J STATISTIKA

Program Studi Statistika

EISSN : 2654-7511
PISSN : 2089-0028

J STATISTIKA

JURNAL ILMIAH TEORI DAN APLIKASI STATISTIKA

Volume 17 | Nomor 1 | 2024

EDITORIAL TEAM

INTRODUCTION

We are delighted to announce the current publication of Volume 17, Number 1 of JStatistika, affiliated with the Statistics Department at PGRI Adi Buana University Surabaya, has been released in July 2024. This particular issue of the JStatistika Scientific Journal features a diverse array of articles addressing a wide spectrum of topics. One of the highlighted articles delves into “Binary Logistic Regression Analysis on the Spread of Dengue Fever in Bali Province; Association of Poverty Categories, Educational Characteristics, and Area of Residence in Indonesia Using a Three-Way Log-Linear Model; Factors Affecting the Resilience Index Food in Papua Province and West Papua Province Using a Spatial Model Approach; Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest; Accelerating SIREKAP Digital Transformation in the 2020 Natuna Regency Election; Forecasting Average Rice Prices at Milling Level According to Quality Using Support Vector Regression; Dynamical Analysis of Mathematical Model of Social Behavior with Law Enforcement and Religious Approaches; The Effect of Motivation, Work Discipline, and Organizational Commitment on Employee Performance at PT. Hebsa Indonesia; Statistical Quality Control (SQC) Method Analysis Regarding Qualty Control of Shoe Products (Case Study of PT-X); Distribution of Soap X in East Java Region with Bhumal Method and Traveling Salesman Problem”

The JStatistika Scientific Journal enthusiastically welcomes and invites contributions in a diverse range of formats, including but not limited to scholarly scientific articles that encompass various facets of statistical science. We eagerly seek research findings, comprehensive reports, insightful case studies, thorough literature reviews, and updates that pertain to the dynamic landscape of statistical science. Our overarching objective is to cultivate a repository of knowledge that is not only current but also invaluable in tackling the ever-evolving and intricate challenges confronting our field. We actively encourage authors to submit their work if it resonates with the most recent advancements and frontiers in statistical science. Our aspiration is to foster an environment where these contributions can flourish, ultimately serving as a wellspring of cutting-edge insights and understanding. We believe that these insights are instrumental in addressing the multifaceted issues that confront us in today's complex world.

Our editorial team extends a warm and inclusive invitation to scientists and scholars from diverse backgrounds and affiliations, including institutions of higher learning and esteemed research organizations. We seek your valuable contributions, whether they be grounded in empirical research results or rooted in rigorous scholarly studies within the expansive domain of statistics and its myriad practical applications. We hold a deep appreciation for the feedback and perspectives of our esteemed readership. Your input not only enriches the discourse but also plays a pivotal role in our continuous efforts to elevate the quality and relevance of the journal. We earnestly value your insights and ideas, recognizing that they are integral to our ongoing pursuit of excellence. Our ultimate vision is for the articles featured in the JStatistika Scientific Journal to transcend the confines of academia and serve as a wellspring of knowledge that benefits not only scholars and researchers but also professionals actively engaged in the diverse realms of statistical science and its multifaceted real-world applications. Through collaborative efforts and a shared commitment to advancing our understanding of statistics, we aim to make a meaningful impact in the broader scientific community and beyond.

Jstatistika has been indexed by Sinta 4 Kemendikbud, Garuda, Google Scholar, Crossref, Worldcat, Scilit, ROAD, Onesearch, Journal Stories, Dimensions, Base, Open Alex, Wikidata, Internet Archieve, Root Indexing, Core, Harvard Library, Universiteit Leiden Library, Semantic Scholar, Open Air Explore, ASCI, Cite Factor, University of Saskatchewan Library, The University of Queensland Library, George University Library and Boston University Library.

Surabaya, July 2024

Editor in Chief

LIST OF CONTENTS

Binary Logistic Regression Analysis on the Spread of Dengue Fever in Bali Province

Luh Putu Safitri Pratiwi (1), I Made Pasek Pradnyana Wijaya (2)

[1]Program Studi Sistem Informasi, ITB STIKOM Bali

Jl. Raya Puputan No. 86 Renon –Denpasar –Bali. Tlpn (0361) 244445

e-mail: putu_safitri@stikom-bali.ac.id(1), pasek_pradnyana@stikom-bali.ac.id(2)

ABSTRAK

Penyebaran demam berdarah melibatkan siklus kompleks antara manusia sebagai inang dan nyamuk sebagai vektor. Gejala demam berdarah dapat bervariasi dari demam ringan hingga bentuk berat yang dapat mengancam jiwa. Salah satu wilayah yang memiliki angka persebaran demam berdarah tertinggi di Provinsi Bali adalah wilayah Denpasar. Penelitian terus dilakukan untuk memahami faktor-faktor yang memengaruhi penyebaran demam berdarah menggunakan metode regresi logistik biner. Regresi logistik biner merupakan model regresi yang sering digunakan dalam pemodelan data kategori, dimana variabel dependen dalam penelitian ini adalah penyebaran kasus demam berdarah dengan angka penyebaran kasus di masing-masing wilayah akan diberikan kategori nol untuk jumlah kasus sedikit dan satu untuk jumlah kasus tinggi. Sehingga dalam penelitian ini dikembangkan strategi yang lebih efektif dalam pengendalian penyakit demam berdarah serta bagaimana model terbaik dari data penyebaran demam berdarah di Provinsi Bali. Hasil yang diperoleh dari penelitian ini adalah uji pengaruh variabel independent terhadap variabel dependen memperlihatkan bahwa variabel jumlah sarana sanitasi layak (X_5) memiliki pengaruh yang signifikan terhadap jumlah penderita demam berdarah di Provinsi Bali yaitu sebesar 0,081.

Kata kunci: Demam Berdarah; Regresi Logistik; Provinsi Bali

ABSTRACT

The spread of dengue fever involves a complex cycle between humans as hosts and mosquitoes as vectors. Symptoms of dengue fever can vary from a mild fever to a severe form that can be life-threatening. One of the areas that has the highest spread of dengue fever in Bali Province is the Denpasar area. Research continues to be carried out to understand the factors that influence the spread of dengue fever using the binary logistic regression method. Binary logistic regression is a regression model that is often used in modeling categorical data, where the dependent variable in this study is the distribution of dengue fever cases with the number of cases spreading in each region being assigned a category of zero for a low number of cases and one for a high number of cases. So in this research a more effective strategy was developed in controlling this disease as well as the best model for data on the spread of dengue fever in Bali Province. The results obtained from this research were a test of the influence of the independent variable on the dependent variable, showing that the variable number of adequate sanitation facilities (X_5) had a significant influence on the number of dengue fever sufferers in Bali Province, namely 0.081.

Keywords: *Dengue Fever; Logistic Regression; Bali Province*

INTRODUCTION

Dengue fever (DHF), also known as dengue hemorrhagic fever, is an infectious disease caused by the dengue virus, which is transmitted by the Aedes aegypti and Aedes albopictus mosquitoes. This disease is a significant public health problem in various tropical and subtropical regions throughout the world [1]. Dengue fever (DHF) has become a global concern because of its serious public health impacts. Indonesia is one of the countries affected by dengue fever, with a number of provinces, including Bali Province, reporting significant cases. Bali Province, with a high population and tourist visits, has its own challenges in controlling this disease. In addition, there are factors such as climate change, urbanization, and population mobility that can influence the spread of the dengue virus in this region [2].

The occurrence of dengue fever cases can be explained by several factors, including the rainy season. During the rainy season, environmental conditions create ideal conditions for the development of Aedes mosquitoes, which spread dengue fever [3]. Previous research shows that factors such as regional neighbourhoods, area size, and the presence of jumantik influence the spread of dengue fever in Denpasar City [4]. A deeper understanding of the pattern of dengue fever spread is key to effective control and prevention efforts. For this reason, it is necessary to use an effective method in predicting the probability of an event occurring with certain independent variables using a method, namely logistic regression [5]. Logistic regression is a mathematical modeling approach used to analyze the relationship of one or several independent variables with a categorical dependent variable that is dichotomous/binary [6]. Research using logistic regression was investigated by [7], based on the results of the analysis it was found that the number of applicants who were declared accepted was 50% of the participants, while the remaining 50% of applicants were declared not accepted. In binary logistic regression modeling, the influencing variables are obtained, namely education (length of last education) and experience (work experience). Other research on binary logistic regression was carried out by [8] regarding the stunting problem in West Java where after modeling there were three variables that had a significant effect at a real level of 0,10, namely complete basic immunization, food management places that met health requirements, and poor people. Similar research was also conducted by [9] regarding maternal knowledge about immunizations on toddlers' health where maternal knowledge about immunizations did not affect toddlers' health in general, seen from the Nagelkerke coefficient of determination of 0,148 or 148%.

With this approach, we can understand the relative influence of various factors on the results we want to predict, as well as identify the most significant factors in influencing the spread of dengue fever in Bali Province.

METHOD

This research took the location of each sub-district in Bali Province, totaling 57 sub-districts. The data used in this research is secondary data obtained from the Bali Province Central Statistics Agency, Bali Province Bappeda and Bali Province Health Service in 2022 [10]. The data contains information about the spread of dengue fever cases and the factors that influence it, which will be used as research variables. The following are the response variables and predictor variables used.

Table 1. Research Variable

Variable	Variable Type	Notation
Number of dengue fever sufferers in categories: 0=low/few numbers of dengue sufferers 1=high number of dengue sufferers	Variable Response	Y
number of drinking water facilities	Variable Predictor	X_1
population density	Variable Predictor	X_2
number of doctors	Variable Predictor	X_3
number of health workers, namely nurses	Variable Predictor	X_4
number of adequate sanitation facilities	Variable Predictor	X_5

Logistic Regression

Binary logistic regression is a logistic regression model with a response variable (Y) on a binary category scale, namely having two categories of values 0 and 1 [11]. Binary logistic regression analysis is used to look for patterns of probability relationships between variables x and p (the probability of events caused by x). The value of the logistic function ranges between 0 and 1.

The following is the function of logistic regression

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \tag{1}$$

$\pi(x) = E(Y|x)$ represents the conditional average for all x values

A transformation for $\pi(x)$ the value called the logit transformation is carried out to obtain the assumption that the log odds ratio value has a linear relationship with x [12]. This transformation will produce a function g(x) which is linear in its parameters [13].

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{2}$$

Another difference between linear regression and logistic regression is the distribution of the response variable. In the linear regression model, the response variable is assumed to be $Y = \pi(x) + \varepsilon$ where ε is the error following a normal distribution with a mean equal to zero and constant variance. But in binary logistic regression, the error value only consists of two possibilities, namely if then $\varepsilon = 1 - \pi(x)$ with chance $\pi(x)$ or if y = 0 then $\varepsilon = -\pi(x)$ with chance $1 - \pi(x)$. So the error has a distribution with mean equal to zero and variance $[\pi(x)(1-\pi(x))]$.

$$\pi(\boldsymbol{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)} \tag{3}$$

So the logit transformation form $\pi(\boldsymbol{x})$ of equation (3) becomes:

$$g(\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \tag{4}$$

$$g(x) = \sum_{j=0}^{k} \beta_j x_j \quad , x_0 = 1$$

Logistic Regression Model Parameter Estimator

To estimate the parameters of the logistic regression model, use the Maximum Likelihood Estimator (MLE) method. Each observation has a distribution function [11]:

The ln likelihood function is as follows:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \ln\prod_{l=1}^{n}\pi(\boldsymbol{x}_l)^{y_l}[1-\pi(\boldsymbol{x}_l)]^{1-y_l}$$

$$= \sum_{l=1}^{n} y_l \boldsymbol{\beta}^T x_l - \left(1+\exp\left(\boldsymbol{\beta}^T x_l\right)\right) \tag{5}$$

Testing Logistic Regression Model Parameters

After getting parameter estimates in a logistic regression model, the next step is to carry out tests to find out which predictor variables have a significant effect on the response variable. The significance test of the β coefficient in the model consists of a simultaneous test and a partial test.

1. Simultaneous Test

This test is used to check the significance of each β coefficient together in the model with the following hypothesis [14]:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_1 : there is at least one $\beta_j \neq 0$, j=1, 2, ..., k

The test statistic used is the test statistic G_{hit} or Likelihood Ratio Test, namely [13].

$$G = -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\prod_{l=1}^{n}\hat{\pi}_l^{y_l}\left(1-\hat{\pi}_l\right)^{(1-y_l)}}\right] \tag{6}$$

where

$$n_1 = \sum_{l=1}^{n} y_l \;;\quad n_0 = \sum_{l=1}^{n}(1-y_l) \;;\quad n = n_0 + n_1$$

if $G_{hit} > X^2{}_{(\alpha,p)}$; p is the number of parameters then Ho is rejected.

2. Partial Test

This test was carried out to find out whether the predictor variables had a significant effect on the response variable individually [15]. The hypothesis used is:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Wald test statistics:

$$W = \frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)} \tag{7}$$

if $|W| > Z_{\alpha/2}$ then Ho is rejected.

RESULT AND DISCUSSION

To determine the characteristics of each sub-district in Bali Province, descriptive statistical analysis was carried out on all the variables studied. These characteristics include exploration of response and predictor variables so that broader information is obtained. The results of the descriptive analysis can be seen in Table 2 and Table 3.

Table 2. Percentage of DHF Categories in Bali Province

Group	Total	Percentage
Low DHF	49	85,96
High DHF	8	14,04

Based on Table 2 above, it can be seen that the majority of sub-districts in Bali are in the dengue fever group with a low/low number of dengue cases, namely 85,96 percent (49 sub-districts). Meanwhile, 14,04 percent (8 sub-districts) are classified as areas that have a high number of dengue cases.

The predictor variables in this study are factors that are thought to influence the dengue fever value in Bali Province. The descriptive analysis resulting from the predictor variables that are thought to have an influence on DHF is presented in the table as follows.

Table 3. Descriptive Statistics of Independent Variables

	Variabel	Mean	StDev	Min	Max
X_1	Number of Drinking Water Facilities	26,60	43,993	1	279
X_2	Population density	1438,33	1723,370	177	8620
X_3	Number of Doctors	10,68	6,362	4	30
X_4	Number of Health Personnel (Nurses)	30,51	19,418	12	114
X_5	Number of Adequate Sanitation Facilities	19137,32	1043,216	753	49828

Table 3 shows that the number of drinking water facilities in Bali Province has an average of 26,60, where Manggis and Kediri Districts have the smallest number of drinking water facilities and Kintamani District has the largest drinking water facilities. The fastest population density is in West Denpasar, while the lowest population density is in West Selemadeg at 177,2. The average number of doctors in Bali is 10,68 with the lowest number being 4 doctors (Payangan, Tembuku and Susut Districts) and the highest being 30 (Abiansemal District). The role of nurses is also no less important in handling dengue fever cases in Bali where the lowest number of nurses is in Sidemen District at 12 people and the highest is in Abiansemal District at 114 people with an average number of nurses of 30,51. Apart from that, another factor that causes the number of dengue fever sufferers to increase is the small number of adequate sanitation facilities, where the average household that has adequate sanitation is 19.137,32. The lowest number of households that have proper sanitation is 753, namely located in Petang District and the highest, namely located in East Denpasar, is 49.828, with the number of sub-districts in Bali Province being 57 sub-districts. Tabanan Regency is the district with the largest number of sub-districts, namely 10 sub-districts.

Model Fit Test

Hosmer and Lemeshow Test to see the suitability or FIT of the model, with the following hypothesis [16].

H0: Model FIT (p value>0,1)
H1: Model does not FIT

Table 4. Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4,461	8	0,813

From table 4, a sig value of 0,813>0.1 is obtained, so H_0 is accepted, which means the FIT model. The FIT model in question is a binary logistic regression model suitable for use for further analysis because there is no significant difference between the predicted classification and the observed classification.

Coefficient of Determination Test

Nagelkerke R Square is a test carried out to find out how much the independent variable is able to explain and influence the dependent variable. The Nagelkerke R Square value varies between 1 and 0. If the value is closer to 1, then the model is considered to have greater goodness of fit. Conversely, if the value is closer to 0, then the model is considered not to be good of fit [17].

Table 5. Coefficient of Determination Test

Step	-2 *Log likelihood*	*Cox & Snell R Square*	*Nagelkerke R Square*
1	25,580[a]	0,304	0,547

From the results of Table 5, the Nagelkerke R Square of 0,547 shows that the ability of the independent variable to explain the dependent variable, namely DHF, is 54,7% and the rest is explained by other variables not studied.

Logistic Regression Model

The Bali Province DBD data used in this research is binary category data, so a Logistic regression model is used. Logistic regression functions to determine the pattern of relationship between the independent variable (X) and the dependent variable (Y) in the form of categories.

a. Simultaneous Testing

The simultaneous formation of a regression model aims to obtain an appropriate and simple model based on factors that are considered to influence dengue fever. According to Vikaliana (2022), to test the influence of the independent variables together on the dependent variable by looking at the significant values of the Omnibus Tests of Model Coefficients table with the hypothesis:

H_0 : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1 : there is at least one $\beta_j \neq 0$

$\alpha = 10\% = 0,1$

The test criterion is to reject H_0 if Sig. <10%

Table 6. Simultaneous Testing

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	20,659	6	0,002
	Block	20,659	6	0,002
	Model	20,659	6	0,002

From table 6, a sig value of 0,02 <0.1 is obtained, so H0 is rejected, which means the number of drinking water facilities (X_1), population density (X_2), number of doctors (X_3), number of health workers, namely nurses (X_4), number of facilities Proper sanitation (X_5) influences the dependent variable, namely dengue fever.

b. Individual Testing

The following are the results of the parameter estimation of the Logistic regression model with. With the following hypothesis.

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

$\alpha = 10\%$

The test criterion is to reject H_0 if Sig. <10%

Table 7. Logistic Regression Model Parameter Estimation

Variable	**B**	**S.E.**	**Wald**	**Sig.**	***Exp*(B)**
X_1	-0,036	0,049	0,556	0,456	0,964
X_2	0,000	0,000	0,261	0,609	1,000
X_3	0,085	0,180	0,225	0,635	1,089
X_4	0,008	0,043	0,035	0,851	1,008
X_5	0,000	0,000	3,036	**0,081**	1,000
Constant	-7,070	2,114	11,187	0,001	0,001

*) Parameter yang signifikan pada α=10%

The results of the test for the influence of the independent variable on the dependent variable in Table 7 are that the variable number of adequate sanitation facilities Meanwhile, the odds ratio or Exp(B) model is selected from only significant variables, so the model is formed as follows.

$$\ln\left(\frac{P(y-1)}{1-P(y-1)}\right) = \alpha + \beta_1 X_1$$

$$= -7,070 + 0,040 X_1$$

From the significant model, it can be interpreted that the odds ratio value or Exp(B) of variable X_5, which is the number of adequate sanitation facilities, is 1,000. This indicates that with an increase in the number of adequate sanitation facilities, it is estimated to have a 1 times greater chance of reducing the number of cases of dengue fever.

Probability Estimation

$$P(y=1)=\frac{e^{\alpha+\beta_1 X_1}}{1+e^{\alpha+\beta_1 X_1}}$$

$$P(y=1)=\frac{e^{-7{,}070+0{,}040X_1}}{1+e^{-7{,}070+0{,}040X_1}}$$

$$\hat{\pi}(\boldsymbol{x})=\frac{\exp(2{,}931x_3-0{,}142x_5+1{,}18x_6)}{1+\exp(2{,}931x_3-0{,}142x_5+1{,}18x_6)}$$

$$\hat{g}(\boldsymbol{x})=-7{,}070+0{,}040X_1$$

Based on the odds ratio value above, it is known that the greater the amount of adequate sanitation, a sub-district will have a higher number of dengue sufferers than low dengue fever, or in other words, the greater the amount of adequate sanitation, the greater the tendency for a sub-district to have low dengue fever.

Based on the Logistic regression model above, it is possible to reduce the dengue fever value for each sub-district in Bali by increasing the amount of adequate sanitation for each sub-district in Bali.

Calculation of Logistic Regression Classification Accuracy

Based on probability calculations, prediction results can be obtained, so that the truth of the logit model can be seen based on the results of classification between predictions and observations.

Table 8. Classification of DHF Results Logistic Regression Model

Observation	Prediction		Accuracy
	Low DHF	High DHF	percentage
Low DHF	48	1	98,0
High DHF	3	5	62,5
Overall Accuracy Percentage			93

Table 8 informs that the number of low category dengue fever is known in 49 sub-districts. The prediction results for areas with the lowest DHF were correctly classified in the low category as many as 48 areas, while 1 area was incorrectly classified (from low to high), namely East Denpasar District with a classification accuracy of 98%. Meanwhile, the number of high DBD categories is 8 sub-districts, where the areas with high DBD change to low DBD are 3 areas, namely Klungkung, Karangasem and South Kuta sub-districts, and the high DHF areas remain in the upper high DBD category in 5 areas with an accuracy percentage of 62,5% and the overall percentage of accuracy is 93%, which means that the percentage of accuracy that the model can predict correctly is 93%.

CONCLUSION

The results obtained from this research are a test of the influence of the independent variable on the dependent variable shows the variable number of adequate sanitation facilities (X_5) has a value of 0,081 which is smaller than the significance level of 0,1, so the decision taken is that H_0 is rejected, which means that the number of adequate sanitation facilities influences the number of dengue fever sufferers in the province of Bali.

So from a significant model it can be interpreted that the value of the odds ratio or Exp(B) variable (X_5), namely the number of adequate sanitation facilities, is 1,000. This indicates that if there is an increase in the number of adequate sanitation facilities, it is estimated that the number of dengue fever will decrease by 1 time.

ACKNOWLEDGMENTS

Thank you to ITB STIKOM Bali for providing research funding grants to individual lecturers for this research so that it can be completed well.

REFERENCE

[1] Ibrahim, A. Irviani, H. Habibi, and E. Latifahanun, "Faktor-Faktor Yang Berhubungan dengan Ketepatan Waktu Pelaksanaan Penyelidikan Epidemiologi Demam Berdarah Dengue (DBD) Tingkat Puskesmas Di Kota Makassar Tahun 2015." ,*Al-Sihah : Public Health Science Journal* 7(2), 2015.

[2] World Health Organization (2022) Dengue and severe dengue. Available at: https://www.who.int/news-room/factsheets/detail/dengue-and-severe-dengue

[3] M. Andani, " Hubungan Pengetahuan Dan Sikap Kepala Keluarga Dengan Pelaksanaan 4m Plus Dalam Mengatasi DBD Di Wilayah Kerja Puskesmas Basuki Rahmad Bengkulu Tahun 2016", Skripsi, Program Studi Ilmu Keperawatan ,Stikes Dehasen Bengkulu,2016.

[4] Malela, "Asuhan Keperawatan pada Anak dengan Dengue Hamorrhagic Fever (DHF) di RSUD dr. Wahid Sudiro", Asuhan Keperawatan Hemorrhagic.

[5] N. M. S. Jayanti, I. W. Sumarjaya, dan M. Susilawati, "Pemodelan Penyebaran Kasus Demam Berdarah Dengue (DBD) di Kota Denpasar dengan Metode SpatialAutoregressive (SAR)", *E-Jurnal Matematika*, Vol. 6 (1), Januari 2017, pp.37-46, 2017

[6] A. C. Kurniawan, A. Maharani, F.E. Zulvia, M. Lusiana, R. Vikaliana, Y. A. Iskandar dan N. Ruswandi, Penerapan Statistik dalam Logistik, Jakarta: Media Sains Indonesia, 2023.

[7] A. I. Sofiyat, A. Tjalla, & Mahdiyah, "Pemodelan Regresi Logistik Biner Terhadap Penerimaan Pegawai Di PT XYZ Jakarta". *Matematika Sains*, 1(1), 1–11, 2023.

[8] S. A. R. Mana, Erfiani, Indahwati, A. Fitrianto, & R. Amelia, "Faktor –Faktor yang Memengaruhi Permasalahan *Stunting* di Jawa Barat Menggunakan Regresi Logistik Biner". *J Statistika,* 15(2), 265 -274, 2022.

[9] R. Oktavianingsih dan W. Pramesti, "Pengetahuan Ibu tentang Imunisasi terhadap Kesehatan Balita di Desa Sumput Kecamatan Driyorejo Kabupaten Gresik". *J Statistika*, 10(2), 27-34, 2017.

[10] Kementerian Kesehatan Republik Indonesia. Profil Kesehatan Indonesia. Profil Kesehatan Provinsi Bali. 2022.

[11] A. Agresti, Categorical Data Analysis, John Willey and Sons, New York,1990.

[12] Zainal M, M., A. J. Rindengan, dan W. Ch. D. Weku, "Penggunaan Association Rule Data Mining Untuk Menentukan Pola Lama Studi Mahasiswa F-MIPA UNSRAT", *Jurnal de Cartesian*, 3(1):1-8, 2014.

[13] D. W. Hosmer, & S. Lemeshow, Applied Logistic Regression, John Wiley and Son, New

York, 2000.

[14] Misna, Rais, L. T. Utami, "Analisis Regresi Logistik Biner untuk Mengklasifikasi Penderita Hipertensi Berdasarkan Merokok di RSU Mokopido Tolitoli" dalam Natural Science : *Journal Of Science And Technology*, Vol 7 (3) : 341-348. Palu : Universitas Tadulako, 2018.

[15] R. D. Haloho & S. I. Santoso, "Analisis profitabilitas pada usaha peternakan sapi perah di Kabupaten Semarang". *Ragam*, 13(1), 2013

[16] Lai, X., & Liu, L, "A simple test procedure in standardizing the power of HosmerLemeshow test in large data sets", *Journal of Statistical Computation and Simulation, 1*. 2018.

[17] I. Ghozali. Aplikasi Analisis Multivariate Dengan Program SPSS. Semarang: Badan Penerbit Universitas Diponegoro, 2011.

Association of Poverty Categories, Educational Characteristics, and Area of Residence in Indonesia Using a Three-Way Log-Linear Model

Ainun Salsabila (1), Erfiani (2), Indahwati (3), Anwar Fitrianto (4), Muftih Alwi Aliu (5)

1,2,3,4,5 Department of Statistics, IPB University

Meranti Street, Dramaga, Bogor, West Java 16680

e-mail: ainunsalsabila@apps.ipb.ac.id(1), erfiani@apps.ipb.ac.id(2), indah.stk@gmail.com(3), anwarstat@gmail.com(4), muftihalwialiu@apps.ipb.ac.id(5)

ABSTRAK

Tabel kontingensi adalah salah satu cara untuk menyajikan data dengan semua variabel kategorik. Analisis yang digunakan untuk memodelkan tabel kontingensi adalah model *log-linear*. Model *log-linear* juga digunakan untuk mengestimasi parameter dan melihat hubungan antar variabel. Tujuan dari penelitian ini adalah ingin memanfaatkan model *log-linear* tiga arah tersebut untuk memodelkan dan melihat hubungan antara variabel kategori kemiskinan, karakteristik pendidikan (tingkat pendidikan dan kemampuan membaca dan menulis) kepala rumah tangga, dan daerah tempat tinggal di Indonesia pada tahun 2023. Penelitian dilakukan dengan membentuk model *log-linear saturated* dan *homogeneous* terlebih dahulu, lalu membandingkan selisih nilai *deviance* dari kedua model tersebut dengan nilai *chi-square* tabel atau dapat juga dengan melihat nilai AIC terkecil untuk menentukan model terbaik. Hasil yang diperoleh adalah model *saturated* yang signifikan. Artinya terdapat hubungan antara variabel kategori kemiskinan, tingkat pendidikan kepala rumah tangga, dan daerah tempat tinggal. Terdapat pula hubungan antara variabel kategori kemiskinan, kemampuan membaca dan menulis kepala rumah tangga, dan daerah tempat tinggal. Selain itu, terdapat kecenderungan kemiskinan yang lebih besar bagi kepala rumah tangga yang memiliki pendidikan sekolah dasar atau lebih rendah, serta tidak dapat membaca dan menulis.

Kata kunci: Tabel Kontingensi; Model *Log-Linear*; Kemiskinan; Karakteristik Pendidikan; Daerah Tempat Tinggal

ABSTRACT

Contingency tables are one way to present data with all categorical variables. The analysis used to model the contingency table is a log-linear model. The log-linear model is also used to estimate parameters and see the association between variables. This research aims to utilize the three-way log-linear model to model and see the association between poverty category variables, educational characteristics (level of education and reading and writing ability) of the head of the household, and area of residence in Indonesia in 2023. Research is done by forming a saturated and homogeneous log-linear model first, then comparing the difference in deviance values from the two models with the table chi-square value or choosing the smallest AIC value to determine the best model. The results obtained are a significant saturated model. This means that there is an association between the poverty category variable, the education level of the head of the household, and the area of residence. There is also an association between the poverty category variable, the reading and writing ability of the head of the household, and the area of residence. In addition, there is a greater tendency for poverty for heads of households who have a primary school education or less and cannot read and write.

Keywords: *Contingency Table; Log-Linear Model; Poverty; Characteristics of Education; Area of Residence*

INTRODUCTION

Often in research, the data used is of the categorical type. One way of presenting categorical data is with a contingency table. The contingency table consists of rows and columns in the form of the category level of each observed categorical variable. One development of contingency table analysis is the log-linear model. The log-linear model is used to model the number of cells in a contingency table and describe the association between categorical variables [1] and measure propensity ratio statistics such as contrast values or odds ratio [2]. In this study, researchers want to model the poverty category, educational characteristics (level of education and reading and writing ability) of the head of the household, and area of residence in Indonesia using a three-way log-linear model, so that we can find out how the association and tendency of the variables. Because poverty and education are very complex problems and have existed for a long time, especially in developing countries like Indonesia. Moreover, we know that poverty and education in urban areas tend to be better than in rural areas. These two things are related to the welfare of the Indonesian people. Therefore, with this modeling, it is hoped that it can help the government to better understand the conditions of poverty and education in urban and rural areas. So, if there is still an association and gap between poverty and education in urban and rural areas, the government can take more appropriate policies to overcome this, so that people in Indonesia will be more prosperous in the future.

The poverty problem in Indonesia itself is a problem that continues to be considered and addressed today. According to Badan Pusat Statistik (BPS), the economic inability to fulfill basic needs (food and non-food) for a decent life as measured in terms of expenditure is considered poverty. Residents are said to be poor if their average per capita expenditure per month is below the poverty line, namely IDR 486,168 [3]. The number of poor people in Indonesia continues to fluctuate, but the decline is not significant. This makes Indonesia still ranked fourth as a poor country in Southeast Asia. Reducing poverty to 6.5% to 7% in 2024 is the national development target stated in the Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 [4]. However, this value is still far from the poverty rate in 2023 which will still be 9.36%.

One of the causes of poverty is low human resources caused by low levels of education. The low levels of education in urban and rural areas are different, where at each higher level of education, more people in urban take that education than people in rural. The reading and writing abilities of people in urban and rural areas are also different, where more people in rural cannot read and write. This happens because there is still a gap or unequal distribution of quality education in urban and rural areas. The number of schools in rural tends to be fewer than in urban areas and access is sometimes not easy, because the locations of schools are sometimes far from residential areas. Facilities, infrastructure, school buildings, and the number and quality of teaching staff in rural are still less than in urban areas. Access to education through technology is also still difficult for schools in rural areas. In addition, the expensive cost of education for higher levels of education makes many people in rural with low incomes prefer not to continue their education.

Research conducted by [5] in his research entitled "Potret Pendidikan di Daerah Terpencil Kampung Manceri Cigudeg Kabupaten Bogor" explains that the low level of education in this rural occurs due to limited teachers and administrative staff, low teacher welfare, lack of facilities and school infrastructure, unequal distribution of education, educational culture and very low economic factors. If someone has difficulty getting an education or has a low level of education, it will also be difficult to get a job. According to research conducted by Astrini, education has a big influence on poverty, because education is a form of investment in human resources where the higher a person's education, the more skills will also increase and encourage work productivity [6]. Low human productivity will result in low income received.

Research conducted by [7] their research entitled "Pengaruh Tingkat Pendidikan Terhadap Kemiskinan di DKI Jakarta" also states that education has a significant effect on poverty levels. The higher a person's education level, the lower the poverty level. A person's higher level of education will increase their understanding, skills, and abilities so that the quality of work productivity will be good. If the quality of work productivity is good, the opportunity to get a decent job will be greater. Decent work will produce higher income, so avoiding poverty and increasing welfare. High people's income will also have an impact on reducing the poverty level of a country.

Apart from that, it is also important for someone to be able to read and write to help improve their quality of life to avoid poverty. If someone can read and write even though they may not have received a formal education, then that person can at least broaden their knowledge and look for ideas to create job opportunities by opening a business as long as they have the will. So, there is still an opportunity to earn income that can help get out of poverty if the business is successful. In contrast to someone who cannot read and write, it will be difficult for that person to develop and improve their quality of life, because their income will be very dependent on other people or only on the harvest. As a result, it will be difficult for this person to escape poverty.

METHOD

This research uses secondary data from the BPS report entitled "Penghitungan dan Analisis Kemiskinan Makro Indonesia 2023" [8]. This data is data from 341,802 households in all provinces in Indonesia for urban and rural areas in March 2023. This research will form two log-linear models with two different contingency tables. As can be seen in Tables 1 and 2 the variables used in each log-linear modeling consist of three variables. All variables used are considered response variables. The data is processed by researchers in the form of totals because the data comes from BPS in the form of percentages. The data structure used to form the first log-linear model is as follows:

Table 1. Data structure for the first log-linear model

		Education Level of Head of Household			
Area of Residence	**Poverty Category**	**Not Completed in Elementary School**	**Elementary School Equivalent**	**Junior High School Equivalent**	**Senior High School Equivalent**
Urban	Poor	14957 (21.88%)	24726 (36.17%)	12380 (18.11%)	14526 (21.25%)
Urban	Not Poor	11064 (10.79%)	23102 (22.53%)	17247 (16.82%)	36556 (35.65%)
Rural	Poor	20194 (29.54%)	27071 (39.60%)	9823 (14.37%)	9605 (14.05%)
	Not Poor	21810 (21.27%)	37356 (36.43%)	17770 (17.33%)	20488 (19.98%)

Meanwhile, the data structure used to form the second log-linear model is as follows:

Table 2. Data structure for the second log-linear model

Area of Residence	Poverty Category	Reading and Writing Ability			
		Latin Letters	Other Letters	Latin and Other Letters	Cannot Read and Write
Urban	Poor	24890 (36.41%)	704 (1.03%)	39328 (57.53%)	3439 (5.03%)
	Not Poor	37048 (36.13%)	513 (0.5%)	62601 (61.05%)	2379 (2.32%)
Rural	Poor	32286 (47.23%)	827 (1.21%)	27426 (40.12%)	7820 (11.44%)
	Not Poor	43416 (42.34%)	810 (0.79%)	52142 (50.85%)	6173 (6.02%)

The stages of data analysis in this research are as follows:

1. Descriptive Statistics

   This stage is used to see a general overview of the data used.

2. Formation and Testing Log-Linear Models

   The log-linear model is a model that can be used to analyze categorical data in contingency tables by modeling the number of observations in each cell for all combinations of category levels of observed categorical variables. In the log-linear model, there is an assumption that all variables are considered as response (dependent) variables, or in other words there is no distinction between independent and dependent variables. This is caused by the log-linear model which shows dependencies between variables [9].

   The link function used by the log-linear model is the natural logarithm [10]. Log-linear assumes the response variable to have a Poisson distribution [11] The log-linear model is an expansion of the natural logarithm of the frequency for each cell, equal to the mean (constant) plus the lambda parameter for each other variable, plus lambda for all interaction effects, both 2-factor, 3-factor interaction effects and interaction effects for higher orders, if there is interaction.

   Testing in the log-linear model was carried out in stages, namely from saturated with homogeneous, homogenous with conditional, conditional with joint independence, and the last joint independence with complete independence model. If in the first stage, a significant model is found to be saturated, then model testing is not continued to the next stage. On the other hand, if the model obtained is homogenous, then model testing will proceed to the next stage. Therefore, researchers will form a saturated and homogenous model first, and then test the two models. Testing the model by comparing the difference between the deviance values of the saturated and homogenous models with chi-square table value.

3. Selection of The Best Model

   At this stage, two models will be compared to select one that is significant in stages from a saturated model to a simple model without interaction. Each model compared is calculated for its deviance value. The best model is determined by the difference between two deviance values which is smaller than the table Chi-Square value or the model with the smallest AIC value. The deviance value can be calculated with the following equation [12]

$$D = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left[y_{ijk}log\left(\frac{y_{ijk}}{n_{ijk}\hat{\pi}_{ijk}}\right) + \left(n_{ijk} - y_{ijk}\right)\log\left(\frac{n_{ijk}-y_{ijk}}{n_{ijk}-n_{ijk}\hat{\pi}_{ijk}}\right)\right] \quad (1)$$

4. Interpretation of Model Coefficient

Interpretation of coefficients for variable parameters without interactions uses odds values and for variable parameters with interactions uses odds ratio. The values of these parameters also will be useful for estimating cell values in the contingency table. Odds are defined as the ratio of the probability that the event will occur to the probability that the event will not occur [13]:

$$Odds = p/(1-p) \tag{2}$$

The odds ratio is a ratio of two odds values. The odds ratio can be obtained [14]:

$$\theta = \exp(\beta_j) \tag{3}$$

The stages of the analysis can be described with the flowchart in Figure 1 as follows:

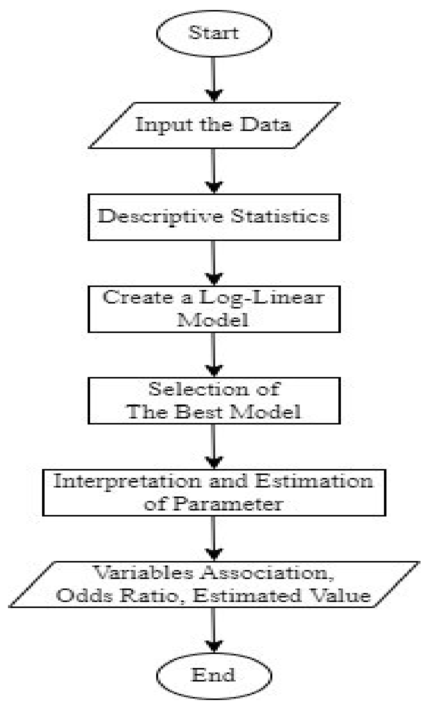


Figure 1. Research Flowchart

RESULT AND DISCUSSION

A. Descriptive Statistics

Based on Table 1, it can be seen that the percentage of heads of poor households who have low education (not completed elementary school or elementary school equivalent) is higher than non-poor households. This shows that the participation rate or interest of heads of households in rural in pursuing higher education is still relatively low compared to heads of households who live in urban. In addition, the number of heads of poor households from all levels of education is greater in rural than in urban areas. The number of households in rural who have not completed elementary school and only graduated from elementary school also shows that there are more heads of poor households.

Economic problems are also the reason why many people in rural do not pursue higher education. Most rural people work as farmers, traders, or fishermen, whose daily income is uncertain and tends to be smaller than that of people in urban. Research conducted by [15] it is explained that "Unpredictable income in Labuan Kepak Village every month and not coming from economically well-off families is the reason why many people in this village only study up to elementary school because higher education is considered expensive". In fact, according to research conducted by [16], if the percentage of the population with higher education increases, the percentage of the poor will decrease. Although more heads of households in urban areas have higher education, the highest level of education is still high school. This can also be caused by the high cost of education at university. Meanwhile, based on Table 2, it can be seen that more heads of households who cannot read and write (Latin letters and/or other letters) are classified as poor than not poor, whether they live in urban or rural. There are more heads of households who cannot read and write in rural than in urban areas. As a result, many households are poor in rural areas.

B. Formation and Testing of Log-Linear Models

1. First Log-Linear Model

Table 3. Deviance Value, Degrees of Freedom, and AIC Model 1

Model	Deviance	DF	AIC	Model	Deviance	DF	AIC
Saturated	0	0	266.85	Homogenous	1867.3	4	2126.2

- Hypothesis:
  H_0: $\lambda_{ijk}^{ABC} = 0$ (There is no three-way interaction / The model formed is a homogenous model)
  H_1: $\lambda_{ijk}^{ABC} \neq 0$ (There is a three-way interaction / The model formed is a saturated model)
- Significance Level:
  $\alpha = 5\%$
- Test Statistics Value:
  ΔD = Deviance homogenous model – Deviance saturated model
  = 1867.3 – 0 = 1867.3
  db = db Deviance homogenous model – db Deviance saturated model
  = 4 – 0 = 4
- Critical Value:
  Reject H_0 if $\Delta D > \chi^2_{0.05;db} = \chi^2_{0.05;4} = 9.48$
- Decision:
  Reject H_0 because ΔD (1867.3) > $\chi^2_{0.05;4}$ (9.48)
- Conclusion:
  With a confidence level of 95%, it can be concluded that there is sufficient evidence to say that there is a three-way interaction or that the model formed is a saturated model.

Because the test results were rejected H_0, the test stopped or did not proceed to other model tests.

2. First Log-Linear Model

Table 4. Deviance Value, Degrees of Freedom, and AIC Model 2

Model	Deviance	DF	AIC	Model	Deviance	DF	AIC
Saturated	0	0	205.64	Homogenous	373.37	3	573.01

- Hypothesis:
  H_0: $\lambda_{ijk}^{ABC} = 0$ (There is no three-way interaction / The model formed is a homogenous model)
  H_1: $\lambda_{ijk}^{ABC} \neq 0$ (There is a three-way interaction / The model formed is a saturated model)
- Significance Level:
  $\alpha = 5\%$
- Test Statistics Value:
  ΔD = Deviance homogenous model – Deviance saturated model
  = 373.37 – 0 = 373.37
  $db = db$ Deviance homogenous model – db Deviance saturated model
  = 3 – 0 = 0
- Critical Value:
  Reject H_0 if $\Delta D > \chi^2_{0.05;db} = \chi^2_{0.05;3} = 7.81$
- Decision:
  Reject H_0 because ΔD (373.37) > $\chi^2_{0.05;3}$ (7.81)
- Conclusion:
  With a confidence level of 95%, it can be concluded that there is sufficient evidence to say that there is a three-way interaction or that the model formed is a saturated model.

Because the test results were rejected H_0, the test stopped or did not proceed to other model tests.

C. Selection of The Best Model

Model testing resulted in the model formed being a three-way interaction model (saturated model). To prove that this model is indeed the best, it can be seen from the smallest AIC value. Based on Tables 3 and 4, it can be seen that the model that has the smallest AIC value is the saturated model, with a value of 266.85 for the first log-linear model and 205.64 for the second log-linear model. Therefore, the best model to be used in this research is the saturated model with the equation: $log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$, where A is the poverty category, B is the characteristics of education, and C is the area of residence.

D. Interpretation of Coefficients in The Best Model

1. First Log-Linear Model

The parameter coefficients λ of the saturated model obtained can be seen at Table 5, where NCES is Not Completed in Elementary School, ES is Elementary School, JHS is Junior High School, and SHS is Senior High School.The model coefficients are as follows:

Table 5. Coefficient and Odds Ration in The Best Model 1

Variable	Coefficient $\boldsymbol{\lambda}$	$\boldsymbol{Exp}$ $(\boldsymbol{\lambda})$	Variable	Coefficient $\boldsymbol{\lambda}$	$\boldsymbol{Exp}$ $(\boldsymbol{\lambda})$
Intercept	8.5403		Poor:SHS	0.3633	1.4382
Poor	-1.1209	0.3260	Poor:Urban	-0.9865	0.3729
NCES	1.4498	4.2623	NCES:Urban	-1.7251	0.1781

ES	1.9879	7.3004	ES:Urban	-1.5270	0.2172
JHS	1.2449	3.4727	JHS:Urban	-1.0763	0.3408
SHS	1.3873	4.0039	SHS:Urban	-0.4675	0.6266
Urban	1.0465	2.8476	Poor:NCES:Urban	1.3650	3.9157
Poor:NCES	1.0440	2.8404	Poor:ES:Urban	1.3765	3.9611
Poor:ES	0.7989	2.2231	Poor:JHS:Urban	1.2478	3.4826
Poor:JHS	0.5282	1.6958	Poor:SHS:Urban	0.8212	2.2732

Based on the odds ratio value in Table 5, it can be explained that the probability the head of the household in Indonesia is classified as poor is 0.3260 times compared to not being poor or it could be said that the probability of the head of the household not being poor is greater. The probability that the head of the household has an elementary school education is 7.3004 times compared to college and this probability is also greater than the probability for other levels of education compared to college. This means that more heads of households in Indonesia only have an elementary school education.

Regardless of the area of residence, the odds of the head of the household being classified as poor (compared to not poor), if he has not completed elementary school is 2.8404 times compared to the same odds if the head of the household is college-educated. This means that the greatest tendency for heads of households to be classified as poor is when they have not completed elementary school. Without paying attention to education level, the odds of the head of the household being classified as poor if he lives in an urban area is 0.3729 times compared to the same odds if the head of the household lives in a rural area. This means that the tendency for poor households to live in urban is smaller than those living in rural.

The odds of the head of the household having a high school education (compared to college) if he lives in an urban area and regardless of poverty category is 0.6266 times compared to the same odds if the head of the household lives in a rural area. This means that heads of households in Indonesia who live in urban are more likely to have a college education than those who live in rural. The odds of the head of a household being classified as poor (compared to not poor), if he has an elementary school education and lives in an urban area is 3.9611 times compared to the same odds if the head of the household is college-educated and lives in a rural area.

The results of this study show that not always living in an urban area will guarantee someone's escape from poverty. This is because life in urban is more difficult than in rural areas where money is the main resource for urban residents to fulfill all life's needs. In contrast to rural areas, even though financial conditions are low, rural residents can still fulfill basic needs such as food by utilizing harvested crops [17]. Apart from that, the expensive cost of living in the urban area can also make it more difficult for someone who lives in the urban area to fulfill their living needs if they don't have a job with a high income, or if they have a large family member.

Apart from that, population density continues to increase in urban but there are fewer employment opportunities, plus low levels of education, making the tendency for poverty in urban areas to be greater. This is because many people find it difficult to earn income to meet their daily needs. Ultimately, many people become beggars, buskers, scavengers, or homeless people.

Population density and poverty in urban areas tend to be high because many rural residents move to cities in the hope of getting decent jobs with high incomes, but their education is low.

2. First Log-Linear Model

The following are the parameter coefficients λ of the saturated model obtained:

Table 6. Coefficient and Odds Ration in The Best Model 2

Variable	Coefficient λ	$Exp\ (\lambda)$	Variable	Coefficient λ	$Exp\ (\lambda)$
Intercept	8.7279		Poor:LOL	-0.8790	0.4152
Poor	0.2365	1.2668	Poor:Urban	0.1320	1.1411
LL	1.9506	7.0332	LL:Urban	0.7949	2.2142
OL	-2.0309	0.1312	OL:Urban	0.4967	1.6434
LOL	2.1338	8.4468	LOL:Urban	1.1363	3.1153
Urban	-0.9535	0.3854	Poor:LL:Urban	-0.2336	0.7917
Poor:LL	-0.5327	0.5870	Poor:OL:Urban	0.1637	1.1779
Poor:OL	-0.2157	0.8060	Poor:LOL:Urban	0.0456	1.0467

Where LL is Latin Letters, OL is Other Letters, and LOL is Latin and Other Letters. Based on Table 6, it can be explained that the probability of the head of a household in Indonesia being classified as poor is 1.2668 times compared to not being poor. The probability that the head of the household can read and write Latin letters and other letters is 8.4468 times compared to not being able to read and write (Latin letters and/or other letters). This means that more heads of households in Indonesia can read and write (Latin letters and others).

If without paying attention to the area of residence, the odds of the head of a poor household (compared to not being poor), if he can read and write (Latin letters and others), is 0.4152 times compared to the same odds if he cannot read and write. This means that if the head of the household cannot read and write, the tendency is to be classified as poor. This is in accordance with the research conducted by [18] that the higher the literacy rate, the poverty rate will decrease, because the better the reading and writing skills, the better the quality and quality of human resources.

The odds of a poor household, living in an urban, are 1.1411 times compared to the same odds if they live in rural. However, without considering poverty, the odds of the head of the household being able to read and write (compared to not being able to read and write) who lives in urban is 3.1153 times compared to the same odds if he lives in rural areas. This means that the facilities and quality of education in rural are still less good than in urban because more heads of households in urban areas can read and write.

Then, if we pay attention to the interaction of the three variables, the odds of the head of the household being classified as poor if he can only read and write Latin letters and lives in urban is 0.7917 times. This means that if the head of a household who lives in an urban can read and write Latin letters, he is less likely to be poor. However, the odds of the head of the household being classified as poor if he can read and write (Latin letters and others) and lives in urban is 1.0467 times. This means that if heads of households who live in urban can read and write (Latin letters and others), they are more likely to be poor. This may be caused by a low level of education or other factors not included in this research because poverty factors are very complex.

E. Parameter Estimation

This stage is used to estimate the μ parameter value (sell value in the contingency table) according to the model obtained.

1. First Log-Linear Model

   Using the values in Table 5, one example of the equation used to estimate the μ_{111} parameter in this study is as follows:

   $log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$

   $log(\mu_{111}) = 8.5403 - 1.1209 + 1.4498 + 1.0465 + 1.0440 - 0.9865 - 1.7251 + 1.3650$

2. Second Log-Linear Model

   Using the values in Table 6, one example of the equation used to estimate the μ_{111} parameter in this study is as follows:

   $log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$

   $log(\mu_{111}) = 8.7279 - 0.2365 + 1.9506 - 0.9535 - 0.5327 - 0.1320 - 0.7949 - 0.2336$

CONCLUSION

Based on the research results discussed, it can be concluded that the poverty category, the education level of the head of the household, and the area of residence are related to each other. Apart from that, the poverty category, the reading and writing ability of the head of the household, and the area of residence are also related to each other. Poverty in Indonesia in 2023 will mostly occur among heads of households who have elementary school education or below and who cannot read and write. This shows that educational characteristics (level of education and ability to read and write) are related to poverty. The results of this research also provide information that the poverty trend of heads of households living in urban is greater than in rural. This shows that life in urban is not always easy, especially if you don't have a job with a high income because you have a low level of education and cannot read and write.

Therefore, the government needs to pay attention to and improve education in Indonesia so that all Indonesian people can obtain higher education and at least not be illiterate so that the opportunity to get a decent job with a high income is much greater. Apart from that, the government must also be able to create sufficient job opportunities for all people who are looking for work. This must be done equally in both urban and rural so that all Indonesian people can avoid poverty and live prosperously.

REFERENCES

[1] S. A. Pujiati, "Penerapan Model Log Linear Untuk Mengetahui Hubungan Antara Status Sekolah, Jumlah Guru Berpendidikan S1 Dan Angka Mengulang Pada Siswa Sd," *J-Statistika*, vol. 4, no. 1, pp. 31–34, 2018.

[2] T. A. I. Satria, N. Imro'ah, and N. M. Huda, "Penerapan Model Log Linier Dalam Menganalisis Tabel Kontingensi Dua Arah," *Buletin Ilmiah Math. Stat. dan Terapannya (BIMASTER)*, vol. 12, no. 6, pp. 545–552, 2023.

[3] A. Zaqiah, M. Triani, and I. Yeni, "Pengaruh Pendidikan, Pengangguran dan Jumlah Penduduk Terhadap Tingkat Kemiskinan di Indonesia," *Jurnal Kajian Ekonomi dan Pembangunan*, vol. 5, no. 3, pp. 33–42, 2023, [Online]. Available: http://ejournal.unp.ac.id/students/index.php/epb/index

[4] Y. M. Hutahaean and J. R. H. Sitorus, "Faktor-Faktor yang Mempengaruhi Kemiskinan Rumah Tangga Bekerja di Pulau Jawa," *Seminar Nasional Official Statistics*, pp. 1165–1176, 2022.

[5] M. Abduh, A. Angga Basiru, M. W. Narayana, N. Safitri, and R. Fauzi, "Potret Pendidikan di Daerah Terpencil Kampung Manceri Cigudeg Kabupaten Bogor," *Jurnal Citizenship Virtues*, vol. 2, no. 1, pp. 291–300, 2022.

[6] Kamaruddin, N. Sudiyarti, Y. Kurnilawan, and R. Rachman, "Pengaruh Tingkat Pendidikan Terhadap Jumlah Penduduk Miskin Di Kabupaten Sumbawa Barat Tahun 2015-2019," *Jurnal Ekonomi dan Bisnis*, vol. 8, no. 2, pp. 98–106, 2020.

[7] R. Susanto and I. Pangesti, "Pengaruh Tingkat Pendidikan Terhadap Kemiskinan Di Dki Jakarta," *Journal of Applied Business and Economic*, vol. 5, no. 4, pp. 340–350, 2019.

[8] Badan Pusat Statistik, "Penghitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2023," vol. 15, 2023.

[9] S. F. Sihotang, "Analisis Model Log Linier Tiga Dimensi Untuk Data Kualitatif Dengan Metode Forward," *Journal of Mathematics Education and Science*, vol. 6, no. 1, pp. 2528–4363, 2020.

[10] G. Yesiltas and I. Paek, "A Log-Linear Modeling Approach for Differential Item Functioning Detection in Polytomously Scored Items," *Educ Psychol Meas*, vol. 80, no. 1, pp. 145–162, Feb. 2020, doi: 10.1177/0013164419853000.

[11] A. A. Akinrefon, R. Emmanuel, and A. Okolo, "Log-Linear Models For Hiv/Aids Prevalence In Adamawa State, Nigeria," *Fudma Journal Of Sciences*, vol. 7, no. 3, pp. 103–109, Jul. 2023, doi: 10.33003/fjs-2023-0703-1804.

[12] E. Priscilla, J. Prinssesa, M. S. J. Aurelia, and E. S. Nugraha, "Log Linear Model on Contingency Table to Analyze Relationship between Age, Income, and Health Insurance Ownership," *Journal of Actuarial, Finance and Risk Management (JAFRM)* , vol. 1, no. 1, pp. 17–26, 2022.

[13] S. H. Park and K. Han, "How to Clearly and Accurately Report Odds Ratio and Hazard Ratio in Diagnostic Research Studies?," *Korean J Radiol*, vol. 23, no. 8, pp. 777–784, Aug. 2022, doi: 10.3348/kjr.2022.0249.

[14] Y. Rahmawati and P. C. Taylor, *Empowering Science and Mathematics for Global Competitiveness*. The Netherlands: CRC Press/Balkema, 2019.

[15] Y. H. Ladaria, J. Lumintang, and C. J. Paat, "Kajian Sosiologi Tentang Tingkat Kesadaran Pendidikan Pada Masyarakat Desa Labuan Kapelak Kecamatan Banggai Selatan Kabupaten Banggai Laut," *Jurnal Ho;istik*, vol. 13, no. 2, pp. 1–15, 2020.

[16] K. Laila Nirmala, W. Pramesti, and F. Fitriani, "Spatial Autoregresive Moving Average Pada Pemodelan Persentase Penduduk Miskin Kabupaten/Kota Di Provinsi Jawa Timur Tahun 2020," *J Statistika*, vol. 15, no. 1, pp. 158–166, 2022, [Online]. Available: www.unipasby.ac.id

[17] M. A. Lasaiba, "Perkotaan dalam Perspektif Kemiskinan, Permukiman Kumuh dan Urban Heat Island (Suatu Telaah Literatur)," *GEOFORUM*, vol. 1, no. 2, pp. 63–72, Dec. 2022, doi: 10.30598/geoforumvol1iss2pp63-72.

[18] I. Made, S. Yoga, I. Komang, T. E. Putra, and R. B. Utomo, "Pengaruh Upah Minimum dan Angka Melek Huruf Terhadap Tingkat Pengangguran dan Kemiskinan di Provinsi Bali," *Jurnal SUTASOMA (Science Teknologi Sosial Humaniora)*, vol. 01, no. 01, 2022, [Online]. Available: https://doi.org/00.00000

J Statistika Vol. 17, No. 1, (2024), Hal. 635 - 644 jurnal.unipasby.ac.id

Factors Affecting the Resilience Index Food in Papua Province and West Papua Province Using a Spatial Model Approach

Dewi Sri Hastuti (1), Safa'at Yulianto (2)

[1,2]Institut Teknologi Statistika dan Bisnis, Muhammadiyah Semarang

Jl. Prof. Dr. Hamka Km 01 No. 17 Tambakaji Ngaliyan

e-mail: dewisrihastuti0@gmail.com(1), safaat.yulianto@itesa.ac.id(2)

ABSTRAK

Indeks Ketahanan Pangan adalah ukuran dari indikator-indikator untuk menghasilkan nilai komposit yang mencerminkan status ketahanan pangan di suatu wilayah. Ketahanan pangan berperan penting dalam pembangunan berkelanjutan, mencakup ketersediaan pangan, pelestarian lingkungan, dan keseimbangan ekonomi, juga sebagai dasar pertumbuhan ekonomi, pencegah kemiskinan, dan ketidaksetaraan. Di Indonesia, dengan perkiraan pertumbuhan penduduk mencapai 430 juta jiwa pada tahun 2050, tantangan dalam memenuhi kebutuhan pangan semakin besar. Komitmen Indonesia terhadap Sustainable Development Goals (SDGs) mencakup upaya mengakhiri kelaparan dan mempromosikan pertanian berkelanjutan. Penelitian ini bertujuan untuk menerapkan analisis regresi spasial pada Provinsi Papua dan Papua Barat untuk menentukan model terbaik dan faktor-faktor signifikan yang mempengaruhi Indeks Ketahanan Pangan di wilayah tersebut guna mengidentifikasi tantangan yang dihadapi wilayah tersebut dalam memperhitungkan ketersediaan pangan masyarakatnya sekaligus membantu dalam menyusun upaya untuk mengatasinya. Digunakan lima variabel prediktor dengan asumsi memiliki pengaruh yang signifikan terhadap Indeks Ketahanan Pangan, penelitian ini mengkaji persamaan regresi spasial dengan pendekatan wilayah SAR, SEM, dan SARMA. Diperoleh hasil yang menunjukkan bahwa model SEM terpilih dengan nilai *p-value* 0.0082581 yang tepat untuk mengidentifikasi ketergantungan efek spasial terhadap Indeks Ketahanan Pangan di Provinsi Papua dan Provinsi Papua Barat. Angka Harapan Hidup Saat Lahir (AHH), Prevalensi Balita Stunting (PBS), Persentase Penduduk Miskin (PPM), Tingkat Pengangguran Terbuka (TPT), dan Rata-Rata Lama Sekolah (RLS) adalah faktor signifikan yang memengaruhi Indeks Ketahanan Pangan di Provinsi Papua dan Provinsi Papua Barat secara spasial.

Kata kunci: Indeks Ketahanan Pangan; Regresi Spasial; Provinsi Papua, Provinsi Papua Barat, Lagrange Multiplier (Error)

ABSTRACT

The Food Security Index is a measure of indicators to produce a composite value that reflects the status of food security in a region. Food security plays an important role in sustainable development, including food availability, environmental preservation and economic balance, as well as being the basis for economic growth, preventing poverty and inequality. In Indonesia, with an estimated population growth of 430 million people in 2050, the challenge of meeting food needs is increasing. Indonesia's commitment to the Sustainable Development Goals (SDGs) includes efforts to end hunger and promote sustainable agriculture. This research aims to apply spatial regression analysis to the Provinces of Papua and West Papua to determine the best model and significant factors that influence the Food Security Index in the region in order to identify the challenges faced by the region in calculating the food availability of its people as well as assist in developing efforts to overcome them. Five predictor variables were used with the assumption that they have a significant influence on the Food Security Index. This research examines the spatial regression equation using the SAR, SEM and SARMA regional approaches. The results obtained showed that the selected SEM model with a p-value of 0.0082581 was appropriate for identifying the dependence of spatial effects on Food Security Index in Papua Province and West Papua Province. Life Expectancy at Birth, *Prevalence of Stunting Toddlers, Percentage of Poor Population, Open Unemployment Rate, and Average Length of Life are significant factors that influence the Food Security Index in Papua Province and West Papua spatially.*

Keywords: *Food Security Index; Spatial Regression; Papua Province, West Papua Province, Lagrange Multiplier (Error)*

INTRODUCTION

Development and food security have strong ties and influence each other in the context of sustainable development. Development that neglects one's own ability to meet the basic needs of its people will lead to high dependence on other countries, which will ultimately eliminate state sovereignty [1] . In the context of sustainable development, food security is not enough just to maintain aspects of food availability, but also to harmonize efforts to preserve the environment and achieve long-term economic balance, building a solid foundation for progress and the welfare of society in the future. Food security is one of many important issues in the development of a country with the agricultural sector as the main food provider, especially for developing countries, because it plays an important role as one of the main development targets as well as the main tool in economic development [2] .

Food security not only acts as a basis for sustainable economic growth with sufficient energy availability and productivity, but also as a preventive measure for poverty and inequality, considering that equitable access to food plays an important role in improving the welfare of the entire community. The historical experience of Indonesia's development confirms that the issue of food security is related to economic stability, especially in the face of inflation, rising costs of living, and national political stability [3] . The importance of food security for national development shows that it is not only necessary to have adequate food availability, but also to pay attention to aspects of safety, quality and fair distribution for society.

The prediction of Indonesia's population growth, which is expected to double with an increase of around 1.5% per year to around 430 million people in 2050, will have a major impact on meeting food needs. As the population increases, demand for food will also continue to increase, because population size has a direct effect on food availability [4] . There is a need for integrated efforts involving the agricultural sector, population policy, and natural resource management to respond to the need for increased food production, diversification of food resources, and the establishment of equitable food distribution policies to maintain food security in line with sustainable population and economic growth.

In the 2030 agenda for Sustainable Development, Indonesia is committed to realizing the targets of "ending hunger, achieving food security, improving nutrition, and promoting sustainable agriculture" [5] . Food Security Index plays an important role in assessing the success of developing food stability in a region, assessing the region's ability to fulfill government responsibilities, and functions as an instrument for determining regional interests and program interventions [6] .

According to the Global Food Security Index , Indonesia's Food Security Index in 2022 is at the level of 60.2, or an increase of 1.7% compared to 2021 which is in 69th place out of 113 countries and below the global average of 62.2. Bali province's Food Security Index is 85.19, the highest value in 2022, while there are two provinces (5.88%), namely Papua and West Papua, which are included in provinces with low Food Security Index, at 37.80 and 45.92 respectively. In general, there appears to be a tendency for provinces with Food Security Index in the very resistant category to be close to provinces with Food Security Index in the very resistant category as well, and this also happens for provinces with Food Security Index in the lower category. Based on this distribution, there are indications that the Food Security Index between neighboring provinces influence each other spatially, so that the condition of the Food Security Index in a province can be related to the condition of the Food Security Index in its surrounding/neighboring provinces.

From the description above, spatial regression analysis will be used in this research to determine the factors that significantly influence the Food Security Index in Papua Province and West Papua Province by considering the influence of location, analyzed by testing spatial effects using spatial dependencies. Spatial dependency explains that the locations of research objects are related or related to each other [7] . In modeling spatial dependencies, several types of models emerge, such as the Spatial Autoregressive Model (SAR) which indicates a relationship between response values at various locations, Spatial Error Model (SEM) which indicates a relationship between error values at various locations, and Spatial Autoregressive Moving Average (SARMA) which indicates the relationship between response values and error values at various locations.

Research related to Food Security Index was previously carried out by Ayu Safitri, Baharuddin, Agusrawati, Bahriddin Abapihi, Ruslan Gusti Ngurah Adhi Wibawa (2022) regarding Modeling of Factors that Influence Food Security Index in Southeast Sulawesi which can be analyzed using the best spatial dependency modeling, namely the Spatial Error Model (SEM) and shows that the variables that influence the Food Security Index include the Percentage of Malnourished and Stunting Toddlers, the Life Expectancy Rate variable, the Human Development Index variable, the GRDP Rate variable, the Rice Production variable, and the Population Number variable [8] . Other research related to Food Security Index was also conducted by Irma Yani Safitri, Muhammad Arif Tiro, and Ruliana (2022) regarding Spatial Regression Analysis to See Factors that Influence Food Security at the Regency Level in South Sulawesi Province, obtaining the best spatial dependency model, Spatial Error Model (SEM) and shows that the variables that influence the Food Security Index include Per Capita Normative Consumption Ratio to Clean Availability, Percentage of Population Living Below the Poverty Line, Percentage of Households with the Proportion of Expenditure on Food More Than 65% of Total Expenditure, Percentage of Households without Access to Electricity, Percentage of Households without Access to Clean Water, Ratio of the Number of Population per Health Personnel to Population Density Level, Percentage of Toddlers with Height Below Standard (*stunting*), Life Expectancy, Average Length Of School For Women Over 15 Years [9].

This research aims to utilize spatial regression analysis to identify the best regression model and determine the variables that have a significant influence on the Food Security Index in Papua Province and West Papua Province.

METHOD

Data

This research will use data sourced from the National Food Agency, the Indonesian Ministry of Health, the Papua Province Statistics Center and the West Papua Province Central Statistics Agency which consists of 41 Regencies/Cities. The variables used include the Food Security Index (Y), Life Expectancy at Birth (X_1), Percentage of Stunting Toddlers (X_2), Percentage of Poor

Population (X_3), Open Unemployment Rate (X_4), and Average Years of Schooling (X_5) according to Regency/City in Papua and West Papua Provinces in 2022 .

Data Processing Procedures

The following are the stages that will be carried out in this research:

1. Conduct descriptive analysis on Food Security Index in Papua Province and West Papua Province in 2022
2. Using correlation analysis of the data, to identify any predictor variables that are related to the response variable.

$$r_{xy} = \frac{n\Sigma_{i=1}^{n} x_i y_i - (\Sigma_{i=1}^{n} x_i)(\Sigma_{i=1}^{n} y_i)}{\sqrt{[n\Sigma_{i=1}^{n} x_i^2 - (\Sigma_{i=1}^{n} x_i)^2][n\Sigma_{i=1}^{n} y_i^2 - (\Sigma_{i=1}^{n} y_i)^2]}} \quad (1)$$

3. Multiple linear regression analysis of Food Security Index in Papua Province and West Papua Province involving 41 Regencies/Cities and 5 predictor variables.

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1n} \\ 1 & X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, dan\ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2)$$

If written in equation form it is as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

4. Durbin-Watson test statistic, if autocorrelation occurs then it is indicated that there is a spatial effect on the residual value.

$$\mathrm{D} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \quad (4)$$

5. Determine the spatial weighting matrix using the Queen Continguity weighting matrix
6. Test the effect of spatial dependence with the Moran's I dependency test to evaluate the level of spatial autocorrelation and produce a Moran's scatterplot to visualize the distribution between locations. If there is spatial autocorrelation, the Lagrange Multiplier (LM) test is continued.

$$Z = \frac{I - E(I)}{\sqrt{\mathrm{Var}(I)}} \approx N(0,1) \quad (5)$$

7. Determine the spatial model using the Lagrange Multiplier (LM) test

$$LM_{lag} = \frac{(\varepsilon^T W y)^2}{s^2 (WX\beta)^T M (WX\beta) + Ts^2)} \quad (6)$$

$$LM_{error} = \frac{(\varepsilon^T W \varepsilon / S^2)^2}{T} \quad (7)$$

8. Determining the right Spatial Regression model. The general form of the spatial regression model is as follows:

$$y = \rho W_y + X\beta + u \quad (8)$$

$$u = \lambda w_u + \varepsilon \sim \varepsilon \sim N(0, \sigma^2 I) \quad (9)$$

9. Interpret the model and formulate conclusions.

RESULTS AND DISCUSSION

The Food Security Index in Papua Province and West Papua Province can be described visually as follows:

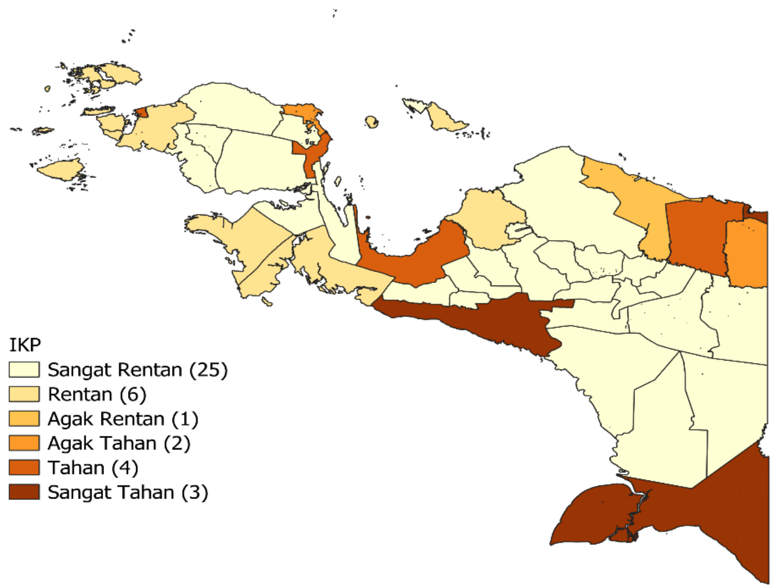


Figure 1. Map of Food Security Index Distribution for Papua Province and West Papua Province

Figure 1 shows the geographical description of Food Security Index in Papua Province and West Papua Province, divided into 6 interval groups. Each interval represents a certain range of values with categories as shown on the map. Apart from that, the colors on the map indicate the Food Security Index level , increasing color intensity on the map indicates a higher Food Security Index value. The highest Food Security Index achievement with a value of 80.55 is in Mimika Regency, while the lowest Food Security Index achievement with a value of 15.66 is in Nduga Regency. Most of the Food Security Index for Regency/City areas are in the very vulnerable category with a total of 25 Regency areas, while Regency/City areas in the very resistant category only cover 2 Regency and 1 City.

Multiple Linear Regression Analysis

The initial step that must be applied to the data before carrying out regression analysis is to carry out correlation analysis first to identify predictor variables that are significant to the response variable. The following correlation results were obtained:

Table 1. Pearson Correlation

Variables	Value	p-value
X_1	0.5596469	0.0001422
X_2	-0.6201429	0.000103
X_3	-0.8339579	1.283e-11
X_4	0.6031419	2.996e-05
X_5	0.826388	2.848e-11

Based on Table 1, the results show that all predictor variables have a linear relationship with Food Security Index because they have a p-value > α(5%), so that all predictor variables are significant and can be used in regression analysis to find their influence on Food Security Index.

Multiple linear regression analysis is carried out with the aim of understanding the relationship pattern of the response variable and a number of predictor variables by evaluating how each predictor variable contributes to changes in the response variable. The following are the results of multiple linear regression estimates:

Table 2. Regression Parameter Estimation

	Estimate	t-value	p-value
Intercept	-41.3413	-0.997	0.325463
X_1	1.5278	2,543	0.015584
X_2	-0.3012	-1,766	0.086181
X_3	-0.9698	-3,926	0.000387
X_4	-1.4181	-1,578	0.123576
X_5	2.9068	2,615	0.013060

In Table 2, the results obtained at a significance level (α) of 0.5 (5%), there are 3 variables that influence the Food Security Index in Papua Province and West Papua Province. The significant variables consist of the Life Expectancy at Birth (X_1), Percentage of Poor Population (X_3), Years of Schooling (X5), because they have a p-value < α(5%). The regression model obtained is formed as follows:

$$\hat{Y} = -25.7868 + 1.3374_{X_1} - 0.3540_{X_2} - 0.9953_{X_3} - 1.2248_{X_4} + 2.7898_{X_5} \quad (10)$$

Next, simultaneous tests and partial tests were carried out on the regression parameters. Simultaneous test results were obtained with a p-value of 3.545e-12 < α(0.05), which means there is enough evidence to say that there must be at least one predictor variable that significantly influences the response variable in the model. Meanwhile, with the partial test it was found that the partial test showed that the Life Expectancy at Birth (X_1), Percentage of Poor Population (X_3), Years of Schooling (X_5) significantly influenced the Food Security Index, because they had a p-value < α (0.05).

Classical Assumption Testing

If the parameter testing is significant and the residual assumptions are met, then the regression model can be said to be good. These assumptions include the following:

1. Normality test

The normality test aims to check that the residuals from a normal distribution have been fulfilled using the Shapiro-Wilk test. It was found that the results of the normality test using the Shapiro-Wilk test, obtained a p-value of 0.3467 > α(0.05), which means that the data is relatively the same as the average, so the residuals are normally distributed.

2. Heteroscedasticity Test

Heteroscedasticity detection was carried out using the Breusch-Pagan test . It was found that from the results of the heteroscedasticity test using the Breusch-Pagan test, the p-value was obtained 0.639 > α(0.05) , which means that the error variance is homoscedasticity.

3. Multicollinearity Test

The multicollinearity test was carried out to determine whether there was a significant relationship between the predictor variables in the regression model. Regression analysis can be carried out if there are no cases of multicollinearity. The existence of multicollinearity can be determined by checking the Variance Inflation Factor (VIF) value.

Table 3. VIF value

Variable	**VIF**
X_1	1.386125
X_2	1.691362
X_3	2.648298
X_4	2.667831
X_5	4.868858

It was found that the VIF value of all predictor variables was <10, which means that in the regression model there were no symptoms of multicollinearity.

4. Autocorrelation Test

The autocorrelation test is used to evaluate the independence of one residual from another. This test uses the Durbin-Watson test. The results of the autocorrelation test using the Durbin-Watson test show a p-value of 0.01374 < α(0.05), which means there is autocorrelation between the residuals which indicates a spatial effect on the response variable.

Spatial Weighting Matrix

There are indications that there is a spatial effect on the response variable, therefore it is necessary to carry out a Moran's I test with the first step being to determine the spatial weighting matrix. This spatial weighting matrix is a positive symmetric matrix with size *n x n*, where *n* is the number of locations observed, and this matrix describes the proximity or relationship between these regions. Information about proximity between regions can be obtained from two relationship conditions, namely neighborhood and distance. The Queen Contiguity Matrix will be used in this research, namely a weighting matrix that takes into account the sides and angles that intersect each other between the observation areas. The spatial weighting matrix is described as follows:

$$W = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \tag{11}$$

The W matrix functions to show the proximity between various observation locations. The elements in the matrix, called W ij, where i represents the row and j represents the column in the matrix, indicate whether two regions are adjacent or not. In this context, the element W ij has the value 1 if region i is close to the observation location j. On the other hand, if region i is not close to the observation location j , then the value of W ij is 0. The standardization of the matrix results used in the model is as follows:

$$W_{ij(std)} = \frac{W_{ij}}{\sum_{i=1}^{n} W_{ij}} \tag{12}$$

Where:

$W_{ij(std)}$ = Standardized weighting matrix elements
W_{ij} = Elements of the weighting matrix

Testing Aspects of Spatial Data

Spatial dependency means that the value of a variable in one region is related to the value in the surrounding region. Moran's I statistic is used to detect this pattern. The Moran Index test results obtained with the Queen Contiguity weighting matrix produced a p-value of 0.01097379 < α(0.05), which means there is spatial dependency between regions. Next, the Lagrange Multiplier (LM) test was carried out to determine the correct model. The following are the test results:

Table 4. Lagrange Multiplier Test

	Value	**p-value**
Lagrange Multiplier (Lag)	1.902430	0.01661
Lagrange Multiplier (Error)	5.736823	0.16781
Lagrange Multiplier SARMA	5.800666	0.05500

The LM test results show that only the Lagrange Multiplier (Error) has a p-value of 0.0166 < α(0.05), which means that there is a spatial dependency of error on the response . Therefore, the appropriate model for Food Security Index in Papua Province and West Papua Province is the Spatial Error Model (SEM).

***Spatial Error* (SEM)**

In the Lagrange Multiplier test results, it was found that there was spatial autocorrelation in the response variable errors. This confirms that analysis using the Spatial Error Model (SEM) is the right step. Parameter estimation in this model was carried out using the Wald test . The following SEM model was obtained:

Table 5. SEM Model Testing Estimation Results

	Estimate	**Standard Error**	**z-value**	**p-value**

λ	0.50832	0.1472	3.4532	0.0082581
Intercept	-80.92208	37.62819	-2.1506	0.031510
X_1	2.22200	0.56445	3.9366	8.266e-05
X_2	-0.37523	0.13926	-2.6944	0.007052
X_3	-0.98015	0.22075	-4.4401	8.991e-06
X_4	-1.34116	0.65885	-2.0356	0.041791
X_5	2.23025	0.97869	2.2788	0.022678

Based on Table 4, the SEM model is produced as follows:

$$\hat{y}_i = 0.50832 \sum_{j=1,i\neq j}^{41} w_{ij}y_j - 80.92208 + 2.22200_{X_1} - 0.37523_{X_2} - 0.98015_{X_3} - 1.34116_{X_4} + 2.23025_{X_5} \qquad (13)$$

Model Interpretation

The results obtained were that the Life Expectancy at Birth (X_1), Percentage of Stunting Toddlers (X_2), Percentage of Poor Population (X_3), Open Unemployment Rate (X_4), and Average Years of Schooling (X_5) significantly influenced the Regency/City Food Security Index in Papua and West Papua Provinces. Furthermore, the lambda coefficient (λ) obtained has a p-value of $0.0082581 < \alpha$ (0.05), which means that the lambda value is significant, indicating that there is an error influence in a Regency/City in Papua and West Papua Provinces which will increase the Food Security Index by 3.4532, which significantly influences the Food Security Index of surrounding districts/cities.

It is known that the Regencies/Cities are correlated with each other as indicated by the lambda coefficient (λ) value of 0.50832, which means the spatial interaction between 41 regencies/cities in Papua and West Papua Provinces which have a regional intersection of 0.50832. If the Life Expectancy at Birth (X_1) variable increases by 1 year, then the Food Security Index will also increase by 2.22200%. If the Percentage of Stunting Toddlers (X_2) variable increases by 1%, the Food Security Index will decrease by 0.37523%. If the Percentage of Poor Population (X_3) variable increases by 1%, then the Food Security Index will decrease by 0.98015%. If the Open Unemployment Rate (X_4) variable increases by 1%, then the Food Security Index will decrease by 1.34116%. If the Average Years of Schooling (X_5) variable increases by 1 year, then the Food Security Index will also increase by 2.23025%. The variables that influence the Food Security Index contribute *an R-Square* of 0.84982 which can be interpreted that the Life Expectancy at Birth (X_1), Percentage of Stunting Toddlers (X_2), Percentage of Poor Population (X_3), Open Unemployment Rate (X_4), and Average Years of Schooling (X_5) variables are able to explain the Food Security Index variable in Papua and West Papua Provinces in 2022 by 84.98% while the other 16.17% is explained by other variables not included in the model.

CONCLUSION

It was concluded that based on the best spatial regression analysis, namely the Spatial Error Model (SEM) with Lagrange Multiplier (LM) testing, the Life Expectancy at Birth (X_1) variable and the Average Years of Schooling (X_5) variable had a positive effect on the Food Security Index in Papua Province and West Papua Province. Meanwhile, Percentage of Stunting Toddlers (X_2) variable, Percentage of Poor Population (X_3) variable, Open Unemployment Rate (X_4) variable have a negative effect on the Food Security Index in Papua and Papua Provinces. These factors have a significant real influence at the significance level of 0.05 (5%).

REFERENCES

[1] Y. Purwaningsih, "Ketahanan Pangan: Situations, Problems, Policies and Community Empowerment," *J. Ekon. Pembang.*, vol. 9, no. 1, pp. 1–27, 2008.

[2] V. V Rumawas, H. Nayoan, and N. Kumayas, "Peran Pemerintah Dalam Mewujudkan Ketahanan Pangan di Kabupaten Minahasa Selatan (Studi Dinas Ketahanan Pangan Minahasa Selatan)," *Governance*, vol. 1, no. 1, pp. 1–8, 2021.

[3] R. Chaireni, D. Agustanto, R. A. Wahyu, and P. Nainggolan, "Ketahanan Pangan Berkelanjutan," *J. Kependud. dan Pembang. Lingkung.*, vol. 1, no. 2, pp. 70–79, 2020, [Online]. Available: jkpl.ppj.unp.ac.id/index.php/JKPL/article/view/13

[4] R. Khairati and R. Syahni, "Respons Permintaan Pangan Terhadap Pertambahan Penduduk Di Sumatera Barat," *J. Pembang. Nagari*, vol. 1, no. 2, p. 19, 2016, doi: 10.30559/jpn.v1i2.5.

[5] J. Cai, E. Ma, J. Lin, L. Liao, and Y. Han, "Exploring global food security pattern from the perspective of spatio-temporal evolution," *J. Geogr. Sci.*, vol. 30, no. 2, pp. 179–196, 2020, doi: 10.1007/s11442-020-1722-y.

[6] A. Dedy Sutrisno, "Kebijakan Sistem Ketahanan Pangan Daerah," *Kebijak. J. Ilmu Adm.*, vol. 13, no. Vol. 13 No. 1, pp. 28–42, 2022, doi: 10.23969/kebijakan.v13i1.4862.

[7] A. Mahmud and E. Pasaribu, "Permodelan Spasial pada Analisis Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka Provinsi Bangka Belitung Tahun 2018," *Eng. Math. Comput. Sci. J.*, vol. 3, no. 2, pp. 47–58, 2021, doi: 10.21512/emacsjournal.v3i2.7034.

[8] A. Safitri, Baharuddin, Agusrawati, B. Abapihi, Ruslan, and G. N. A. Wibawa, "Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Ketahanan Pangan di Sulawesi Tenggara dengan Spatial Error Model (SEM)," *Semin. Nas. Stat. Aktuaria I*, 2022.

[9] I. Y. Safitri, M. A. Tiro, and Ruliana, "Spatial Regression Analysis to See Factors Affecting Food Security at District Level in South Sulawesi Province," *ARRUS J. Math. Appl. Sci.*, vol. 2, no. 2, pp. 60–72, 2022, doi: 10.35877/mathscience740.

[10] P. R. Arum, Y. Fitri, and P. R. Arum, "Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Berdasarkan Kabupaten/Kota Di Jawa Tengah," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 1, pp. 424–433, 2023, doi: 10.36456/jstat.vol16.no1.a7079.

[11] P. Rismawati Arum, R. Putra Gautama, I. Fitriani, and F. Naza Nurvahyani, "Identifying Factors that Influence Life Expectancy in Central Java Using Spatial Regression Models," *J Stat.*, vol. 16, no. 2, pp. 605–613, 2023.

Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest

Ariska Fitriyana Ningrum (1), Ihsan Fathoni Amri (2)

[1,2]Program Studi Sains Data, Universitas Muhammadiyah Semarang

Jln. Kedungmundu No 18, Kedungmundu, Kec. Tembalang, Kota Semarang

e-mail: ariskafitriyana@unimus.ac.id(1), ihsanfathoni@unimus.ac.id(2)

ABSTRAK

Masalah stunting penting untuk diselesaikan, karena berpotensi mengganggu potensi sumber daya manusia dan berkaitan dengan tingkat kesehatan, bahkan kematian anak. Pemerintah Indonesia menargetkan angka stunting turun menjadi 14 persen pada tahun 2024 melalui program percepatan penurunan stunting sebagai upaya meningkatkan status gizi masyarakat dan juga menurunkan prevalensi stunting atau balita pendek. Memahami sentimen publik terhadap inisiatif stunting sangat penting bagi para pembuat kebijakan dan pemangku kepentingan untuk merancang intervensi yang efektif dan mengalokasikan sumber daya secara efisien. Pada penelitian ini dilakukan klasifikasi pada sentiment positif dan negatif menggunakan algoritma random forest. Data yang digunakan adalah data komentar pada salah satu laman media sosial yaitu twitter mengenai sentiment masyarakat terhadap penanganan kasus stunting di Indonesia. Tahapan pertama pada penelitian ini setelah didapatkan sebuah data yaitu dilakukan prepocessing data. Tahapan preprocessing data dalam analisis sentimen berguna untuk membersihkan dan menormalkan teks, menghilangkan kata-kata tidak relevan, serta mempersiapkan data agar algoritma dapat menganalisis sentimen dengan lebih akurat dan efisien. Selanjutnya hasil data yang sudah di prepocessing diberikan label 0 untuk positif dan 1 untuk label negatif. Klasifikasi terhadap sentiment positif dan negatif ini dilakukan menggunakan random forest dan menghasilkan nilai akurasi sebesar 97,5%. Model ini sudah baik, namun kami menyarankan untuk mencoba algoritma lain dalam penelitian selanjutnya.

Kata kunci: Analisis Sentiment, Random Forest, Stunting

ABSTRACT

The problem of stunting is important to solve, as it has the potential to disrupt human resource potential and is linked to health outcomes and even child mortality. The Indonesian government targets the stunting rate to drop to 14 percent by 2024 through an accelerated stunting reduction program as an effort to improve the nutritional status of the community and also reduce the prevalence of stunting or short toddlers. Understanding public sentiment towards stunting initiatives is essential for policy makers and stakeholders to design effective interventions and allocate resources efficiently. In this research, classification of positive and negative sentiment is carried out using the random forest algorithm. The data used is comment data on one of the social media pages, namely Twitter, regarding public sentiment towards handling stunting cases in Indonesia. The first stage in this research after obtaining a data is data preprocessing. The data preprocessing stage in sentiment analysis is useful for cleaning and normalizing text, removing irrelevant words, and preparing data so that algorithms can analyze sentiment more accurately and efficiently. Furthermore, the results of the preprocessed data are labeled 0 for positive and 1 for negative labels. The classification of positive and negative sentiment was done using random forest and resulted in an accuracy value of 97.5%. This model is good, but we suggest trying other algorithms in future research.

Keywords: *Sentiment Analyst, Random Forest, Stunting*

INTRODUCTION

Stunting is a condition in children who experience growth disorders, so that the height and weight of children are not normal due to problems of nutritional deficiencies for a long time [1]. The problem of stunting in Indonesia is still quite large in the health sector today. According to the World Health Organization (WHO), as many as 22% or around 149.2 million children in the world under the age of five were recorded as stunted in 2020. Indonesia's position on the prevalence of stunting in the world is ranked 115 out of 151 countries. Meanwhile, in Southeast Asia, Indonesia is ranked second at 31.8% after Timor Leste at 48.8%. The third is Laos at 30.2%, the fourth is Cambodia at 29.9%, and the fifth is the Philippines at 29.9% [2] Stunting is caused by health problems, environmental factors and health services received by children. Genetic factors do not significantly affect stunting. Lack of nutrition in the fetus is the biggest cause of stunting in children. The first 1000 days of a child's life (1000 HPK) is the starting point for making important conclusions on long-term growth Thus, ineffective parenting and diet can increase the chance of stunting. Mental disorders and hypertension in mothers also affect the behavior and practices of nutrition in children. Limited access to health and sanitation services exacerbates the stunting conditions that occur in Indonesia such as lack of clean water, unclean latrines, and so on [3] Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers [4].

Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers[5] Understanding public sentiment towards stunting initiatives is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. Several previous studies have analyzed stunting predictions using the random forest algorithm which resulted in a classification accuracy value of 90.7%. [6]. Another study on social media analysis with the topic of stunting in Indonesia was conducted where the results showed that negative sentiment dominated by 60.6%, positive sentiment by 31.5%, and neutral by 7.9% [7]. In addition, this study shows that 'children', 'decline', 'numbers', 'prevention', and 'nutrition' are words that often appear in stunting [7]. Another study comparing SVM and random forest algorithms for the classification of stunting disease. the results show that the random forest algorithm provides higher accuracy of 88.2% compared to SVM of 65.6% [8].

The background of this study is based on the need to analyze public sentiment regarding the handling of stunting in Indonesia, which is a significant public health problem. The data used are the results of positive and negative reviews from the public on social media such as twitter regarding the handling of stunting cases in Indonesia. Some previous studies have also analyzed using comment data on twitter such as research on Sentiment Analysis of Twitter Netizens on the News of VAT on Basic Food and Education Services with Social Network Analysis and Naive

Bayes Classifier Approaches with data obtained as many as 4090 tweets [9]. While other studies have also conducted sentiment analysis of twitter users regarding online transportation service users [10].

The method used in this research is Random Forest. Random Forest was chosen because of its superior ability to handle complex and varied data, and provide accurate results in classification and prediction. This method is suitable for sentiment analysis because it is able to overcome overfitting, works well on large and irregular datasets, and provides a good interpretation of the features that affect the results. Random Forest is one of the state-of-the-art methods in machine learning that consists of a number of independently trained decision trees and the results are combined to produce more accurate and stable predictions. In the context of sentiment analysis, Random Forest can handle variations in language expression and identify relevant patterns from unstructured text data. In addition, Random Forest's ability to handle data with many features is very useful in sentiment analysis involving various emotional aspects and public opinions related to stunting treatment [11].

The state of the art of the Random Forest method shows that this technique has been successfully applied in various domains, including text and sentiment analysis, with satisfactory results. Previous studies have shown that Random Forest often outperforms other methods such as logistic regression and Support Vector Machines (SVM) in terms of accuracy and robustness to noise in the data. This makes Random Forest an appropriate choice for this research in an effort to understand and measure public sentiment towards stunting response efforts in Indonesia.

METHOD

In this study, a sentiment analysis about stunting in Indonesia was conducted. The remaining data will be analyzed using the random forest algorithm to check the accuracy of the sentiment results. The following are the stages of sentiment analysis research on stunting using the random forest algorithm

1. Data Collection

The data used in this study is the result of crawling Twitter data related to positive and negative responses to stunting conditions in Indonesia. The data collection process was carried out over a one-month period, starting from January 1, 2024 to January 31, 2024, to get a current picture of public sentiment on the issue of stunting. In the crawling process, certain filters and keywords related to infant stunting were used. The keywords used included "stunting", "child growth", and "child nutrition", as well as other keyword variations related to stunting and child health in Indonesia. In applying a series of filters, including language settings (Bahasa Indonesia) and geographical location (Indonesia) were used to ensure that the data captured was specifically related to responses to stunting conditions in Indonesia. This resulted in a total of 4601 comments showing both positive and negative opinions. Labeling the data into positive and negative categories was done through a manual process by the researcher, where each opinion was classified based on the sentiment expressed towards stunting conditions. This assessment is based on the context within each opinion, with the aim of gaining an accurate understanding of public sentiment.

2. Prepocessing Data

The first stage of the system is preprocessing. This stage involves several processes including Case Folding, Tokenization, Normalization, and Stemming. Case Folding is a task of splitting review text into smaller units called tokens or terms [12]. For infant stunting cases, what is done before and after case folding is, for example, "Breastfeeding mothers must have good nutrition" becomes "breastfeeding mothers must have good nutrition". Next is Tokenizing, in this process the separation is carried out on each word that makes up a document. In general, each word is identified or separated from other words by space characters, so the tokenizing process relies on space characters in the document to perform word separation [13] In sentiment analysis of stunting cases, what is done is to present the number of tokens generated from a review or comment. For example, from the sentence "breastfeeding mothers should have good nutrition", the tokens generated are "mother", "breastfeeding", "should", "nutrition", "which", "good". Normalization (Stopword Removal) process Removes special characters, numbers, and stopwords (common words) from each token. In the case of sentiment analysis, it shows a list of stopwords used and examples of text before and after stopword removal. For example, from "mother", "breastfeeding", "should", "nutrition", "which", "good", after removing the stopwords "should", "which", then "mother", "breastfeeding", "nutrition", "good" remains. This research also uses Stemming techniques which aim to find the base word, by removing all affixes that are fused to the word.[14] In Indonesian, this usually involves the removal of prefixes, suffixes or infixes. As an example of words before and after the stemming process, for example, "menyusui" can be reduced to "susu".

3. Sentiment Analyst Using Random Forest

The last stage is sentiment classification. Each review will be classified into positive or negative category. In this study, we employ random forest for the classification task. Random forest algorithm is a supervised classification algorithm. It is an ensemble learning technique based on decision tree algorithm [15]. Random Forest Algorithm is the advancement of Classification and Regression Tree (CART) method with the implementation of bootstrap aggregating (bagging) and random feature selection. Procedure of random forest algorithm on the data of n observations and p predictor [16]

a. Random samples of size n are drawn with the possibility of obtaining the same data (with replacement). This phase is called bootstrap.
b. Using the bootstrap samples, the tree is grown until the maximum size is reached, which is done without pruning. At each node, the random feature selection is used to determine the split, which m number of variables randomly sampled as candidates at each split must be $m \ll p$, at which point, the best node will be chosen based on m number of variables available for splitting [17]
c. Repeat stage 1 and 2 for k times to generate a forest that consists of k trees.
Breiman and Cutler suggests to observe the error OOB when

$$m = \left(\frac{1}{2}\left|\sqrt{p}\right|, \left|\sqrt{p}\right|, 2\left|\sqrt{p}\right|\right) \qquad (1)$$

where *p* is the total variable and the number of *k* is small, then m with the smallest error OOB will be chosen.

In order to determine the split used as root node/node, Gini Index is used in Random Forest method. The formula of Gini index can be described as following:

$$Gini = 1 - \sum_{i=1}^{k} p_i^2 \quad (2)$$

where:

p_i : probability of an attribute being classified to class i.

k : total number of attributes being classified to a particular class.

The number of k suggested to apply in bagging is k = 50 which will provide satisfied results for classification [18]

RESULT AND DISCUSSION

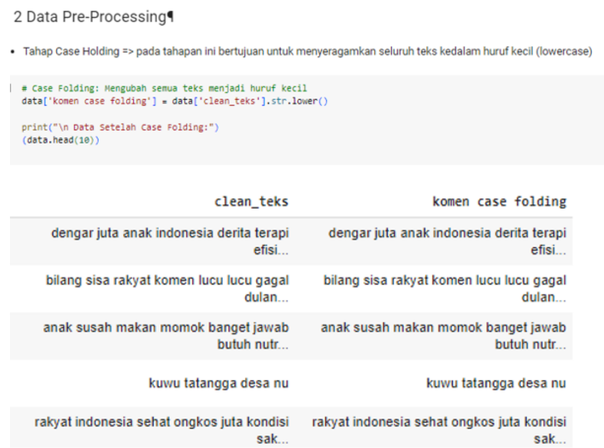
In this study, the process of labeling public opinion data on stunting in Indonesia, obtained from social media through crawling, was carried out using a manual approach. Initially, a portion of the data was manually labeled by the research team to form the training dataset. Each opinion was classified into two categories: 0 for negative sentiment and 1 for positive sentiment. This manual labeling process is important to ensure that the data used accurately reflects public sentiment towards stunting and to generate a valid classification model.

```
Label                                     komen bersih
    0  dengar juta anak indonesia derita terapi efisi...
    0  bilang sisa rakyat komen lucu lucu gagal dulan...
    0  anak susah makan momok banget butuh nutrisi an...
    0                             kuwu tatangga desa nu
    0  rakyat indonesia sehat ongkos juta kondisi sak...
  ...                                              ...
    0  hamil minum puskesmas kasih gratis takut anemi...
    0  iya cupu cupu paham google search baca jurnal ...
    0  banteng jawa barat komitmen entas pdiperjuanga...
    0                                  bahan becandaan
    0  tingkat sabar setip tissue hadap anak emosi mu...
```

Figure 1. Review data that has been labelled

After obtaining a dataset of community reviews on stunting handling, data preprocessing is carried out with various stages such as *case folding, tokenizing, normalization, and stemming*. Data preprocessing is important because it helps to optimally clean, prepare, and organize data before analysis, thereby improving the accuracy, reliability, and interpretation of results from the model or analysis technique used.

The first stage, namely preprocessing, aims to homogenize all text into lowercase letters. The results of the case folding process are as follows.



2 Data Pre-Processing¶

- Tahap Case Holding => pada tahapan ini bertujuan untuk menyeragamkan seluruh teks kedalam huruf kecil (lowercase)

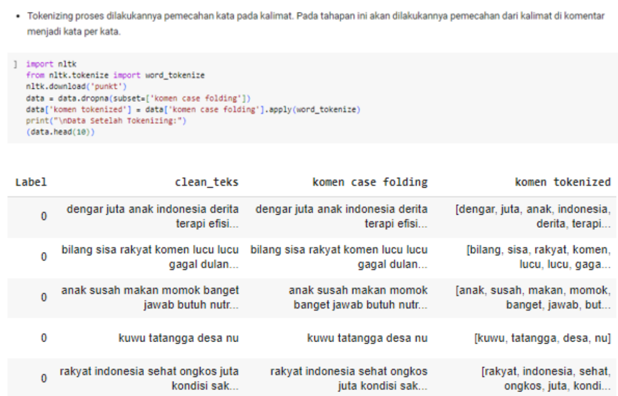
```
# Case Folding: Mengubah semua teks menjadi huruf kecil
data['komen case folding'] = data['clean_teks'].str.lower()

print("\n Data Setelah Case Folding:")
(data.head(10))
```

clean_teks	komen case folding
dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...
bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...
anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...
kuwu tatangga desa nu	kuwu tatangga desa nu
rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...

Figure 2. Case Holding Stages Result

Next, the tokenizing process is carried out. At this stage, the sentence in the comment will be broken down into words. The results of the tokenizing process are presented as in the following figure.



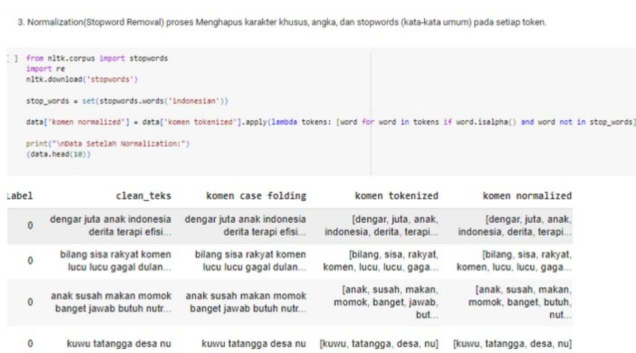
- Tokenizing proses dilakukannya pemecahan kata pada kalimat. Pada tahapan ini akan dilakukannya pemecahan dari kalimat di komentar menjadi kata per kata.

```
import nltk
from nltk.tokenize import word_tokenize
nltk.download('punkt')
data = data.dropna(subset=['komen case folding'])
data['komen tokenized'] = data['komen case folding'].apply(word_tokenize)
print("\nData Setelah Tokenizing:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 3. Tokenizing Stages Result

The next preprocessing is to perform normalization to change the values of a dataset so that they have a uniform scale. The main purpose is to ensure that variables with different value ranges have equal influence when used in the analysis. The results of data normalization are as follows.



3. Normalization(Stopword Removal) proses Menghapus karakter khusus, angka, dan stopwords (kata-kata umum) pada setiap token.

```
from nltk.corpus import stopwords
import re
nltk.download('stopwords')

stop_words = set(stopwords.words('indonesian'))

data['komen normalized'] = data['komen tokenized'].apply(lambda tokens: [word for word in tokens if word.isalpha() and word not in stop_words]

print("\nData Setelah Normalization:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]

Figure 4. Normalization Stages Result

The last step in data preprocessing is stemming. This aims to find the base word, by removing all affixes that are fused to the word. The results of stemming performed for sentiment analysis are as follows.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
data['komen stemming'] = data['komen normalized'].apply(lambda tokens: [stemmer.stem(word) for word in tokens])

print("\nData Setelah Stemming:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized	komen stemming
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 5 Stemming Performed for sentiment analysis

Furthermore, the results of data preprocessing that have been carried out can be seen visually regarding positive and negative opinions. Visualization aims to display the words that appear most or most often in a sentiment. Wordcloud this time describes each sentiment, the more often a word is used when giving a review, the larger the size of the word displayed on the wordcloud visualization. The following figure shows the visualization results for positive and negative sentiments

Figure 6. Visualization Positive Sentiment

Based on the figure above, it can be seen that in the positive sentiment there are several words that stand out such as, "balanced nutrition,", "helps reduce", "Welfare Growth" and several other words which indicate that the public's response to handling stunting in Indonesia has helped reduce stunting rates, provide balanced nutrition to children and can foster community welfare.

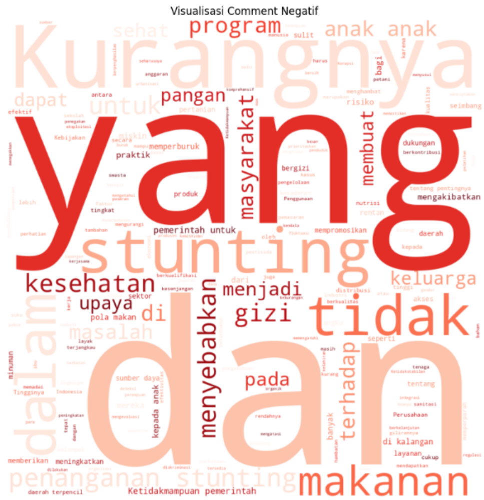


Figure 7. Visualization Negative Sentiment

Based on the figure above, it can be seen that in the negative sentiment regarding the handling of stunting in Indonesia, there are prominent words such as "lack of", "government inability", "gap" and several other words that indicate that the handling of stunting is still lacking, there is a social gap and the lack of government in handling cases of stunting in the society. Next, Classification on positive and negative comment data that has been done feature extraction using TF-IDF function is to convert text into a numerical vector representation based on the frequency of occurrence of words in it, taking into account the TF-IDF weight of each word. The classification metric evaluation calculations that we use are confution matrix, F1 measure and accuracy. The following are the results of the random forest performance evaluation

Table 1 Confution Matrix

		Actual	
		Negative	Positive
Prediction	Negative	893	0
	Positive	23	5

Based on table 1, which is related to the confusion matrix of prediction results, TP = 5, FP = 23, TN = 893, FN = 0, Total data = 258. The calculation results are obtained from the following calculations:

Table 2. Precision, recall, and f1-score

	Precision	Recall	f1-score
Negative	0.97	1.00	0.99
Positive	1.00	0.18	0.30

The performance generated by the random forest algorithm provides considerable accuracy, which is 97.50%, indicating that this model can classify data has a very good indication, and produces precision on Label 0 (negative comments) of 97% and recall of 100%, the results obtained are very high, and F1 score of 99%, indicating a high balance of precision and Recall. Meanwhile for precision on Label 1 (Positive comments) of 100%, and recall of 18% and the result for f1-score is 30%.

CONCLUSION

This study shows that public sentiment towards the handling of stunting cases in Indonesia can be divided into positive and negative based on the analysis of 4601 comments from Twitter social media. The results show that positive responses include the view that the handling of stunting has succeeded in reducing stunting rates, providing balanced nutrition to children, and potentially improving the general welfare of society. On the other hand, negative responses include dissatisfaction with the effectiveness of stunting handling, the existence of unresolved social inequalities, and the lack of effort from the government in handling stunting cases in the community.

This research continued with the classification of comment data based on sentiment using TF-IDF feature extraction. This method is important because it converts text into a numerical vector representation, where the TF-IDF weight of each word gives an idea of the importance of the word in determining positive or negative sentiment. Through this classification, it is possible to identify and categorize the sentiments present in the text data, enabling a deeper understanding of the public's views on stunting in Indonesia.

We performed sentiment analysis using random forest algorithm and achieved about 97.5% accuracy. I would recommend you to try using some other machine learning algorithms such as LSTM or KNN and see if you can get better results.

ACKNOWLEDGMENTS

This work was supported by Data Science Program, Faculty of Science and Agricultural Teknologi, Univercity of Muhammadiyah Semarang.

REFERENCE

[1] H. Rahman, M. Rahmah, and Nur Saribulan, "UPAYA PENANGANAN STUNTING DI INDONESIA," *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, vol. VIII, no. 01, pp. 44–59, 2023.

[2] W. H. S. W. H. Organization, " Monitoring Health for the SDGs, World Health Organization," 2022.

[3] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[4] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, "Aplikasi Naive Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[5] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, “Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, Jan. 2024, doi: 10.57152/malcom.v4i1.1074.

[6] A. A. Reza1 and M. S. Rohman2, “JITE (Journal of Informatics and Telecommunication Engineering) Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization,” *JITE*, vol. 7, no. 2, 2024, doi: 10.31289/jite.v7i2.10628.

[7] A. Muhaimin *et al.*, “Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia,” *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[8] M. Banurea, D. Betaria Hutagaol, and O. Sihombing, “KLASIFIKASI PENYAKIT STUNTING DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST,” *Jurnal TEKINKOM*, vol. 6, no. 2, 2023, doi: 10.37600/tekinkom.v6i2.927.

[9] J. A. Nursiyono and C. Chotimah, “Analisis Sentimen Netizen Twitter terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan Social Network Analysis dan Naive Bayes Classifier,” *Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 14, pp. 52–58, 2021.

[10] “Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications With Noise (DBSCAN) dan K-Means,” *Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 15, pp. 184–194, 2022.

[11] Stephenie, B. Warsito, and A. Prahutama, “Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method,” in *E3S Web of Conferences*, EDP Sciences, Nov. 2020. doi: 10.1051/e3sconf/202020216006.

[12] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, “Indonesian news classification using naïve bayes and two-phase feature selection model,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 3, pp. 610–615, Dec. 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.

[13] Bahrawi, “Sentiment Analysis Using Random Forest Algorithm Online Social Media Based,” Makassar, Dec. 2019.

[14] G. Khanvilkar and D. Vora, “Sentiment analysis for product recommendation using random forest,” *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 87–89, 2018, doi: 10.14419/ijet.v7i3.3.14492.

[15] L. Breiman, “Random Forests,” 2001.

[16] Leo Breiman and A. Cutler, “ Manual on Setting Up, Using, and Understanding Random Forest 4.0,” 2003.

[17] M. A. Fauzi, “Random forest approach fo sentiment analysis in Indonesian language,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.

[18] I. Afdhal *et al.*, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.

Accelerating SIREKAP Digital Transformation in the 2020 Natuna Regency Election

Bagus Aji Kuncoro [(1)], Supratiwi [(2)], Fitriyah [(3)]

[1,2,3]Faculty Of Social Science and Political Science, Universitas Diponegoro,

Jl. Prof. Soedarto, SH., Tembalang, Semarang 50275
e-mail: bagoesadjikoentjoro@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk menganalisis implementasi transformasi digital dalam penggunaan Sirekap pada Pilkada 2020 Kabupaten Natuna. Jenis penelitian ini adalah kualitatif, dengan menggunakan strategi penelitian studi kasus tunggal, yang melibatkan satu individu dalam satu perusahaan atau kantor, yaitu kantor KPU Kabupaten Natuna. Desain penelitian yang digunakan adalah penelitian kualitatif. Percepatan transformasi digital sangat penting dalam proses operasional akuisisi dan sebagai alat pendukung sekaligus meminimalisir risiko dari tahapan awal pemilihan, hari pemilihan, dan pasca pemilihan. (pasca pemilu). Penerapan Sirekap pada pemilihan kepala daerah Bupati dan Wakil Bupati Kabupaten Natuna tahun 2022 meningkatkan transparansi dan akuntabilitas untuk meningkatkan kepercayaan masyarakat terhadap hasil perhitungan pemilu. Aplikasi Sirekap membuat waktu kerja KPU menjadi lebih efektif dibandingkan dengan perhitungan manual. KPU, dengan hadirnya aplikasi Sirekap, juga membuat informasi yang disebarkan ke masyarakat tidak kalah cepat dengan lembaga survei karena daerah dapat memantau data yang masuk di tempat pemungutan suara secara langsung. Aplikasi Sirekap juga memiliki tingkat ketelitian yang tinggi dan meminimalisir tingkat kesalahan perhitungan suara

Kata kunci*:* Sirekap, Transformasi Digital, Pemilihan Umum

ABSTRACT

The aim of this research is to analyze the implementation of digital transformation in the use of Sirekap in Pilkada 2020 district of Natuna. This type of research is qualitative, using a single case study research strategy, which involves an individual in one company or office, namely the Natuna Regency KPU office. The research design used is qualitative research. The acceleration of digital transformation is essential in the operational process of acquisition and as a support tool while minimizing risks from the early stages of the elections, the election day, and the post-election period. (post-election). The over-implementation of the Sirekap in the election of the head of the Bupati district and the Deputy Bupati District of Natuna district in 2022 increases transparency and accountability to increase public confidence in the results of the election calculations. The Sirekap application makes the working time of the KPU more effective than manual calculations. The KPU, with the presence of the Sireap application, also makes the information disseminated to the public no less rapidly than the survey agency because the region can monitor the data entered in the place of direct voting. The Sirekap application also has a high level of rigor and minimizes the error rate of voting calculation

Keywords: *Sirekap, Digital Transformation, General Election*

INTRODUCTION

Covid-19 has been declared a world pandemic by WHO. Furthermore, due to the increase in cases and the spread of disease between regions, the government issued Government Regulation Number 21 of 2020 concerning Large-Scale National Restrictions to accelerate the handling of Corona Virus Disease 2019 (COVID-19). These large-scale national restrictions impact all aspects of life, including the economy, business, work, worship, education, political activities, and others [1]. One of the affected political activities is holding general elections, which is a complex and challenging task to keep the polls democratic, with integrity, healthy, and safe [2].

The General Election Commission (KPU), in the emergency of the Covid-19 pandemic, seeks good election governance through accelerating digital transformation into all stages of electoral operations (pre-election, election day, and post-election) [3]. The acceleration of digital transformation in elections uses digital-based technology to facilitate the performance of the KPU, considering that it is currently developing the Industrial Revolution 4.0. Acceleration of digital transformation in elections as a performance support and to minimize risks in the election and post-election processes. The International Institute IDEA notes that there are several advantages of implementing election digitalization, namely faster vote counting and tabulation, more accurate results, efficient handling, improved ballot display, increased comfort for voters, increased voter participation, alignment with community needs, prevention of fraud at polling stations, increased accessibility, multilingual services, cost savings, and minimizing fraud.

The massive acceleration of digital transformation in a pandemic situation aligns with the digital transformation roadmap at the KPU. Some electoral frameworks have designed, developed, and introduced digitalization methods in election governance. Based on the experience of the death and illness of KPPS officers in the 2019 Simultaneous Elections due to fatigue with a heavy workload during vote counting at polling stations, Sirekap election digitalization is prepared for four things, namely preventing overtime and fatigue of election officers, reducing the workload of election administrators, minimizing election stages and processes, and simplifying and accelerating stages [4]. The acceleration of digitisation in the 2020 local elections also minimizes the risk of spreading and expanding the spread of Covid-19 due to election activities [5]. This acceleration of digitization did not go smoothly. The results of Suri & Yuneva's research found that digitalization efforts at all stages of the election have not been carried out thoroughly due to administrative constraints/technical aspects in addition to non-technical factors, namely related to stakeholders in the election [6].

Sirekap is a new product that was used for the first time in the 2020 Simultaneous Regional Elections. PKPU No.19 of 2020 concerning Recapitulation of Vote Counting Results and Determination of Election Results defines Sirekap as an information technology-based application tool as a means of publication of vote counting results and recapitulation of vote counting results as well as a tool in the implementation of recapitulation of election vote counting results. Sirekap replaces the Counting Information System (Situng) used in previous elections, whose function is limited to the publication of election results. Meanwhile, Sirekap is not only a means of publication but also functions as a tool in vote counting and tiered vote recapitulation (PPK-KPU) [7]. The researcher will analyze how the digital transformation of Sirekap (electronic vote recapitulation information system) in the 2020 Pilkada in Natuna Regency, Riau Islands Province. Natuna

Regency and Anambas Regency are two regions in Riau Islands Province, the northernmost islands in the Karimata Strait.

METHOD

This type of research is qualitative, using a single case study research strategy, which involves an individual in one company or office, namely the Natuna Regency KPU office. The research design used is qualitative research. The location of this research is in Natuna Regency, Riau Islands Province. The technique of determining informants in this study uses purposive sampling, which is a data collection technique by the criteria set and has the required criteria according to the sample with the aim of the quality or quality of an object. Data collection techniques using in-depth interviews and documentation.

RESULT

Description of Geographical Situation

The land area of Natuna Regency is 1,978.29 km2. The capital of Natuna Regency is Ranai City. The sub-district with the largest area is the North Bunguran sub-district. This sub-district consists of 8 villages: West Kelarik, Kelarik, North Kelarik, Kelarik Air Mali, Teluk Buton, Belakang Gunung, West Seluan and Gunung Durian. The capital of the North Bunguran sub-district is Kelarik village. Natuna Regency originally consisted of 12 sub-districts. On 10 December 2014, 3 new sub-districts were formed, bringing the total number of sub-districts in the Regency to 15. Of these 15 sub-districts, there are 70 villages and six sub-districts. These 6 villages are Sabang Barat, Sedanau, Ranai, Ranai Darat, Bandarsyah and Serasan. Meanwhile, in 2020, a new sub-district was formed in the Bunguran Timur sub-district, namely the Batu Hitam sub-district.

Natuna Regency consists of islands. According to data obtained from the local government, there are 159 islands. Of all the sub-districts, Serasan has the largest number of islands at 31 (19.5 percent of the total islands). Based on the topography, the area of Natuna Regency is generally hilly and mountainous, but lowlands and slopes are also found on the coast. This is natural because the region is an archipelago surrounded by ocean. Therefore, the height of the sub-districts above sea level (DPL) ranges from 58 to 980 meters. Subi is the sub-district with the lowest elevation, reaching only 58 meters at the highest point.

Meanwhile, the sub-district with the highest elevation is Bunguran Timur Laut. This is because there is a mountain in this sub-district with a height of 980 meters above sea level (DPL). The distance from the sub-district capital to the capital of Natuna Regency is quite far and varied. Because the capital of Natuna Regency is in the Bunguran Timur sub-district, the distance between the two is very close. Meanwhile, the sub-district with the furthest straight distance to the capital of Natuna Regency is Serasan Timur, reaching 200.35 km. This sub-district is the furthest from Bunguran Island and is quite close to West Kalimantan Province.

Description of the Demographic Condition of Voters

Understanding the demographic characteristics of voters is very important to know voter preferences to get as many votes as possible. The number of voters on the final voter list (DPT) determined by the General Election Commission (KPU) in 2020 is 52,896. This reflects the large number of residents with voting rights in 2020. Approximately 65 percent of Natuna residents have voting rights with various inherent socio-demographic characteristics. Millennials (Y) born in 1981-1996 will have the most voters, approximately 70 percent of the total voters. The next generation with the largest voters is Generation X (born 1964-1980), which is 60 percent, and Generation Z (born 1997-2007), which reaches 30 percent. Although Generation Z is now the majority of Natuna's population, not everyone has the right to vote, while elderly voters (age 60 and over) make up 40 percent of the electorate. The number of older voters is slightly below Generation X and has a higher turnout than the younger generation. There are still more young voters than senior voters. In addition to generations and ages, the composition of the population according to education is also important to be taken into account. Based on census data, it is known that potential voters for the 2020 elections are predominantly those with low education. Nearly 80 percent of voters have the highest high school education level, but only 20 percent of candidates have the lowest undergraduate education degree. (S-1 dan D-4). The KPU should formulate a strategy to encourage election participation in areas with a low population, given that various literature concludes that the lower voter education, the lower the tendency to participate.

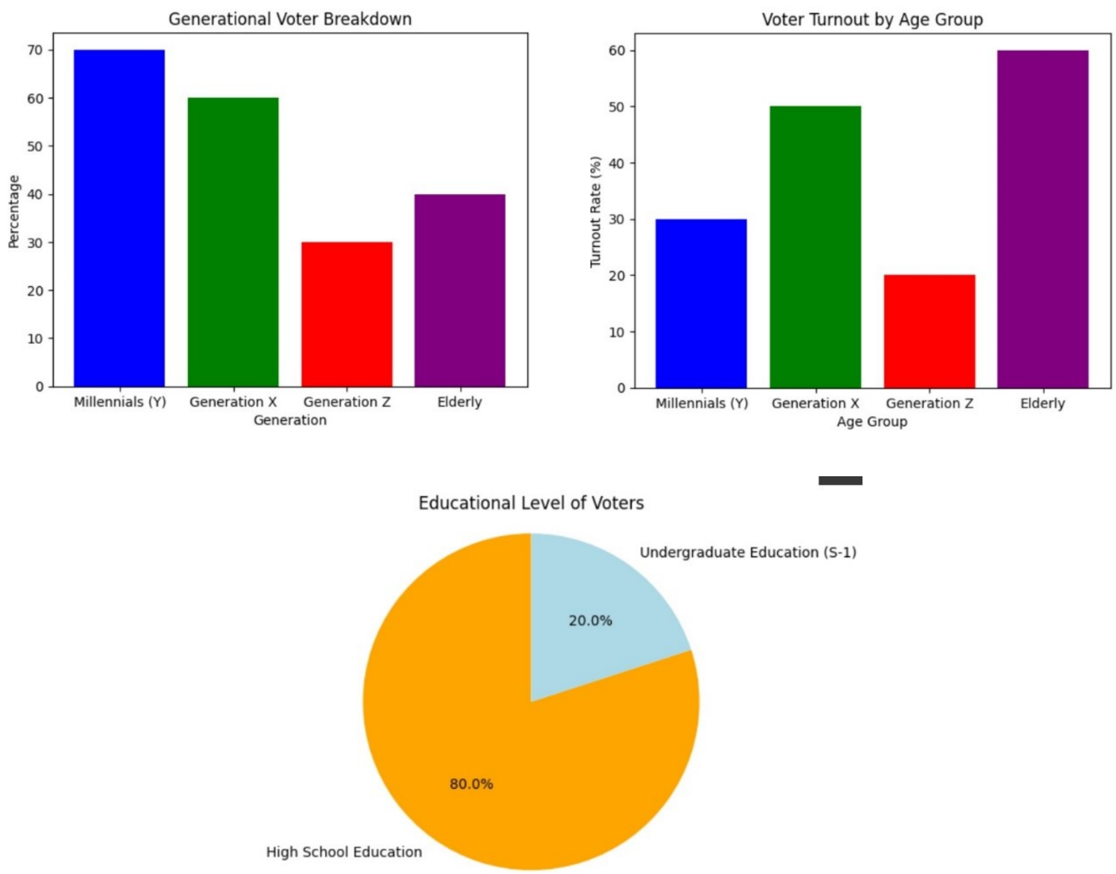


Figure 1. Diagrams of Description of the Demographic Condition of Voters

Political participation dynamics in the district of Natuna

The general election (election) is the only legitimate institution that legitimizes the power of a democratic government. In managing a democratic order of government, there have been some changes in the electoral aspects, from electoral governance to the electoral system to the implementation of electoral law. The electoral scheme is not just a routine political agenda. The government's efforts to establish an effective and consolidated government order depend on the quality of the elections. The electoral political landscape heavily influences the dynamics of voter perception. In every election, the political landscape is generally different and may have something in common, even if it won't be the same.

Table 1. Strategic Objective Performance Measurement

No	Target Performance Indicators	Target	Realisation	Achievements
1	Voter participation rate	75	77,78	103,71

SIREKAP Overview

General Overview The Election of Governors, Chancellors, and Mayors is a Law regulating the mechanisms of manually counting and recapitulating the votes of elections and using electronic voting systems regulated by the Regulations of the KPU. Based on these provisions, KPU has developed a tool to support performance accountability in the implementation of the stages of counting votes and recapitulation of voting results, as well as establishing the results of the Single Elections of the Year 2020.

Sirekap is an application device based on information technology that publishes the Vote Count results and recapitulates the Voice Count results. It is also a tool for implementing the recapitalization of the results. The number of voters participating in the 2020 National Election was 86.3 percent, according to gender: male voters: 23.125 and female votes: 23.057, making the total 46.182, while the number of disabled voters is male voters: 152 and feminine voting: 159 Total: 311. The total number of valid votes in the 2020 election was 44,980 votes. The overall number of invalid votes was 1,202 votes for the 2020 elections. The number of votes used in the National Elections and the 2020 National Election was 46,182.

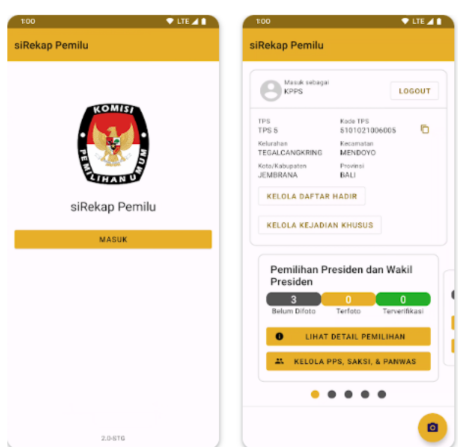


Figure 2. SIREKAP App

General Overview:

The poll on the 2020 Natuna Bupati Election and Deputy Bupati was held on December 9, 2020, from 07.00 PM to 1.00 PM and was conducted safely, peacefully, orderly, and conducive. Submission of the Recapitulation Result Certificate from the KPPS to the PPS on the same day, the KPS has the recapitulation of the voting count to PPS. In the election of the governor and Vice-governing, deputy and deputy deputy, and mayor and mayor, there are some differences in the type of form used. The information system for counting votes has also been updated, which was previously used in the 2019 *Sistem Informasi Perhitungan Suara* (SITUNG) ; at the 2020 Uniform Election, the General Electoral Commission introduced Sirekap. The atmosphere at the polling stations required to implement strict health protocols also differs from the previous elections. Here's what's new in the 2020 simultaneous elections.

Digital Transformation in Sirekap Usage at Pilkada 2020 in Natuna District

In the voting phase in Indonesia, e-voting is aimed not only to minimize the spread of the coronavirus but also to help the organizers' workload accelerate the process of recording votes. A feasibility study is needed before e-voting can be applied to heterogeneous Indonesian societies. The method of e-voting itself is divided into three: e-counting or optical scanning system in which the voice paper is specially made scanned by the optic of the scanner machine; Electronic Direct Recording System (DRE), voters give voting rights to a computer, a touch screen, or an electronic voice panel where the recording of the vote is stored in memory in the TPS and can then be sent to the center online or offline; and Internet Voting, voters can give the right to vote on a computer or touch screen device that is connected to the Internet network. The voice delivered will be directly recorded centrally. This method requires a reliable data communication network and security.

E-voting is permitted as long as it meets the cumulative requirements. Its legal basis is MK Decision No. 147/Law-VII/2009 [8]. Other alternative voting methods can be postal or electronic voting or proxy voting, in which voting rights are delegated to a trusted person, possibly from a lower-risk group. The voting system in the electoral process is inclusive and safe for vulnerable age groups (i.e., over 60) or ethnic minorities that are highly susceptible to disease.

Technical Guidance on Recording and Counting of Voice and Use of Sirekap

Technical guidance activities voting, counting, and recapitulation, as well as the use of Sirekap in Pilkada Year 2020, took place for 6 (six) days from November 19, 20, and 22 for wave one at RM Gerai Beach Natuna and November 23, 24, 25 2020, at Ballroom Natuna Hotel, Ranai, Natuna District. The contestant is the Technical Division of Natuna District. This technical guidance activity aims to submit material on the selection ceremony, calculation, and use of Sirekap in the election of Bupati and Deputy Bupati district of Natuna in 2020.

The event was opened by the Chief of the KPU of Natuna District and Mr. Risno, a Natuna County KPU Technical Maintenance Division member, who conducted the material exhibition. The next session was conducted live practice, where we filled out Form C-Results-KWK and Simulation of Use of Application Sirekap.

Simulation of Summing, Voice Counting, and Use of Sirekap

Simulation of polling, counting of votes, and use of the Sirekap application to the Prefectural Election Commission of Natuna District took place for one day on November 21, 2020, at the Village Kaki Ball Course of Harapan Jaya, Bunguran Tengah, Natuna County. The Simulation Activity is aimed at strengthening the Section Selection Committee's capacity to provide technical guidance in selecting, counting votes, and using Sirekap for KPPS. The Activity was opened by the Chief of the KPU of Natuna District and continued with the simulation of the mobile Sire Kap application.

DISCUSSION

Among other things, the surplus of elections increases public confidence in the transparency of the elections, makes the election more efficient, and minimizes the error rate in counting election results. Transparency in electoral governance is the openness of electoral authorities' rules and procedures, results, and processes. It is considered to build public confidence, enhance the dignity of policymakers, and facilitate accountability [9]. Furthermore, applying transparency can help the KPU identify any violations of ownership, weaknesses of competence, and favoritism towards a particular political group, as well as improve the credibility of the KPU [10]. KPU stated that prevention is essential to prevent misconduct or electoral corruption. As expressed by [11]. The emphasis on preventing corruption needs to be more focused than the repression. Legal sanctions for electoral organizers need to be imposed as a form of prevention of electoral misconduct [12].

In addition to legal sanctions, transparency is considered a means of preventing electoral misconduct. According to Abidin, transparency can also be understood as " information relating to the organization is readily available and freely accessible by those affected by the policies implemented by the organization. In addition, sufficient information about the agency's performance is available and presented in an easy-to-understand form or medium [13]. In addition to the preventive aspects, a systemic solution is needed or applied throughout the region. Research results suggest that to improve accuracy in counting votes and making news of events and certificates of voting gains, e-recapitulation is needed to count votes at voting stations (TPS). With the presence of Sirekap, KPU is not inferior to Survey Institutions, which means the speed and volume of KPU data are wider than those of survey agencies that only perform sampling or data spotted. So that people will have more confidence and support for the Sirekap program [14].

The Recapitulation System (SIRAP) application that was already in use at the time of counting votes and the recapitulation of the 2020 simultaneous elections is considered successful in assisting the organizer's performance [15], [16]. In addition to counting votes, this instrument is also used to publish accurate and accountable recapitulations of votes. Implementation of Sirekap in Pilkada 2020 District of Natuna 228 Research results show there are still weaknesses in implementing Sirekap for electoral calculations. Not all KPPS members can operate the Android used for Sirekap. It's because not all KPPS members understand the technological aspects; some are still not familiar with the use of technology. The technological understanding of KPPS members varies. The ability to adapt to technology cannot be denied; implementing the Sirekap application is the first. Hence, the time to implement the technical guidance is very short, requiring a fast adaptive ability to use the application [17]. Device availability for KPPS operators having a communication device whose specifications meet the minimum standards of this application

CONCLUSION

The acceleration of digital transformation is essential in the operational process of acquisition and as a support tool while minimizing risks from the early stages of the elections, the election day, and the post-election period. (post-election).

The over-implementation of the Sirekap in the election of the head of the Bupati district and the Deputy Bupati District of Natuna district in 2022 increases transparency and accountability to increase public confidence in the results of the election calculations. The Sirekap application makes the working time of the KPU more effective than manual calculations. The KPU, with the presence of the Sireap application, also makes the information disseminated to the public no less rapidly than the survey agency because the region can monitor the data entered in the place of direct voting. The Sirekap application also has a high level of rigor and minimizes the error rate of voting calculation.

REFERENCE

[1] Belawati, Tian & Nizam. (2020). Potret Pendidikan Tinggi di Masa Covid-19. Jakarta: Direktorat Jenderal Pendidikan Tinggi.

[2] Rengganis, V.M.S., Hadi Ismail Sidiki, Fajar Saputra, & Winengku Damarjati. (2021). Problematika Partisipasi Pemilih Penyandang Disabilitas Dalam Pemilihan Serentak Lanjutan 2020. *Electoral Governance Jurnal Tata Kelola Pemilu Indonesia*, *3*(1), 116–137.

[3] Suri, E. W., & Yuneva, Y. (2021). Akselerasi Transformasi Digital pada Tata Kelola Pemilu di Kota Bengkulu. Mimbar: Jurnal Penelitian Sosial Dan Politik, 10(2), 172–181.

[4] Mpesau, A. (2024). Transformasi Elektronika Digital dalam Penghitungan dan Rekapitulasi Suara Pemilu/Pilkada: Analisis Eksistensi Sistem di Persidangan Perselisihan Hasil di Mahkamah Konstitusi. Jurnal Ilmu Manajemen Sosial Humaniora (JIMSH), 6(1), 21-29.

[5] Hardiyanti, M., Pratama, P. A., Saputra, A. D., & Sholehah, M. M. A. (2022). Urgensi Sistem E-Voting dan SIREKAP dalam Penyelenggaraan PEMILU 2024. Journal Equitable, 7(2), 249-271.

[6] Sumaryana, Y. (2014). Sistem Informasi Penghitungan Suara (SITUNG) Pemilihan Bupati Dan Wakil Bupati Di Komisi Pemilihan Umum Kabupaten Ciamis. Jurnal Teknik Informatika (JUTEKIN), 2(1).

[7] Putra, Y. F., & Hasan, E. (2019). Efektifitas Teknologi Informasi (It) Situng Pada Praktek Pelaksanaan Pilkada Di Kota Banda Aceh Tahun 2017. Jurnal Ilmiah Mahasiswa Fakultas Ilmu Sosial & Ilmu Politik, 4(4).

[8] Ruslan, Rosady. (2012). Metode Penelitian Public Relations dan Komunikasi. Jakarta: Rajawali Press.

[9] Cheema, S. G., & Maguire, L. (2002, December). Democracy, governance and development: A conceptual framework. In 4th Global Forum on Re-Inventing Government (pp. 27-33).

[10] Lovita, A. J., & Adnan, M. F. (2021). Implementasi Prinsip Transparansi Dan Imparsialitas Dalam Pelaksanaan Pilkada Pada Masa Pandemi Covid-19 Di Kabupaten Siak, Riau. JISIP (Jurnal Ilmu Sosial dan Pendidikan), 5(4).

[11] Badjuri, A. (2011). Peranan Komisi Pemberantasan Korupsi (KPK) sebagai lembaga anti korupsi di indonesia. Jurnal Bisnis dan Ekonomi, 18(1).

[12] Wibawa, S. (2022). Upaya Komisi Pemilihan Umum Dalam Meningkatkan Partisipasi Masyarakat Pada Pemilihan Umum (Pemilu) Dan Pemilihan Kepala Daerah (Pilkada) 2024 Mendatang. Serunai: Jurnal Ilmiah Ilmu Pendidikan, 8(2).

[13] Daniel, T. C. K., Pangemanan, F. N., & Sampe, S. (2022). Implementasi Kebijakan New Normal Dalam Pilkada Tahun 2020 Di Kabupaten Minahasa Selatan. GOVERNANCE, 2(1).

[14] Inzana, N., Maulana, A. A., & Sari, P. M. (2024). Inovasi Sirekap dalam Meningkatkan Partisipasi Politik. Jurnal Administrasi Pemerintahan Desa, 5(2).

[15] Juliana, J. (2013). Sistem Informasi Rekapitulasi Perolehan Suara Studi Kasus Kpu Kabupaten Inhil. Sistemasi: Jurnal Sistem Informasi, 2(2), 48-57.

[16] Zurqoni, & Rahman, F. . (2019). Aplikasi Statistik Proses Kontrol Untuk Evaluasi Proses Pembelajaran Menggunakan Diagram Kontrol X¯dan R . J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika, 12(2), 30–35. https://doi.org/10.36456/jstat.vol12.no2.a2226

[17] Fernanda, J. W., Luthifiana, V. ., & Akhyar, M. K. . (2022). Analisis Partial Least Square Structural Equation Model (PLS-SEM) untuk Pemodelan Penerimaan Sistem Jaringan Informasi Bersama Antar Sekolah (JIBAS). J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika, 15(2), 292–297. https://doi.org/10.36456/jstat.vol15.no2.a6436

Forecasting Average Rice Prices at Milling Level According to Quality Using Support Vector Regression

Fauziah Roshafara

Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Islam Bandung

Jln Ranggagading No.8, Tamansari, Kecamatan Bandung Wetan, Kota Bandung, Jawa Barat
e-mail: fauziahroshafara@unisba.ac.id

ABSTRAK

Indonesia merupakan negara agraris yang mayoritas penduduknya berprofesi sebagai petani dan salah satu komoditas besar yang dihasilkan sekaligus sebagai makan utaman adalah beras. Harga beras yang fluktuatif akan berpengaruh terhadap daya beli Masyarakat, sehingga perlu upaya untuk menyiapkan kebijakan peningkatan daya beli masyarakat terhadap beras melalui peramalan. Penelitian ini menggunakan SVR untuk pemodelan harga rata-rata beras menggunakan 114 dataset yang diperoleh dari Januari 2013 hingga Juni 2023, kemudian mengevaluasi kinerjanya menggunakan *Mean Absoute Percetage Error* (MAPE). Model terbaik terbentuk dari kernel linier dengan parameter $\varepsilon = 0.078$ dan $C = 3.1$. Model tersebut menghasilkan nilai MAPE terkecil yaitu 2.32% pada data pengujian dan 1,2% pada data pelatihan yang juga kurang dari 10% artinya kinerja model untuk meramalkan harga rata-rata beras sangat tinggi.
Kata kunci: peramalan; *support vector regression*; rata-rata harga beras

ABSTRACT

Indonesia is an agricultural country where the majority of the population work as farmers and one of the humongous commodities produced is rice. Rice is a very important commodity for the Indonesian people, because it is the main food of them. This is why rice production in Indonesia is the big concern to the government, including of the average rice prices at milling level. The fluctuative of the rice prices will be affect to the purchasing power of the people. One of the efforts that can be made to prepare a policy to increase people's purchasing power of the rice is by forecasting. This study used SVR to modeling the average rice prices using 114 datasets obtained from January 2013 to June 2023, then evaluating its performance using Mean Absoute Percetage Error (MAPE). The best model formed from a linear kernel with parameters $\varepsilon = 0.078$ and $C = 3.1$. The model produced the smallest MAPE value of 2.32% in testing data and 1.2% in training data which also less than 10% meaning that the performance of the model to forecast the average price of rice is very high.
Keywords: *Forecasting; Support Vector Regression; Average Rice Prices*

INTRODUCTION

Indonesia is an agricultural country where the majority of the population work as farmers and one of the humongous commodities produced is rice. Rice is a very important commodity for the Indonesian people, because it is the main food of them. According to United States Department of Agriculture (USDA), Indonesia is the top fourth rice consuming contry after China, India and Bangladesh. From the same source also Indonesia is the top third rice producing countries from 2017 to 2022. This is why rice production in Indonesia is the big concern to the government, including of the average rice prices at milling level. The fluctuative of the rice prices will be affect to the purchasing power of the people [13]. One of the efforts that can be made to prepare a policy to increase people's purchasing power of the rice is by forecasting [4].

Forecasting is an activity to predict future events using time series data [10]. The basic core of the forecasting process is to predict future events based on patterns of events that have occurred in the past. The purpose of forecasting is as a source of information about conditions that will occur, so the suitable action can be taken in dealing with these conditions. One of the time series models that often used for forecasting is the autoregressive integrated moving average (ARIMA) [3]. However there is requirement that need to be fulfilled, the data have to stationary to the mean and variance [17]. If the time series are non-stationary, then it needs to be transformed first to be able to obtained the model used for forecasting.

In recent years, there has been growing interest in using machine learning techniques to forecast in many sectors. The advantage of the machine learning model with other conventional time series models is that it can be applied to both linear and non-linear data. In the tourism and hospitality sector machine learning models can be used for forecasting international tourist arrivals and allso the daily room rates [16], [2]. Artificial Neural Network (ANN) can also be applied in predicting active family planning participants through government channel [6]. Another machine learning models that can be used for forecasting is support vector regression (SVR). There are several studies that use SVR for modelling and predicting wetland rice production [11], [1] which the data is non-linear patterns. On the studies about forecasting electricity consumption comparing Neural Networks and Support Vector Regression by Oğcu, et al shows that the SVR give the better performance in the forecasting accuracy than the NN [11].

This study aims to forecast the average price of rice using 114 datasets obtained from January 2013 to June 2023 based on BPS (Badan Pusat Statistik) official website. The data indicated that it was non stationary to the average due to seasonal effects. Therefore, SVR is more suitable to used to overcome this problem where the data does not need to be transformed first to produce a model used for forecasting. After obtaining the results by the optimum parameter, the performance is then evaluated using the Mean Absoute Percentage Error (MAPE).

METHOD

Support Vector Regression is the application of machine learning concepts to classify regression models using the Support Vector Machine method [9]. The goal of SVR is to find the $f(y)$ function as a hyperplane in the form of a regression function that corresponds to all input data by making the smallest possible error (ε) [14]. Suppose there is a $y(t)$ time series datasets, where $t = \{0,1,2,\dots,N-1\}$. Then by using regression analysis the prediction function in linear and non-linear regression can be written as follows:

$$\widehat{Y_t} = (\boldsymbol{w} \cdot y) + b \tag{1}$$

$$\widehat{Y_t} = \big(\boldsymbol{w} \cdot \varphi(y)\big) + b \tag{2}$$

where $\boldsymbol{w}$ is the weighting vector, $\varphi(y)$ is a function that maps to a higher dimensional space and b is biased. Then, to maximize the hyperplane can be done by minimizing the regularized risk function defined by equation (3).

$$R_{reg}(f) = R_{emp}(f) + \frac{\lambda}{2}\|\boldsymbol{w}\|^2 \quad (3)$$

where,

$$R_{emp}(f) = \frac{1}{N}\sum_{t=1}^{N} L(\gamma(t) - f(y(t), \boldsymbol{w})) \quad (4)$$

with λ is a constant used to reduce overfitting of the data and minimize the adverse effects of generalization, i is the time series index $t = \{0,1,2, \dots, N-1\}$, $\gamma(t)$ is the actual data of the predicted value that to be found and $L(.)$ is the loss function [15].

Loss function is a function that shows the relationship between errors and how errors are penalized [5]. The most commonly used of loss function is the ε-insensitive loss function. To obtain the optimal weight by minimizing regularized risk, a quadratic equation is formed using the ε-insensitive loss function as follows:

$$\min\left\{\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{t=1}^{N} L(\gamma(t) - f(y(t), \boldsymbol{w}))\right\} \quad (5)$$

where

$$L(\gamma(t), f(y(t), \boldsymbol{w})) = \begin{cases} |\gamma(t) - f(y(t), \boldsymbol{w})| - \varepsilon, & |\gamma(t) - f(y(t), \boldsymbol{w})| \geq \epsilon \\ 0 \quad , & other. \end{cases} \quad (6)$$

The C constant is a normalizing factor of sum by $1/N$ and ε is a measure of the precision used to approximate the function. Solving for optimal weights and refractive values uses the Lagrange multiplier and kernel function, so the prediction function of the regression can be explicitly written as follows:

$$\widehat{Y}_t = \sum_{t=1}^{N}(\alpha_t - \alpha_t^*)\langle y, y(t)\rangle + b \quad (7)$$

where α_i is a Lagrange multiplier and $\langle y, y(t)\rangle$ is a kernel function. To perform non-linear regression using SVR, it is necessary to map the $y(t)$ input space into higher-dimensional features $\varphi(y(t))$. Then a kernel function that satisfies *Mercer's condition* can be generated as:

$$k(y, y') = \langle\varphi(y), \varphi(y')\rangle \quad (8)$$

There are several kernel functions that can be used to solve (8) equation, namely:

1. Linear Kernel Function: $K(y_i, y) = y^T y$
2. Polynomial Kernel Function: $K(y_i, y) = (y^T y + 1)^d$
3. Radial Basis Fuction *(RBF)* Kernel Function: $K(y_i, y) = exp\left(-\frac{\|y_i - y\|^2}{2\gamma^2}\right)$

While a method that can be used to evaluate the effectiveness of a forecast is the Mean Absoute Percetage Error (MAPE). Error percentage measurement has the advantage of being scale independent and is often used to compare the performance of a forecast on different datasets [8]. Performance measurement using MAPE produces a percentage value. The smaller the MAPE value, the better the level of accuracy of a forecast. The calculation of MAPE can be written as follows:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\% \quad (9)$$

Y_t is actual value of sales in the t-month, $\widehat{Y}_t$ is value of sales forecasting results in the t-month and N is an amount of data. Based on Lewis [8], MAPE values can be interpreted in four categories, highly accurate (<10%), good (10 – 20%), reasonable (20 – 50%) and not accurate (>50%).

RESULT AND DISCUSSION

According from BPS (Badan Pusat Statistik) official website the average rice prices at milling level differ based on its quality, respectively premium, medium and low quality. This research use the average rice prices at milling level with the medium quality obtained from January 2013 to June 2023, and then forecast for the next six months using support vector regression.

The time series plot of the average rice praces can be shown in Figure 1. It can be seen that the data in non-stasionary. There is an increasing trend and also the variance of the data is not constant meaning the regular time series methods that needs stationary of the mean and variance of the data can not be used without making transformation and differencing to remove an unconstant variance and trend. Therefore support vector regression can be used as an alternative method at the data with those problem.

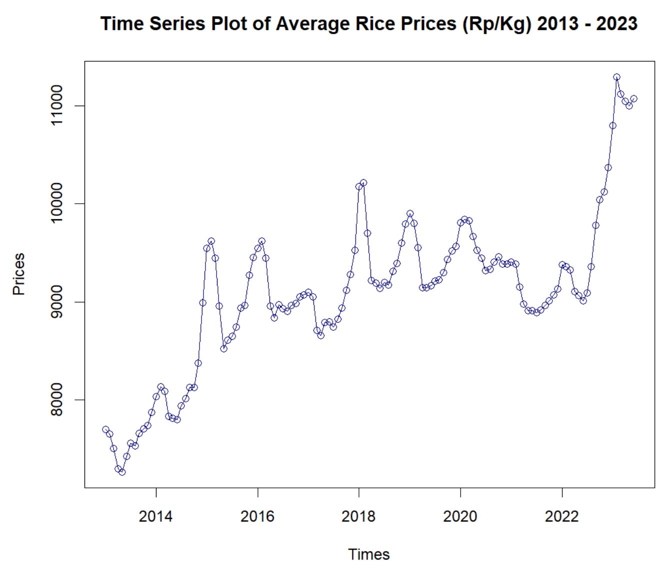


Figure 1. The Series Plot of The Average Rice Prices at Milling Levels

The time series plot of the average rice prices from Figure 1 is not clear whether there is a seasonality or not. To reveal this issue, needs to be plotted by the periods of 12 months in Figure 2. It is very clear that there is seasonality in the average of rice prices which almost every year almost have the same pattern of price shifting in the each of months.

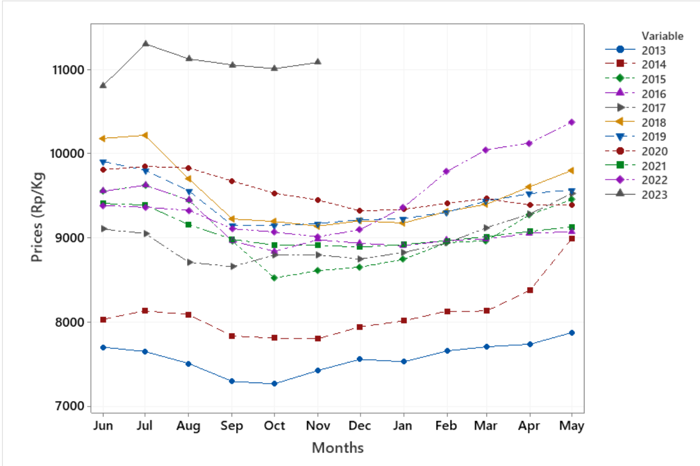


Figure 2. The Average Rice Prices of Each Months

The traditional univariate time series methods for forecasting will only use the average rice prices level as the variable. However in the literature of the forecasting using SVR always used multivariate data where the average rice prices as dependent varible and other variables related to it as independent variables. Here, in this research, total three inputs were used where all derived from the average rice prices data as output. These inputs consist of 'lag 1', 'lag 2', and 'seasonal index'.

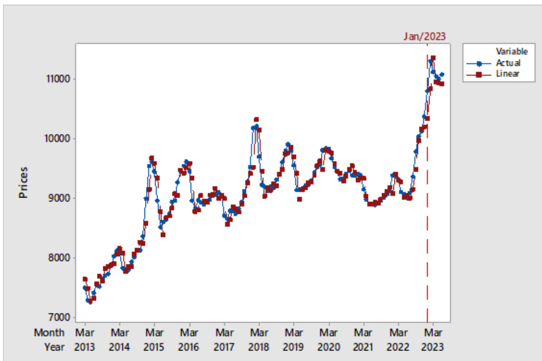
Before modeling the data, it is necessary to do partition of the datasets into training and testing datasets. Initially, the model is trained according to the training data from January 2013 to December 2022 which consist 108 months of the datasets. The 6 months rest of the datasets would be used to evaluate the forecasting performance of the model formed from January 2023 until June 2023. The testing datasets were used for assessment of the model.

The kernel function used in this study is a linear kernel. For comparison, polynomial kernel functions are used to find out the best model results. The best SVR modeling results are generated from optimal parameter values. The selection of optimal parameters is carried out using the grid search method. The combination of cost parameters used is between 1 to 100 with a difference in value of 1, epsilon parameters are used between 0.001 to 0.1 with a difference in values of 0.001. While the parameter of degree for the polynomial kernel used combination between 1 to 15 with a difference value of 1. In order to measure the performance of the model the Mean Absolute Percentage of Error (MAPE) is used following of the (1.8) equation.

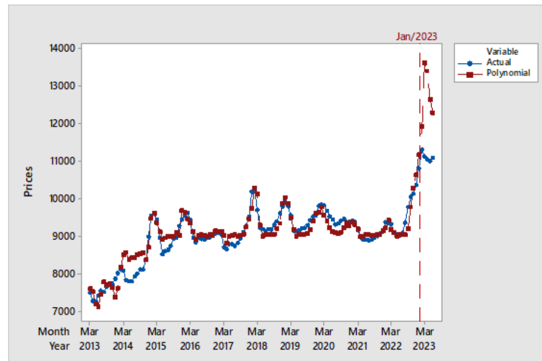
Table 1. Forecasting Performance

Kernel	Parameter				MAPE	
	Epsilon	Cost	Degree	Gamma	Testing	Training
Linear	0.080	3.0	-	-	2.330887	1.200998
	0.078	3.0	-	-	2.317659	1.200759
	0.078	**3.1**	**-**	**-**	**2.315519**	**1.200605**
Polynomial	0.100	1.0	1	-	2.375195	1.208691
	0.037	1.5	2	-	63.43985	3.977249
	0.098	13.6	3	-	11.17574	2.030829
Radial	0.069	1	-	0.1996	17.52425	1.11822
	0.069	15	-	0.1996	16.73884	0.834056
	0.001	15	-	0.1596	18.20053	0.776371

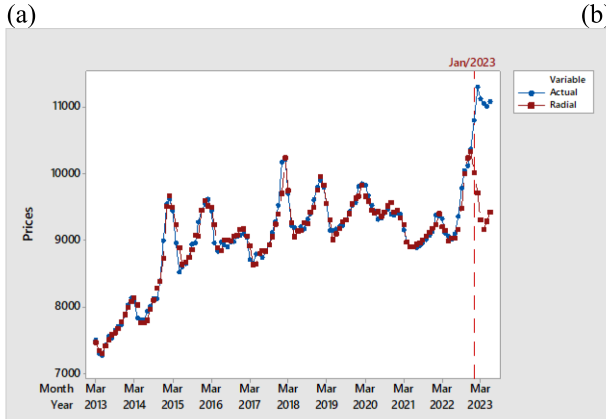
The result of the forecasting performance of the several models using different kernels and its parameter shown in Table 2.1. The smallest MAPE score from testing data obtained from the model with a linear kernel using epsilon and cost parameter of 0.078 and 3.1 respectively. The smallest MAPE score for the polynomial kernel found in the parameter degree is 1 and the score is near with the MAPE score of linear kernel. Meaning that the model is following the linear form. This is evidenced by the smallest MAPE scores found in model with linear kernel resulting from the epsilon parameter and cost parameter of 0.078 and 3.1.



(a)



(b)



(c)

Figure 3. Comparison of Actual Time Series Plots with Forecasting Results using SVR in (a) Linear Kernel (b) Polynimial Kernel (c) Radial Kernel

Forecasting results on the test data using linear kernels in the Figure 3 (a) shows a value of the rice prices closest to its actual value compared to using polynomial kernels and radial kernels. The results of forecasting using polynomial and radial kernels show that they are able to model on training data well but when evaluated using test data the model is not able to predict well the price of rice, which is called overfitting. This is not happend when forecasting using the linear kernel shown in Figure 3 (a) where the forecasting results in both the training data and the test data display values close to the actual values. Then the model with linear kernel and parameters ε = 0.078; C = 3.1 can be said to be the best model. Then the SVR-linear model (ε = 0.078; C = 3.1) is formed as follows:

$$\hat{Y}_t = 0.002176049 + 1.319091Y_{t-1} - 0.3605231Y_{t-2} + 0.07345937I_s \qquad (10)$$

Equation (10) above can be explained that if there is an increase in rice prices by 1 rupiah in the previous month. the price of rice I t-month will also increase by 1.32 rupiahs. Meanwhile. if there is an increase in rice prices in the previous two months. the price of rice in the t-month will decrease by 0.36 rupiahs. As for the seasonal index where the seasonal period occurs every month as shown in figure (2), meaning that for every increase in one month period. the price of rice in the t-month will increase by 0.07 rupiahs.

CONCLUSION

Based on the results of the research that has been done above. it can be concluded on forecasting the average prices of rice using support vector regression formed the best model from a linear kernel with parameters ε = 0.078 and C = 3.1. The model produced the smallest MAPE value of 2.32% in testing data and 1.2% in training data. where this model was able to capture data patterns without ovefitting. The MAPE value of the best model is also already less than 10% which according to Lewis [8] meaning that the performance of the model to forecast the average price of rice is very high. Then this model can be used for forecast the average prices of rice for next period.

REFERENCE

[1] Alkaff, M., Khatimi, H., Puspita, W., & Sari, Y. “Modelling and predicting wetland rice production using support vector regression.” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*. vol. 17. no.2. pp. 819-825. 2019.

[2] Andariesta, D. T., & Wasesa, M. “Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: a multisource Internet data approach.” *Journal of Tourism Futures*. 2022.

[3] De Gooijer, J.G. & Hyndman, R.J. “25 Years of Time Series Forecasting.” *International journal of forecasting*. vol. 22. no. 3. pp. 443-473. 2006.

[4] Firdaus, Y.A., Ngatini, N. & Wijaya, S.U. “Pemodelan Regresi Data Panel Harga Beras di Wilayah Indonesia Bagian Barat.” *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, vo.16. no. 2, pp. 486-498. 2023

[5] Gunn, S. R. “Support vector machines for classification and regression.” *ISIS Technical Report*. vol. 14. no. 1. pp. 5-16. 1998.

[6] Hayadi, B. H., Sudipa, I. G. I., & Windarto, A. P. “Model peramalan pada peserta KB aktif jalur pemerintahan menggunakan Artificial Neural Network Back-propagation.” *Matrik: Jurnal Managemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 21. no.1, pp 11-20. 2021.

[7] Hou, X., Zhu, S., Xia, L., & Wu, G. “Stock price prediction based on Grey Relational Analysis and support vector regression.” *Chinese Control And Decision Conference (CCDC), IEEE*. 2018. pp. 2509-2513.

[8] Lewis, C. D. *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann. 1982.

[9] Lu, C. J., Lee, T. S., & Chiu, C. C. “Financial time series forecasting using independent component analysis and support vector regression.” *Decision support systems*. vol. 47. no. 2. pp. 115-125. 2009.

[10] Montgomery, D.C., Jennings, C.L. & Kulahci, M. *Introduction to Time Series Analysis and Forecasting*, Second Edition. John Wiley & Sons Inc. 2015.

[11] Oğcu, G., Demirel, O. F., & Zaim, S. “Forecasting electricity consumption with neural networks and support vector regression.” *Procedia-Social and Behavioral Sciences*. 2012. pp. 1576-1585.

[12] Paidipati, K.K., Chesneau, C., Nayana, B.M., Kumar, K.R., Polisetty, K. & Kurangi, C. “Prediction of rice cultivation in india—support vector regression approach with various kernels for non-linear patterns.” *AgriEngineering*. vol. 3. no. 2. pp. 182-198. 2021.

[13] Resnia, R. “Fluktuasi Harga Bahan Pangan Pokok (Bapok) dan Daya Beli Kelompok Masyarakat Berpendapatan Rendah.” *Buletin Ilmiah Litbang Perdagangan*, vol. 6. no.2. pp. 169-188. 2012.
[14] Smola, A. J. & Schölkopf, B. “A tutorial on support vector regression.” *Statistics and computing*. vol. 14. no.3. pp. 199-222. 2004.
[15] Sapankevych, N. & Sankar, R. “Time series prediction using support vector machines: A survey.” *IEEE Computational Intelligence Magazine*. vol. 4. no.2, pp.24–38. 2009.
[16] Wiryawanto, Tiara Melati Putri, Zuyyina Hawani, & Muhamad Attar Ramadhani. "Comparison of Support Vector Machine (SVM) and Autoregressive Integrated Moving Average (ARIMA) Methods for Predicting Air Quality Using Python and KNIME." *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*. vol. 16. no. 1. pp. 384-394. 2023.
[17] Zheng, T., Liu, S., Chen, Z., Qiao, Y., & Law, R. “Forecasting daily room rates on the basis of an LSTM model in difficult times of Hong Kong: evidence from online distribution channels on the hotel industry.” *Sustainability*. vol. 12. no. 18. pp. 7334. 2020.

Dynamical Analysis of Mathematical Model of Social Behavior with Law Enforcement and Religious Approaches

Nailul Izzati (1), Jati Widyo Leksono (2), Nanndo Yannuansa (3)

[1,2,3]Teknik Elektro, Fakultas Teknik, Universitas Hasyim Asy'ari Tebuireng Jombang

Jl. Irian Jaya No. 55 Tebuireng, Diwek, Jombang

e-mail: nailulizzati@unhasy.ac.id(1), jatiwidyoleksono@unhasy.ac.id(2), nanndoyannuansa@unhasy.ac.id(3)

ABSTRAK

Model matematika dapat digunakan untuk menggambarkan berbagai macam fenomena, salah satunya adalah fenomena sosial. Salah satu fenomena sosial yang menarik untuk dikaji adalah tentang tindak kriminal. Dengan membagi total populasi menjadi beberapa bagian berdasarkan status perilaku sosialnya, dapat dibangun model matematika untuk menggambarkan dinamika sosial. Dinamika sosial tersebut diketahui dengan melakukan analisis dinamik dan simulasi numerik. Dalam penelitian ini dilakukan simulasi numerik menggunakan software Maple 2022 dan metode Runge-Kutta. Berdasarkan hasil analisis dinamik dan simulasi numerik, diketahui bahwa dengan menerapkan penegakan hukum dan pendekatan keagamaan, tindak kriminal dalam suatu populasi dapat dikurangi bahkan dihilangkan.

Kata kunci: Model matematika; perilaku sosial; tindak kriminal; penegakan hukum; pendekatan keagamaan.

ABSTRACT

Mathematical models could be used to describe various phenomena, one of which is social phenomena. One of the interesting social phenomena to study is criminal behavior. By dividing the total population into several compartments based on their social behavior status, a mathematical model could be constructed to depict social dynamics. These social dynamics could be understood through dynamical analysis and numerical simulation. In this study, numerical simulations were conducted using Maple 2022 software and the Runge-Kutta method. Based on the results of dynamical analysis and numerical simulations, it is known that by implementing law enforcement and religious approaches, criminal activities within a population could be reduced or even eliminated.

Keywords: *Criminal behavior; law enforcement; mathematical model; religious approaches; social behavior.*

INTRODUCTION

According to the Oxford Language Dictionary, crime means an act or omission that constitutes a punishable offense under the law, or can also be interpreted as illegal activity, or an action or activity that, although not illegal, is considered morally wrong, shameful, or incorrect [1]. Meanwhile, according to the Oxford Dictionary of Sociology, crime is considered as a violation that extends beyond personal and public realms, breaking rules or laws that are prohibited, subject to legitimate punishment or sanctions, and requiring intervention by public authorities (state or local bodies). Ideally, institutions execute a formal system to handle crimes and employ representative officers (such as the police) to act on behalf of these institutions. In terms of law and jurisprudence, being guilty of a criminal act typically involves malicious intent or deliberate negligence, although there are exceptions in law. If conscious intent is proven absent (such as in cases involving children or the mentally ill), the offense is not considered a crime and may not incur regular punishment (though some forms of detention or therapeutic care may follow) [2].

Criminal behavior, broadly speaking, is a deviant behavior. Such behavior certainly disrupts social order. Therefore, law enforcement is necessary to prevent and overcome it. From a mathematical perspective, social behaviors could be modeled. Former researches that discuss mathematical model of criminal behavior includes: Misra (2014), who developed a model examining the impact of police strength in controlling crime within a population of varying sizes [3]; Abbas et al. (2017), who constructed a simpler model to depict the dynamics between two factions, where law enforcement is a factor in reducing the criminal population [4]. This model was further developed in subsequent research. González-Parra et al. (2018) considered the law enforcement process in more detail [5], Tripathi et al. (2021) incorporated Holling type II response function into their model [6], Kumar and Abbas (2023) explored age structure [7], and Arora et al. (2023) utilized a fractional-order model to describe the dynamics of criminal activity spread [8].

In addition to law enforcement, prevention of criminal acts can be achieved through religious approaches. In practice, one of the functions of religion in society is as a form of social control [9]. The religious approaches play a crucial role, both directly and indirectly, in preventing criminal behavior [10]. Considering that religious activities are expected to prevent individuals from committing evil and wrongful acts. A study mentioned that religious activities significantly influence social behavior due to the inherent values of goodness they promote [11].

In this study, we discuss a mathematical model of social behavior with law enforcement and religious approaches. The model is constructed by modifying previous mathematical models of social behavior in former research. Besides model modification, we also discuss dynamical analysis and numerical simulations of the modified model.

METHOD

There are six steps undertaken in this study, the steps are shown in Figure 1. The first step is literature review. In this step, data supporting and related to criminal behavior are collected, including the influence of law enforcement and religious approaches on such behavior. The literature review also identifies the mathematical model that be used as a reference for constructing the mathematical model of social behavior with law enforcement and religious approaches. The second step is modifying the mathematical model obtained from the first step. In the second step, new state variables and parameters are constructed to depict the mathematical model of social behavior with law enforcement and religious approaches. The third step is dynamical analysis of the modified model. This step aims to determine equilibrium points and eigenvalues of the system describing the model. The Routh-Hurwitz criterion is used to assess the stability characteristics of the obtained equilibrium points. The fourth step is numerical simulation. In numerical simulations, calculations are performed using Maple 2022 to verify the analytical results obtained in the previous steps, including equilibrium points, eigenvalues, and stability characteristics. The fourth-order Runge-Kutta method is used in numerical simulations to plot population dynamics graphs in the model of social behavior with law enforcement and religious approaches. The fifth step is drawing conclusions from the results obtained.

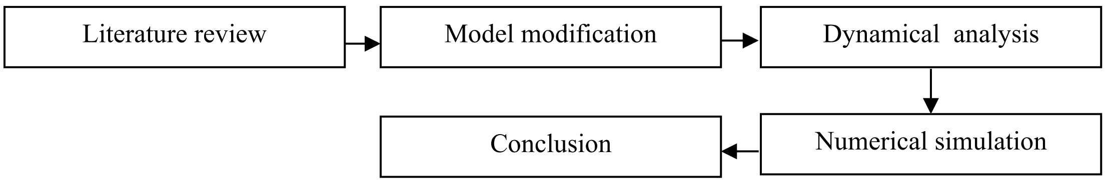


Figure 1. Research flow diagram

RESULT AND DISCUSSION

The mathematical model of social behavior with law enforcement and religious approaches builds upon the model developed by Abbas et al. [4]. In addition to the non-criminal population (N_p) and criminal population (C_p), another compartment is considered, namely the religious population denoted as R_p. Interactions between individuals from the criminal population and the religious population cause the criminal-minded individuals to change their status to be included in the non-criminal population. While, interactions between individuals from the non-criminal population and the religious population cause the non-criminal individuals to change their status to be included in the religious population. This is illustrated by the compartment diagram in Figure 2.

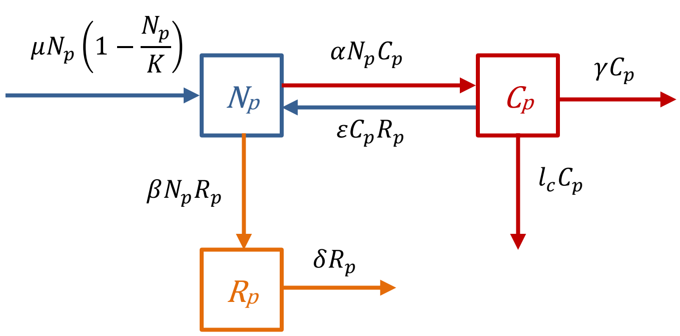


Figure 2. Compartment diagram of the mathematical model of social behavior with law enforcement and religious approaches

The compartment diagram in Figure 2 illustrates that the dynamics of the three populations in the model are affected by:

1. The growth of the non-criminal population is based on the logistic model $\mu N_p\ (1-\frac{N_p}{K})$. Additionally, the non-criminal population increases due to interactions between the criminal population and the religious population, $\varepsilon C_p R_p$. Conversely, interactions between the non-criminal population and the criminal population decrease the non-criminal population by $\alpha N_p C_p$. Similarly, interactions between the non-criminal population and the religious population decrease the non-criminal population by $\beta N_p R_p$.
2. The growth of the criminal population is influenced by interactions between the non-criminal population and the criminal population, which increases the criminal population by $\alpha N_p C_p$. Conversely, it decreases due to interactions between the criminal population and the religious population, $\varepsilon C_p R_p$. The criminal population also decreases due to law enforcement measures $l_c C_p$ and natural deaths denoted by γC_p.
3. The growth of the religious population is influenced by interactions between the non-criminal population and the religious population, which increases it by $\beta N_p R_p$. Conversely, it decreases due to natural deaths in the religious population, denoted by δR_p.

The compartment diagram in Figure 2 could be mathematically expressed by the system of ordinary differential equations (1)-(3).

$$\frac{dN_p}{dt} = \mu N_p\left(1-\frac{N_p}{K}\right) - \alpha N_p C_p - \beta N_p\, R_p \tag{1}$$

$$\frac{dC_p}{dt} = -\gamma C_p + \alpha N_p C_p - l_c C_p - \varepsilon C_p R_p \tag{2}$$

$$\frac{dR_p}{dt} = \beta N_p R_p - \delta R_p \tag{3}$$

Where μ is the natural growth rate, K is the carrying capacity of the environment, α is the criminality rate, β is the rate of interaction between the non-criminal population and the religious population, ε is the rate of interaction between the criminal population and the religious population, γ is the death rate of the criminal population, l_c is the law enforcement rate, and δ is the death rate of the religious population.

By solving $\frac{dN_p}{dt}=0, \frac{dC_p}{dt}=0, \frac{dR_p}{dt}=0$, equilibrium points of the system (1)-(3) are obtained. There are five equilibrium points: the trivial equilibrium point $E_1=(0{,}0{,}0)$, equilibrium points with no criminal community $E_2=(K,0{,}0)$ and $E_4=\left(\frac{\delta}{\beta},0,\frac{\mu(\beta K-\delta)}{\beta^2 K}\right)$, equilibrium points with no religious community $E_3=\left(\frac{\gamma+l_c}{\alpha},\frac{\mu(\alpha K-\gamma-l_c)}{\alpha^2 K},0\right)$, and equilibrium point $E_5=\left(\frac{\delta}{\beta},\frac{\delta\left(\mu\beta K\varepsilon-\mu\delta\varepsilon+\gamma\beta^2 K-\alpha\beta\delta K+\beta^2 l_c K\right)}{\beta^2 K\varepsilon(\gamma+l_c)},\frac{-\gamma\beta+\alpha\delta-\beta l_c}{\varepsilon\beta}\right)$. Using The Next Generation method to determine the basic reproduction number, R_0 is obtained as $R_0=\frac{\alpha\delta\beta K+\varepsilon\mu\delta}{\beta K(\gamma\beta+l_c\beta+\varepsilon\mu)}$.

The Jacobian matrix of the system is

$$J = \begin{bmatrix} \mu - \frac{2\mu N_p}{K} - \alpha C_p - \beta R_p & -\alpha N_p + \varepsilon R_p & -\beta N_p + \varepsilon C_p \\ \alpha C_p & -\gamma + \alpha N_p - l_c - \varepsilon R_p & -\varepsilon C_p \\ \beta R_p & 0 & \beta N_p - \delta \end{bmatrix}. \quad (4)$$

By substituting equilibrium points into the Jacobian matrix and solving $|J_{E_i} - \lambda I| = 0$ for $i = 1,2,3,4,5$, the characteristic equations of each equilibrium point are determined. For equilibrium points E_1, E_2, E_3, and E_4, eigenvalues and stability properties could be obtained. However, for equilibrium point E_5, due to the algebraically length characteristic equation, stability is determined using the Routh-Hurwitz criterion. The results, shown in Table 1, with parameters a, b, c, and d are as follows.

$$a = \mu^2\gamma^2 + \mu^2 l_c^2 + 2\mu^2\gamma l_c + 4\alpha\mu K\gamma^2 + 4\alpha\mu K l_c^2 + 8\alpha\mu K\gamma l_c - 4\mu\alpha^2 K^2\gamma - 4\mu\alpha^2 K^2 l_c. \quad (5)$$

$$b = \mu\varepsilon\beta^2\gamma K + \mu\varepsilon\beta^2 l_c K - 2\mu\delta\varepsilon\beta\gamma - 2\mu\delta\varepsilon\beta l_c - \alpha\mu\delta\varepsilon\beta K + \alpha\mu\delta^2\varepsilon - \alpha\delta\gamma\beta^2 K + \alpha^2\delta^2\beta K - \alpha\delta l_c\beta^2 K + \gamma^2\beta^3 K + \gamma l_c\beta^3 K - \alpha\delta\beta^2\gamma K - \alpha\delta\beta^2 l_c K + \beta^3\gamma l_c K + \beta^3 l_c^2 K. \quad (6)$$

$$c = \frac{-bc_1(\gamma + l_c) + bc_2\left(\frac{-\beta\gamma + \alpha\delta - \beta l_c}{\varepsilon}\right) - c_3(\gamma + l_c)\frac{-\beta\gamma + \alpha\delta - \beta l_c}{\varepsilon}}{b}. \quad (7)$$

$$c_1 = \frac{\alpha\varepsilon\mu\delta\beta K - \alpha\varepsilon\mu\delta^2 + \alpha\delta\gamma\beta^2 K - \alpha^2\delta^2\beta K + \alpha\delta l_c\beta^2 K}{\varepsilon\beta^2(\gamma + l_c)K}. \quad (8)$$

$$c_2 = \frac{-\delta\beta^2\gamma K - \delta\beta^2 l_c K + \varepsilon\mu\delta\beta K - \varepsilon\mu\delta^2 + \delta\gamma\beta^2 K - \alpha\delta^2\beta K + \delta l_c\beta^2 K}{\beta^2(\gamma + l_c)K}. \quad (9)$$

$$c_3 = \frac{-\varepsilon\mu\delta\beta K + \varepsilon\mu\delta^2 - \delta\gamma\beta^2 K + \alpha\delta^2\beta K - \beta^2\delta l_c K}{\beta^2(\gamma + l_c)K}. \quad (10)$$

$$d = -c_3(\gamma + l_c)\frac{-\beta\gamma + \alpha\delta - \beta l_c}{\varepsilon}. \quad (11)$$

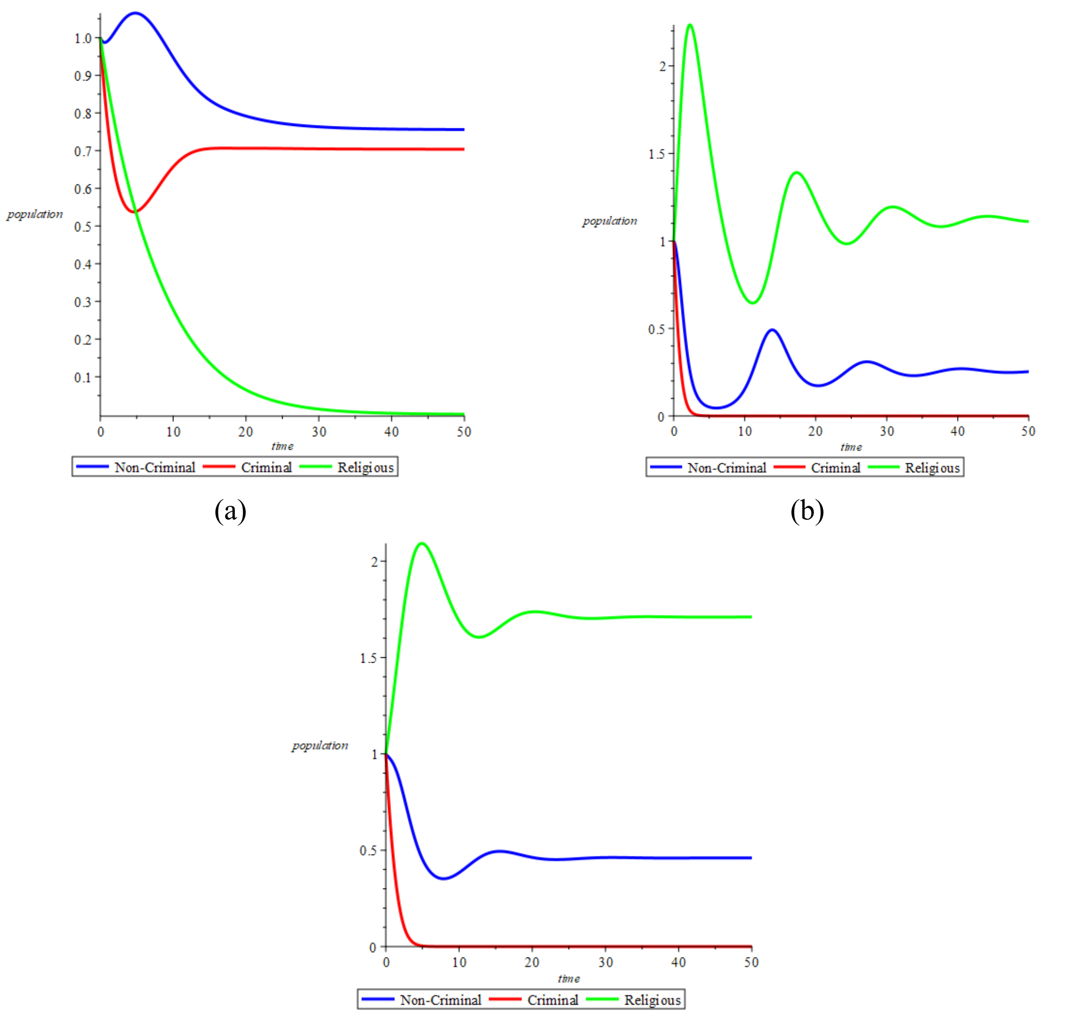
Table 1. Eigen value and stability of the equilibriums

Equilibrium	Eigen value	Stability
E_1	$\lambda_1 = \mu, \lambda_2 = -\gamma - l_c, \lambda_3 = -\delta.$	Unstable for all $K > 0$.
E_2	$\lambda_1 = -\mu, \lambda_2 = \alpha K - \gamma - l_c, \lambda_3 = \beta K - \delta.$	Stable if $\alpha K < \gamma + l_c$ and $\beta K < \delta$.
E_3	$\lambda_1 = \frac{\beta(\gamma + l_c)}{\alpha} - \delta, \lambda_2 = \frac{-\mu\gamma - \mu l_c + \sqrt{a}}{2\alpha K}, \lambda_3 = \frac{-\mu\gamma - \mu l_c - \sqrt{a}}{2\alpha K}.$	Stable if $\frac{\beta(\gamma + l_c)}{\alpha} < \delta$ and $\gamma + l_c < \alpha K$.
E_4	$\lambda_1 = \frac{\alpha\delta\beta K - \varepsilon\mu\beta K - \gamma\beta^2 K - l_c\beta^2 K + \varepsilon\mu\delta}{\beta^2 K}, \lambda_2 = \frac{-\mu\delta + \sqrt{\mu^2\delta^2 - 4\beta^2\delta\mu K^2 + 4\beta\delta^2\mu K}}{2\beta K}, \lambda_3 = \frac{-\mu\delta - \sqrt{\mu^2\delta^2 - 4\beta^2\delta\mu K^2 + 4\beta\delta^2\mu K}}{2\beta K}.$	Stable if $R_0 < 1$ and $\beta K > \delta$.
E_5	Not analyzed.	Stable if $b < 0$, $c < 0$, and $d < 0$.

In numerical simulations to depict population dynamics over time, parameter values used are $\mu = 1.2$, $\gamma = 0.23$, $l_c = 0.45$ [6], $\varepsilon = 0.5$, and $\delta = 0.23$. Meanwhile, for parameters α, β, and K, various values are employed to observe their effects on the system's solutions. Numerical simulations are conducted with three different cases: Case A where $\alpha > \beta$, Case B where $\alpha < \beta$, and Case C where $\alpha = \beta$. Each case is simulated with three different values of K, i.e. $K = 1.6$, $K = 16$, and $K = 160$. The comparison of values obtained from numerical simulations is presented in Tables 2, 3, and 4. The behavior of the system's solutions is illustrated in Figure 3, Figure 4, and Figure 5.

Table 2. The results of numerical simulation I ($K = 1.6$)

Parameter	Case A	Case B	Case C
α	0.9	0.1	0.5
β	0.1	0.9	0.5
Feasible equilibrium(s)	E_1, E_2, E_3	E_1, E_2, E_4	E_1, E_2, E_3, E_4
R_0	1.601	0.098	0.306
Stable equilibrium(s)	$E_3(0.756,0.704,0)$	$E_4(0.256,0,1.120)$	$E_4(0.460,0,1.710)$



(a) (b)

(c)

Figure 3. Population dynamics in the mathematical model of social behavior with law enforcement and religious approaches with $K = 1.6$, (a) $\alpha = 0.9$, $\beta = 0.1$, (b) $\alpha = 0.1$, $\beta = 0.9$, (c) $\alpha = 0.5$, $\beta = 0.5$

In numerical simulation I, the value $K = 1.6$ is used. In Table 2, it is observed that when $\alpha = 0.9 > \beta = 0.1$, the basic reproduction number $R_0 > 1$, and the numerical solution converges to equilibrium point E_3. When $\alpha = 0.1 < \beta = 0.9$, and $\alpha = \beta = 0.5$, the basic reproduction number $R_0 < 1$, and the numerical solution converges to equilibrium point E_4. With initial values $N_p(0) = 1$, $C_p(0) = 1$, and $R_p(0) = 1$, the behavior of the numerical solution in Figure 3 is obtained. Figure 3.a shows that over time, the numerical solution converges to $E_3(0.756, 0.704, 0)$. This indicates that when the criminality rate is greater than the interaction rate between the non-criminal population and the religious population, over time, the system that initially with $N_p = C_p = R_p = 1$ evolves to $N_p = 0.756$, $C_p = 0.704$, and $R_p = 0$. In this case, eventually, the religious population becomes extinct, and criminal activities persist in the system.

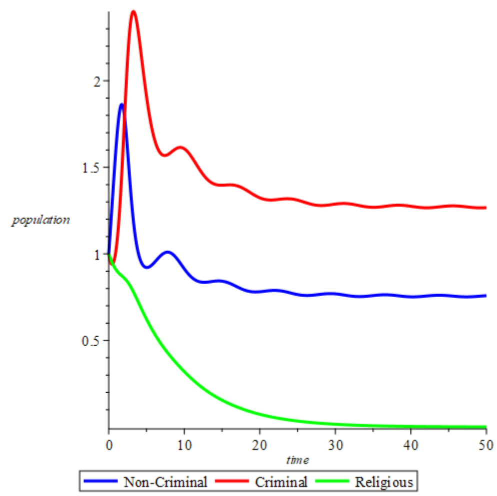
Figure 3.b shows that the numerical solution converges to $E_4(0.256, 0, 1.120)$. This means that when the rate of interaction between the non-criminal population and the religious population is greater than the criminality rate, over time, the criminal population will eventually become extinct. Meanwhile, Figure 3.c demonstrates that the numerical solution converges to $E_4(0.460, 0, 1.710)$. This indicates a similar outcome to Figure 3.b, with the difference lying in the time required to reach the equilibrium point free from criminal activities, E_4. Figure 3.b takes longer to converge compared to Figure 3.c. Based on the comparison of Case A, Case B, and Case C in numerical simulation I, it is evident that the ratio between α (criminality rate) and β (rate of interaction between the non-criminal population and the religious population) influences the dynamics occurring within the system.

In numerical simulation II and III, the carrying capacity values were increased compared to numerical simulation I, specifically to $K = 16$ and $K = 160$. The results obtained in simulations II and III are similar. The difference lies in the fact that with larger values of K compared to before, it is known from the results shown in Tables 3 and 4 that equilibrium points E_4 and E_5 are feasible in Case A, and equilibrium point E_3 is feasible in Case B. In Case A, it is found that E_3 and E_4 are locally stable. To observe this, the differences in the graphs shown in Figure 4 and Figure 5 can be noted.

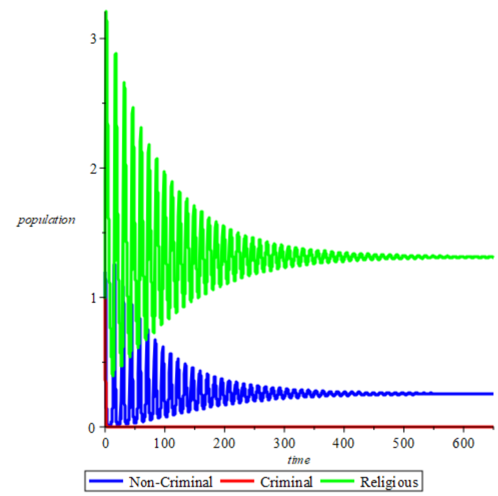
Table 3. The results of numerical simulation II ($K = 16$)

Parameter	Case A	Case B	Case C
α	0.9	0.1	0.5
β	0.1	0.9	0.5
Feasible equilibrium(s)	E_1, E_2, E_3, E_4, E_5	E_1, E_2, E_3, E_4	E_1, E_2, E_3, E_4
R_0	0.439	0.027	0.141
Stable equilibrium(s)	$E_3(0.756,1.270,0)$ and $E_4(2.3,0,10.275)$	$E_4(0.256,0,1.312)$	$E_4(0.460,0,2.331)$

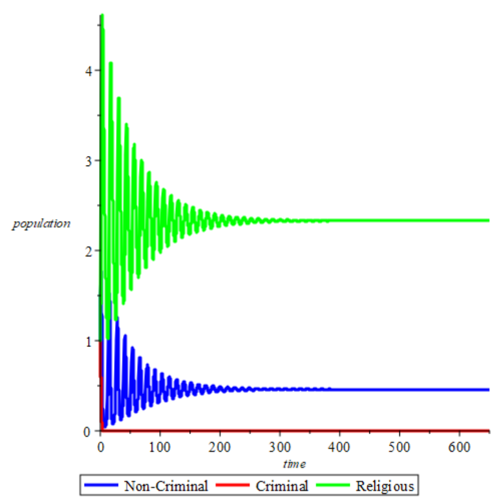
In Figure 4.a, with $K = 16$, it can be seen that if the initial values are $N_p(0) = C_p(0) = R_p(0) = 1$, over time, the numerical solution in Case A converges to equilibrium point $E_3(0.756, 1.270, 0)$. This means that the religious population will become extinct while the criminal population persists in the system. However, if the initial values are taken as $N_p(0) = C_p(0) = 1$, $R_p(0) = 3$, in Case A (see Figure 5.a), the numerical solution converges to equilibrium point $E_4(2.3, 0, 10.275)$. This indicates that the criminal population becomes extinct, while the religious population remains. This shows that if the initial values are sufficiently close to equilibrium point E_3, the numerical solution will converge to E_3, whereas if the initial values are sufficiently close to equilibrium point E_4, the numerical solution will converge to E_4.



(a)



(b)



(c)

Figure 4. Population dynamics in the model with $K = 16$, $N_p(0) = C_p(0) = R_p(0) = 1$ (a) $\alpha = 0.9$, $\beta = 0.1$, (b) $\alpha = 0.1$, $\beta = 0.9$, (c) $\alpha = 0.5$, $\beta = 0.5$

In other words, when the environmental capacity is increased, with the criminality rate greater than the rate of interaction between the non-criminal population and the religious population, the initial amount of the religious population plays a role in determining the behavior of the numerical solution. Meanwhile, in Figure 4.b and Figure 5.b, both show that the numerical solution converges to $E_4(0.256, 0, 1.312)$. Similarly, in Figure 4.c and Figure 5.c, both converge to $E_4(0.460, 0, 2.331)$. Similar results are obtained when $K = 160$ (see Table 4).

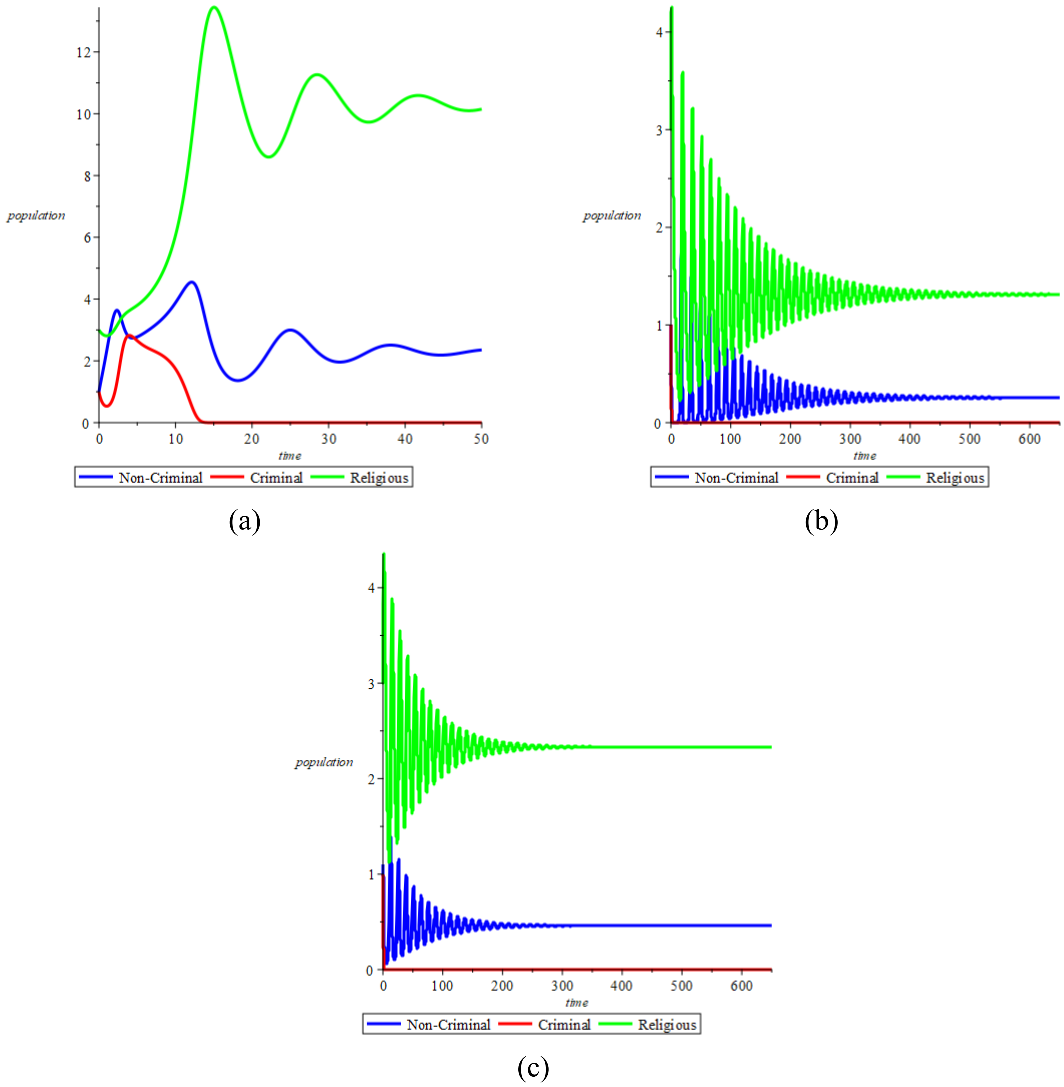


Figure 5. Population dynamics in the model with $K = 16$, $N_p(0) = C_p(0) = 1$, $R_p(0) = 3$ (a) $\alpha = 0.9$, $\beta = 0.1$, (b) $\alpha = 0.1$, $\beta = 0.9$, (c) $\alpha = 0.5$, $\beta = 0.5$

Table 4. The results of numerical simulation III ($K = 160$)

Parameter	Case A	Case B	Case C
α	0.9	0.1	0.5
β	0.1	0.9	0.5
Feasible equilibrium(s)	E_1, E_2, E_3, E_4, E_5	E_1, E_2, E_3, E_4	E_1, E_2, E_3, E_4
R_0	0.323	0.019	0.124
Stable equilibrium(s)	$E_3(1.327,0.756,0)$ and $E_4(2.3,0,11.828)$	$E_4(0.256,0,1.331)$	$E_4(0.460,0,2.393)$

CONCLUSION

Based on the results and discussion, two conclusions can be drawn. First, the value of the criminality rate and the rate of interaction between the non-criminal population and the religious population influence the dynamics within the system. If the criminality rate is greater than the rate of interaction with the religious population, the religious population may face extinction. Conversely, if the interaction rate with the religious population is greater than the criminality rate, then the criminal population may become extinct. Second, there are changes in the behavior of the solution of dynamic system as shown in the numerical simulations when the carrying capacity is increased. The changes of the solution's behavior as a parameter value varies indicate bifurcation in the system. However, bifurcation is not discussed in this study. Further studies could explore bifurcation occurring in models of social interaction between criminal and non-criminal behavior. In addition, a study mentioned the correlation of criminal acts with factors such as age, gender, and level of education [12]. This can certainly be used as the basis for developing models in future research.

REFERENCE

[1] Google Dictionary, "Crime Dictionary Definition," 06 December 2023. [Online]. Available: https://languages.oup.com/google-dictionary-en/. [Accessed 06 December 2023].

[2] J. Scott, "A Dictionary of Sociology," 06 December 2023. [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780199683581.001.0001/acref-9780199683581-e-441. [Accessed 06 December 2023].

[3] A. K. Misra, "Modeling the effect of police deterrence on the prevalence of crime in the society," *Applied Mathematics and Computation*, vol. 237, pp. 531-545, 2014.

[4] S. Abbas, J. P. Tripathi and A. A. Neha, "Dynamical analysis of a model of social behaviour: Criminal vs non-criminal population," *Chaos, Solitons and Fractals*, vol. 98, pp. 121-129, 2017.

[5] G. González-Parra, B. Chen-Charpentier and H. V. Kojouharov, "Mathematical modeling of crime as a social epidemic," *Journal of Interdisciplinary Mathematics*, vol. 21, no. 3, pp. 623-643, 2018.

[6] J. P. Tripathi, S. Bugalia, K. Burdak and S. Abbas, "Dynamical analysis and effect of law enforcement in a social interarction model," *Physica A*, vol. 567, p. 125725, 2021.

[7] M. Kumar and S. Abbas, "Modelling and prevention of crime using age-structure and law enforcement," *Journal of Mathematical Analysis and Applications*, vol. 519, no. 2, p. 126849, 2023.

[8] S. Arora, T. Mathur and K. Tiwari, "A fractional-order model to study the dynamics of the spread of crime," *Journal of Computational and Applied Mathematics*, vol. 426, p. 115102, 2023.

[9] Mulyadi, "Agama dan Pengaruhnya dalam Kehidupan," *Jurnal Tarbiyah Al-Awlad*, vol. 6, no. 2, pp. 556-564, 2016.

[10] K. Ihsan, "Faktor Penyebab Anak Melakukan Tindakan Kriminal (Studi Kasus Lembaga Pemasyarakatan Pekanbaru Kelas II B)," *JOM FISIP*, vol. 3, no. 2, pp. 1-15, 2016.

[11] A. S. Latifi, Suklani and Suteja, "Pengaruh Aktivitas Keagamaan Terhadap Perilaku Sosial Remaja Usia 13-18 Tahun di Blok 1 Desa Gembongan Induk Kecamatan Babakan Kabupaten Cirebon," *Al-Tarbawi Al-Haditsah: Jurnal Pendidikan Islam*, vol. 3, no. 1, pp. 154-169, 2018.

[12] W. Pramesti, "Analisis Korespondensi Untuk Mengetahui Keterkaitan Tindak Pidana Dengan Usia, Jenis Kelamin, Tingkat Pendidikan, Pekerjaan Dan Alasan Melakukan Tindak Pidana," *J-Statistika,* vol. 4, no. 2, pp. 41-54, 2012.

The Effect of Motivation, Work Discipline, and Organizational Commitment on Employee Performance at PT. Hebsa Indonesia

Afina Safarah (1), Endang Iryanti (2)

[1]Universitas Pembangunan Nasional "Veteran" Jawa Timur

Jl. Rungkut Madya No. 1, Gn. Anyar, kec. Gn. Anyar, Surabaya, Jawa Timur
e-mail: afinsafarah@gmail.com(1)

ABSTRAK

Sumber daya manusia adalah aset penting dalam sebuah organisasi, dikarenakan memiliki peranan subjek sebagai pelaksana kebijakan dan kegiatan operasional perusahaan. Perusahaan yang memiliki sumber daya manusia yang baik akan memberikan hasil optimal untuk kinerja yang diberikan kepada perusahaan. Sangat penting bagi perusahaan untuk mematangkan perencanaan yang telah dibuat sehingga dapat meningkatkan produktifitas kinerja karyawan. Dilakukannya penelitian ini bertujuan untuk mengetahui pengaruh motivasi, disiplin kerja, komitmen organisasi terhadap kinerja karyawan pada PT. Hebsa Indonesia dengan melibatkan seluruh karyawan yang berjumlah 35 karyawan. Teknik pengambilan sampel pada penelitian ini yaitu menggunakan teknik sampling jenuh. Metode penelitian yang digunakan pada penelitian ini yaitu dengan menggunakan metode kuantitatif. Teknik analisa yang digunakan dalam penelitian ini menggunakan bantuan program perangkat lunak SmartPLS 3 yaitu Partial Least Square (PLS). hasil penelitian ini menunjukkan bahwa motivasi tidak memiliki pengaruh terhadap kinerja karyawan PT. Hebsa Indonesia, sedangkan disiplin kerja dan komitmen organisasi memiliki pengaruh terhadap kinerja karyawan PT. Hebsa Indonesia.
Kata kunci: Komitmen Organisasi; Disiplin Kerja; Motivasi

ABSTRACT

Human resources are a crucial asset in an organization, as they play a key role in implementing policies and operational activities. Companies with good human resources will achieve optimal results in their performance. It is essential for organizations to refine their planning to enhance employee productivity. This research aims to examine the impact of motivation, work discipline, and organizational commitment on employee performance at PT. Hebsa Indonesia, involving all 35 employees. The sampling technique used in this study is saturated sampling. The research method employed is quantitative. Data analysis was conducted using SmartPLS 3 software with the Partial Least Square (PLS) approach. The results indicate that motivation does not have an effect on employee performance at PT. Hebsa Indonesia, while work discipline and organizational commitment do influence employee performance.
Keywords: *Work Discipline; Organizational Commitment; Motivation*

INTRODUCTION

Human resources are an important asset in an organization or company, because human resources have a role as the subject of implementing company policies and operational activities. An organization or company that has human resources with good performance can provide optimal results for a company. So it is very important for a company to finalize the plans that have been made so that it can increase employee performance productivity.

In managing, organizing and utilizing employees in a company so that they can function productively to achieve company goals, good human resources are needed. Company resources need to be managed professionally to create a balance between employee needs and the demands and capabilities of the company organization.

Therefore, to create good and efficient human resources, supporting factors are needed, including motivation. According to (Fajrina & Kustini, 2022), motivation is the urge to move someone or the desire to give all their energy for a goal. With motivation, employees can be inspired to do more in utilizing their energy and thoughts to achieve company goals. When these needs are met, satisfaction will arise and employee performance will increase.

Apart from motivation, work discipline is one of the factors that can influence employee performance. According to Siswanto (2001) in (Kartikasari & Irbayuni, 2021) Discipline is an attitude of respecting, obeying, following established rules, both written and unwritten, and being willing to carry them out and not avoiding sanctions.

Another factor that can influence employee performance is organizational commitment. According to Luhan (2006) in (Angraini, dkk, 2021) organizational commitment is an attitude that shows employee loyalty and is a person's ongoing process of expressing his concern for the success of the organization.

PT Hebsa Indonesia is a company that operates in the service sector, one of which provides planning and engineering consulting services, where the company must carry out orders according to what consumers want. Based on observations and interviews conducted by the author, obstacles were found related to employee performance levels that were less than optimal. This problem can be seen based on the PT target achievement data report. Hebsa Indonesia has experienced a decline over the last 4 years. Indicated due to a person's lack of motivation in doing work, this can happen if the facilities provided by the company to employees to support their work are still lacking. Apart from that, the rewards given are not what employees expect.

Apart from that, the decline in employee performance can be seen through the high percentage of employee absenteeism data at PT. Hebsa Indonesia in the last 4 years. This is related to work discipline problems of employees PT. Hebsa Indonesia is still lacking such as absence from work. The high number of employees who do not arrive on time and who complete tasks according to the deadline set by the company are not completed. This can cause delays in the work assigned, resulting in the company experiencing a decrease in turnover or the targets achieved cannot be met.

Apart from that, several things that make organizational commitment in a company not high enough are that employees do not feel proud to be part of the company and do not use all their abilities in their work, which can affect the resulting performance. Apart from that, there are still some employees who lack a sense of loyalty to the company.

This research uses SEM PLS with a PLS approach in forming a structural model that is applied to the case of employee performance, where endogenous latent variables are used is employee performance, and exogenous latent variables which influence the endogenous variables of motivation, work discipline and organizational commitment.

Structural Equetion Modeling (SEM) is a statistical analysis technique that has the ability to analyze relationship patterns between latent variables and their indicators. SEM with the Partial Least Square (PLS) approach is a powerful analysis method, because it does not require many assumptions and the sample size does not have to be large Wibisono, dkk (2015)

METHOD

The data collection method used in this research uses library research by collecting some information related to the problem in the research. This information was obtained through journals, articles, or other sources, both print and electronic. Apart from that, it also uses a data collection method which is carried out using a questionnaire which is distributed to 35 employees at PT. Hebsa Indonesia.

The variable measurement scale used in this research is the original scale or what is usually called a rating scale. Regarding the scoring technique in this research questionnaire using the Likert scale technique, according to [5] the Likert scale is a scale used to measure the attitudes, opinions and perceptions of a person or group of people about social phenomena.

There are five points in the response scale used:

- Score 1
- Score 2
- Score 3
- Score 4
- Score 5

The analysis technique used in this research uses partial least squares (PLS) with the help of SmartPLS 3 software. PLS is a more appropriate approach for prediction purposes, especially in conditions where the indicators are formative. With latent variables in the form of a linear combination of indicators, predictions of the values of the variables can be easily obtained, so that predictions of the latent variables that they influence can also be easily made [6].

The steps for PLS structural equation modeling with software are as follows:

- designing a structural model (inner model)
- designing a measurement model (outer model)
- construct path diagrams
- Convert path diagrams to systems of equations
- Estimation: Road coefficient, r loading, and weight
- Evaluation of goodness of fit
- Hypothesis test

RESULT AND DISCUSSION

Outer Loading

Assessment of the Outer model (measurement model) by looking at the outer loading factor, discriminant validity and composite reliability and construct.

Table 1 Outer Loading (Mean, STDEV, T-Values)

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)
X1.1 <- Motivation (X1)	0.700	0.624	0.253	2.766
X1.2 <- Motivation (X1)	0.892	0.877	0.205	4.354
X1.3 <- Motivation (X1)	0.828	0.781	0.183	4.523
X2.1 <- Work Discipline (X2)	0.728	0.710	0.129	5.643
X2.2 <- Work Discipline (X2)	0.903	0.891	0.104	8.660
X2.3 <- Work Discipline (X2)	0.814	0.804	0.086	9.420
X2.4 <- Work Discipline (X2)	0.884	0.875	0.080	11.072
X3.1 <- Organizational Commitment (X3)	0.740	0.743	0.161	4.607
X3.2 <- Organizational Commitment (X3)	0.852	0.809	0.186	4.585
X3.3 <- Organizational Commitment (X3)	0.836	0.774	0.238	3.515
Y.1 <- Employee Performance (Y)	0.782	0.734	0.173	4.511
Y.2 <- Employee Performance (Y)	0.784	0.741	0.205	3.831
Y.3 <- Employee Performance (Y)	0.825	0.797	0.182	4.527

From the table above, the validity of indicators is measured by looking at the Factor Loading Value of the variable to the indicator. It is said that the validity is sufficient if it is greater than 0.5 and/or the T-Statistic value is greater than 1.96 (Z value at $\alpha = 0.05$). Factor Loading is a correlation between an indicator and a variable. If it is greater than 0.5, it is considered that its validity is met. Likewise, if the T-Statistic value is greater than 1.96, then its significance is met.

Based on the outer loading table above, all reflective indicators on the variables Motivation (X1), Work Discipline (X2), Commitment (X3), and Employee Performance (Y), show factor loadings (original sample) greater than 0.50 and or significant (T-Statistic value more than Z value $\alpha = 0.05$

Table 2. Average Variance Extracted (AVE), and Composite Reliability, and R-Square

	AVE	Composite Reliability	R-Square
Work Disicpline (X2)	0,697403	0,901541	
Performance Employee (Y)	0,635982	0,839694	0,727829
Commitment (X3)	0,657270	0,851457	
Motivation (X1)	0,656857	0,850466	

The next measurement model is Average Variance Extracted (AVE), which shows how much indicator variance is contained in the latent variable. If the convergent AVE value is greater than

0.5, it indicates that there is good validity for the latent variable. For reflective indicators, the AVE value of each construct (variable) must be greater than 0.5 for the model to be considered good.

While Composite reliability is an index that shows the extent to which a measuring instrument can be trusted to be relied upon. Construct reliability testing is carried out by looking at the composite reliability value of each construct. Reliability refers to the consistency of a measuring instrument in assessing the same symptoms. A construct is declared reliable if it has a composite reliability value above 0.7 Fernanda (2022).

Based on the research results above, the AVE test results for the Motivation variable (X1) are 0.697403, the Work Discipline variable (X2) is 0.635982, the Commitment variable (X3) is 0.657270, and Employee Performance (Y) is 0.656857, These four variables show a value of more than 0.5, so overall the variables in this study can be said to have good validity.

While the Composite Reliability test results show that the Motivation variable (X1) is 0.850466 the Work Discipline variable (X2) is 0.901541, the Commitment variable (X3) is 0.851457, and Employee Performance (Y) is 0.839694, These four variables show Composite Reliability values above 0.70, so it can be said that all variables in this study are reliable.

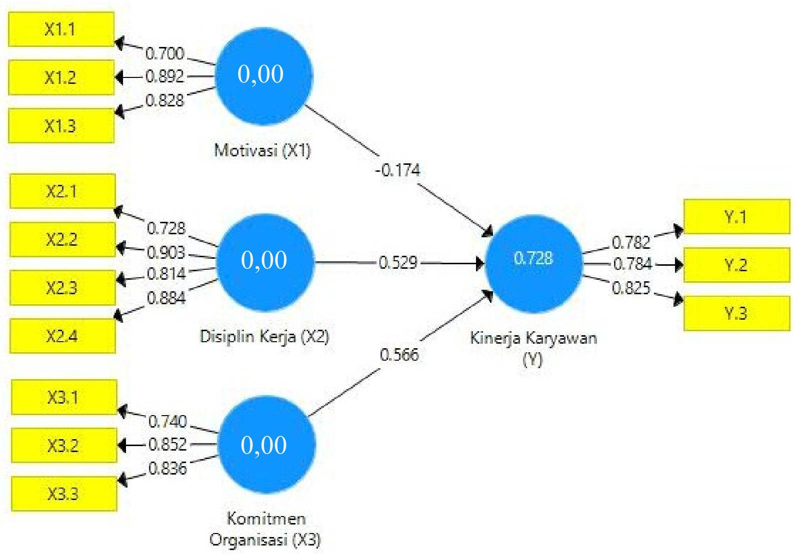


Figure 1 Outer Model with Factor Loading, Patch Coefficient dan R-Square

From the PLS output image above, you can see the magnitude of the factor loading value for each indicator which is located above the arrow between the variables and indicators, you can also see the magnitude of the path coefficients which are above the arrow line between the exogenous variables and the endogenous variables. Apart from that, you can also see the size of the R-Square which is right within the circle of the endogenous variable (Employee Performance variable).

Tabel 3 Path Coefficients (Mean, STDEV, T-Values)

	Path Coefficients (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STERR|)	P Values	Keterangan
Motivation (X1) -> Performance Employee(Y)	-0,174340	-0,160843	0,229092	0,761005	0,447	**Tidak Signifikan (Negatif)**
Work Discipline (X2)-> Performance Employee (Y)	0,565770	0,559437	0,130106	4,348521	0,012	**Signifikan (Positif)**
Commitment (X3) -> Work Employee (Y)	0,528891	0,519397	0,210203	2,516092	0,000	**Signifikan (Positif)**

From the table above it can be concluded that the hypothesis states:

- H1: It is suspected that motivation has a positive effect on the performance of PT employees. Hebsa Indonesia is unacceptable, with a path coefficient of -0.174340, and a T-statistic value of 0.761005 < 1.96 (T-table value of Zα = 0.05), or P-Value of 0.447 > 0.05 , with Insignificant (Negative) results.
- H2: It is suspected that work discipline influences the performance of PT employees. Hebsa Indonesia is acceptable, with a path coefficient of 0.565770, and a T-statistic value of 4.348521 > 1.96 (T-table value of Zα = 0.05), or P-Value of 0.012 < 0.05, with Significant (Positive) results.
- H3: It is suspected that organizational commitment influences the performance of PT employees. Hebsa Indonesia is acceptable, with a path coefficient of 0.528891, and a T-statistic value of 2.516092 > 1.96 (T-table value of Zα = 0.05), or P-Value of 0.012 < 0.05, with Significant (Positive) results.

DISCUSSION

1. The Effect of Motivation on Employee Performance

Based on the results of research that has been carried out, the results show that motivation does not have a positive effect on the performance of PT employees. Hebsa Indonesia so it does not match the proposed hypothesis. Strengthened by the results with a path coefficient of -0.174340, and a T-statistic value of 0.761005 < 1.96 (T-table value of Zα = 0.05), or P-Value of 0.447 > 0.05, with results Not Significant (Negative).

Based on descriptive analysis of motivation variables, the indicator with the highest percentage is good supervision conditions. Lack of good supervision in the form of direction and guidance when employees work is not optimal so it can reduce employee morale to improve their performance at PT. Hebsa Indonesia. Lack of good supervision can have a negative effect on employee productivity. On the other hand, if the company pays attention to good supervision by providing maximum work direction and guidance, employees may be motivated in carrying out their duties and responsibilities, which can have a positive impact on the performance of PT employees. Hebsa Indonesia.

This research is in line with research conducted by Hidayat (2021). The results obtained from this research indicate that motivation does not have a positive effect on employee performance. Apart from that, this research is also in line with research conducted by Sari et al (2020). The results of this research show that motivation does not have a positive effect on employee performance.

2. The Effect of Work Discipline on Employee Performance

Based on the results of the research that has been carried out, the results obtained are in accordance with the proposed hypothesis that work discipline has a positive effect on employee performance at PT. Hebsa Indonesia. Strengthened by the path coefficient result of 0.565770, and the T-statistic value of 4.348521 > 1.96 (T-table value of $Z\alpha = 0.05$), or P-Value 0.012 < 0.05, with significant results (Positive).

Based on descriptive analysis of work discipline variables, the indicator with the highest percentage is compliance with other regulations in the company. The attitude of complying with the rules regarding what can be done and what cannot be done, such as dressing and behaving in the workplace, reflects employee discipline in carrying out their duties, so that it can make it easier for the company to achieve the goals that have been set.

This research is in line with research conducted by Darmadi (2020) The results obtained from this research indicate that work discipline has a positive effect on employee performance. Apart from that, it is also strengthened by research conducted by Ircham & Iryanti (2022) and Hidayat (2020). The results obtained from this research show that work discipline has a positive effect on employee performance.

3. The Effect of Organizational Commitment on Employee Performance

Based on the results of the research that has been carried out, the results obtained are in accordance with the proposed hypothesis that organizational commitment has a positive effect on performance at PT. Hebsa Indonesia. Strengthened by a path coefficient of 0.528891, and a T-statistic value of 2.516092 > 1.96 (T-table value of $Z\alpha = 0.05$), or a P-Value of 0.012 < 0.05, with significant (positive) results.

Based on descriptive analysis regarding organizational commitment variables, the indicator with the highest percentage is employee willingness. Employees want to always give their best abilities at work so as to produce the best work results. This desire encourages employees to be active in operational activities, which ultimately helps them demonstrate professionalism in carrying out their duties at PT. Hebsa Indonesia.

This research is also in line with previous research conducted by Ircham & Iryanti (2022). The results obtained from this research are that organizational commitment has a significant effect on employee performance. Apart from that, it is also strengthened by research conducted by Frimayasa & Lawu (2020). The results of this research show that organizational commitment has a positive effect on employee performance.

CONCLUSION

Based on the results of tests carried out by researchers using PLS analysis related to the influence of motivation, work discipline, organizational commitment on the performance of PT employees. Hebsa Indonesia, can be concluded as follows:

1) Motivation does not have a significant effect on employee performance at PT. Hebsa Indonesia. Therefore, companies need to continue to strive to provide motivation to employees so that their performance is maximized so that it will benefit the company.
2) Work discipline has a significant or positive effect on the performance of PT employees. Hebsa Indonesia. By increasing employee discipline, maximum performance will be provided to the company.
3) Organizational commitment has a significant effect on employee performance at PT. Hebsa Indonesia. The higher the level of organizational commitment, the stronger the emotional bond between employees, which makes employees want to always give the best abilities they have, thereby potentially increasing employee performance.

REFERENCE

[1] A. N. Fajrina and Kustini, "Dampak Motivasi, Lingkungan Kerja Dan Kompensasi Terhadap Kepuasan Kerja PT. Puma Logistics International," *Nusant. J. Ilmu Pengetah. Sos.*, vol. 7, no. 2, pp. 408–420, 2022, [Online]. Available: http://jurnal.um-tapsel.ac.id/index.php/nusantara/index

[2] Divya Putri Kartikasari and S. Irbayuni, "Pengaruh Disiplin, Keselamatan dan Kesehatan Kerja (K3) Terhadap Kinerja Karyawan Produksi Divisi Extruder PT. X Sidoarjo," *J. Ilmu Adm. dan Manaj.*, vol. 4, no. 1, pp. 53–60, 2021.

[3] R. Angraini, A. Parawangi, and N. Mustari, "PENGARUH KOMITMEN ORGANISASI TERHADAP KINERJA PEGAWAI DI BADAN KEPEGAWAIAN DAN DIKLAT DAERAH KABUPATEN ENREKANG," 2021. [Online]. Available: https://journal.unismuh.ac.id/index.php/kimap/index

[4] A. Wibisono, M. Anwar, and I. Kirono, "Stuctural Equation Modeling Partial Least Square (Sem Pls) Untuk Mengetahui Kinerja Karyawan Pada Pt. Dempo Laser Metalindo Surabaya," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 7, no. 1, 2015, doi: 10.36456/jstat.vol7.no1.a184.

[5] Sugiyono, *Metodologi Penelitian Kuantitatif, Kualitatif dan R & D*. 2014.

[6] W. Abdillah and J. Hartono, "abdillah jogiyanto 2015.pdf." 2015.

[7] J. W. Fernanda, V. Luthifiana, and M. K. Akhyar, "Analisis Partial Least Square Structural Equation Model (PLS-SEM) untuk Pemodelan Penerimaan Sistem Jaringan Informasi Bersama Antar Sekolah (JIBAS)," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 292–297, 2022, doi: 10.36456/jstat.vol15.no2.a6436.

[8] R. Hidayat, "Pengaruh Motivasi, Kompetensi Dan Disiplin Kerja Terhadap Kinerja," *Widya Cipta J. Sekr. dan Manaj.*, vol. 5, no. 1, pp. 16–23, 2021, doi: 10.31294/widyacipta.v5i1.8838.

[9] A. Sari, F. Zamzam, and H. Syamsudin, "Pengaruh Kepemimpinan, Kompensasi, dan Motivasi terhadap Kinerja Karyawan," *J. Nas. Manaj. Pemasar. SDM*, vol. 1, no. 2, pp. 1–18, 2020, doi: 10.47747/jnmpsdm.v1i2.91.

[10] Darmadi, "Pengaruh Lingkungan Kerja dan Disiplin Kerja Terhadap Kinerja Karyawan Pada Indomaret Cabang Kelapa Dua Gading Serpong Kabupaten Tangerang PRODI MAGISTER MANAJEMEN & FORKAMMA UNPAM," *J. Ilm. Manaj. Forkamma)*, vol. 3, no. 3, 2020.

[11] M. A. Ircham and E. Iryanti, “Pengaruh Motivasi, Disiplin Kerja dan Komitmen Organisasi terhadap Kinerja Karyawan PT. Multipack Thamrin Jaya Pasuruan,” *SEIKO J. Manag. \&* ..., no. c, pp. 538–547, 2022, doi: 10.37531/sejaman.v5i2.2414.

[12] A. Frimayasa and S. H. Lawu, “PENGARUH KOMITMEN ORGANISASI DAN HUMAN CAPITAL TERHADAP KINERJA PADA KARYAWAN PT. FRISIAN FLAG,” 2020.

Statistical Quality Control (SQC) Method Analysis Regarding Qualty Control of Shoe Products (Case Study of PT-X)

Dona Rizqi Aulia (1), Taswati Nova Wijayaningrum (2)

[1,2]Faculty of Science and Technology, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang

Jl. Prof Dr. Hamka KM 01 Ngaliyan, Semarang Barat
e-mail: donarizqi0@gmail.com (1), taswati.nova@itesa.ac.id (2)

ABSTRAK

Industri sepatu berperan penting dalam perekonomian dengan produksi massal, menciptakan persaingan ketat. Perusahaan perlu merencanakan produksi secara matang untuk memenuhi permintaan pasar. Observasi menunjukkan adanya cacat produksi yang menurunkan kinerja dan menimbulkan kerugian, terutama dalam distribusi. Penelitian ini menggunakan Analisis P-Chart untuk memantau proporsi produk cacat, bertujuan mendeteksi kesalahan sebelum penjualan. Penerapan metode pengendalian kualitas melalui Statistical Quality Control (SQC) diusulkan. Diagram Pareto mengidentifikasi jenis kerusakan utama: over cementing (35,9%), open bonding (27,7%), overlay (24,1%), damage material metal (6,3%), dan damage material (6,1%). Analisis P-Chart menemukan titik di luar batas kendali pada data ke-15 dan ke-18, yang dihapus untuk stabilisasi sampel. Analisis fishbone membantu mengidentifikasi penyebab masalah. Over cementing, open bonding, dan overlay disebabkan oleh faktor manusia, material, dan mesin. Damage material dan metal material juga disebabkan oleh faktor manusia, material, dan mesin, serta faktor metode. Penelitian ini menunjukkan pentingnya penerapan kontrol kualitas untuk mengurangi cacat produksi dan meningkatkan kinerja perusahaan.
Kata kunci: Pengendalian Mutu; Metode Pengendalian Mutu Statistik (SQC)

ABSTRACT

The shoe industry plays an important role in the economy with mass production, creating fierce competition. Companies need to plan production carefully to meet market demand. Observations show that there are production defects that reduce performance and cause losses, especially in distribution. This research uses P-Chart Analysis to compare the proportion of defective products, aiming to detect errors before sale. The application of a quality control method through Statistical Quality Control (SQC) is proposed. The Pareto diagram identifies the main types of damage: over cementing (35.9%), open bonding (27.7%), overlay (24.1%), metal material damage (6.3%), and material damage (6.1 %). P-Chart analysis found points outside the control limits in the 15th and 18th data, which were removed for sample stabilization. Fishbone analysis helps identify the cause of the problem. Over cementing, open bonding, and overlay are caused by human, material, and machine factors. Damage to metal materials and materials is also caused by human, material and machine factors, as well as method factors. This research shows the importance of implementing quality control to reduce production defects and improve company performance.
Keywords: *Quality Control; Statistical Quality Control Method (SQC)*

INTRODUCTION

The shoes industry in Indonesia has experienced significant development. According to the Director General of Small and Medium Industries (IKM) of the Ministry of Industry, Euis Saedah, the shoe manufacturing industry is one of the sectors experiencing rapid growth in Indonesia. Currently, the shoe industry plays an important role in the country's economic growth by being able to produce shoes in large quantities. With so many shoe companies developing in Indonesia, competition between companies to win market share is increasing. Therefore, every company needs to plan their production carefully to meet market demand.

Production planning is the most important thing in the company, because in the production planning process the company will determine how many products they have to produce, on time for completion, the capacity of available resources ranging from workers, raw materials, to machine capacity so that market demand can be met appropriatel garety. In the production process, production planning and control will be carried out so that optimal production costs are obtained. To achieve company goals as well as production control, the aim is to utilize limited production resources appropriately, especially in an effort to meet consumer demand and create profits for the company [1].

According to [2]The success of a company does not only depend on the amount of income it earns, but is also built on efficient, effective and good processes to survive in increasingly tight business competition. In the face of intensive competition in the modern business world, companies need to increase their productivity by managing superior production systems, optimizing the use of resources, and improving the quality of their products. Continuously developing technology and rapid market dynamics in the manufacturing industry require that companies can meet consumer expectations with high quality products that comply with established standards. Factory operations can run efficiently and effectively if quality control is implemented well to reduce the number of defective products and ensure the achievement of the desired quality standards. Despite a surge in demand from consumers, competition in the market is not getting any easier. This is proven by the emergence of new factories with large production capacities. Therefore, factories must be able to produce high quality products in order to compete with similar companies in this increasingly fierce market.

Quality is a component that can become the company's basic model so that it can survive as a superior company and be able to compete in any era. The quality management system cannot be separated from the implementation of quality within the company. Product quality can be measured from its dimensions. (According to David A. Garvin in 1987), there were 8 (eight) dimensions of product quality including: Performance, Features, Reliability, Conformity, Durability, Ease of Service, Aesthetics and Perceived Quality [3].

Currently, competition in trade, both in industry and manufacturing, is very tight. Every company has quality standards that must be maintained to keep their products in demand by consumers. Apart from that, these companies must also continue to maintain and improve the quality of their products in order to compete with other competitors in the market.

According to [4]the crucial factor in producing quality products is part of the production process itself. Effective quality control is very important in efforts to improve product quality and reduce the number of defects. The causes of product defects vary, including human factors, machines, raw materials, work methods, and work environment. Therefore, to prevent product defects, increase customer satisfaction, and build trust, companies need to implement methods that support quality improvement. This will help the company to remain competitive in the highly competitive global market.

According to Feingenbaum in [5] quality control is a series of techniques and actions planned to achieve, maintain and improve the quality of products and services, so that they comply with predetermined standards and meet consumer satisfaction. The main objective of quality control is to ensure that the quality of the products or services produced is in accordance with established standards.Quality control measures must be carried out continuously and continuously. The quality control process can be carried out through the application of the PDCA (Plan – Do – Check – Act) cycle introduced by Dr. W. Edwards Deming, a well-known quality expert from the United States [6].

Based on observations made by researchers, there are defects that still occur in shoe production. After identification, shoes are classified into three grades: A-Grade are shoes that are ready to be sold with good quality, B-Grade have defects or damage to the material, and C-Grade are shoes that are not ready to be sold because they have significant defects or not. meet the standards. The presence of defective products has a direct impact on reducing company performance and causing losses, especially in the distribution process. To overcome this problem and reduce the level of product defects, this research proposes implementing a quality control method using *Statistical Quality Control* (SQC).

P-Chart is a tool that can be used for statistical process control. P- Chart was chosen to be used, because quality control is attribute in nature. The P-Chart shows changes in data over time, with the inclusion of maximum and minimum limits which are the boundaries of the control area. P-Chart has the advantage that it can help control packaging defects and can provide information about when and where companies need to make quality improvements [7]

The control method uses *Statistical Quality Control* (SQC) which is a system designed to maintain consistent quality standards for production results. By implementing quality control and using statistical methods, it is hoped that it can have a significant impact on the final quality of the product so that it meets company standards and is cost efficient. Production quality must be controlled and improved continuously. If the production method is not optimal, the company needs to carry out supervision or quality control which aims to ensure that the products produced comply with the established standards. P-Chart, or Proportion Chart, is a tool used in Statistical Quality Control (SQC) to monitor the proportion of defective units in a sample from the production process. P-Charts are very useful in situations where a product or service can be classified as defective or non-defective

P-Chart is usually used to describe the proportion of production that does not meet requirements. If you use data that varies or is different in size, then the upper control limit and lower control limit of the P-chart will not be flat. P-charts can also be used for data that has the same or not sample subgroup sizes. Calculations carried out in making the P control map include: calculation of defect proportion values, calculation of the Upper Control Limit (UCL) and Lower Control Limit (LCL). UCL and LCL are used to make it easier to monitor the quality produced, and to determine the quality produced in accordance with standards [8]

In this research, researchers used P-Chart Analysis to monitor the proportion or percentage of damaged products from finished production. The main goal is to detect errors before the product reaches the sales stage. Based on the problems above, the author is interested in conducting research on *Quality Control* with the title " ***Statistical Quality Control (SQC)*** **Method Analysis of Shoe Product Quality Control".**

METHOD

Data Source

This research uses secondary data obtained from PT – X, an industrial company that produces shoes in December 2023.

Research Variables

The variables used in this research are the results of rejected shoe production at Company X in the period December 2023. In processing the data obtained, PT. X identifies three types of damage to shoes, which are divided into three categories:

a. A-Grade Shoes: Shoes that pass QC checks (release).
b. B-Grade Shoes: Shoes that are defective in material, such as:
- Shoe production does not match the sample.
- Damage to materials or defects due to processing or exposure to machinery (material damage).
- Open bonding on the upper (the part of the shoe that covers the entire upper foot) and outsole (bottom of the shoe), as well as uneven application of upper and outsole glue (Open bonding).
- Too much glue on the upper part that exceeds 3mm (Over cement).
- Slope on shoes (Overlay).
- There is metal material in the shoes (Metal material).
- Defects in materials that can be repaired. If it cannot be repaired, the shoes will be sold at half price to the supplier itself.

c. C-Grad e Shoes: Shoes that do not meet the SOP, have fatal damage, and must be destroyed. PT. The Shoe making model at PT. X must be in accordance with the PO that has been determined or planned according to the PO.

Flow Diagram

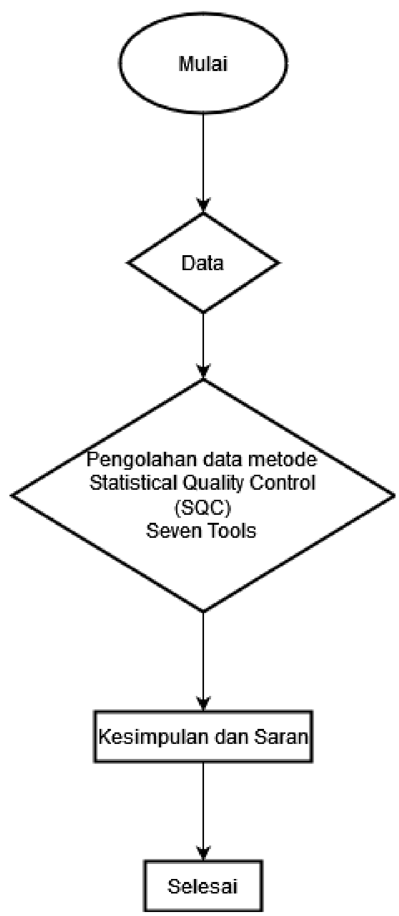


Figure 1. Research Flow Diagram

The flow diagram describes the data processing process using the Statistical Quality Control (SQC) method with Seven Tools. The process begins with collecting relevant data, followed by processing the data using seven main SQC tools: Check Sheet, Control Chart, Histogram, Pareto Chart, Cause-and-Effect Diagram, Scatter Diagram, and Flow Chart. After data processing is complete, the results of the analysis are used to draw conclusions and provide suggestions for improvement. The final stage of this process is completion, where all conclusions and recommendations are concluded and corrective actions are proposed to improve the quality of the process or product [9].

RESULT AND DISCUSSION

Based on the research above, there are several steps in carrying out quality control using the *Statistical Quality Control method*. The first step is to create and fill out a check sheet which is useful for simplifying the data collection process and for identifying problems that occur based on the type or cause.

Check Sheet

Table 1. Check Sheet

No	Date	Production	Type of Damage					Number of defects
			Open Boanding	Over Semen	Overlay	Damage Material	Metal Material	
1	1 Desember 2023	1200	25	37	25	7	5	99
2	4 Desember 2023	1190	40	35	30	12	7	124
3	5 Desember 2023	1130	34	35	32	8	12	121
4	6 Desember 2023	1105	37	47	20	10	6	120
5	7 Desember 2023	980	46	20	18	4	5	93
6	8 Desember 2023	1100	33	48	25	13	3	122
7	11 Desember 2023	950	36	40	28	8	10	122
8	12 Desember 2023	1080	34	30	20	5	8	97
9	13Desember 2023	1025	22	35	24	3	5	89
10	14 Desember 2023	1115	41	50	32	2	16	141
11	15 Desember 2023	970	27	42	28	13	2	112
12	18 Desember 2023	950	37	34	35	5	7	118
13	19 Desember 2023	1150	29	37	31	7	5	109
14	20 Desember 2023	1105	24	29	20	6	7	86
15	21 Desember 2023	1030	33	61	43	4	9	150
16	22 Desember 2023	940	13	40	30	6	11	100
17	26 Desember 2023	1200	19	45	32	8	8	112
18	27 Desember 2023	1215	27	46	25	3	4	105
19	28 Desember 2023	1130	45	59	31	8	6	149
20	29 Desember 2023	1160	33	52	22	7	8	122
Total		21725	635	822	551	139	144	2291

From the *check sheet table* aboved there are 5 types of defects or types of damage, consisting of 635 *Open Bonding units,* 822 *Over Cement units,* 551 *Overlay units,* 139 *Damage Materials, and* 144 *Metal Materials.*

Histogram

After the check sheet is created, the next step is to create a histogram. Histograms are useful for making it easier to see the types of damage that occur most frequently.

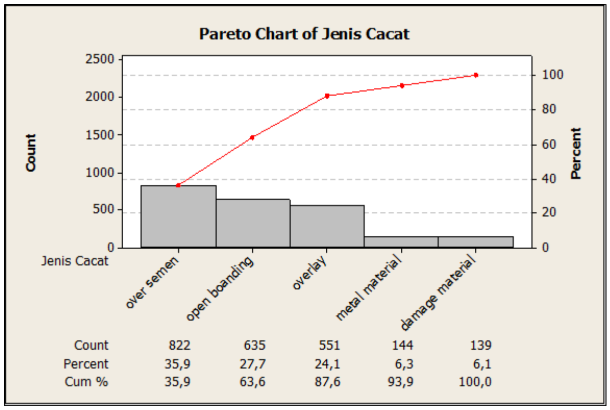


Figure 2. Pareto Diagram for Types of Shoe Product Damage

Based on the Pareto diagram above, it can be seen that the type of damage that occurs most frequently is Over Cement with a percentage reaching 35.9%. The second most frequently occurring damage is Open Bonding with a percentage of 27.7%. Furthermore, damage to Overlay reached 24.1%, damage to Metal Material was 6.3%, and damage to Damage Material was 6.1%. From this diagram, it can be concluded that Over Cement defects are the highest, so it is a major concern for the shoe industry to be careful about this type of defect.

Control Chart

A control chart is a graphic method used to evaluate whether a process or product is within statistical quality control limits or not. The goal is to monitor and control process variability so as to identify problems and produce necessary quality improvements [10]. The steps in creating a P control chart:

a. Calculating Damage Percentage

$$P = \frac{np}{n} \tag{1}$$

$$P = \frac{2290}{21725}$$

$$= 0.1054$$

b. Calculating the center line *(CL)*

$$CL = \overline{P} = \frac{\sum np}{\sum n} \tag{2}$$

$$= \frac{2290}{21725}$$

$$= 0.1054$$

c. Upper *control limit (UCL)*

$$P = \bar{P} + 3\sqrt{\frac{\bar{P}(1-\bar{p}}{n}} \tag{3}$$

$$= 0.1054 + 3\sqrt{\frac{0,1054\ (1-0,1054)}{21725}}$$

$$= 0.1054 + 0.01$$

$$= 0,1117$$

d. Calculating the lower control limit *Lower Control Limit (LCL)*

$$LCL = \bar{P} - 3\sqrt{\frac{\bar{p}\,(1-\bar{p}}{n}} \quad (4)$$

$$= 0.1054 - 3\sqrt{\frac{0{,}1054\;(1-0{,}1054)}{21725}}$$

$$= 0.1054 - 0.01$$

$$= 0.0992$$

Based on the calculations above, the complete P Control Map calculation results can be made for December 2023 as can be seen in the following table:

Table 2. Control Limits P with the help of Microsoft Excel

proporsi	CL	UCL	LCL
0,08	0,1054	0,1117	0,0992
0,10	0,1054	0,1117	0,0992
0,11	0,1054	0,1117	0,0992
0,11	0,1054	0,1117	0,0992
0,09	0,1054	0,1117	0,0992
0,11	0,1054	0,1117	0,0992
0,13	0,1054	0,1117	0,0992
0,09	0,1054	0,1117	0,0992
0,09	0,1054	0,1117	0,0992
0,13	0,1054	0,1117	0,0992
0,12	0,1054	0,1117	0,0992
0,12	0,1054	0,1117	0,0992
0,09	0,1054	0,1117	0,0992
0,08	0,1054	0,1117	0,0992
0,15	0,1054	0,1117	0,0992
0,11	0,1054	0,1117	0,0992
0,09	0,1054	0,1117	0,0992
0,09	0,1054	0,1117	0,0992
0,13	0,1054	0,1117	0,0992
0,11	0,1054	0,1117	0,0992

After calculating CL, UCL, and LCL in the table above, a P control chart (P-Chart) can be created using the help of Minitab which can be seen in the following image:

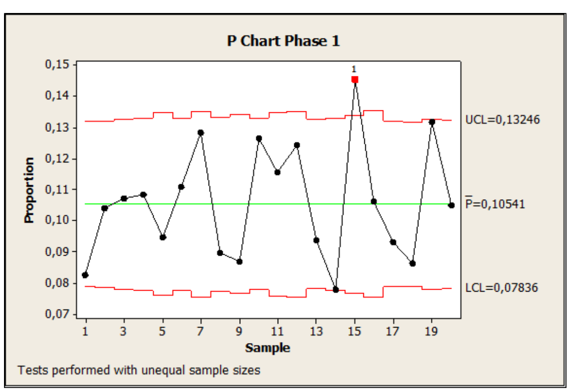


Figure 3. P Control Chart Graph (P – Chart)

From the graphic image above, it can be seen that at point 15, namely on December 21 2023, there is data that is outside the control limits. From the results of observations, the cause of the data leaving the control limits was due to a lack of accuracy on the part of employees and differences in the materials received, which were different from usual. Once the cause is known, the 15th data is deleted, and the updated control chart can be seen in the following image:

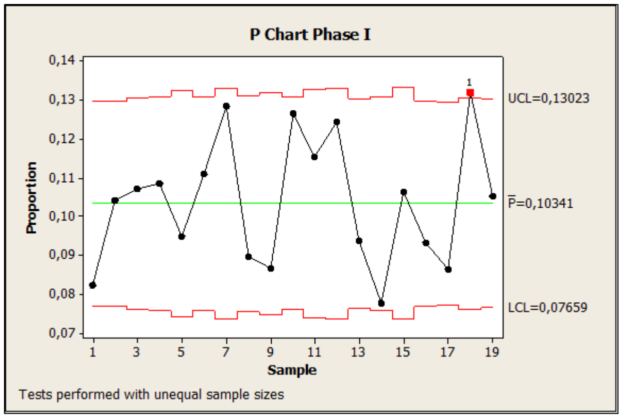


Figure 4. P Control Chart (P – Chart) After the 15th Data Deletion

From the graphic image above, it seems that there is still one data that exceeds the control limits. This happened at the 18th point on December 28, 2023. From the observations, it was discovered that the cause of the data leaving the control limits was because the operator was carrying out work in a hurry because he was tired and wanted to rest quickly. Therefore, the data at the 18th point is deleted, and the updated control chart can be seen in the following figure:

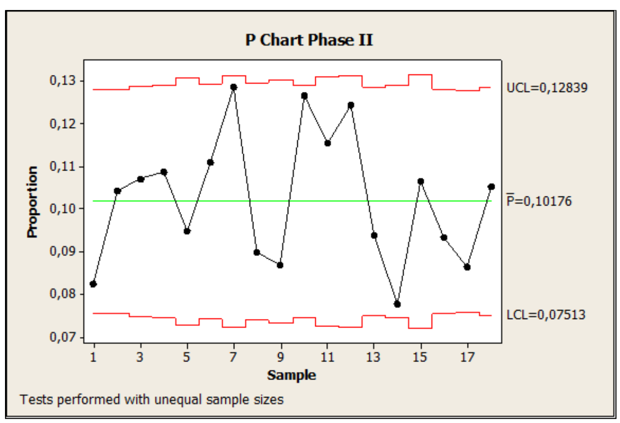


Figure 5. P Control Chart (P – Chart) After the 18th Data Deletion

From the graphic image above, it can be seen that there are no lines on the graph that exceed the control limits or *are out of control*, which means that all samples are within the accepted area. This shows that the sample has normal behavior or a stable condition.

Cause and Effect Diagram

The use of cause-and-effect diagrams is to help solve the problems faced by linking the causes and the factors that influence them. In a company, of course, there are applicable SOPs

(Standard Operational Procedures) , which will determine whether the implementation has been effective or not. This will be proven by a cause and effect diagram caused by several types of failure or defective products including those caused by *Over Lay* (slope), *Damage Material, Metal Material* (metal material), *Over Cementing* (lots of glue), and *Open Boanding* (bonding). open).

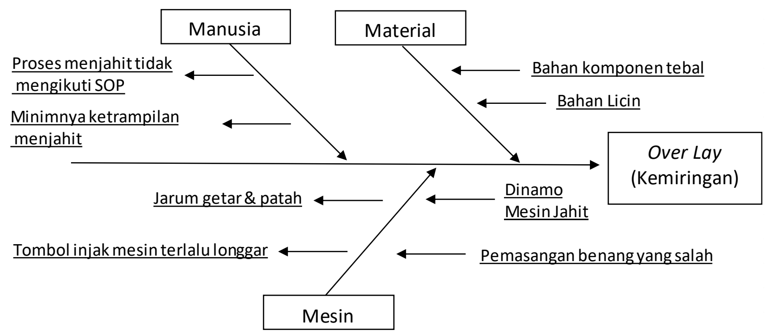


Figure 6. Fishbone (cause and effect) Diagram of Shoe Production Based on Over Lay

It can be seen in Figure 6 Fishbone Diagram for Shoe Production that there are three factors that cause defective products, namely human factors, material factors and machine factors.

1. Human Factors
   It was found that labor factors that cause defective products in shoe production include sewing processes that do not follow SOPs, lack of sewing skills, operators being in a hurry and not being careful.
2. Material Factors
   Problems were found with the component materials that arrived, because the materials that arrived were not as usual, such as thick materials and slippery materials, making it difficult for operators to sew.
3. Machine Factor
   Problems were found in machine maintenance that were not checked regularly and on a schedule, resulting in decreased performance of the machine, such as sewing machine dynamos, needles vibrating and breaking, machine push buttons that were too loose, and incorrect thread installation.

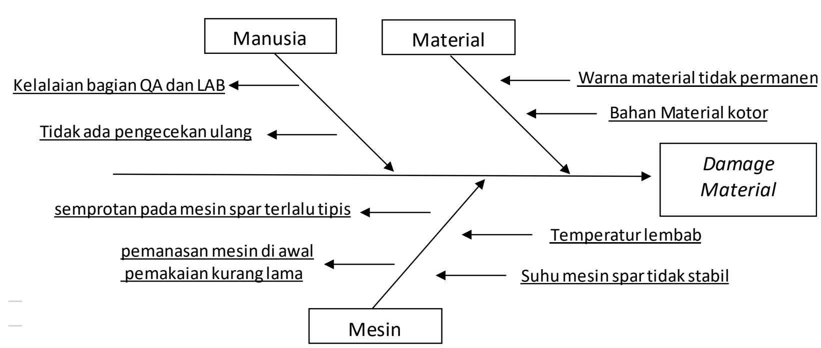


Figure 7. Fishbone (cause and effect) Diagram of Shoe Production Based on Damage Material

It can be seen in Figure 7 Shoe Production Fishbone Diagram that there are three factors that cause defective products, namely human factors, material factors and machine factors.

1. Human Factors
   It was found that there was negligence on the part of QA and LAB and there was no re-checking of the goods used.
2. Material Factors
   Problems were found with dirty materials and non-permanent material colors
3. Machine Factor
   Problems were found in engine maintenance that were not checked, such as damp temperatures, unstable spar engine temperatures, engine heating at the start of use that was not long enough, and the spray on the spar engine was too thin.

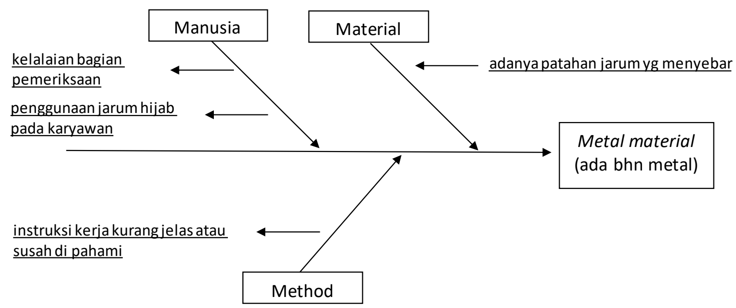


Figure 8. Fishbone (cause and effect) Diagram of Shoe Production Based on Metal Material

It can be seen in Figure 8 Shoe Production Fishbone Diagram that there are three factors that cause defective products, namely human factors, material factors and method factors.

1. Human Factors
   Factors found were negligence on the part of the inspection, and the use of hijab needles on employees.
2. Material Factors
   The problem was discovered because there was a widespread needle fracture.
3. Method Factor
   Problems found, namely work instructions that are unclear or difficult to understand.

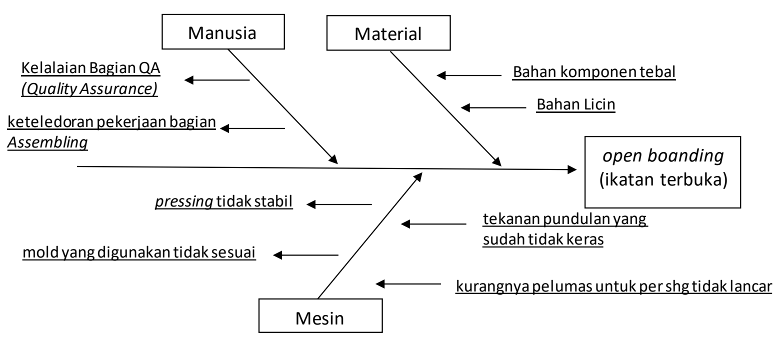


Figure 9. Fishbone (cause and effect) Diagram of Shoe Production Based on Open Boanding

It can be seen in Figure 9 Shoe Production Fishbone Diagram that there are three factors that cause defective products, namely human factors, material factors and machine factors.

1. Human Factors
   Factors found were negligence in the QA (Quality Assurance) department, and negligence in the work of the Assembling department.

2. Material Factors
   Problems were found with the materials, namely thick component materials and slippery materials.
3. Machine Factor
   The problems found were that the pressing machine was unstable, the mold used was not suitable, there was a lack of lubricant for the spring so it did not run smoothly, and the pressure on the bending was no longer firm.

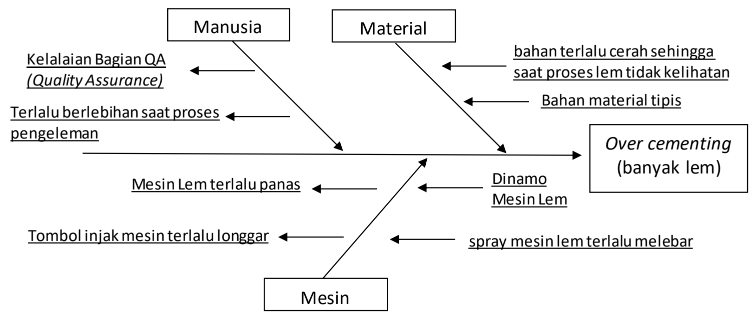


Figure 10. Fishbone (cause and effect) Diagram of Shoe Production Based on Over Cementing

It can be seen in Figure 10 Shoe Production Fishbone Diagram that there are three factors that cause defective products, namely human factors, material factors and machine factors.

1. Human Factors
   Factors found were negligence by the QA *(Quality Assurance) department* , and operators who were too excessive during the gluing process.
2. Material Factors
   Problems were found with the material, namely the material was too bright so that during the gluing process it exceeded the limit, and the material was thin.
3. Machine Factor
   The problems found were that the glue machine was too hot, the engine push button was too loose, the glue machine dynamo was not double checked, and the glue machine spray was too wide, causing the glue to spread everywhere.

CONCLUSION

Based on the data that has been analyzed along with the discussion described in the previous chapter regarding the results of the shoe production process at PT – X, the following conclusions can be drawn:

1. Based on the Pareto diagram, it can be concluded that the type of damage that often occurs in the shoe production process is *over cement* with a percentage reaching 35.9%, then there is *open boanding* reaching 27.7%, then there is *overlay* with a total damage reaching 24.1%. Next, *metal material* damage was 6.3%, and *material damage* was 6.1%.
2. From the P-Chart analysis it can be concluded that there are points that are outside the control limits (Out of Control), once the cause is known then the data that causes the out of control is recorded. Remove it and process it again using Minitab. Once it is known that there are no lines on the graph that exceed the control limits or are out of control , this means that all samples are in the accepted area. This shows that the sample has normal behavior or a stable condition.

ACKNOWLEDGMENTS

In this research, we would like to thank the Muhammadiyah Semarang Institute of Statistics and Business Technology for the support and resources provided. We would also like to thank PT - X for the publication of data and datasets which greatly enriched our research results. The support and cooperation of all parties is very valuable for the success of this research.

REFERENCE

[1] V. Kartikasari, "Perencanaan produksi menggunakan metode goal programming di PT. X," 2017.

[2] K. Husein and R. Rochmoeljati, "Meminimasi Cacat Produk Bogie Tipe S2E-9C Menggunakan Metode Statistical Quality Control (Sqc) Dan Failure Mode Effect Analysis (Fmea) Pada Pt Xyz," *Juminten*, vol. 2, no. 2, pp. 168–179, 2021, doi: 10.33005/juminten.v2i2.250.

[3] E. P. Saputra, "Perencanaan Peningkatan Kualitas Pada Sepatu Model Junior Di Departemen Assembling (Studi Kasus: Pt. Primarindo Asia Infrastructre, Tbk)," 2019. [Online]. Available: http://repository.unpas.ac.id/id/eprint/41117

[4] A. Y. Bagaskoro, M. Yusuf, and P. Wisnubroto, "Analisis Faktor Penyebab produk Cacat Pakaian Dengan Statistical Quality Control dan Failure Mode and Effect Analysis (FMEA) di CV. Yusuf & CO," *J. Rekavasi*, vol. 8, no. 1, pp. 44–51, 2020.

[5] N. A. R. Yurin Febria Suci, Yuki Novia Nasution, "Penggunaan Metode Seven New Quality Tools dan Metode DMAIC Six Sigma Pada Penerapan Pengendalian Kualitas Produk (Studi Kasus : Roti Durian Panglima Produksi PT. Panglima Roqiiqu Group Samarinda)," *Matriks Tek. Sipil*, vol. 17, no. 1, pp. 314–319, 2017.

[6] M. N. Nasution, *MANAJEMEN MUTU TERPADU*. 2015.

[7] E. Muslimah and T. Keriswanto, "Pengendalian Kualitas Kain Denim Dt 650 Pada Departemen Weaving Menggunakan P-Chart," *Simp. Nas. RAPI*, pp. 167–171, 2010, [Online]. Available: https://publikasiilmiah.ums.ac.id/handle/11617/6599

[8] Ahadya Silka Fajaranie and A. N. Khairi, "Pengamatan Cacat Kemasan Pada Produk Mie Kering Menggunakan Peta Kendali Dan Diagram Fishbone Di Perusahaan Produsen Mie Kering Semarang, Jawa Tengah," *J. Pengolah. Pangan*, vol. 7, no. 1, pp. 7–13, 2022, doi: 10.31970/pangan.v7i1.69.

[9] D. W. Astriyani, T. Aspiranti, and N. Koesdiningsih, "Prosiding Manajemen ISSN: 2460-6545," *Hery*, vol. 2014, no. 3, pp. 278–284, 2012.

[10] M. E. Setiabudi, P. Vitasari, and T. Priyasmanu, "Analisis Pengendalian Kualitas Untuk Menurunkan Jumlah Produk Cacat Dengan Metode Statistical Quality Control Pada Umkm. Waris Shoes," *J. Valtech*, vol. 3, no. 2, pp. 211–218, 2020, [Online]. Available: https://ejournal.itn.ac.id/index.php/valtech/article/view/2734

Distribution of Soap X in East Java Region with Bhumal Method and Traveling Salesman Problem

Hery Murnawan (1), Surani (2), Dony Mahardhika (3)

[1]Universitas 17 Agustus 1945 Surabaya

Jalan Semolowaru 45, Surabaya 60118, East Java, Indonesia

e-mail: herymurnawan@untag-sby.ac.id (1), surani1379@gmail.com (2), dhic.elc@gmail.com (3)

ABSTRAK

Proses pendistribusian produk merupakan suatu upaya produsen untuk menyalurkan produknya kepada konsumen dengan sistem yang telah terdaftar dan terprogram. Pendistribusian produk kepada konsumen harus dilakukan secara cermat dan hati-hati. Sebab hal tersebut dapat menjadi salah satu kendala bagi perusahaan dalam memenangkan persaingan dengan perusahaan lain apabila kecepatan pendistribusiannya tidak tinggi. Cepat atau lambatnya pendistribusian produk kepada konsumen salah satunya bergantung pada kedekatan antara gudang penyalur dengan pasar (konsumen). Tujuan dari penyajian dan analisis data ini adalah untuk lebih memahami bagaimana cara membuat data, menganalisis, dan membuat simpulan yang dapat menentukan suatu keputusan terhadap suatu proses logistik secara umum dan produk sabun batangan dengan optimasi biaya transportasi menggunakan metode bhumal dan Traveling Salesman Problem.

Kata kunci: Logistik; Metode Bhumal; Penjadwalan

ABSTRACT

The product distribution process is an effort by producers to deliver their products to consumers with a registered and programmed system. Product distribution to consumers must be handled carefully and carefully. Because this can be one of the obstacles for companies in winning competition with other companies if the distribution speed is not high. The speed or slowness of product distribution to consumers, one of which depends on the proximity between the dealer's warehouse and the market (consumers). The purpose of this data presentation and analysis is to better understand how to make data, analyse, and make conclusions that can determine a decision on a logistics process in general and bar soap products with transportation cost optimisation using the bhumal method and Traveling Salesman Problem.

Keywords: *Logistics; Method Bhumal; Scheduling*

INTRODUCTION

Distribution of soap, a seemingly mundane process, encompasses a sophisticated and multifaceted network involving manufacturing, logistics, marketing, and sales. This process begins at the manufacturing facilities where raw materials are transformed into finished soap products through a series of chemical reactions and mechanical processes. Once production is complete, the soaps are packaged and prepared for distribution. The journey from the factory to the consumer's hands is a complex orchestration of activities requiring precision and efficiency to ensure that the product maintains its quality and integrity [1].

Logistics play a critical role in the distribution of soap. This involves the coordination of transportation, warehousing, and inventory management. Efficient logistics ensure that the soap is delivered from the manufacturer to various distribution centers, retailers, and ultimately to consumers [2]. Advanced logistics systems utilize sophisticated software and technology to track shipments, manage inventory levels, and optimize routes for delivery trucks. This not only reduces costs but also minimizes the environmental impact by reducing fuel consumption and emissions.

Marketing strategies are also crucial in the distribution of soap. Companies invest heavily in advertising and promotional activities to create brand awareness and drive consumer demand [3]. This includes traditional marketing channels such as television and print advertisements, as well as digital marketing strategies like social media campaigns and influencer partnerships [4]. Effective marketing ensures that consumers are aware of the soap brands available in the market and are persuaded to purchase them. Additionally, packaging design and product placement in stores are key elements that influence consumer purchasing decisions [5].

Sales channels for soap distribution can vary widely, including supermarkets, convenience stores, online platforms, and specialty shops. Each of these channels has unique characteristics and requires different approaches for effective distribution [6]. For instance, supermarkets may demand large volumes of product with specific shelf placement requirements, while online platforms require efficient order fulfillment and delivery systems. Understanding the nuances of each sales channel is essential for companies to successfully distribute their soap products and meet consumer demand [7].

Finally, the distribution of soap is influenced by regulatory and environmental considerations. Manufacturers and distributors must adhere to regulations governing product safety, labeling, and environmental impact. This includes compliance with local and international standards for ingredients, packaging, and waste disposal. Sustainable practices are increasingly important as consumers become more environmentally conscious. Companies are adopting eco-friendly packaging, reducing plastic use, and implementing recycling programs to minimize their environmental footprint. Balancing these regulatory and environmental factors with the need for efficient and cost-effective distribution is a complex but necessary aspect of the soap distribution process [8].

To determine efficient and effective distribution methods in the distribution process, it is essential to conduct a comprehensive analysis of various logistical factors [9]. This begins with assessing the current distribution network, including transportation routes, warehousing locations, and inventory management practices. By employing advanced analytical tools such as Geographic Information Systems (GIS) and supply chain modeling software, businesses can identify bottlenecks, optimize delivery routes, and strategically position warehouses to minimize transit times and costs. Additionally, leveraging data analytics to predict demand patterns allows for more accurate inventory forecasting, reducing the risk of overstocking or stockouts, thereby enhancing overall efficiency [10], [11].

Equally crucial is the integration of modern technology and automation in distribution processes. Implementing technologies such as Radio Frequency Identification (RFID) for real-time tracking, automated storage and retrieval systems (AS/RS), and advanced robotics for order fulfillment can significantly streamline operations. [12], [13] These technologies not only improve accuracy and speed but also reduce labor costs and minimize human error. Furthermore, adopting a multi-channel distribution strategy that encompasses both traditional retail and e-commerce platforms ensures a broader market reach and meets diverse consumer preferences. Regular performance monitoring through key performance indicators (KPIs) such as delivery lead times, order accuracy rates, and customer satisfaction levels is imperative for continuous improvement and maintaining an effective distribution network.

METHOD

Bhumal Method

The Bhumal Method, as depicted in the image, appears to be a specialized approach or framework used within a specific context, likely within the fields of logistics, distribution, or a related area. While the exact details of the method are not discernible from the image alone, such methods typically encompass a structured set of principles or steps designed to optimize certain processes. For instance, in logistics and distribution, a method like this might focus on streamlining supply chain operations, enhancing efficiency, and reducing costs through systematic analysis and strategic planning[8].

To fully understand and implement the Bhumal Method, one would need to delve into its specific components and steps, which could involve various analytical techniques, optimization models, and best practices tailored to the particular industry or application it addresses. These might include aspects such as demand forecasting, inventory management, transportation logistics, and performance metrics. By adhering to the structured guidelines of the Bhumal Method, organizations can achieve more efficient and effective distribution strategies, ultimately leading to improved operational performance and competitive advantage in their respective markets.

The Traveling Salesman Problem (TSP)

The Traveling Salesman Problem (TSP) is a classic optimization problem in the field of combinatorial optimization and operations research. It involves finding the shortest possible route for a salesman to visit a given set of cities exactly once and return to the starting city. Despite its seemingly straightforward premise, TSP is notoriously complex because the number of possible routes increases factorially with the number of cities, making it computationally challenging to solve as the problem size grows. The TSP is classified as an NP-hard problem, meaning that no efficient solution algorithm is known that can solve all instances of the problem optimally in polynomial time.

Solutions to the TSP have significant practical implications in various domains such as logistics, manufacturing, and transportation planning. Heuristic and approximation algorithms, such as the nearest neighbor, genetic algorithms, and simulated annealing, are often employed to find near-optimal solutions within a reasonable timeframe. Exact methods like branch and bound, dynamic programming, and integer linear programming can provide optimal solutions for smaller instances. Advances in computational techniques and increased computational power continue to push the boundaries of what can be solved in practical scenarios, but the TSP remains a benchmark problem for testing new optimization algorithms and methodologies in applied mathematics and computer science.

RESULT AND DISCUSSION

Retail location

Table 1. Retail locations

No	Retail locations	Demand Retail
1	Blitar City	274.889
2	Blitar Regency	26.102
3	Kediri Regency	333.490
4	Kediri City	37.809
5	Malang Regency	756.889
6	Malang City	127.077
7	Lumajang Regency	217.980
8	Probolinggo City	256.049
9	Probolinggo Regency	9.807
10	Pasuruan City	325.206
11	Pasuruan Regency	12.807
12	Sidoarjo Regency	718.960
13	Mojokerto City	182.643
14	Mojokerto Regency	5.417
15	Jombang Regency	255.718
16	Nganjuk Regency	191.122
17	Madiun City	61.211
18	Madiun Regency	6.044
19	Ngawi Regency	79.325
20	Lamongan Regency	416.082
21	Gresik Regency	242.037
22	Surabaya City	839.172

Warehouse locations

Table 2. Warehouse locations

No	Warehouse locations	Address
1	Probolinggo City	Anggrek No.11, Pilang, Kec. Kademangan, Kota Probolinggo, Jawa Timur 67221
2	Sidoarjo Regency	Lingkar Timur No.4, Prasungtani, Prasung, Kec. Buduran, Kabupaten Sidoarjo, Jawa Timur 61252
3	Kediri City	Kertosono-Kediri No.134, Putih, Gampengrejo, Kediri Regency, East Java 64182
4	Madiun City	Basuki Rahmad No.91, Tawangrejo, Kec. Kartoharjo, Kota Madiun, Jawa Timur 63123

Shipping Cost

For delivery to Retail, Colt Diesel Double (CDD) is a small capacity truck with six tyres. Four rear tyres and two front tyres with a maximum capacity of four tonnes. CDD trucks are a type of delivery truck that is divided into several types, namely: CDD box standard is up to four metres long with a width and height of up to two metres. The carrying capacity of the CDD Box Standard truck itself is up to 4 tonnes. This CDD Box Standard truck also has a protective box made of aluminium with a volume of 16 cbm. The rental price is IDR 3,000,000 for a distance of 359 km or IDR 8,400/km.

Table 3. Distance warehouse to retail

Retail	1	2	3	4	5	6	7	8	9	10	11
Total Requirement (1=96 pcs)	2863	272	3474	394	7884	1324	2271	2667	102	3388	133
Shipping Cost	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400
Distance WH1	168,0	156,0	159,0	179,0	101,0	92,0	51,0	3,2	5,0	35,0	38,0
Distance WH2	145,0	144,0	120,0	114	93,0	79,0	130,0	83,0	82,0	84,0	45,0
Distance WH3	51,0	45,0	29,0	7,6	105,0	106,0	214,0	167,0	166,0	168,0	129,0
Distance WH4	125,0	111,0	103,0	81	178,0	171,0	261,0	215,0	213,0	216,0	176,0

Retail	12	13	14	15	16	17	18	19	20	21	22
Total Requirement (1=96 pcs)	7489	1903	56	2664	1991	638	63	826	4334	2521	8741
Shipping Cost	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400	Rp8.400
Distance WH1	79,0	100,0	100,0	125,0	169,0	218,0	216,0	231,0	139,0	113,0	92,0
Distance WH2	5,0	39	45,0	71,0	113,0	163,0	160,0	175,0	63,0	37,0	16,0
Distance WH3	109,0	69	66,0	42,0	35,0	82,0	80,0	95,0	106,0	125,0	115,0
Distance WH4	156,0	120	113,0	89,0	45,0	6,0	4,1	32,0	150,0	169,0	162,0

Factory Transport Costs to Retail

Delivery of products to the Warehouse is planned using a Wingbox Truck. Dimensions of the box: Length: 950cm, Width: 235cm, Height: 235cm and Carrying capacity: 50 cubic meters. SMB Trans provides rental price starting from IDR 2,500,000 for Surabaya-Malang (24,000/km).

Table 4. Factory Transport Costs to Retail

Warehouse	1	2	3	4
Distance	90 km	15 km	113 km	161 km
Factory	Rp 108	Rp 18	Rp 135	Rp 193

Warehouse to Retail Transportation Cost

Table 5. Warehouse to Retail Transportation Cost

	1	2	3	4	5	6	7	8	9	10	11
WH 1	492,8	4819,5	384,5	3817,8	107,6	583,8	188,7	10,1	411,1	86,8	2392,7
WH 2	425,4	4448,8	290,2	2431,4	99,1	501,3	480,9	261,4	6742,6	208,3	2833,5
WH 3	149,6	1390,2	70,1	162,1	111,9	672,6	791,7	525,9	13649,7	416,6	8122,6
WH 4	366,7	3429,3	249,1	1727,6	189,6	1085,1	965,6	677,1	17514,4	535,6	11081,9

	12	13	14	15	16	17	18	19	20	21	22
WH 1	88,6	441,5	14887,7	394,2	713,1	2872,0	28820,0	2348,3	269,4	376,5	88,4
WH 2	5,6	172,2	6699,5	223,9	476,8	2147,4	21348,2	1779,0	122,1	123,3	15,4
WH 3	122,3	304,6	9825,9	132,4	147,7	1080,3	10674,1	965,7	205,4	416,5	110,5
WH 4	175,0	529,8	16823,1	280,7	189,9	79,0	547,0	325,3	290,7	563,1	155,7

Distribution Cost Optimisation Analysis Using the Traveling Salesman Problem Method

From the results of the bhumal method allocation, the number of warehouses that must be served by the factory and the number of retailers that each warehouse must serve by saving using the Traveling Salesman Problem method.

1. **Fuso Box**
   Delivery of products to the Warehouse is planned using a Wingbox Truck.
   Dimensions of the box:
   Length: 660 cm
   Width: 235 cm
   Height: 235 cm Carrying capacity: 12 tonnes
   Cardboard dimensions:
   Length: 34 cm --- 600 : 34 = 17,6
   Width: 17 cm --- 235 : 17 = 13,8
   Height: 15 cm --- 235 : 15 = 15,67
   1 Cardboard: 10.56 kg

Number of boxes that can be transported 3,315 boxes
Total transport capacity 3,315 x 10.56 = 35,005 kg (35 tonnes)
Because the capacity exceeds the carrying capacity, the number of boxes transported is 1136 boxes with a load of (11,996 kg / 12 tonnes).

2. Truck CDD Long

Delivery of products to the Warehouse is planned using a Wingbox Truck.
Dimensions of the box:
Length: 50 cm
Width: 200 cm
Height: 210 cm Carrying capacity: 5.2 tonnes
Cardboard dimensions:
Length: 34 cm --- 530 : 34 = 15,5
Width: 17 cm --- 200 : 17 = 11,7
Height: 15 cm --- 210 : 15 = 14
1 Cardboard: 10.56 kg
Number of boxes that can be transported 2310 boxes
Total transport capacity 2310 x 10.56 = 24,303.2 kg (23 tonnes)
Because the capacity exceeds the transport capacity, the number of boxes transported is 492 boxes with a load of (5,200 kg / 5.2 tonnes).

3. CDE Box

Delivery of products to the Warehouse is planned using a Wingbox Truck.
Dimensions of the box: Length: 320cm
Width: 170cm
Height: 170cm Carrying capacity: 2500
Pallet Dimensions:
Length: 34 cm --- 320 : 34 = 9,4
Width: 17 cm --- 170 : 17 = 10
Height: 15 cm --- 170 : 15 = 11,3
1 Cardboard: 10.56 kg
Total transport capacity 1,062 x 10.56 = 11,216 kg (12 tonnes)
Because the capacity exceeds the carrying capacity, the number of boxes transported is 236 boxes with a load of (2,492 kg / 2.4 tonnes).

CONCLUSION

In soap distribution, the Bhumal Method offers a structured framework to optimize supply chain efficiency by leveraging strategic planning and advanced analytical techniques for inventory management, demand forecasting, and transportation logistics. When integrated with the Traveling Salesman Problem (TSP), it further refines distribution by determining the most efficient delivery routes to minimize travel distance and time. The TSP, renowned for its complexity, addresses the challenge of visiting multiple locations in the shortest possible path, ensuring cost-effective and timely deliveries. Together, these methodologies enhance operational performance, reduce costs, and promote sustainability, thereby ensuring optimal product availability and superior customer satisfaction in the competitive soap market.

REFERENCE

[1] P. A. Pangastuti and H. Murnawan, "Optimalisasi Supply Chain Management Guna Mengatasi Tantangan Persediaan Pada Industri Tape Di Raja Tape Bondowoso," *J. Serambi Eng.*, vol. 9, no. 2, pp. 8983–8991, 2024.

[2] N. F. Aryansah and H. Murnawan, "Usulan Perancangan Tata Letak Gudang dengan Menggunakan Metode Class Based Storage," *J. Tek. Ind. Terintegrasi*, vol. 7, no. 2, pp. 1018–1026, 2024.

[3] A. Ismiyah and P. E. D. K. Wati, "Perancangan Tata Letak Gudang pada Distriubtor Unicharm Menggunakan Metode Class Based Storage," *J. Surya Tek.*, vol. 11, no. 1, 2024.

[4] E. A. Utomo and P. E. D. K. Wati, "Optimasi Rute Distribusi Biskuit Ubm Menggunakan Metode Saving Matrix untuk Meminimasi Biaya Distribusi pada PT. SJA," *J. Tek. Ind. Terintegrasi*, vol. 7, no. 2, pp. 1185–1194, 2024.

[5] W. W. Pradana and W. Widiasih, "Penjadwalan Waktu Preventive Maintenance untuk Meningkatkan Kinerja Mesin Frais dan Bubut di PT. Isumi," *J. Ind. View*, vol. 5, no. 1, pp. 1–11, 2023.

[6] S. M. Khoiroh, "Pengembangan Algoritma Non Delay Pada Kasus Penjadwalan Non-Permutation Hybrid Flowshop Untuk Minimasi Mean Flowtime," *J. Tek. Ind.*, vol. 19, no. 2, pp. 148–156, 2018.

[7] A. N. Rao, D. E. Johnson, B. Sivaprasad, J. K. Ladha, and A. M. Mortimer, "Weed management in direct-seeded rice," *Adv. Agron.*, vol. 93, pp. 153–255, 2007.

[8] J. Lassila, T. Kaipia, J. Haakana, and J. Partanen, "Concept of strategic planning in electricity distribution business," *Int. J. Energy Sect. Manag.*, vol. 5, no. 4, pp. 447–470, 2011.

[9] R. K. Shukla, D. Garg, and A. Agarwal, "Understanding of supply chain: A literature review," *Int. J. Eng. Sci. Technol.*, vol. 3, no. 3, pp. 2059–2072, 2011.

[10] Y. H. Lee and S. H. Kim, "Production–distribution planning in supply chain considering capacity constraints," *Comput. Ind. Eng.*, vol. 43, no. 1–2, pp. 169–190, 2002.

[11] M. Ahsan, G. Kusuma, and S. Alfarizi, "Optimasi Produk Plastik pendekatan Taguchi Mixed Level pada Faktor Interaksi Injeksi Molding," *J Stat. J. Ilm. Teor. Dan Apl. Stat.*, vol. 16, no. 1, pp. 371–383, 2023.

[12] D. Sulistiowati, M. S. Syahrul, and I. Rina, "Forecasting PT Triputra Agro Persada Tbk (TAPG) Share Prices Using Multivariate Time Series Analysis," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 2, pp. 594–605, 2023.

[13] P. R. Arum, R. P. Gautama, I. Fitriani, and F. Naza, "Identifying Factors that Influence Life Expectancy in Central Java Using Spatial Regression Models," *J. Ilm. Toeri dan Apl. Stat.*, vol. 16, no. 2, pp. 606–613, 2023.