Analysis of Factors Influencing the COVID-19 Mortality Rate in Indonesia using Zero Inflated Negative Binomial Model

Maria Susan Anggreainy[1], Abdullah M. Illyasu[2] , Hanif Musyaffa[3], Florence Helena Kansil[4]

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480[1, 3, 4]

Tokyo Institute of Technology, Yokohama 226-8502, Japan[2]

Abstract—This research aims to create a model, analyze the factors that influence the COVID-19 mortality rate in Indonesia. There are five independent variables and one dependent variable used in the research. The independent variables used are the percentage of poor people, the percentage of households using shared toilet facilities, the percentage of households using wood as the main fuel for cooking, the percentage of the population whose drinking water source comes from pumped water and the percentage of population who have health insurance from private insurance. While the dependent variable used is the Annual Parasite incidence COVID-19. The results obtained are as follows. First, a Zero-Inflated Negative Binomial regression model was obtained for the case of COVID-19 morbidity where this model could overcome overdispersion and excess zero values in observations. Second, there are 4 independent variables that have a significant effect on the count model and there is no independent variable that has a significant effect on the Zero inflation model. Third, a web application is produced that can display the Zero-Inflated Negative Binomial regression model (ZINB).

Keywords—*Mortality rate; overdispersion; zero-inflated negative binomial; Poisson regression; correlation*

I. INTRODUCTION

Health is very important for every human being. With a healthy body, the soul will be good, and the mind will be in balance. Having a healthy body and soul can support human activities without any obstacles. Steps that need to be taken to maintain health include exercising, getting appropriate nutrition coverage and guarding yourself from habits that damage the body. Public awareness is needed to carry out a healthy lifestyle to avoid disease. COVID-19 has become a health emergency in the past one month since the first case in Indonesia was reported in Depok in March 2020. The problem that Indonesia is currently facing is that the government is still not effective in conducting testing and COVID-19 cases and there is no enforcement of rules regarding social distancing and mobility [1].

Research on the analysis of factors that influence COVID-19 in Indonesia is carried out using several different methods. The Zero-Inflated method used by the researcher in this research is the Zero-Inflated Negative Binomial method. This research was conducted to model the COVID-19 mortality rate using the web-based Zero-Inflated Negative Binomial method. The Zero-Inflated Negative Binomial method is used because

there is overdispersion in the Poisson regression model and the data used has an excessive value of 0 [2]. Overdispersion is a condition where the value of the variance is greater than the average [3]. One of the causes of overdispersion is the number of observations that are zero in the dependent variable or variable Y. Zero-Inflated Negative Binomial Regression (ZINB) is a model formed from a mixed distribution of Poisson gamma [3]. If y_i is a discrete independent random variable with $i= 1, 2, n$ the null value of the observation is assumed to appear in two appropriate ways for separate states. The first state called zero state occurs with probability pi and produces only zero observations, while the second state is called Negative Binomial state occurs with probability (1 - pi) and has a Negative Binomial distribution with mean μ, with $0 \le pi \le 1$ [4]. Estimated ZINB regression parameters using the Maximum Likelihood Estimation (MLE) method with the EM (Expectation Maximization) and Newton Raphson Algorithm procedures [5]. This method is usually used to estimate the parameters of a model whose density function is known.

The use of the Zero-Inflated Negative Binomial (ZINB) method has been carried out by several researchers [6][7][8]. In research conducted concluded that the ZINB regression model is more appropriate to be used to model data on the number of maternal deaths in Bali Province which contains many zero values and experiences overdispersion [9]. In another research conducted and concluded that the Poisson regression model does not meet the overdispersion hypothesis, so another model is used, the proposed model is the Zero-Inflated Poisson (ZIP) model, but there is still overdispersion on the ZIP models [10].

To overcome this problem, the ZINB model and the hurdle negative binomial (HNB) model are used [11]. The Akaike Information Criterion (AIC) value of the ZINB model is lower than the value of the HNB model [12]. This shows that the ZINB model is best used in data on the incidence of diptheria in Indonesia [13]. The HNB model can control zero values and overdispersion, just like the ZINB model [14]. However, in the data on the incidence of diphtheria in Indonesia, the ZINB model is more suitable to control the value and over-dispersion of the data concludes that the appropriate model for the frequency of traveling during the last six months in South Tapanuli regency, North Sumatra, for the March 2016 period is the Zero-Inflated Negative Binomial model [15].

The independent variables used in this research were the percentage of poor population, the percentage of households using shared toilet facilities, the percentage of the population whose drinking water source came from pump water and the percentage of the population who had health insurance from private insurance. The dependent variable used in this research is the COVID-19 mortality rate. Based on the data obtained, there are areas where the COVID-19 mortality rate is zero. This causes the data used to be overdispersion or the average value and variance are different. Therefore, the formulation of the problem in this research is:

1) Create a Zero-Inflated Negative Binomial model for the COVID-19 mortality rate in Indonesia.

2) Analyzing the Factors Affecting the COVID-19 mortality rate in Indonesia.

II. MATERIALS AND METHODS

A. Research Variable

The type of data used in this research is secondary data. The variables used in this research consisted of the dependent variable and the independent variable (can be seen in Table I). Measurement scale of the variables is ratio.

TABLE I. EXPLANATION OF THE VARIABLES USED

Variable Name	Variable type.	Definition
COVID-19 mortality rate	Dependent	API is the morbidity rate per 1000 population at risk in one year.
Percentage of poor people	Independent	Population whose average monthly per capita expenditure is below the poverty line. The unit used is Percent (%).
Percentage of households that use shared toilet facilities	Independent	MCK stands for bathing, washing, and toileting is one of the public facilities that are shared by several families for the purpose of bathing, washing, and defecating in certain residential locations which are considered to have a fairly dense population and low level of economic capacity. The unit used is Percent (%).
Percentage of households that use wood as the main fuel for cooking	Independent	Until now, household energy needs in rural areas are still supported by firewood and agricultural waste. In rural areas, especially in remote areas, people still use more than 60% of their energy needs from firewood or biomass. The unit used is Percent (%).
Percentage of population whose drinking water source comes from pumped water	Independent	Pump water is water that comes from pumping from a water source in the ground, then distributed into existing water pipes in the house or in the water tank. The unit used is Percent (%).
Percentage of population who have health insurance from private insurance.	Independent	Health Insurance is an insurance that provides insurance to the insured to replace any medical expenses which include hospital treatment costs, surgery costs and drug costs. The unit used is Percent (%).

B. Research Steps

The data analysis technique used in this research is descriptive analysis, Poisson regression and Zero-Inflated Negative Binomial. The following is a more detailed explanation of the steps taken by researchers in completing this research.

1) Perform secondary data collection.

2) Perform data processing so that the data used.

3) Determine the type of variable from each variable used.

Conducting descriptive analysis by calculating the Mean, Median and standard deviation values. In addition, it determines the maximum and minimum values of each variable.

The following is an explanation of the steps taken by researchers in making this Zero-Inflated Negative Binomial model (can be seen in Fig. 1):
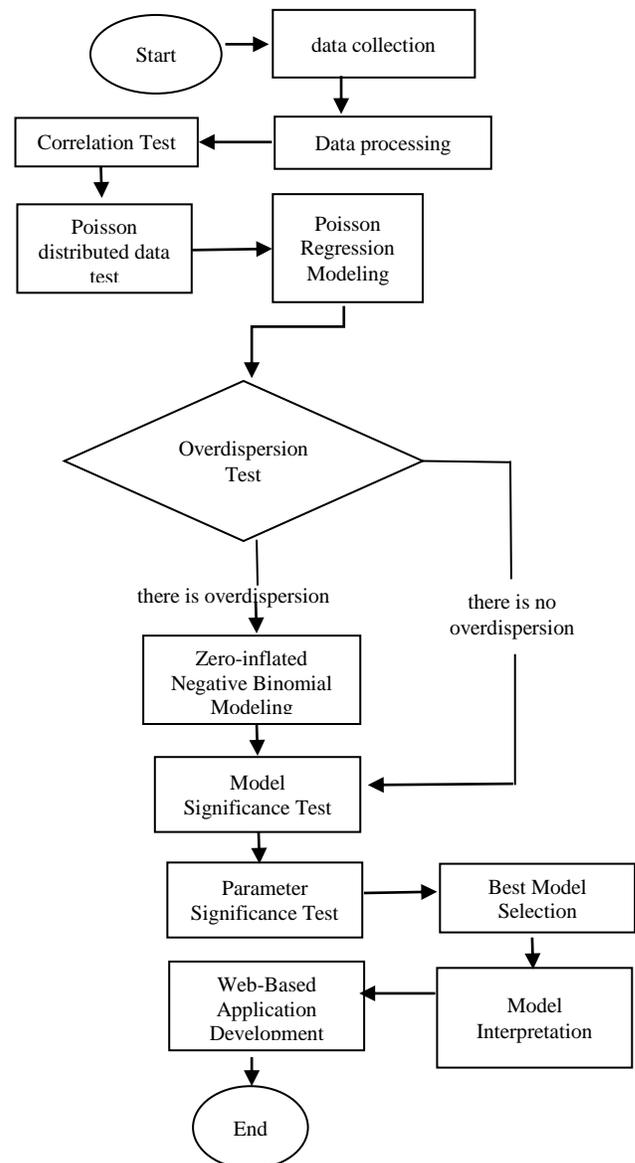
Fig. 1. Flowchart of Research Steps.

1) Perform the Spearman rank correlation test to find out whether there is a relationship between the dependent variable and the independent variable with the following steps:

- Specifies H_0 and H_1.

- Calculate ρ_{count} according to equation.

Equation (1) if there is no the same rank.

$$\rho_{calculate} = 1 - \frac{6 \sum d^2}{n(n^2-1)} \tag{1}$$

Equation (2) If there is the same rank.

$$\rho_{calculate} = \frac{\frac{\sum (R_x - \overline{R_x})(R_y - \overline{R_y})}{n}}{\sqrt{\frac{\sum (R_x - \overline{R_x})^2}{n}} \sqrt{\frac{\sum (R_y - \overline{R_y})^2}{n}}} \tag{2}$$

- ρ_{count} = Spearman Correlation Coefficient.

- d = Rank difference for each data.

- n = Number of Samples or data.

- R_x = Rating for each data on Variable X.

- $\overline{R}_x$ = Average Rating for each data on Variable X.

- R_y = Rank for each data on Variable Y.

- $\overline{R}_y$ = Average Rating for each data on variable Y.

    - H_0: $\rho 0$ which means there is no correlation betweenvariables X and Y.

    - H_1 : $\rho \neq 0$ which means there is a correlation betweenthe variables X and Y.

- Determine the $\rho_{(table)}$.

- Comparing the value of ρ_{count} with the value of $-\rho_{(table)}$.

- Draw conclusions based on the results of the comparison.

2) Determine the estimated regression coefficient by using the Maximum Likelihood Estimation method and Newton Raphson iteration [16].

3) Create a Poisson Regression model using Equation (3).

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} , i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, \tag{3}$$

4) where the coefficients used are the coefficients obtained is step 3.

5) Create a Poisson Regression model using Equation (3).

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} , i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, p \tag{3}$$

where the coefficients used are the coefficients obtained in step 3.

6) Perform the Overdispersion test on Poisson regression with the following steps [17]:

- Specifies H_0 and H_1.

- Determine the value of Deviance (θ_1) and Pearson Chi-Squared (θ_2).

- Compares the values of θ_1 and θ_2 with 1.

- Draw conclusions based on the results of the comparison.

7) If Overdispersion is found, then the model made is Zero Inflated Negative Binomial.

C. Analyzing

After the previous steps have been taken, the steps taken to analyze the factors that influence the COVID-19 mortality rate are as follows:

1) Determine the estimated regression coefficient by using the Maximum Likelihood Estimation method and Newton Raphson iteration.

2) Create a Zero-Inflated Negative Binomial model using Equations (4).

$$logit \,\hat{\pi}_i = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j x_{ij} , i = 1,2,\ldots, n \text{ and } j = 1,2,\ldots, p \tag{4}$$

p : number of independent variables.

n : number of observations.

$\hat{\beta}$: estimated ZINB regression model parameters.

$\hat{\gamma}$: estimated ZINB regression model parameters.

3) Conducting the test simultaneously with the following steps:

   a) Specifies H_0 and H_1.

   b) Calculating the value of the G test statistic according to Equation (5).

$$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \tag{5}$$

The test criteria is reject H_0 if $W > t(\frac{\alpha}{2}; n - 1)$. Denial H_0 at the level of significance α means that a certain j-th independent variable has a significant contribution to the dependent variable Y.

   c) Determining the value $X^2_{(\alpha, n-k-1)}$.

   d) Comparing the value of the G test statistic with the value of $X^2_{(\alpha, n-k-1)}$.

   e) Drawing conclusions based on comparison results.

4) Conduct a parameter significance test for each of the independent variables in the former model to find out which variables influence the COVID-19 mortality rate. The test steps are as follows:

a) Specifies H_0 and H_1.

b) Calculating the value of the Wald test statistic according to Equation (5).

c) Looking for value $t(\frac{\alpha}{2}; n-1)$.

d) Comparing the statistical value of the Wald test of each independent variable with the value of $t(\frac{\alpha}{2}; n-1)$.

e) Determining the independent variables that affect the dependent variable based on the comparison results.

5) Choose the best model based on Akaike Information Criterion (AIC) based on Equation (6).

$$AIC = 2k - 2\ln(\hat{L}) \tag{6}$$

k = Number of parameter estimates in the model $\hat{L}$= The maximum value of the model Likelihood function

6) Make an interpretation of the model that has been made.

III. RESULT AND DISCUSSION

In this research, a web-based application was made. Users must register their personal data first as a complete account. After the account has been successfully created, the user will be directed to the Login page to input the registered Username and Password. If the Username and Password match, the User can access the features contained in the application. After successful login, the user will be directed to the home menu. In the home menu there is a Navigation bar that contains the features found in the application. Users can see the results of predictions of mortality rates that have been done before or Users can make new interpretations of COVID-19 mortality rates based on existing variables. After the user has finished using the application, the user can log out of the account by pressing the logout button. Use cases in this research can be seen in Fig. 2.
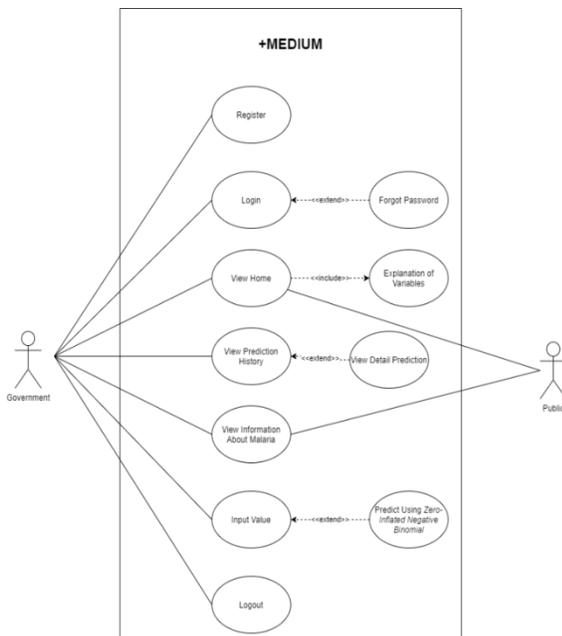
Fig. 2. Use Case.

A. Correlation Analysis

The independent variables tested were the percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), the percentage of the population whose drinking water source came from pumps (X_4), the percentage of the population who had health insurance from Private Insurance (X_5), Area Height (X_6), Number of Hospitals (X_7), Number of Doctors (X_8) and Percentage of Households with Ground Floor (X_9). For example, the correlation value of X1 and Y is -0.4570287, then the interpretation of the correlation value is that the relationship between the percentage of the poor and the COVID-19 mortality rate is in the moderate category and contradicts each other where if the percentage of the poor increases, then the mortality rate will decrease (can be seen in Table II).

Based on Table I, the correlation value of several independent variables is greater than $_{(0.05,29)}$ which means that several independent variables have a significant effect on the dependent variable. The influential variable is Percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), the percentage of the population whose drinking water source came from pumps (X_4), and the percentage of the population who had health insurance from Private Insurance (X_5). Therefore, the variables Area Height (X_6), Number of Hospitals (X_7), Number of Doctors (X_8) and Percentage of Households with Ground Floor (X_9) are excluded from the modeling to be carried out.

B. Poisson Distributed Data Test

From the test, the results of the D_n test statistic value of 0.4286. These results were compared with the statistical value of the Kolmogorov-Smirnov test (0.05;7) which was worth 0.48343. Because the statistical value of the D_n test is smaller than the statistical value of the Kolmogorov-Smirnov test, it can be concluded that H_0 a failed to reject or the data came from a population that followed the Poisson distribution. Based on the conclusion, the regression model that will be made is Poisson regression modeling.

TABLE II. SPEARMAN CORRELATION TEST

Variable	ρ	Direction	$\rho(0.05,29)$	Conclusion
$X1$	0.4570287	Negative	0.368	There is a relationship
$X2$	0.435146	Positive	0.368	There is a relationship
$X3$	0.3951663	Negative	0.368	There is a relationship
$X4$	0.4761548	Positive	0.368	There is a relationship
$X5$	0.4286245	Positive	0.368	There is a relationship
$X6$	0.290077569	Negative	0.368	No connection
$X7$	0.292051614	Positive	0.368	No connection
$X8$	0.146538363	Positive	0.368	No connection
$X9$	0.196191728	Negative	0.368	No connection

TABLE III.   POISSON REGRESSION MODELING RESULTS

Variable	Coefficient	Standard Error	Test Statistic Value	Conclusion
Intercept	3.772434	0.318566	11.842	Significant
X_1	-0.004478	0.013024	- 0.344	Not Significant
X_2	0.078039	0.021383	3.650	Significant
X_3	-0.034350	0.003827	- 8.975	Significant
X_4	0.085689	0.011711	7.317	Significant
X_5	-1.014189	0.201304	- 5.038	Significant

Table III shows the coefficient values, standard error, test statistics and conclusions from each Poisson regression variable. The influential variable is Percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), the percentage of the population whose drinking water source came from pumps (X_4), and the percentage of the population who had health insurance from

Private Insurance (X_5). The model formed based on the values shown in Table III is as follows:

$$\ln \hat{\mu_i} = 3.772434 - 0.004478 * X_1 + 0.078039 * X_2 - 0.034350 * X_3 + 0.085689 * X_4 - 1.014189 * X_5.$$

C. Zero-Inflated Negative Binomial Modeling

The Zero-Inflated Negative Binomial (ZINB) regression model is a regression model that can be used to model data with the dependent variable having a Poisson distribution, many observations that are zero in the dependent variable and overdispersion occurs.

Table IV displays the coefficient values, Standard error, test statistics, and conclusions from each of the Zero-Inflated Negative Binomial regression variables. Based on the table, it is known that the percentage of the population who has health insurance from private insurance does not have a significant effect on both models. So that the Zero-Inflated Negative Binomial modeling was carried out without using the variable Percentage of the population who had health insurance from private insurance. The results of modeling can be seen in Table V.

TABLE IV.   RESULTS OF THE ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODEL

Variable	Coefficient	Standard Error	Wald Test Statistic Value	Conclusion
β_0	2.26342	0.33687	6.719	Significant
β_1	0.48196	0.06754	7.136	Significant Influence on the Dependent Variable for the Count Model.
β_2	0.12367	0.05760	2.147	Significant Influence on the Dependent Variable for the Count Model.
β_3	-0.24037	0.04078	-5.894	Significant Influence on the Dependent Variable for the Count Model.
β_4	0.13436	0.02281	5.891	Significant Influence on the Dependent Variable for the Count Model.
β_5	-0.49147	0.48896	-1.005	No Significant Effect on the Dependent Variable for the Count Model.
γ_0	-15.2554	5294.8945	-0.003	Not Significant
γ_1	2.0205	709.9964	0.003	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_2	-13.0143	3093.7702	-0.004	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_3	-0.2961	287.7473	-0.001	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_4	-1.2309	2178.5769	-0.001	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_5	34.4325	8887.1670	0.004	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.

TABLE V.   ZERO-INFLATED NEGATIVE BINOMIAL MODELING RESULTS WITHOUT VARIABLE PERCENTAGE OF POPULATION WHO HAVE HEALTH INSURANCE FROM PRIVATE INSURANCE

Variable	Coefficient	Standard Error	Wald Test Statistic Value	Conclusion
β_0	2.13703	0.32859	6.504	Significant
β_1	0.52405	0.06850	7.650	Significant Influence on the Dependent Variable for the Count Model.
β_2	0.14371	0.05981	2.403	Significant Influence on the Dependent Variable for the Count Model.
β_3	-0.26860	0.04007	-6.704	Significant Influence on the Dependent Variable for the Count Model.
β_4	0.15242	0.01937	7.870	Significant Influence on the Dependent Variable for the Count Model.
γ_0	-27.395	24062.983	-0.001	Not Significant
γ_1	7.081	2213.102	0.003	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_2	-17.144	697.278	-0.025	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_3	-2.119	617.880	-0.003	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.
γ_4	-9.382	2751.073	-0.003	No Significant Effect on the Dependent Variable for the Zero-Inflation Model.

The model formed can be as follows:

a) Count Model Coefficient

$$\hat{\mu} = exp(\ 2.13703\ +\ 0.52405\ *\ X_1\ +\ 0.14371\ *\ X_2\ -\ 0.26860\ *\ X_3\ +\ 0.15242\ *\ X_4)$$

b) Zero-Inflation Model Coefficient

$$\hat{\pi} = \frac{exp(-27.395 + 7.081 * X1\ -17.144 * X2\ - 2.119 * X3\ -9.382 * X4\ )}{1 + exp(-27.395 + 7.081 * X1\ -17.144 * X2\ - 2.119 * X3\ -9.382 * X4)}$$

D. Best Model Selection

The best model selection is done by looking at the AIC (Akaike Information Criterion) value. The selection of the best model is done by comparing the 2 models that have been formed, namely the Poisson Regression model and the Zero-Inflated Negative Binomial Model. The AIC values of the two models can be seen in Table VI.

TABLE VI. COMPARISON OF AIC VALUES

Regression Model	AIC Values
Regresi Poisson	817.4561
Zero-Inflated Negative Binomial (Full Model)	65.1253
Zero-Inflated Negative Binomial (Without X₅)	62.1329

The AIC value in Table VI shows that the lowest AIC value is the Zero-Inflated Negative Binomial model without a Variable Percentage of the population who has health insurance from private insurance. These variables were excluded because they had no significant effect on the Count Model and Zero-Inflation Model. Therefore, the Zero-Inflated Negative Binomial modeling was carried out without using the Variable Percentage of the population who had health insurance from private insurance.

E. Model Interpretation

The ZINB model is used to deal with overdispersion in the Poisson Regression model. The ZINB model is divided into two components, namely the count model for and the zero inflation model for. The interpretation of the model formed from ZINB is based on the odd ratio value as seen from the exp (β) value.

1) The interpretation of the count model coefficient is as follows :

a) The constant is 2.13703, meaning that if the variables are Percentage of poor people X_1 , Percentage of households using shared MCK facilities (X_2), Percentage of households using wood as the main fuel for cooking (X_3), and Percentage of population whose drinking water source comes from Pump Water $((X_4)$, is zero, then the COVID-19 mortality rate is worth exp(2.13703) = 8.474232.

b) The coefficient of X_2 is 0.52405, meaning that every 1 percent increase in the percentage of poor people (X_2), will increase the COVID-19 mortality rate by exp (0.52405) =

1.688854 times the original COVID-19 mortality rate, if other variables are constant.

c) The coefficient of X_3 is -0.26860, meaning that every 1 percent increase in the percentage of households that use wood as the main fuel for cooking (X_3), it will reduce the COVID-19 mortality rate by exp (-0.26860) = 0.764449 times the original COVID-19 mortality rate, if the variable is else constant value.

d) The X_4 coefficient is 0.15242, meaning that every 1 percent increase in the Percentage of the Population whose drinking water source comes from Pump Water (X_4) will increase the COVID-19 mortality rate by exp (0.15242) = 1.164649 times the original COVID-19 mortality rate, if other variables are constant.

2) The interpretation of the zero inflation model coefficient is as follows:

a) The constant is -27.395 , meaning that if the variables are Percentage of poor people (X_1), Percentage of households using shared MCK facilities (X_2), Percentage of households using wood as the main fuel for cooking (X_3), and Percentage of population whose drinking water source comes from from the Air Pump (X_4), is zero, then the value of the COVID-19 mortality rate is exp(- 27.395) = 1.266E-12.

b) The coefficient of X_1 is 7.081, meaning that every 1 percent increase in the percentage of poor people (X_1), will increase the chance of the COVID-19 mortality rate to zero by exp (7.081) = 1189,157 times, if other variables are constant.

c) The coefficient of X_2 is -17,144, meaning that every 1 percent increase in the percentage of households using shared MCK facilities (X_2), will reduce the chance of the COVID-19 mortality rate to zero by exp (-17,144) = 3.58E-08 times, if other variables are constant.

d) The coefficient of X_3 is -2.119, meaning that every 1 percent increase in the percentage of households that use wood as the main fuel for cooking (X_3), it will reduce the chance of the COVID-19 mortality rate to zero by exp (-2.119) = 0.1201 times, if other variables are worth constant.

e) The coefficient of X_4, is -9.382, meaning that every 1 percent increase in the Percentage of the Population whose drinking water source comes from Pump Water (X_4) will reduce the chance of the COVID-19 mortality rate to zero by exp (-9.382) = 8.42E-05 times, if other variables are worth constant.

F. Parameter Significance Test Results

Based on Table VII, it can be concluded that in the count model there are 4 variables that have a significant effect on the COVID-19 mortality rate. In the zero-inflation model, there are no independent variables that affect the COVID-19 mortality rate. Based on the two models, it can be concluded that the variables used are not appropriate for the zero-inflation model.

TABLE VII. PARAMETER SIGNIFICANCE TEST RESULTS

Variable Name	Count Model Coefficient	*Zero-Inflation Coefficient*
Percentage of poor people (X_1)	Significant	Not significant
Percentage of households using shared toilet facilities (X_2)	Significant	Not significant
Percentage of households using wood as the main fuel for cooking (X_3)	Significant	Not significant
Percentage of Population whose drinking water source comes from Pump Water (X_4)	Significant	Not significant

IV. CONCLUSION

The conclusions obtained from this research are as follows the factors that influence the COVID-19 mortality rate in the count model are percentage of poor people (X_1), the percentage of households using shared toilet facilities (X_2), the percentage of households using wood as the main fuel for cooking (X_3), and the percentage of the population whose drinking water source came from pumps (X_4). In the Zero-Inflation model, there are no factors that affect the COVID-19 mortality rate. so that the ZINB regression model used is the count model. Third, based on the evaluation of user satisfaction, the designed application has been able to help predict COVID-19 mortality and assist in providing information and insight to the public about COVID-19.

REFERENCES

[1] Anggreainy, M.S., Musyaffa, H. Illyasu, A.M, "COVID-19 patient mortality risk classification using Linear Regression and Exponential Smooting Methods". 2021 International Seminar on Machine Learning, Optimization, and Data Science, ISMODE 2021, 2022, pp. 173-177.

[2] Mekuanint Simeneh Workie., Abebaw Gedef Azene. "Bayesian Zero-Inflated Regression Model with application to under five child mortality". Journal of Big Data volume 8, Article number : 4 (2021).

[3] Muggy David Cristian Ginzel, Asep Saefuddin, Erfiani, Determination of Extracurricular Activities for High School Students Based on Multiple Intelligence , Xplore: Journal of Statistics: Vol. 9 No. 1 (2020).

[4] Moh. Irfan Agus Saputro., M.Farhan Qudratullah. "Estimation of zero-inflated negative binomial regression parameters using the maximum likelihood method (case study : factors affectiong infant mortality in Wonogiri in 2015)". Proceeding International Conference on Science and Engineering (ICSE).2021-02-28.

[5] Shakhawat Hossain., Hatem A. Howlader. "Estimation Technique for Regression Model with Zero-inflated Poisson Data". International Journal of Statistics and Probability; Vol.4, No. 4; 2015.

[6] Preisser JS, Das K, Long DL, Stamm JW. A marginalized zero-inflated negative binomial regression model with overall exposure effects. The University of North Carolina at Chapel Hill Department of Biostatistics Technical Report Series. 2014;43.

[7] Pornpop Saengthong, Ampai Thongteeraparp, Winai Bodhisuwan."The zero inflated negative binomial – Crack distribution: Some properties and parameter estimation".Journal of Science and Technology · November 2015.

[8] Long DL, Preisser JS, Herring AH, Golin CE. A Marginalized Zero-Inflated Poisson Regression Model with Random Effects. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* (published online April 30, 2015). DOI: 10.1111/rssc.12104.

[9] Made Susilawati, I Gusti Ayu Made Srinadi. "Perbandingan Regresi Zero Inflated Poisson (Zip) Dan Regresi Zero Inflated Negative Binomial (Zinb) Pada Data Overdispersion (Studi Kasus: Angka Kematian Ibu di Provinsi Bali). E-Jurnal Matematika 5(4):133. November 2016. DOI:10.24843/MTK.2016.v05.i04.p132.

[10] Puspita Rahayu, Kusman Sadik, Indahwati."Overdispersion study of poisson and zero-inflated poisson regression for some characteristics of the data on lamda, n, p"/ International Journal of Advances in Intelligent Informatics 2(3):140. November 2016. DOI:10.26555/ijain.v2i3.73.

[11] Drajat Indra Purnama. "Comparison of Zero Inflated Poisson (ZIP) Regression, Zero Inflated Negative Binomial Regression (ZINB) and Binomial Negative Hurdle Regression (HNB) to Model Daily Cigarette Consumption Data for Adult Population in Indonesia". Jurnal Matematika Statistika dan Komputasi 17(3):357-369. May 2021. DOI:10.20956/j.v17i3.12278.

[12] Osuji G. A., Okoro C. N., Obubu M., Obiora-Ilouno H. O."Effect of Akaike Information Criterion on Model Selection in Analyzing Auto-crash Variables".Effect of Akaike Information Criterion on Model Selection in Analyzing Auto-crash Variables. January 2016.

[13] Mulya Rahma Karyanti, Erni Juwita Nelwan, Iqbal Zein Assyidiqie, Hindra Irawan Satari, Sri Rezeki Hadinegoro. "Diphtheria Epidemiology in Indonesia during 2010-2017".Acta medica Indonesiana. 1 Jul 2019.

[14] Cindy Xin Feng. "A comparison of zero-inflated and hurdle models for modeling zero-inflated count data". Journal of Statistical Distributions and Applications volume 8, Article number: 8 (2021).

[15] Naufal Rifki Fauzan;Daniel Alexander;Muhammad Siraz Hafizh;Maria Susan Anggreainy."Covid-19 Vaccine Tweets - Sentiment Analysis". 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI). DOI: 10.1109/ICCSAI53272.2021.9609712.

[16] Raden Aurelius Andhika Viadinugroho. "Calculate Maximum Likelihood Estimator with Newton-Raphson Method using R". Towards Data Science. 1 Maret 2021.

[17] Dewi Retno Sari Saputro, Ade Susanti, and Nafisa Berliana Indah Pratiwi."The handling of overdispersion on Poisson regression model with the generalized Poisson regression model". AIP Conference Proceedings Volume 2326, Issue 1, https://doi.org/10.1063/5.0040330.