based on the results of the analysis of the grouping of provinces in Indonesia on the value of non-oil and gas exports in 2016 - 2020, the best similarity distance measurement was obtained, namely the DTW distance and the optimal method, namely centroid linkage based on a cophenetic coefficient value of 0.92 with the good cluster category based on a silhouette coefficient value of 0.60 . So that three clusters were obtained, namely high, medium and low export values, including 24 provinces with low export values, 9 provinces with medium export values, and 1 province with high export values.

Keywords : *cluster time series, dynamic time warping, euclid, hierarchical clustering, export.*

1. PENDAHULUAN

Indonesia merupakan negara yang sedang melakukan segala upaya dalam mengembangkan pembangunan disegala sektor, salah satunya yaitu sektor ekonomi. Perkembangan ekonomi suatu

Negara tidak bisa lepas dari perekonomian global. Hubungan ekonomi antar negara sangat penting dan berpengaruh terhadap perkembangan ekonomi negara tersebut. Salah satu kegiatan yang dapat meningkatkan perekonomian suatu negara yaitu

kegiatan ekspor. Ekspor adalah perdagangan dengan cara mengeluarkan barang dari dalam ke luar wilayah pabean Indonesia dengan memenuhi ketentuan yang berlaku (Feriyanto A, 2015). Selain dapat meningkatkan perekonomian negara ekspor juga bisa menambah devisa negara yang dibutuhkan untuk membayar hutang luar negeri.

Salah satu komoditas ekspor yang sangat berperan di Indonesia yaitu ekspor non migas atau yang bukan berasal dari pengolahan minyak dan gas, seperti hasil pertanian, perkebunan dan lain – lain. Penentu ekspor yang penting adalah kemampuan suatu negara untuk memproduksi barang yang kompetitif di pasar luar negeri. Artinya, kualitas dan harga barang ekspor minimal harus sama baiknya dengan barang yang dijual di pasar luar negeri. Selera masyarakat di luar negeri tentang barang yang dapat diekspor dari suatu negara memegang peranan yang sangat penting dalam menentukan ekspor negara tersebut (Tumanggor, 2019).

Namun, tidak semua daerah di Indonesia memiliki potensi yang sama untuk melakukan kegiatan ekspor tersebut. Oleh karena itu peneliti melakukan analisis pengelompokan provinsi di Indonesia berdasarkan nilai ekspor non migas tahun 2016 – 2020 menjadi tiga kelompok yaitu tinggi, sedang dan rendah menggunakan *cluster time series*. Pengelompokan provinsi di Indonesia berdasarkan nilai ekspor non migas tahun 2016 – 2020 dapat menggunakan analisis *Cluster time series* yang merupakan pengelompokan pada data runtun waktu. Metode analisis *cluster* yang digunakan yaitu metode *complete linkage* dan *centroid linkage* serta pengukuran jarak *Euclidean* dan *Dynamic Time Warping* (DTW). Metode *complete linkage* digunakan karena pengelompokan dilakukan berdasarkan jarak terbesar antar objek, sedangkan penggunaan *centroid linkage* karena pada metode tersebut data *outlier* tidak berpengaruh signifikan (Silvi, 2018). Penggunaan pengukuran jarak *euclidean* sangat umum digunakan pada pengelompokan sedangkan DTW lebih realistis digunakan untuk mengukur suatu pola dalam hal ini data deret waktu (Ardiansyah, 2014). Sehingga kedua jarak dan metode tersebut dibandingkan guna mengetahui hasil pengelompokan terbaik untuk nilai ekspor non migas di Indoensia pada tahun 2016 - 2020.

Beberapa penelitian terkait juga pernah dilakukan diantaranya oleh Putu Widya Adnyani & Robinson Sihombing (2021) dengan menggunakan jarak *Euclid* dan DTW dan metode *complete linkage* untuk mengelompokkan provinsi di Indonesia berdasarkan nilai PDRB. Hasil penelitian tersebut

menunjukkan bahwa pengukuran jarak DTW lebih baik dan menghasilkan sebanyak 6 *cluster* dengan 8 Provinsi pada *cluster* pertama, 4 Provinsi pada *cluster* kedua, 3 Provinsi pada *cluster* ketiga, 15 Provinsi pada *cluster* keempat, 3 Provinsi pada *cluster* kelima dan 1 provinsi pada *cluster* ke enam. Penelitian serupa juga dilakukan oleh Munthe (2019) yang melakukan *cluster time series* pada provinsi di Indonesia berdasarkan nilai produksi padi. Analisis tersebut menghasilkan tiga kelompok, dimana ketiga kelompok tersebut menghasilkan nilai koefisien *silhouette* sebesar 0,64 yang artinya pengelompokan tersebut termasuk kedalam kategori *good cluster*.

Berdasarkan latar belakang tersebut peneliti tertarik untuk melakukan pengelompokan runtun waktu dengan membandingkan dua metode yaitu *complete linkage* dan *centroid linkage* serta pengukuran jarak *Euclidean* dan *Dynamic Time Warping* (DTW) pada studi kasus nilai ekspor non migas tahun 2016 – 2020.

2. METODE PENELITIAN

2.1 Analisis Cluster Time Series

Analisis *cluster* adalah teknik untuk menemukan kelompok dalam suatu kumpulan data, dengan tujuan agar data dalam satu kelompok sangat mirip dan berbeda secara signifikan dengan kelompok lainnya (Putu Widya Adnyani & Robinson Sihombing, 2021). *Time series clustering* adalah salah satu jenis algoritma *clustering* yang dibuat untuk menangani data dinamis. *Time series clustering* bisa digunakan sebagai algoritma utama, sebagai tahap lanjutan, ataupun sebagai tahap *preprocessing* (Shirkhorshidi et al., 2015). Dalam perkembangan pengelompokan data deret waktu, banyak teknik yang dikembangkan, seperti penggunaan jarak pengelompokan yang sesuai dengan karakteristik data deret waktu (Wijaya & Ngatini, 2020).

2.2. Metode Analisis Cluster

Jika ditinjau dari metode pengelompokannya, analisis klaster memiliki dua metode yang dapat digunakan yaitu metode hirarki dan non-hirarki. Metode pengelompokan hirarki, merupakan metode pengelompokan yang akan membentuk sebuah pohon atau dendogram berdasarkan kemiripannya. Cara pengelompokan hirarki pada umumnya ada dua yaitu *agglomerative* (pemusatan) dan *divisive* (penyebaran). Adapun pada metode *agglomerative* terdapat beberapa algoritma yaitu seperti *single linkage*, *complete linkage*, *average linkage*, *ward method* dan *centroid linkage*.

Proses analisis pada penelitian ini menggunakan metode *centroid linkage* dan akan dibandingkan dengan metode *complete linkage*. *Centroid linkage* merupakan rata – rata dari semua objek dalam *cluster*, dengan menggunakan metode *centroid linkage* jika terbentuk *cluster* baru maka terjadi perhitungan ulang centroid sampai terbentuk *cluster* tetap (Rahmawati et al., 2013). Metode tersebut dipilih karena pada metode tersebut data *outlier* tidak berpengaruh. Kelebihan dari metode ini adalah *outlier* tidak memiliki pengaruh yang signifikan dibandingkan dengan metode lainnya (Silvi, 2018). *Complete linkage* merupakan metode pengelompokan *agglomerative* berdasarkan jarak terbesar antar objek (Dani et al., 2019).

Hasil klaster dikatakan baik jika memiliki ciri – ciri sebagai berikut (Halim & Widodo, 2017):

1. *Within cluster* yaitu jika anggota pada klaster yang sama memiliki homogenitas atau kesamaan yang tinggi.
2. *Between cluster* yaitu jika antar klaster yang satu dengan klaster yang lainnya memiliki perbedaan atau heterogenitas yang tinggi.

2.3. Pengukuran Kemiripan

a. *Euclidean Distance*

Euclidean distance adalah salah satu jenis pengukuran jarak dalam analisis *cluster* yang mengukur jarak dari objek data ke pusat *cluster*. Jarak *Euclidean* adalah jarak geometris antara dua objek data. Semakin kecil jaraknya, semakin mirip objek datanya (Dani et al., 2019).

b. *Dynamic Time Warping Distance*

Dynamic Time Warping (DTW) adalah algoritma untuk menemukan jalur *warp* yang optimal antara dua data deret waktu sehingga hasilnya adalah kumpulan nilai jalur *warp* dan jarak antara dua data deret waktu. Jarak DTW adalah ukuran ketidaksamaan yang melibatkan perkiraan tanpa model tertentu dan merupakan jarak terkecil antara dua pasang titik mengingat kemungkinan migrasi titik (Putu Widya Adnyani & Robinson Sihombing, 2021)

2.4 Validitas Cluster

a. Koefisien Korelasi *Cophenetic*

Koefisien korelasi *cophenetic* adalah koefisien korelasi antara elemen asli dari matriks ketidakmiripan dengan elemen yang dihasilkan oleh matriks *cophenetic* (da Silva & Dias, 2013).

b. Koefisien *Silhouette*

Objek dapat dimasukkan dalam sebanyak *k* kelompok. Batasan kelompok yang dapat dibentuk

adalah $k \leq n$ dengan $k = 1$ dan $k = n$ dikecualikan. Sehingga dapat disimpulkan bahwa $2 \leq k \leq (n - 1)$. Adapun kriteria ketepatan dan kualitas pengelompokan berdasarkan nilai koefisien *Silhouette* dapat dinyatakan seperti pada tabel berikut menurut Kaufman dan Rausseeuw dalam (Dewi dan Pramita., 2019):

Tabel 1. Kategori Penggerombolan Berdasarkan Koefisien *Silhouette*

Koefisien <i>Silhouette</i>	Kategori Penggerombolan
0,71-1,00	<i>Strong Cluster</i>
0,51-0,70	<i>Good Cluster</i>
0,26-0,50	<i>Weak Cluster</i>
0,00-0,25	<i>Bad Cluster</i>

2.5 Data dan Langkah -langkah Analisis

Data yang digunakan dalam penelitian ini yaitu data sekunder nilai Ekspor Non Migas 34 Provinsi di Indonesia tahun 2016 – 2020 dan diperoleh dari *website* resmi Statistik Kementerian Perdagangan Republik Indonesia (<https://hero.kemendag.go.id/>)

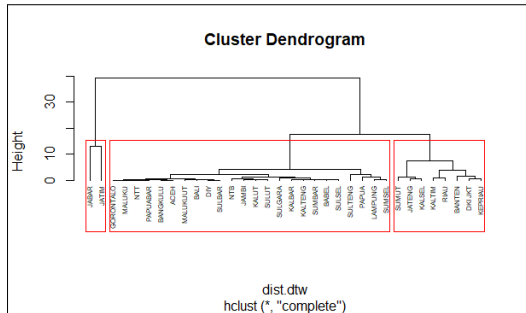
Metode yang digunakan untuk mengelompokkan data runtun waktu 34 provinsi di Indonesia berdasarkan nilai Ekspor Non Migas pada tahun 2016 – 2020 yaitu *complete linkage* dan *centroid linkage* dengan pengukuran jarak *Euclid* dan *Dynamic Time Warping (DTW)*. Dengan penentuan jarak dan algoritma yang optimal menggunakan nilai koefisien *cophenetic* serta menguji ketepatan kelompok menggunakan koefisien *silhouette*. Adapun langkah-langkahnya sebagai berikut :

1. Melakukan analisis deskriptif atau melakukan eksplorasi data ekspor non migas tahun 2016-2020 guna mengetahui gambaran umum mengenai data tersebut.
2. Melakukan standarisasi data jika ditemukan adanya data *outlier* maupun data yang memiliki satuan yang berbeda.
3. Selanjutnya yaitu melakukan analisis *cluster* menggunakan metode *complete linkage* dan *centroid linkage* dengan menggunakan $k=3$ dengan matriks jarak pengukuran *euclidean* dan DTW.
4. Menentukan metode dan pengukuran jarak terbaik menggunakan nilai korelasi *cophenetic*, dan melakukan uji kualitas *cluster* yang terbentuk.

3. HASIL DAN PEMBAHASAN

3.1 Pengelompokan menggunakan *complete linkage* dan *centroid linkage* menggunakan jarak DTW

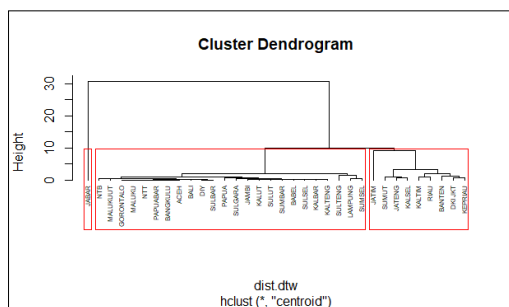
a. *Complete Linkage*



Gambar 1. Complete Linkage Jarak DTW

Berdasarkan dendrogram di atas terlihat bahwa terbentuk tiga kelompok dengan masing masing *cluster* pertama terdiri dari 24 Provinsi diantaranya Gorontalo, Maluku, NTT, Papua Barat, Bengkulu, Aceh, Maluku Utara, Bali, DIY, Sulawesi Barat, NTB, Jambi, Kalimantan Utara, Sulawesi Utara, Sulawesi Tenggara, Kalimantan Barat, Kalimantan Tengah, Sumatera Barat, Bangka Belitung, Sulawesi Selatan, Sulawesi Tengah, Papua, Lampung dan Sumatera Selatan. Kemudian *cluster* kedua terdiri dari 8 Provinsi Sumatra Utara, Jawa Tengah, Kalimantan Selatan, Kalimantan Timur, Riau, Banten, DKI Jakarta, Kepulauan Riau. Dan *cluster* ketiga terdiri dari 2 Provinsi yaitu Jawa Barat dan Jawa Timur.

b. *Centroid Linkage*



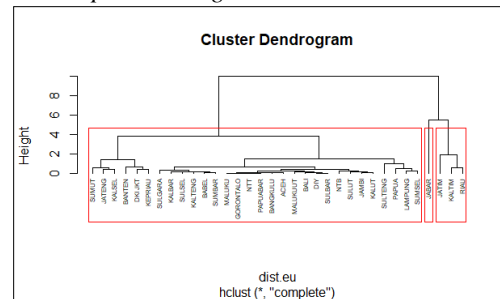
Gambar 2. Centroid Linkage Jarak DTW

Adapun hasil pengelompokan menggunakan metode *centroid linkage* dengan menggunakan jarak kemiripan DTW menghasilkan tiga *cluster* dengan *cluster* pertama terdiri dari 24 Provinsi yaitu NTB, Maluku Utara, Gorontalo, Maluku, NTT, Papua Barat, Bengkulu, Aceh, Bali, DIY, Sulawesi Barat,

Papua, Sulawesi Tenggara, Jambi, Kalimantan Utara, Sulawesi Utara, Sumatra Barat, Bangka Belitung, Sulawesi Selatan, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tenggara, Lampung, Sumatera Selatan. *Cluster* kedua terdiri dari 9 Provinsi Jawa Timur, Sumatra Utara, Jawa Tengah, Kalimantan Selatan, Kalimantan Timur, Riau, Banten, DKI Jakarta, Kepulauan Riau. Dan *cluster* ketiga terdiri dari 1 Provinsi yaitu Jawa Barat.

3.2 Pengelompokan menggunakan *complete linkage* dan *centroid linkage* menggunakan jarak euclidean

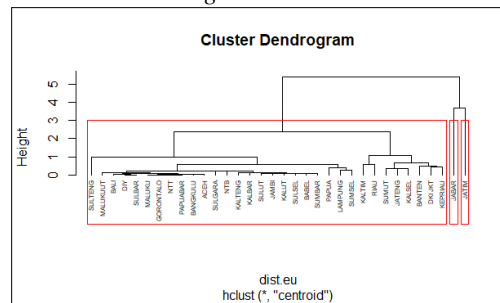
a. *Complete Linkage*



Gambar 3. Complete Linkage Jarak Euclidean

Pada pengelompokan menggunakan pengukuran jarak *euclidean* dengan metode *complete linkage* dengan tiga *cluster* didapatkan sebanyak 30 Provinsi pada *cluster* pertama yaitu Provinsi NTB, Maluku Utara, Gorontalo, Maluku, NTT, Papua Barat, Bengkulu, Aceh, Bali, DIY, Sulawesi Barat, Papua, Sulawesi Tenggara, Jambi, Kalimantan Utara, Sulawesi Utara, Sumatra Barat, Bangka Belitung, Sulawesi Selatan, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tenggara, Lampung, Sumatra Selatan, Sumatra Utara, Jawa Tengah, Kalimantan Selatan, Banten, DKI Jakarta, Kepulauan Riau. 1 Provinsi pada *cluster* kedua yaitu provinsi Jawa Barat, dan 3 Provinsi pada *cluster* ketiga yaitu Provinsi Jawa Timur, Kalimantan Timur, Riau.

a. *Centroid Linkage*



Gambar 4. Centroid Linkage Jarak Euclidean

Metode pengelompokan yang terakhir yaitu metode *centroid linkage* dengan menggunakan pengukuran jarak *euclidean* dan dihasilkan anggota kelompok pada *cluster* pertama yaitu ada 33 Provinsi yaitu NTB, Maluku Utara, Gorontalo, Maluku, NTT, Papua Barat, Bengkulu, Aceh, Bali, DIY, Sulawesi Barat, Papua, Sulawesi Tenggara, Jambi, Kalimantan Utara, Sulawesi Utara, Sumatra Barat, Bangka Belitung, Sulawesi Selatan, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tenggara, Lampung, Sumatra Selatan, Sumatra Utara, Jawa Tengah, Kalimantan Selatan, Banten, DKI Jakarta, Kepulauan Riau, Kalimantan Timur, Riau. Dan untuk *cluster* dua dan tiga terdiri dari 1 Provinsi yaitu provinsi Jawa Barat dan Jawa Timur secara berturut-turut.

3.3 Validasi Cluster

3.3.1 Koefisien Cophenetic

Tabel 2. Perbandingan Nilai Korelasi *Cophenetic*

Jarak	Algoritma	
	<i>Complete</i>	<i>Centroid</i>
DTW	0,912	0,92
<i>Euclid</i>	0,848	0,916

Berdasarkan pada tabel di atas, terlihat bahwa setiap jarak dengan algoritma yang berbeda mempunyai nilai koefisien korelasi *cophenetic* yang berbeda. Nilai koefisien korelasi *cophenetic* berkisar antara -1 sampai 1, dimana jarak pengukuran dalam proses klaster dikatakan cukup baik jika nilai koefisien korelasi *cophenetic*nya mendekati 1.

Dapat disimpulkan bahwa proses pengelompokan Provinsi di Indonesia berdasarkan nilai Ekspor Non migas tahun 2016-2020, jarak pengukuran kemiripan terbaik yaitu *Dynamic Time Warping* (DTW) terlihat dari nilai korelasinya yang lebih besar di kedua algoritma pengelompokan digunakan.

Adapun metode yang paling optimal dalam pengelompokan Provinsi di Indonesia berdasarkan nilai ekspor nonmigas tahun 2016-2020 adalah metode *centroid linkage*, karena memiliki nilai koefisien korelasi *cophenetic* terbesar dari pada algoritma *complete linkage* yaitu sebesar 0,92.

3.3.2 Koefisien Silhouette

Setelah mendapatkan jarak pengukuran kemiripan terbaik yaitu jarak *dynamic time warping* (DTW) serta metode terbaik yaitu *Centroid Linkage*, selanjutnya dilakukan analisis untuk menguji apakah jumlah klaster sama dengan 3 sudah

representatif atau dapat mewakili dalam proses klaster *centroid linkage*. Adapun uji yang digunakan untuk menguji kualitas klaster pada penelitian ini yaitu metode koefisien *silhouette*. Berdasarkan nilai koefisien *silhouette* pada jumlah *cluster* tiga yaitu sebesar 0.60, dimana nilai tersebut menunjukkan bahwa pengelompokan menggunakan metode *centroid linkage* dengan pengukuran jarak kemiripan DTW dengan tiga kategori yaitu “tinggi”, “sedang”, dan “rendah” termasuk kedalam *good cluster*.

4. KESIMPULAN DAN SARAN

Analisis *clustering time series* pada nilai ekspor non migas tahun 2016-2020 menghasilkan beberapa kesimpulan sebagai berikut :

1. Penggunaan metode pengukuran jarak kemiripan *Dynamic Time Warping* pada analisis pengelompokan data *time series* lebih baik daripada pengukuran jarak *euclidean*, dan metode pengelompokan terbaik yaitu metode *centroid linkage* dilihat dari nilai koefisien korelasi *cophenetic*-nya yang lebih tinggi yaitu sebesar 0.92.
2. Adapun untuk pengelompokan provinsi di Indonesia berdasarkan nilai ekspor non migas tahun 2016 hingga 2020 dengan mengelompokkannya menjadi tiga kategori yaitu tinggi, sedang dan rendah menggunakan *centroid linkage* dan jarak DTW termasuk kedalam kategori *good cluster* berdasarkan nilai koefisien *silhouette* sebesar 0.60.
3. Berdasarkan metode terbaik *centroid linkage* dan jarak DTW ada 1 Provinsi masuk kedalam kategori tinggi yaitu Jawa Barat, 9 Provinsi pada kategori sedang diantaranya Jawa Timur, Sumatra Utara, Jawa Tengah, Kalimantan Selatan, Kalimantan Timur, Riau, Banten, DKI Jakarta, Kepulauan Riau dan 24 provinsi termasuk kedalam kategori rendah yaitu NTB, Maluku Utara, Gorontalo, Maluku, NTT, Papua Barat, Bengkulu, Aceh, Bali, DIY, Sulawesi Barat, Papua, Sulawesi Tenggara, Jambi, Kalimantan Utara, Sulawesi Utara, Sumatra Barat, Bangka Belitung, Sulawesi Selatan, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tenggara, dan Lampung,

DAFTAR PUSTAKA

Ardiansyah, M. (2014). Penggunaan Jarak Dynamic Time Warping (DTW) Pada Analisis Cluster Data Deret Waktu (Studi Kasus Pada Dana Pihak Ketiga Provinsi Se- Indonesia). *Jurnal Mahasiswa Statistik*, 2(4), 277–280.

- da Silva, A. R., & Dias, C. T. dos S. (2013). A cophenetic correlation coefficient for tocher's method. *Pesquisa Agropecuaria Brasileira*, 48(6), 589–596. <https://doi.org/10.1590/S0100-204X2013000600003>
- Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1(2), 64–78. <https://doi.org/10.34312/jjom.v1i2.2354>
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi dan Informatika*, 9(3), 102–109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Feriyanto, A. (2015). Perdagangan Internasional “Kupas Tuntas Prosedur Ekspor Impor”. *Yogyakarta: PT. Pustaka Baru*.
- Halim, N. N., & Widodo, E. (2017). Clustering dampak gempa bumi di indonesia menggunakan kohonen self organizing maps. *Prosiding SI MaNIS (Seminar Nasional Integrasi Matematika dan Nilai Islami)*, 1(1), 188–194. <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>
- Kementerian Perdagangan Republik Indonesia, (April 27, 2021), *Citing Internet sources URL* <https://hero.kemendag.go.id/>
- Manso, P. M., Vilar, J. A., & Montero, M. P. (2020). Package ‘TSclust’
- Munthe, A. D. (2019). Penerapan Clustering Time Series Untuk Menggerombolkan Provinsi Di Indonesia Berdasarkan Nilai Produksi Padi. *Jurnal Litbang Sukowati: Media Penelitian dan Pengembangan*, 2(2), 11. <https://doi.org/10.32630/sukowati.v2i2.61>
- Putu Widya Adnyani, L., & Robinson Sihombing, P. (2021). Analisis Cluster Time Series Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Nilai PDRB. *Lppm Bina Bangsa*, 1(1), 47–54. <http://bayesian.lppmbinabangsa.id/index.php/home>
- Rahmawati, L., Abadyo, & Lestari, T. E. (2013). *Analisis Kelompok Dengan Menggunakan*
- Metode Hierarki Untuk Pengelompokan Kabupaten/Kota Di Jawa Timur Berdasar Indikator Kesehatan*. Universitas Negeri Malang.
- Shirkhorshidi, A. S., Aghabozorgi, S., & Ying Wah, T. (2015). A Comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS ONE*, 10(12), 1–20. <https://doi.org/10.1371/journal.pone.0144059>
- Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika “MANTIK,”* 4(1), 22–31. <https://doi.org/10.15642/mantik.2018.4.1.22-31>
- Tumanggor, H. (2019). *Analisis Pengaruh Investasi, Ekspor Dan Konsumsi Terhadap Pdrb Sumatera Utara Tahun 2000-2017*. <http://repository.uhn.ac.id/handle/123456789/3618>
- Wijaya, S. U., & Ngatini, N. N. (2020). Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan Clustering Time Series. *Limits: Journal of Mathematics and Its Applications*, 17(1), 51. <https://doi.org/10.12962/limits.v17i1.5994>

Analisis *Partial Least Square Structural Equation Model* (PLS-SEM) untuk Pemodelan Penerimaan Sistem Jaringan Informasi Bersama Antar Sekolah (JIBAS)

Jerhi Wahyu Fernanda⁽¹⁾, Vira Luthifiana⁽²⁾, M. Khoiril Akhyar⁽³⁾

^{1,2,3} Institut Agama Islam Negeri (IAIN) Kediri

Jl. Sunan Ampel No.7, Ngronggo, Kec. Kota, Kota Kediri, Jawa Timur 64127

e-mail: fernanda.jerhi@iainkediri.ac.id, viralutfiana21@gmail.com, dan akhyar@iainkediri.ac.id

ABSTRAK

Sistem Jaringan Informasi Bersama Antar Sekolah (JIBAS) merupakan sistem yang digunakan dalam pengelolaan institusi pendidikan. Dalam prakteknya, evaluasi sistem ini harus dilakukan untuk mengetahui tingkat penerimaan pengguna terhadap teknologi tersebut. Evaluasi menggunakan konsep *Technology Acceptance Model* (TAM). Penelitian ini dilakukan untuk mengetahui variabel-variabel yang memiliki hubungan signifikan terhadap penerimaan aplikasi JIBAS menggunakan metode *Partial Least Square Structural Equation Model* (PLS-SEM). Total sampel penelitian ini sebanyak 51 responden yang meliputi guru maupun staff administrasi yang pernah berinteraksi dengan JIBAS. Variabel penelitian terdiri dari *Perceived ease of use* (PEOU), *Perceived usefulness* (PU), *Behavioral intention to use* (BI), dan *Actual System Use* (AC). Analisis PLS-SEM menghasilkan kesimpulan PEOU memiliki hubungan yang signifikan terhadap variabel BI dan BI memiliki hubungan yang signifikan terhadap AC. Kesimpulan yang didapatkan dari hasil analisis adalah peningkatan kemudahan aplikasi JIBAS (PEOU) akan berdampak kepada peningkatan perilaku penggunaan (BI) aplikasi JIBAS. Peningkatan perilaku penggunaan aplikasi akan berdampak juga terhadap peningkatan penggunaan (AC) aplikasi JIBAS dalam pengelolaan institusi pendidikan.

Kata kunci : *JIBAS, TAM, Partial Least Square Structural Equation Model*

ABSTRACT

JIBAS is a system used in the management of educational institutions. In practice, an evaluation of this system must be carried out to determine the level of user acceptance of the technology. Evaluation uses the concept of Technology Acceptance Model (TAM). The purpose of this research is to determine the variables that have a significant relationship with the acceptance of the JIBAS using Partial Least Square Structural Equation Model (PLS-SEM) method . The total sample for this study was 51 respondents consisting of educators and educational staff who had interacted with JIBAS. The research variables consist of Perceived ease of use (PEOU), Perceived usefulness (PU), Behavioral intention to use (BI), and Actual System Use (AC). PLS-SEM analysis concludes that PEOU has a significant relationship to the BI variable and BI has a significant relationship to AC. The conclusion obtained from the results of the analysis is that increasing the ease of application of JIBAS (PEOU) will have an impact on increasing usage behavior (BI) of the JIBAS application. Increasing the behavior of using the application will also have an impact on increasing the use (AC) of the JIBAS application in the management of educational institutions.

Keywords : *JIBAS, TAM, Partial Least Square Structural Equation Model.*

1. PENDAHULUAN

Penggunaan teknologi informasi di bidang pendidikan telah mencakup pada pengelolaan administrasi pada instansi pendidikan. Manajemen data akademik siswa, pengelolaan data keuangan institusi pendidikan, manajemen data perpustakaan, dan lainnya dapat dilakukan dengan mudah menggunakan teknologi informasi (Widyastono,

2017). Salah satu teknologi yang telah diterapkan dalam pengelolaan pendidikan adalah aplikasi yang dikenal dengan Jaringan Informasi. Bersama Antar Sekolah atau dikenal dengan istilah JIBAS.

JIBAS merupakan sistem yang membantu pengguna dalam pengelolaan informasi data di instansi pendidikan. Data dan informasi yang dikelola meliputi data akademik siswa, akuntansi

keuangan institusi pendidikan, manajemen data perpustakaan (Natalia & Supriyanto, 2021). Implementasi teknologi JIBAS dilapangan akan mendapatkan permasalahan baik dari segi sistem maupun dari segi pengguna. Dari segi pengguna, masalah yang dihadapi adalah tingkat penerimaan dari para pengguna terhadap JIBAS. Penerimaan terhadap suatu teknologi merupakan dapat dijadikan sebagai indikator bahwa pengguna menerima teknologi tersebut dan dapat digunakan sebagai suatu *tools* yang mendukung pekerjaan (Pramiyati et al., 2019)

Kerangka teori yang digunakan sebagai dasar model pengukuran tingkat penerimaan terhadap teknologi pada penelitian ini adalah *Technology Accepted Model* (TAM). TAM dapat didefinisikan sebagai konsep untuk menganalisis faktor-faktor yang berhubungan dengan kapabilitas pengguna dalam menggunakan suatu teknologi. Teori ini diperkenalkan oleh Davis pada tahun 1985, dan juga dapat digunakan untuk memprediksi penggunaan dari suatu sistem (Karouw et al., 2019). TAM disusun oleh beberapa konstruk, dengan konstruk utama terdiri dari konstruk untuk mengukur manfaat dan kemudahan suatu teknologi (Rahmawati & Narsa, 2019).

Penerapan TAM untuk mengukur tingkat penerimaan teknologi di bidang pendidikan telah dilakukan peneliti lain. Hidayah dan Fernanda (2021) menerapkan TAM 3 dan PLS-SEM untuk mengetahui variabel-variabel yang berhubungan terhadap penerimaan teknologi aplikasi *google classroom* oleh mahasiswa dalam proses pembelajaran online. Hasil penelitiannya menghasilkan kesimpulan bahwa kemudahan aplikasi yang digunakan, dan kebermanfaat aplikasi *google classroom* berhubungan signifikan terhadap perilaku penggunaan teknologi. Perilaku teknologi ini juga memiliki hubungan yang signifikan terhadap penggunaan aplikasi *google classroom* oleh mahasiswa dalam pembelajaran online.

Sukainah dan Putra (2019) melakukan penelitian tentang variabel yang berpengaruh terhadap penerimaan mahasiswa terhadap aplikasi Kelase di program studi pendidikan teknologi pertanian terhadap menggunakan konsep model TAM. Hasil penelitiannya menghasilkan kesimpulan bahwa Aplikasi yang semakin mudah dan semakin bermanfaat, akan berdampak kepada peningkatan penerimaan mahasiswa tersebut terhadap aplikasi yang digunakan untuk pembelajaran. Wiyono dan Ruyani (2019) melakukan penelitian tentang faktor-faktor yang berpengaruh terhadap perilaku belajar mahasiswa menggunakan pendekatan *PLS-SEM*. Hasil yang

yang didapatkan adalah perilaku belajar siswa dipengaruhi oleh lingkungan keluarga, lingkungan kampus dan motivasi belajar siswa.

PLS-SEM adalah teknik dalam *Structural Equation Modelling* yang didasarkan pendekatan iteratif yang memaksimalkan varian yang dijelaskan dari setiap variabel endogen. PLS-SEM memiliki beberapa keunggulan dibandingkan metode SEM berbasis covariance. PLS-SEM dapat digunakan untuk menganalisis data yang tidak dapat memenuhi asumsi kenormalan data, atau data penelitian sedikit. Metode ini juga dapat digunakan untuk menganalisis variabel kontruk yang bersifat reflektif (Hair et al., 2014). Lin *et al* (2019) juga menjelaskan bahwa PLS-SEM lebih efisien untuk analisis data dengan sampel kecil dibandingkan metode *Covarian Based SEM* (CB-SEM). Berdasarkan uraian sebelumnya, penelitian ini bertujuan untuk mengetahui faktor-faktor yang memiliki hubungan signifikan terhadap penerimaan aplikasi JIBAS menggunakan *Partial Least Square Structural Equation Model* (PLS-SEM).

2. METODE PENELITIAN

Data penelitian diperoleh melalui survei kepada 51 responden yang terdiri dari tenaga pendidik dan tenaga kependidikan yang pernah berinteraksi atau menggunakan aplikasi JIBAS. Instrument dalam penelitian ini menggunakan kuesioner untuk mengukur tingkat penerimaan aplikasi JIBAS.

Kuesioner mengadopsi dari konsep penerimaan teknologi menggunakan konsep TAM 3 yang dikembangkan oleh Davis 1989 (Adriansyah et al., 2016). Indikator-indikator pertanyaan yang digunakan sudah disesuaikan oleh peneliti supaya dapat mengukur penerimaan aplikasi JIBAS. Skala pengukuran yang digunakan menggunakan skala likert 1- 4.

Variabel dalam penelitian ini terdiri dari 4 konstruk atau konstruk. Konstruk merupakan variabel yang tidak ada dalam dataset dan diukur melalui indikator-indikator turunannya (pertanyaan) (Fan et al., 2016). Setiap variabel konstruk disusun oleh beberapa indikator pertanyaan seperti yang tersaji pada tabel 1.

Tabel 1. Indikator-indikator penyusun setiap konstruk

Konstruk	Indikator
<i>Perceived usefulness</i> (PU)	PU1: Ketika menggunakan aplikasi tersebut, kinerja saya menjadi lebih efisien

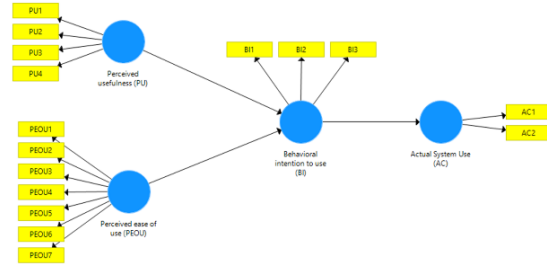
PU2: Ketika menggunakan aplikasi tersebut, kinerja saya menjadi lebih efektif
 PU3: Ketika menggunakan aplikasi tersebut, saya dapat mengakses informasi dan data yang saya butuhkan
 PU4: Saya dapat mengakses aplikasi tersebut diberbagai lokasi lokasi

Perceived ease of use (PEOU)
 PEOU1: Aplikasi mudah digunakan (easy to use)
 PEOU2: Aplikasi memiliki flexibility yang sangat baik
 PEOU3: User interface untuk mengambil informasi aka-demik mudah dioperasikan
 PEOU4: *User interface* untuk menjelaskan informasi aka-demik mudah dioperasikan
 PEOU5: *User interface* memberikan kenyamanan ketika digunakan oleh pengguna
 PEOU6: Aplikasi jarang mengalami kejadian tidak merespon
 PEOU7: *Proses penyalinan data dalam aplikasi mudah dioperasikan*

Behavioral intention to use (BI)
 BI1: Ketika bekerja, saya selalu menggunakan aplikasi JIBAS
 BI2: Saya lebih suka menggunakan JIBAS untuk proses menemukan data dan informasi yang terkait akademik
 BI3: Pada penyampaian informasi kepada stakeholder internal (guru, staff, peserta didik) dan stakeholder eksternal (wali siswa) saya lebih suka memakai JIBAS

Actual System Use (AC)
 AC1: Saya memiliki minat untuk meningkatkan penggunaan JIBAS dalam penyelesaian pekerjaan.
 AC2: Saya memberikan saran kepada orang lain yang sama-sama bekerja di instansi pendidikan untuk menggunakan JIBAS dalam pengelolaan data dan informasi

Setiap konstruk memiliki hubungan dengan konstruk yang lain. Kerangka teori *Technology Accepted Model* (TAM) yang menjelaskan hubungan antar konstruk pada penelitian ini disajikan pada gambar 1 di bawah ini.



Gambar 1. Kerangka teori hubungan antar konstruk menggunakan *Technology Accepted Model* (TAM) 3

Hipotesis yang dibangun berdasarkan kerangka teori pada gambar 1 adalah

- H₁ : Ada hubungan yang signifikan antara *Perceived usefulness* terhadap *Behavioral intention to use*
- H₂ : Ada hubungan yang signifikan antara *ease of use* terhadap *Behavioral intention to use*
- H₃ : Ada hubungan yang signifikan antara *Behavioral intention to use* terhadap *Actual System Use*

3. HASIL DAN PEMBAHASAN

Nilai rata-rata setiap indikator yang menyusun konstruk, akan memberikan informasi tentang penerimaan pengguna terhadap aplikasi JIBAS, Hasil nilai rata-rata setiap indikator yang menyusun konstruk disajikan pada Tabel 2.

Tabel 2. Rata-rata jawaban setiap indikator penyusun konstruk.

Konstruk	Indikator	Rata-rata
<i>Actual System Use (AC)</i>	AC1	3.06
	AC2	3.14
<i>Behavioral intention to use (BI)</i>	BI1	2.96
	BI2	3.16
	BI3	3.04
<i>Perceived ease of use (PEOU)</i>	PEOU1	3.25
	PEOU2	2.98
	PEOU3	3.18
	PEOU4	3.18
	PEOU5	3.27
	PEOU6	2.61
	PEOU7	3.02

	PU1	3.24
Perceived usefulness (PU)	PU2	3.22
	PU3	3.20
	PU4	2.57

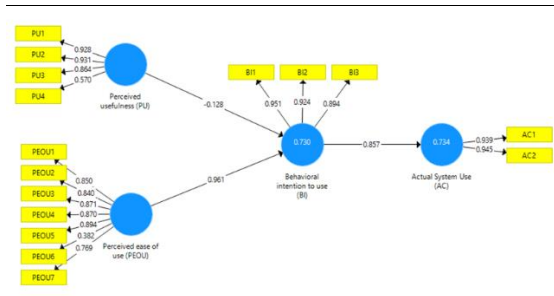
Actual System Use (AC) merupakan penggunaan nyata responden dalam menggunakan JIBAS. Variabel ini terdiri dari dua indikator pertanyaan. Indikator pertanyaan kedua (AC2) yaitu pertanyaan yang menyatakan responden memberikan saran kepada orang lain yang sama-sama bekerja di instansi pendidikan untuk menggunakan JIBAS dalam pengelolaan data dan informasi memiliki rata-rata sebesar 3,14. Rata-rata skor ini lebih tinggi daripada rata-rata jawaban responden untuk pertanyaan yang pertama.

Behaviour intention to use disusun oleh 3 pertanyaan. Pertanyaan yang memiliki rata-rata penilaian tertinggi adalah pertanyaan saya lebih suka menggunakan JIBAS untuk proses menemukan data dan informasi yang terkait akademik. Pertanyaan yang memiliki rata-rata penilaian paling rendah adalah pertanyaan yang menyatakan bahwa responden ketika bekerja, responden selalu menggunakan aplikasi JIBAS

Perceived ease of use (PEOU) dapat digambarkan sebagai kemudahan dalam menggunakan aplikasi. PEOU disusun dari 7 indikator pertanyaan. Indikator pertanyaan yang memiliki rata-rata tertinggi yaitu pada pertanyaan nomor 1 yaitu aplikasi mudah digunakan (easy to use) dengan rata-rata sebesar 3,25. Sedangkan indikator pertanyaan yang mendapatkan penilaian terendah dari responden adalah Aplikasi jarang mengalami kejadian tidak merespon. Informasi ini memberikan gambaran bahwa aplikasi ini masih memerlukan pengembangan karena dalam penerapannya masih sering eror sehingga sering terjadi kesulitan dalam pengaksesan aplikasi.

Perceived usefulness (PU) merupakan konstruk untuk mengukur manfaat yang diberikan oleh aplikasi dalam mendukung pekerjaan yang dilakukan. Pada penelitian ini, variabel PU disusun oleh 4 indikator pertanyaan. Tabel 2 memberikan informasi nilai rata-rata tertinggi penilaian responden terletak pada pertanyaan ke 1 yang menyatakan bahwa penggunaan Ketika menggunakan aplikasi tersebut, kinerja saya menjadi lebih efisien dengan rata-rata sebesar 3,24. Sedangkan indikator yang memiliki rata-rata terendah penilaian terendah adalah pertanyaan ke 4 yang isinya menyatakan tentang Saya dapat mengakses aplikasi tersebut diberbagai lokasi dengan rata-rata skor sebesar 2,57. Berdasarkan

Tabel 2, dapat diketahui bahwa indikator yang memiliki rata-rata paling rendah adalah pada konstruk PU pada pertanyaan Saya dapat mengakses aplikasi tersebut diberbagai lokasi dengan rata-rata skor sebesar 2,57. Informasi ini sangat bermanfaat, dimana responden mengharapkan bahwa aplikasi ini dapat dikembangkan supaya dapat diakses di mana saja atau dikembangkan menjadi aplikasi berbasis web atau bisa juga dikembangkan menjadi apps yang dapat diinstall di handphone.



Gambar 2. Hasil analisis penerimaan aplikasi JIBAS menggunakan PLS-SEM

Analisis statistika selanjutnya adalah melakukan analisis terhadap penerimaan aplikasi JIBAS menggunakan konsep TAM 3 seperti yang disajikan pada Gambar 1. Analisis statistika menggunakan PLS-SEM karena jumlah sampel ini hanya terdiri dari 51 responden. Hasil analisis PLS-SEM dengan bantuan software smartpls disajikan pada Gambar 2 di atas.

Tabel 3. Nilai outer loading setiap indikator

Konstr	Indik	$\hat{\lambda}_{jk}$	$se(\hat{\lambda}_{jk})$	$t = \frac{\hat{\lambda}_{jk}}{se(\hat{\lambda}_{jk})}$	Kesi
uk	ator				mpul
Actual System Use (AC)	AC1	0.94	0.03	36.23	Signifikan
	AC2	0.95	0.02	45.52	Signifikan
Behavioral intention to use (BI)	BI1	0.95	0.02	53.52	Signifikan
	BI2	0.92	0.03	34.49	Signifikan
	BI3	0.89	0.05	19.16	Signifikan
Perceived ease of use (PEOU)	PEOU1	0.85	0.06	13.89	Signifikan
	PEOU2	0.84	0.05	15.39	Signifikan
	PEOU3	0.87	0.06	14.85	Signifikan

PEOU4	0.87	0.05	18.01	Signifikan	
PEOU5	0.89	0.05	18.67	Signifikan	
PEOU6	0.38	0.19	2.01	Signifikan	
PEOU7	0.77	0.09	8.95	Signifikan	
Perceived usefulness (PU)	PU1	0.93	0.03	31.02	Signifikan
	PU2	0.93	0.03	36.34	Signifikan
	PU3	0.86	0.04	19.34	Signifikan
	PU4	0.57	0.16	3.53	Signifikan

Setiap konstruk pada penelitian ini disusun oleh beberapa indikator-indikator pertanyaan. Indikator-indikator pertanyaan harus valid. Uji validitas dalam model PLS-SEM dilakukan dengan membandingkan nilai t hitung dari nilai *outer loading* setiap pertanyaan dengan nilai 1,96 (t tabel). Berdasarkan hasil pada tabel 2 diketahui bahwa seluruh indikator-indikator pertanyaan yang menyusun konstruk valid karena memiliki nilai t hitung lebih besar dari 1,96 pada α sebesar 0,05.

Uji reliabilitas konstruk dilakukan dengan melihat nilai composite reliability setiap konstruk. Konstruk dinyatakan reliabel apabila memiliki nilai reliabilitas kompositnya diatas 0,7. Tabel 4 memberikan informasi nilai composite reliability setiap konstruk atau variabel latent. Berdasarkan tabel tersebut, dapat disimpulkan bahwa semua konstruk yang digunakan dalam penelitian ini memiliki reliabilitas yang baik.

Tabel 4. Nilai composite reliability setiap konstruk

Konstruk	Composite Reliability
AC	0.940
BI	0.945
PEOU	0.922
PU	0.900

Model struktural pada penelitian ini menjelaskan hubungan antar konstruk yang terdiri dari variabel AC, BI, PEOU, dan PU. Hasil pengujian model struktural beserta nilai signifikansinya disajikan pada tabel 4 di bawah ini.

Tabel 5. Nilai koefisien dan signifikansi model struktural

Hubungan antar konstruk	Koefisien	t statistic	signifikansi
BI -> AC	0.857	14.175	0.000
PEOU -> BI	0.961	6.642	0.000
PU -> BI	-0.128	0.812	0.417

Tabel 5 memberikan informasi nilai signifikansi model struktural berdasarkan gambar 1 dan digunakan untuk menjawab hipotesis pada penelitian ini. Dua variabel dinyatakan memiliki hubungan signifikan apabila nilai signifikansi yang dihasilkan kurang dari alpha (0,05). Berdasarkan tabel 5 diketahui bahwa variabel PEOU berhubungan signifikan terhadap BI. Pada tabel tersebut juga memberikan informasi bahwa variabel BI memiliki hubungan yang signifikan terhadap AC Sedangkan variabel PU tidak memiliki hubungan signifikan terhadap BI karena nilai signifikansi yang lebih besar dari 0,05.

Persamaan matematika berdasarkan hasil pada tabel 4 adalah

$$BI = 0,961 PEOU \tag{1}$$

$$AC = 0,857 BI \tag{2}$$

Model 1 merupakan persamaan antara BI dengan PEOU. Koefisien pada model ini sebesar 0,961. Informasi dari nilai koefisien ini adalah peningkatan keragaman (standar deviasi) pada kemudahan dalam penggunaan aplikasi JIBAS (PEOU) sebesar satu satuan, akan berdampak kepada keragaman BI sebesar 0,961 satuan standar deviasi. Sedangkan pada model 2, diketahui bahwa koefisien sebesar 0,857. Koefisien ini memberikan informasi bahwa peningkatan standar deviasi sebesar satu satuan pada *Behavioral intention to use*, akan diikuti peningkatan standar deviasi *Actual System Use* sebesar 0,857 satuan.

Evaluasi model PLS-SEM untuk penerimaan aplikasi JIBAS dapat ditinjau dari *goodness of fit* atau nilai R square yang dihasilkan. Nilai R square pada model ini disajikan pada tabel

Tabel 6. R square model struktural

Konstruk	R Square
AC	0.734
BI	0.730

Nilai R square pada variabel AC sebesar 0,734. Nilai ini memiliki makna bahwa model perimaan teknologi mampu menjelaskan variasi pada variabel AC sebesar 73,4%. Nilai R square pada variabel BI sebesar 0,73, yang artinya model mampu menjelaskan keragaman data pada variabel BI sebesar 73%.

4. KESIMPULAN DAN SARAN

Kesimpulan yang dihasilkan adalah variabel *Perceived ease of use* atau kemudahan dalam penggunaan aplikasi atau teknologi berhubungan signifikan terhadap perilaku pengguna dalam menggunakan teknologi tersebut (*Behavioral intention to use*). Variabel *Behavioral intention to use* (BI) juga memiliki hubungan yang signifikan terhadap *Actual System Use* (AC). Peningkatan kemudahan aplikasi JIBAS (PEOU) akan berdampak kepadan peningkatan perilaku penggunaan (BI) aplikasi JIBAS. Peningkatan perilaku penggunaan aplikasi akan berdampak juga terhadap peningkatan penggunaan (AC) aplikasi JIBAS dalam pengelolaan institusi pendidikan.

DAFTAR PUSTAKA

- Adriansyah, N., Syaifulah, & Jazman, M. (2016). Analisa Penerimaan Dan Penggunaan Teknologi Informasi E-Vote menggunakan Technology Acceptance Model (TAM). *Rekayasa Dan Manajemen Sistem Informasi*, 2(2).
- Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes*, 5(1). <https://doi.org/10.1186/s13717-016-0063-3>
- Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), 106–121. <https://doi.org/10.1108/EBR-10-2013-0128>
- Hidayah, N., & Fernanda, J. W. (2021). Analisis Penerimaan Aplikasi Pembelajaran Online Menggunakan Technology Acceptance Model 3 Dan Partial Least Square Structural Equation Model (Pls-Sem). *Factor M*, 3(2), 161–172. <https://doi.org/10.30762/factor-m.v3i2.3097>
- Karouw, S., Narasiang, B., & Litouw, J. (2019). Analisis Pengukuran Model Penerimaan Teknologi OpenSID di Area Rural dan Urban. *CogITO Smart Journal*, 5(2), 191–202. <https://doi.org/10.31154/cogito.v5i2.175.191-202>
- Lin, H., Lee, M., Liang, J., & Chang, H. (2019). *modeling in e-learning research*. 0(0), 1–19. <https://doi.org/10.1111/bjet.12890>
- Natalia, D. N., & Supriyanto, S. (2021). Implementasi Jaringan Informasi Bersama Antar Sekolah (JIBAS) sebagai Media Humas. *Inspirasi Manajemen Pendidikan*, 9(5), 1078–1087. <https://ejournal.unesa.ac.id/index.php/inspirasi-manajemen-pendidikan/article/view/42331%0Ahttps://ejournal.unesa.ac.id/index.php/inspirasi-manajemen-pendidikan/article/view/42331/37161>
- Pramiyati, T., Jayanta, J., & Mahfud, H. (2019). Analisis Tingkat Penerimaan Pengguna Terhadap SIMBUMIL (Hasil Survey Penerimaan SIMBUMIL di Puskesmas Mandalawangi). *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, 5(1), 61–67. <https://doi.org/10.23917/khif.v5i1.7328>
- Rahmawati, R. N., & Narsa, I. M. (2019). Penggunaan e-learning dengan Technology Acceptance Model (TAM). *Jurnal Inovasi Teknologi Pendidikan*, 6(2), 127–136. <https://doi.org/10.21831/jitp.v6i2.26232>
- Sukainah, A., & Putra, R. P. (2019). Analisis Tingkat Penerimaan Mahasiswa Pendidikan Teknologi Pertanian Terhadap Kelase Menggunakan Technology Acceptance Model. *Indonesian Journal of Educational Studies*, 22(1). <https://doi.org/10.26858/ijes.v22i1.9349>
- Widyastono, H. (2017). Pemanfaatan Teknologi Informasi Dan Komunikasi Pada Sekolah Menengah Pertama Negeri Akreditasi a Di Provinsi Jawa Timur. *Jurnal Kwangsan*, 5(1), 21. <https://doi.org/10.31800/jtpk.v5n1.p21--38>
- Wiyono, D., & Ruyani, N. A. (2019). *Factors Affecting the Learning Behaviors of Students Using Partial Least Square-Structural Equation Modeling (Pls-Sem) Analysis Faktor-Faktor yang Mempengaruhi Perilaku Belajar Mahasiswa dengan Menggunakan Pendekatan Partial Least Square-Structural Equati*. III(2), 127–144.

JStatistika : Jurnal Ilmiah Teori Dan Aplikasi Statistika ber ISSN online 2654-7511 dan ISSN cetak 2089-0028,

JStatistika merupakan media penerbitan karya ilmiah berupa literasi hasil penelitian ataupun kajian tentang teori, metodologi serta aplikasi dalam bidang ilmu statistika dan data sains. **JStatistika** diterbitkan oleh Program Studi Statistika Fakultas Sains dan Teknologi Universitas PGRI Adi Buana Surabaya dan terbit dua kali dalam setahun setiap bulan Juli dan Desember.

PUBLISHED BY :

**PRODI STATISTIKA
FAKULTAS SAINS DAN TEKNOLOGI**

