Analysis The Influence of Climate Factors With COVID-19 Recovery Rates in DKI Jakarta

Dimas Teguh Ramadhan[1*], Shofiyah Calista Ricadonna[2], Syalwa Rastia Kaldi[3]

[1, 2, 3] *Study Program of Actuarial Science, School of Business, President University, 17550, Indonesia*
[*] *Corresponding author: dimas.ramadhan @student.president.ac.id*

Abstract— *COVID-19* is a disease caused by the SARS-CoV-2 virus. The ideal air for the growth of this virus is at a temperature of around 8-10°C with a relative humidity of 60-90%. From March 2020 to December 2021, there were 7,709,036 cases of recovering positive *COVID-19* patients. One of the factors thought to affect the rate of recovery from *COVID-19* is climate factors, including the average temperature, average relative humidity of the air, rainfall, length of time. solar radiation, and average wind speed. The cure rate for *COVID-19* in DKI Jakarta tends to increase from May 2020 to July 2020 and decrease from July 2020 to December 2020. This research was carried out with a spatial analysis using the Poisson regression, negative binomial and generalized Poisson regression methods. Based on the AIC value model, the most suitable model for this model is the Generalized Poisson Regression. From this study, it was concluded that the climatic factors that had a significant effect on the recovery rate of *COVID-19* in DKI Jakarta were the average temperature, the average relative humidity of the air, the duration of sunlight, and the average wind speed.

Keywords— *Poisson regression method, DKI Jakarta, COVID-19, Climate Factor*

I. INTRODUCTION

The World Health Organization (WHO) reports that there have been new cases of influenza in Wuhan City, Hubei Province, China [1]. On January 7, 2020, the Chinese government said the new strain was coronavirus or *COVID-19* [2]. Severe and critical clinical symptoms of *COVID-19* patients tend to be similar to those of *SARS* and *MERS* [3]. Signs and symptoms that arise when infected with *COVID-19* are fever, cough, and shortness of breath or are similar to symptoms of acute respiratory distress in general. The average incubation period is 5-6 days, with a maximum time of 14 days. In severe cases of *COVID-19*, it can cause kidney failure, acute respiratory syndrome, and even death [4]. Then, WHO decided the current *COVID-19* issue was a pandemic on Wednesday, March 11, 2020 [5]. The journey of the epidemic is determined by several factors, including demographics and the environment, many of which have unknown correlations to *COVID-19*. Environmental conditions, including climatic factors, are known to impact the transmission and survival of viruses that cause respiratory diseases such as influenza and SARS. Climatic factors can affect disease life, such as protozoa, viruses, bacteria, and others that are very small in size and do not have a thermostatic mechanism, so the temperature and the fluid level of disease agents are directly determined by local climatic conditions [6]. Several studies conducted outside Indonesia found that climatic factors such as rainfall, wind speed, temperature, and humidity correlate with *COVID-19* cases [7]. Meanwhile, several studies in Indonesia found that temperature and wind speed correlated with *COVID-19* cases [8].

In this study, the *COVID-19* recovery data indicated that the Poisson Regression Model was a nonlinear regression model used to analyze discrete data (counts). Some of the characteristics of experiments that follow a Poisson distribution are events in many population members with low probability (rare events). Depending on certain time intervals, these events fall into the counting process or are included in the stochastic process environment. The data indicate that Poisson regression can be used to model climate factors on the rate of *COVID-19* recovery in DKI Jakarta. The author chose to examine this case because research on changing climate factors with *COVID-19* in Indonesia is limited. Based on the background above, the research purpose is to find out the climatic factors that influence the recovery rate of COVID-19 sufferers.

II. LITERATURE REVIEW

A. Poisson Distribution

A random variable Y is defined as having a Poisson distribution if the density (probability function) is given [9].

$$f_Y(y) = \begin{cases} \dfrac{e^{-\lambda}\lambda^y}{y!} \\ 0, y \quad another \end{cases} \tag{1}$$

A Poisson process has the following properties:

1. The number of events that occur in a specific time interval or area is independent of the number of events in another time interval or area.
2. The probability of an event occurring in a minimal time interval or area is proportional to the length of the time interval or the area. It does not depend on the number of events outside this time interval or area.
3. The probability of more than one occurrence in a very small time interval or area is negligible.

A random variable *Y* of discrete type will follow a Poisson distribution if μ it is the average of an event per unit of time and *t* is a certain period. The average of *y* becomes μt. The following equation gives the probability of occurrence of event *y* in time period *t*:

$$P(y;\mu) = \frac{[\exp{(\mu t)}[\mu t]^y}{y!}, y = 0,1,2,\dots; \mu > 0 \tag{2}$$

If the time interval of events is the same, then the probability distribution function for the random variable Poisson *Y* with μ becomes:

$$P(y;\mu) = \frac{[\exp(-\mu)]\mu^y}{y!}, y = 0,1,2,\dots; \mu > 0 \tag{3}$$

$$\mathrm{E(Y)} = \mu,\ var\ (Y) = \ \mu \tag{4}$$

B. Multicollinearity

Before performing a Poisson regression analysis, the first thing to do is to detect the presence of multicollinearity, which can be seen from the VIF (Variance Inflation Factors) value. The equation can be written [10]:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{5}$$

Where R_j^2 is the value of the coefficient of determination. A VIF value of more than 10 indicates that the predictor variables are correlated with other predictor variables. Meanwhile, the VIF value of less than 10 indicates that there is no autocorrelation between the predictor variables. If there is multicollinearity, the solution to overcome this is by removing the insignificant predictor variable (dropping variable) and regressing the significant predictor variables.

C. Overall Test

The overall test aims to determine the independent variables that have a significant effect on the dependent variable at each location. Wald's test hypothesis [11]:

$H_0: \beta_k(u_i v_i) = 0$ (There is no significant effect between the independent variables on the dependent variable).

$H_1: \beta_k(u_i v_i) \neq 0$ with k=1,2, …, p (There is a significant influence between the independent variables on the dependent variable).

Critical Area

Reject H_0 if Pr(>z) less than α, with α = 0.05

D. Partial Test

Partial test uses the Maximum Likelihood Ratio Test (MLRT) method where the test hypothesis is as follows:

$H_0\ : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

$H_1\ :$ at least $\beta_1 \neq 0$; $\quad j = 1,2,\dots,p$

Critical Area

Reject H_0 if *Chi-Square* less than deviance

E. Poisson Regression

Poisson regression is used to analyze discrete data, usually used to express the number or number of events in a period, such as a number or rate of an event at a certain location per unit of time. For example, the number of disease sufferers in an area per year and the number of bacterial colonies in an area. Poisson regression extends the general regression model for response variables that have an exponential distribution. Poisson regression analyzes count data (discrete type or count data). In Poisson regression, it is assumed that the response variable (*Y*) has a Poisson distribution, and there is no multicollinearity between each predictor variable (*X*). These variables have non-negative integer values and are asymmetrical distribution. The Poisson regression model is obtained from the Poisson distribution by defining the parameter as a function of the covariate (predictor) variable. In Poisson regression, the data are arranged in a cross-tabulation structure (cross-section data) consisting of n independent observations, and the ith observation is (y_i, x_i) the dependent variable (response), *Y* is a scalar quantity which states the number of occurrences. An event or occurrence, and X is a vector of predictor variables. Note that the mean and variance of the poisson random variables are the same and are related as [12]:

$$E(x) = \lambda, Var(x) = \lambda \tag{6}$$

Poisson regression modeling is a modeling of the expected value of the response variable

$$E[Y|Xi] = \mu\,(Xi\,,\beta) \tag{7}$$

$$XI = [x1i\,, \ldots, xki] \tag{8}$$

Where K represents the number of predictors. μ (mean Poisson distribution) is always +, so the function μ (X_i,β) is such that the predictor is linear:

$$\eta = \beta o + \beta 1X1 + \ldots + \beta kXk \tag{9}$$

can map each real value to a positive real value.

F. Poisson Process

The Poisson process is an event that occurs when the counting process has a time interval and comes from an exponential distribution. The Poisson process is divided into two, namely, the homogeneous and non-homogeneous Poisson processes. If the exponential distributions have the same parameter values, it is called a homogeneous Poisson process. Otherwise, it is called a non-homogeneous Poisson process. The Poisson process is a simple stochastic process and is widely used for modeling the time at which arrivals enter a system. A process of calculating { $N(t)$; $t \geq 0$ } is said to be a Poisson process with a rate (parameter) λ [13]:

1. $N(0) = 0$,
2. Independent increments,
3. The probability of having k events in the time interval t.

$$P_k(t) = P(N(t+s) - N(s) = k) = (\lambda t)^k \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0,1 \tag{10}$$

A Poisson process is said to be homogeneous if its intensity is constant and is denoted by λ. A Poisson process is said to be non-homogeneous if its intensity is a function of time and is denoted by $\lambda(t)$ which is a function of t.

A non-homogeneous Poisson process with variations in arrival time λ (t) is defined as the process of calculating $\{N(t); t > 0\}$ which has the property of independent increment. A process of calculating $\{N(t);\ t \geq 0\}$ is said to be a non-homogeneous Poisson process with an intensity function $\lambda(t)$ if:

1. $N(0) = 0$,
2. Independent increments,
3. $P(N(t + h) - N(t) = 1) = \lambda(t)h + o(h)$,
4. $P(N(t + h) - N(t) \geq 2) = o(h)$.

with,

$$N(t, t + \delta) = N(t + \delta) - N(t) \tag{11}$$

The non-homogeneous Poisson process does not have the property of stationary increments. Suppose a time interval is divided into parts of length and if the probability of arrival at every increment of some value is fixed,

$$p = \delta\, \lambda(t),\ 0 \tag{12}$$

then (t) is defined as:

$$m(t) = \int_0^t \lambda(\tau) d(\tau) \tag{13}$$

where (t) is the mean value function of the non-homogeneous Poisson process. For each:

$$0 \leq s < t, N(t) - N(S) \tag{14}$$

is a Poisson process with the mean:

$$m(t) - m(s) = \int_s^t \lambda(x) dx \tag{15}$$

G. Poisson Regression Model

The Poisson Regression Model is expressed in the following form [14]:

$$E(Yi) = \mu i \tag{16}$$

$$\begin{aligned} Yi &= E(Yi) + \varepsilon_i\,,\ i=1,2,\ldots,n \\ &= \mu_{i+}\varepsilon_i \end{aligned} \tag{17}$$

μ_i assumed with independent variable $x_{1,}x_2, \ldots, x_{p-1}$. By using the notation $\mu(X_i$,β) to show the relationship function of the mean response variable (μ_i) by independent variable X_i with i=1,2,….,n and β is the value of the regression coefficient, then the Poisson regression model can also be written:

$$Y_i = \mu(X_i, \beta) + \varepsilon_i\,, \quad i = 1,2, \ldots, n \tag{18}$$

There are several functions commonly used in the Poisson regression model, namely:

$$\hat{\mu} = \mu(X_i, \beta) = \begin{cases} X_i\,\beta \\ \exp\,(X_i\beta) \\ log_e(X_i\beta) \end{cases} \tag{19}$$

H. Poisson Regression Model Parameter Estimation

To determine the estimated parameters for Poisson regression, Maximum Likelihood Estimation (MLE) was chosen as the computational method. This method maximizes the likelihood function value to find the best parameter estimate. For example: n Poisson random variable y_i, where $i = 1,2,\dots,n$ are taken independently[15]

$$Fi\ (yi) = \pi^{yi}_{i}(1-\pi_i)^{1-y_i},\ i\ =\ 1,2,\dots,n \tag{20}$$

Then the probability function of this distribution is:

$$L(y,\mu) = \prod_{i=1}^{n} f(y_i,\mu)$$

$$L(y,\mu) = \prod_{i=1}^{n} \left\{\frac{\mu^{y_i}e^{-\mu}}{y_i!}\right\}$$

$$L(y,\mu) = \frac{\left\{\prod_{i=1}^{n}\mu^{\sum_{i=1}^{n} y_i}\right\}\exp\,(-\sum_{i=1}^{n}\mu)}{\prod_{i=1}^{n} y_i!} \tag{21}$$

Substitute μ with:

$$exp\big(\beta_0 + \textstyle\sum_{j=1}^{k}\beta_j X_{ij}\big) \tag{22}$$

then the function can be written like:

$$L(y,\beta) = \frac{\left\{\prod_{i=1}^{n}\big(\exp\big(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\big)\big)^{\sum_{i=1}^{n} y_i}\right\}\exp\left(-\sum_{i=1}^{n}\left(exp\big(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\big)\right)\right)}{\prod_{i=1}^{n} y_i!}$$

$$\ln L(y,\beta) = \ln\left[\frac{\left\{\prod_{i=1}^{n}\left(exp\big(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\big)\right)^{\sum_{i=1}^{n} y_i}\right\}\exp\big(-\sum_{i=1}^{n}\big(\exp\big(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\big)\big)\big)}{\prod_{i=1}^{n} y_i!}\right]$$

$$= \ln\left[\frac{\big(\exp\big(\beta_0+\sum_{j=1}^{k}\beta_1 x_{11}\big)\big)^{y_1}\dots\big(\exp\big(\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}\big)\big)^{y_n}\exp\big(-\sum_{i=1}^{n}\big(\exp\big(\beta_0+\sum_{j=1}^{k}\beta_j x_{ij}\big)\big)\big)}{\{(y_1!).(y_2!)\dots(y_i!)\}}\right]$$

$$\log L(y,\beta) = \sum_{i=1}^{n} y_i \log\left(\exp\left(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right)\right) - \sum_{i=1}^{n}\left(\exp\left(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right)\right) - \sum_{i=1}^{n}\log y_1! \tag{23}$$

I. Overdispersion

The Poisson regression model experiences overdispersion because the enumeration data has a more significant variance than the average. This will affect the standard error value, which causes underestimation, and the results become invalid. Overdispersion can be written as[16]:

$$Var(Y) > E(Y)$$

Overdispersion can be indicated by two methods, namely the deviance value and Pearson chi-squares. Overdispersion in the data occurs when both values are >1.

1) *Deviance*

$$\Phi 1\ =\ \frac{D2}{db};\ D2\ =\ 2\sum_{i=1}^{n}\{\ yi\ In\ (\frac{yi}{\mu i}) - (yi\ -\ \mu i)\} \tag{24}$$

Information:
$Db = n\text{-}k$
k = parameter
n = many observations
D^2 = *Deviance*

2) Pearson Chi-Squares

$$\Phi 2 \;=\; \frac{x}{db}\;;\; X\,2 = \sum_{i=1}^{n} \frac{(yi - \mu i)2}{Var\;(yi)} \tag{25}$$

Information:
Db = n-k
k = parameter
n = many observations
x^2 = *Pearson chi squares*

J. Negative Binomial

In analyzing this data, overdispersion occurs, finding data whose variance is greater than the average. Poisson regression applied to overdispersion data will underestimate the standard error value. The approach taken to handle this can use Negative Binomial regression. Applying the Generalized Linear Model (GLM), the Negative Binomial distribution has three random components, a systematic component, and a connecting function. Variables as Negative Binomial generated by Poisson Gamma. The gamma distribution can accommodate overdispersion in Poisson regression because there is no equidispersion condition in the context of the application. The negative binomial distribution is expressed as follows [17] :

$$f(y;\mu,\alpha) = \frac{r(y_i + 1/\alpha)}{r(y_i+1)r(1/\alpha)}\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}\left(1 - \frac{1}{1+\alpha\mu}\right)^{y_i} \tag{26}$$

In the event of an equidisperse condition:

$$\begin{aligned} f(y;\mu,\alpha) \qquad &= \frac{r_{(y_i+r)}}{r_{(y_i+1)}r_{(r)}}\left(\frac{1}{1+\frac{\mu}{r}}\right)^{r}\left(\frac{r\mu}{1+r\mu}\right)^{y_i} \\ &= \frac{r_{(y_i+r)}}{y!\,r_{(r)}}\left(\frac{1}{1+\frac{\mu}{r}}\right)^{r}\left(\frac{r\mu}{1+r\mu}\right)^{y_i} \\ &= \frac{r_{(y_i+r)}}{r_{(r)(r+\mu)^{y_i}}}\frac{\mu^{y_i}}{y!}\frac{1}{\left(1+\frac{\mu}{r}\right)^{r}} \\ &= 1\frac{\mu^{y_i}}{y!}\frac{1}{e^{\mu}} \\ &= \frac{\mu^{y_i}}{y!}e^{-\mu},\;\; with\; r = \frac{1}{\alpha} = \infty \end{aligned} \tag{27}$$

K. Generalized Poisson Regression

The Poisson regression model requires equidispersion. It must meet the assumption that the variance value of the response variable is the same as the average value. However, assumptions are often violated from equidispersion itself, namely overdispersion or underdispersion. Overdispersion occurs when the variance value is greater than the average value, while underdispersion occurs when the variance value is smaller than the average value. Overdispersion or underdispersion can cause inefficient parameter estimates obtained. Poisson regression can be fatal in model interpretation because it can estimate the standard error parameter that is too low. To overcome this, it is necessary to use a research method using Generalized Poisson Regression (GPR) to overcome the overdispersion phenomenon in the Poisson regression case.

Estimating GPR Model Parameters Using the MLE Method [18]:

1. Establish the Likelihood function:

$$L(\beta,\alpha,y) = \prod_{i=1}^{n} P(y_i,\beta,\alpha)$$

$$(\beta,\alpha,y) = \prod_{i=1}^{n}\left(\frac{e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)}}{1+\alpha e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)}}\right)^{y_i}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!}\exp\left(-\frac{e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)(1+\alpha y_i)}}{1+\alpha e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)}}\right) \tag{28}$$

2. Forming the log likelihood function from the likelihood function that has been obtained.

$$L(\mu i\ ,\ yi\ ) = L = \ln L(\beta,\ \alpha,\ y)$$

$$L = \sum_{i=1}^{n}\left\{y_i\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right) - y_i \ln 1 + \alpha e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)} + (y_i-1)\ln(1+\alpha y_i) - \frac{e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)(1+\alpha y_i)}}{1+\alpha e^{\left(\beta_0+\sum_{j=1}^{5}\beta_j x_{ij}\right)}}\right\} - \ln(y_i!) \quad (29)$$

3. Differentiating the obtained log-likelihood equation.

L. Maximum Likelihood Method

The likelihood function is defined as the joint probability function of X_1, X_2,..,X_n which can be considered as a function of θ. For example, the likelihood function:

$$L(\theta) = f(x_1;\theta)f(x_2;\theta)\ \dots f(x_n;\theta);\ \theta\epsilon\Omega \quad (30)$$

$$= \prod_{i=1}^{n} f(x_i;\theta) \quad (31)$$

The maximum likelihood estimate that maximizes the $\hat{\theta}$ function is called the maximum likelihood estimate of θ. The value of that maximizes l (θ) can be obtained by finding a solution to the equation:

$$\frac{\partial l(\theta)}{\partial\theta} = 0 \quad (32)$$

M. Goodness of Fit (GoF)

Testing the model's suitability to determine whether the model used is appropriate or not with the data under study. Testing the suitability of this model is done by using the goodness of fit test. The goodness of fit test aims to conclude the distribution of the population. A random sample is selected from the population concerned; then, the sample information is used to test the validity of the population. This test is based on how well the goodness of fit is between the frequency of observations obtained by the sample data and the expected frequency obtained from the hypothesized distribution. This case will show whether the data is by the model obtained using the R-Studio software.

Testing the model by looking at the goodness of fit test value using the following hypothesis:

$H_0 =$ The model obtained is by the observed data.

$H_1 =$ The model obtained does not match the observed data.

With a significant degree $\alpha = 0{,}05$. H_0 criteria accepted if the Pearson Chi-Square value is greater than α.

N. Akaike's Information Criterion (AIC)

The Akaike Information Criterion (AIC) was developed by Akaike to estimate the expected Kullback Leibler information between the model generating the data and a fitted candidate model. The AIC is widely used in model selection because it is an estimate of the expected Kullback-Leibler information of a fitted model. AIC is a biased estimator, given by

$$\text{AIC} = n(log\hat{\sigma}^2+1) + 2p \quad (33)$$

Where p and $\hat{\sigma}^2$ are the number of parameter and the variance of the subsets model, respectively. AIC involves the selection of subsets which minimize equation. AIC is an asymptotically unbiased estimate. AIC is a sample estimate of expected entropy, or expected cross-entropy.

III. RESEARCH METHODOLOGY

A. Type of Research

This study uses quantitative and secondary data to develop and analyze numerical data from the data obtained. The data was obtained from the DKI Jakarta Covid-19 data and the Meteorology, Climatology and Geophysics Agency (BMKG) Kemayoran unit, which contains five variables, namely the number of recoveries from COVID-19, temperature, humidity, wind speed, rainfall, and duration of sunshine.*ARIMA Box-Jenkins Flowchart.*

B. Data Preparation

The data used is the number of people recovering from COVID-19 [19]. In addition, temperature, humidity, wind speed, rainfall, and duration of solar radiation were monitored from the BMKG Kemayoran station, Central Jakarta, from March 2020 [20] to December 2021 [21].

TABEL 1
LIST OF DATA

Month	Y	X_1	X_2	X_3	X_4	X_5
	Recovery rate in DKI Jakarta	Temperature (C)	Humidity (%)	Wind Velocity (m/second)	Rainfall (mm)	Sun Exposure (%)
20-Mar	49	28,6	79,43	1,55	220,7	52
20-Apr	363	29,08	78,41	1,37	182,8	57
20-May	1690	29,58	75,86	1,23	50,4	49
20-Jun	4410	29,49	74,7	1,26	21,1	57
20-Jul	6696	28,93	71,54	1,36	12,1	57
20-Aug	17330	29,12	70,89	1,45	101	68
20-Sept	29782	29,29	70,97	1,49	151,9	79
20-Oct	34114	28,83	74,51	1,51	208,3	69
20-Nov	29644	29	75,46	1,42	87,3	59
20-Des	40803	28,15	75,56	2,11	134,7	40
21-Jan	77188	27,3	81,8	3,1	332,8	23
21-Feb	81823	27,4	83	3,8	604,4	46
21-Mar	45115	28,3	78,3	3,2	244,1	14
21-Apr	25932	28,7	76,1	2,9	213,9	62
21-May	17135	29,4	76	2,6	203,6	60
21-Jun	52869	28,5	79	2,5	79,1	46
21-Jul	319725	28,8	72,8	2,4	35,8	58
21-Aug	45529	29	72,1	3,1	79,1	61
21-Sept	12344	29,2	72,7	2,7	35,8	76
21-Oct	4480	29,2	73,5	2,6	182,1	68
21-Nov	2908	28,5	76,3	3,1	134,1	59
21-Dec	1293	28,2	79,6	2,5	171,6	38

C. Variable Operational Definition

Operational is the determination of the construct or trait to be studied so that it becomes a measurable variable. The operational definition describes the particular way used to research and operate the construct, making it possible for other researchers to replicate the measurement in the same way or develop a better way of measuring the construct. While the research variables are everything in any form determined by the researcher to be studied so that information is obtained about it, then conclusions are drawn. The variables in the study consisted of the independent variable (X) and the dependent variable (Y).

1. Independent Variable (X)
   The independent variable (X) is often the stimulus, predictor, and antecedent variable. The independent variable is a variable that cannot be influenced by other variables and is marked with the letter X to make it easier for researchers to identify. The independent variable in this study is the variable of the influence of climate on the healing of COVID-19.
2. Dependent Variable (Y)
   The dependent variable is a variable that other variables can influence. It can be interpreted that the variable has a dependence on other variables and is marked with the letter (Y) to make it easier for researchers to identify. The dependent variable is the variable that is influenced or that becomes the result because of the independent variable (Sugiyono, 2016: 39). In this study, the dependent variable in question is the recovery variable from COVID-19.

D. Research Procedure

The steps in this research are as follows.

1) Data Sources
   According to Sugiyono, data sources directly provide data to data collectors (Sugiyono, 2014: 3). Based on the source, the data is divided into two, namely:
   a. Primary data

Primary data is data obtained from direct empirical research to direct actors or those involved directly by using data collection techniques.

b. Secondary data

Secondary data is data obtained from other parties or research results from other parties. Sources of data used in research conducted by the author is secondary data sources. Secondary data is data obtained from other parties or research results from other parties. Collecting data related to the number of people recovering from COVID-19, temperature, humidity, wind speed, rainfall, and sunshine duration in DKI Jakarta, Central Jakarta, from March 2020 to December 2021.

2) Data Analysis

Data analysis simplifies data into a form that is easier to interpret. Data analysis is an interpretation of research aimed at answering research questions to reveal certain social phenomena.

E. Flowchart

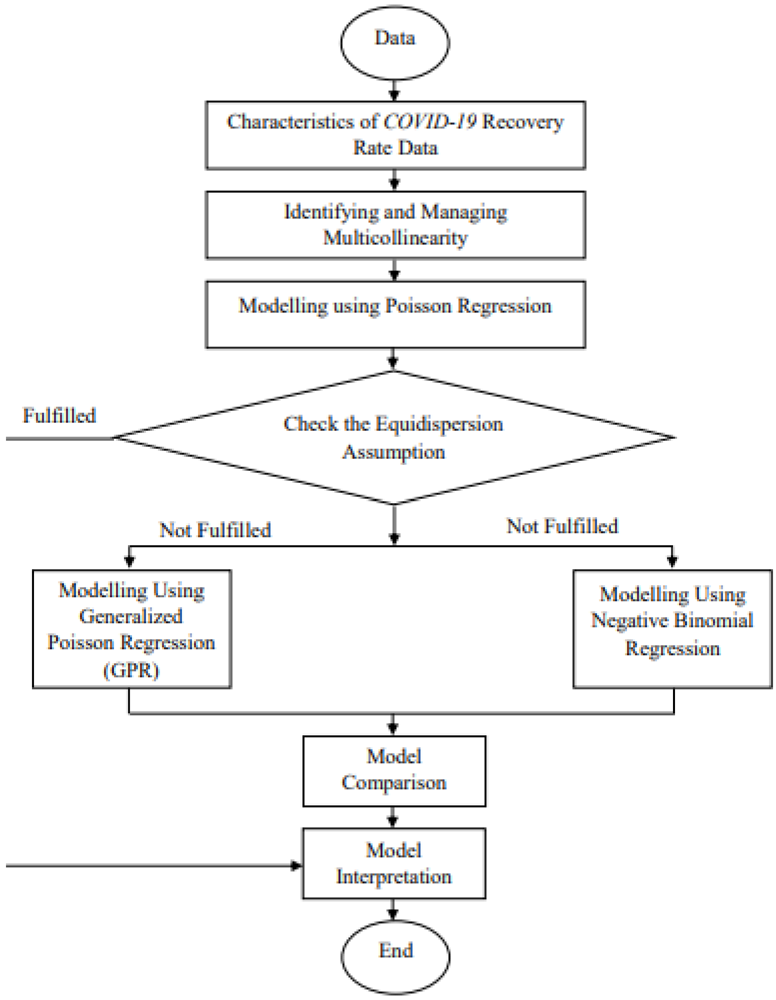

Figure 1. Poisson Regression Flowchart

IV. RESULT AND ANALYSIS

A. Characteristics of COVID-19 Recovery Rate Data in DKI Jakarta

DKI Jakarta Province, which is the capital city of Indonesia, is located between 6^0 12' latitude and 106^0 48' east longitude, which is bordered by the province of West Java in the south and east, the western and northern borders are the provinces of Banten and the Java Sea, respectively. DKI Jakarta Province is recorded to have an area of 662.33 km^2 which administratively consists of 6 regencies/cities with no less than 110 islands.

There were 865.297 cases of Covid 19 in DKI Jakarta from March 2020 to December 2021. However, the number of positive cases is also accompanied by a cure rate. The following graph also presents a graph of the recovery rate in DKI Jakarta from March 2020 to December 2021.

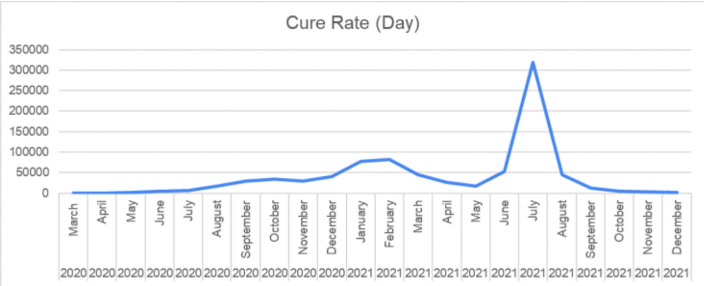

Figure 2. Cure Rate Graph of DKI Jakarta March 2020-Desember 2021

From this data, it can be seen that the highest recovery rate for *COVID-19* patients was in July, but after that, there was a decline until December 2021.

TABLE 1
CURE RATE GRAPH OF DKI JAKARTA MARCH 2020-DESEMBER 2021

Variable	Y
Variance	4521699235
Average	38691.90909
Maximum	319725
Minimum	49

Table 2 shows that the average recovery rate in DKI Jakarta from March 2020 to December 2021 is 38.692 people. Based on the table, it can also be seen that there is an indication of overdispersion because the value of variance is greater than the average value, but this will be confirmed through generalized linear model analysis at a later stage.

B. Characteristics of Influential Factors
The characteristics of the factors that influence the recovery rate of *COVID-19* patients in DKI Jakarta can also be known through descriptive statistics in the form of numbers and visuals.

TABLE 3
CHARACTERISTICS OF THE *COVID-19* RECOVERY RATE FACTORS IN DKI JAKARTA

Variable	Mean	Variance	Minimum	Maximum
Y	38691.9	4521699235	49	319725
X_1	28.7532	0.36702273	27.3	29.58
X_2	75.8423	11.6821517	70.89	83
X_3	2.23864	0.62085043	1.23	3.8
X_4	158.486	1683.2593	12.1	604.4
X_5	54	244.225406	14	79

Based on table 3, variables X_2, X_4, and X_5, there are a significant difference between the minimum, maximum and average values. This is because Indonesia has two seasons during the observation period, namely the dry and the rainy, so there is a big difference.
Another characteristic that needs to be known in this analysis is the relationship between the *COVID-19* cure rate factor. The Pearson and Spearman correlation is one indicator that can be used.

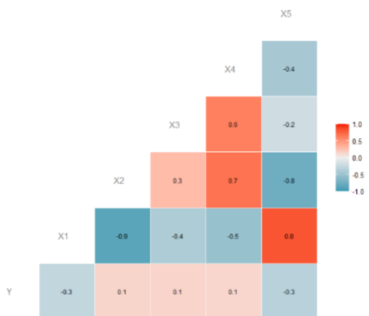
Figure 3. Spearman Correlation

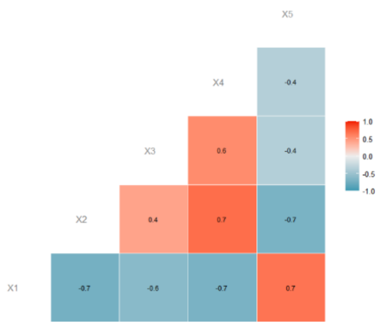
Figure 4. Pearson Correlation

Based on the two correlation tests above, the relationship with a negative value describes the opposite situation. If one variable is more significant in value, the variable will be smaller. As for the positive relationship, it states a linear condition. If one variable gets bigger, the other variables are also more significant.

At this stage the amelia packages will be used to detect missing data. From this missing map, it can be concluded that there is no missing value in the data set.

3.2 Multicollinearity Test

The classic assumption that must be met in the regression test is the absence of multicollinearity. Multicollinearity test can be done by paying attention to the Variance Inflation Factor (VIF) value. In the multicollinearity test, if the VIF value is greater than 10, there is an indication of multicollinearity in the variable concerned. Here are the VIF values for the five variables used:

TABLE 4
VIF VALUE

Variance	VIF Value
X_1	6.193936
X_2	8.66479
X_3	3.580341
X_4	6.131253
X_5	2.545279

Table 4 shows that the VIF value of all variables is less than 10, so it can be concluded that there is no indication of multicollinearity in the variables used, and the multicollinearity assumption is met.

C. Overall Test

The next step is a simultaneous test to know whether there are significant independent variables. The hypothesis in the simultaneous test is as follows.

H_0 : *there is no significant independent variable*

H_1 : *there is at least one significant independent variable*

TABLE 5
OVERALL TEST RESULTS

Variable	Df	Deviance Residual	Df Residual	Deviance
Null	1		21	14000488
X_1	1	121835	20	1278953
X_2	1	304501	19	974152
X_3	1	88092	18	886060
X_4	1	23846	17	862214
X_5	1	2322	16	859892

The value of the chi-square table (p=1, 0.05) is 3,841, so based on table 4, the deviance value is greater than the chi-square value, and the hypothesis H_0 is rejected. So, it can be concluded that there is at least one significant variable.

D. Partial Test of Poisson Regression Model

Poisson regression is performed to obtain a model with a response variable in the form of count data. The purpose of the partial test is to determine whether the independent variables partially affect the dependent variable. In this test, the hypotheses used are:

$H_0 : \beta_j = 0$

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

With critical area

Reject H_0 if $\Pr(> z)$ *less than* α, *with* $\alpha = 0.05$

Based on testing using R software with a significant level of 0.05, the following results were obtained:

TABLE 6
POISSON REGRESSION PARAMETER ESTIMATION

Variable	Estimation	Standard Error	Z-Value	Pr(>|z)	Decision	Information
X_0	9.47E+01	1.68E-01	564.64	<2e-16		
X_1	-2.20E+00	4.81E-03	-457.55	<2e-16	*Reject H_0*	Significant
X_2	-2.99E-02	8.03E-04	-371.98	<2e-16	*Reject H_0*	Significant
X_3	6.21E-01	1.97E-03	316.21	<2e-16	*Reject H_0*	Significant
X_4	-2.53E-03	1.60E-05	-157.74	<2e-16	*Reject H_0*	Significant
X_5	5.95E-03	1.26E-04	47.28	<2e-16	*Reject H_0*	Significant

Table 6 shows that all variables have a P-value greater than a significant level of 0.05, so it can be concluded that all variables significantly affect the model. The model obtained is as follows:

$$\hat{\mu} = \exp(94.7 - 2.2X_1 - 0.0299X_2 + 0.621X_3 - 0.0025X_4 + 0.0059X_5) \quad (34)$$

From the model above, it can be explained that the number of cure rates for covid patients in Jakarta will decrease exponentially by exp (2.20) = 9.025013 if the temperature (X_1) increases by one unit provided that the other independent variable is constant. This also applies to variables X_2, X_3, X_4, and X_5. The number of recovered numbers will decrease exponentially by exp (0.0299) = 1.0303 if the air humidity (X_2) decreases by one unit, provided that the other independent variables are constant. The number of recovered patients will increase exponentially by exp (6.5889) = 1.8608 if the wind speed (X_3) is increased by one unit. The number of recovered patients will also decrease exponenttially by exp (0.0025) = 1.0025 if the variable (X_4) is increased by one unit, provided that the other independent variables are constant. The number of recovered numbers will increase exponentially exp (0.0059) = 1.00591 if duration of sun exposure (X_5) is increased by one unit.

E. Poisson Regression Equidispersion Assumption Test

Some assumptions must be met in Poisson regression. Namely, the mean and variance values of the response variables have the same value as a condition for the fulfillment of the equidispersion assumption. In testing the equidispersion assumption, it can also be seen from the quotient of the deviance value to the degrees of freedom that must be 1

TABLE 7
POISSON RESULT OF EQUIDISPERSION ASSUMPTION VALUE

	Value	df	Value/df
Deviance	859892	16	53743.25

Based on table 7, it can be concluded that the entire model cannot be used because it does not meet the equidispersion assumption.

F. Negative Binomial Regression Partial Test

Negative binomial regression is used to overcome the oversdispersion in Poisson regression. The results of parameter estimation for the negative binomial regression model can be seen in the following table:

TABLE 8
TOP 5 MODEL ASSUMPTION

Variable	Significant Parameters	AIC	Dispersion Parameter
X_3	β_0, β_3	503.9649	0.5846
X_1, X_2	$\beta_0, \beta_1, \beta_2$	499.5845	0.7195
X_1, X_2, X_3	$\beta_0, \beta_1, \beta_2$	499.2627	0.7776
X_1, X_2, X_3, X_4	$\beta_0, \beta_1, \beta_2, \beta_3$	500.6915	0.7927
X_1, X_2, X_3, X_4, X_5	$\beta_0, \beta_1, \beta_2, \beta_3$	502.5942	0.7953

Based on the table above, model 4 has a small AIC value and there are 3 variables that have significance value.

TABLE 9
NEGATIVE BINOMIAL PARTIAL TEST RESULT

Parameter	Estimation	Standard Error	Z-Value	Pr(>|z)	Decision	
β_o	90.15225	25.837913	3.489	0.000485	*Reject H0*	Significant
β_1	-1.97159	0.700187	-2.816	0.004866	*Reject H0*	Significant
β_2	-0.32769	0.115508	-2.837	0.004555	*Reject H0*	Significant
β_3	0.812067	0.409625	1.982	0.047427	*Reject H0*	Significant
β_4	-0.00191	0.003005	-0.635	0.525544	*Accept H0*	Unsignificant

Table 9 shows the parameter estimation and the z value for partially testing the significance of the negative binomial regression parameter. The hypothesis used is that the H_0 parameter does not significantly affect the model. The significant level used is 0.05, so based on table 4.7, there are three significant parameters, namely β_1, β_2, β_3 , which is the variables of temperature, humidity, and wind speed. Here is the obtained model

$$\hat{\mu} = \exp(90.15225 - 1.97159X_1 - 0.32769X_2 + 0.812067X_3) \quad (35)$$

Or it can also be expressed in linear form as follows:

$$\ln(\hat{\mu}) = 90.15225 - 1.97159X_1 - 0.32769X_2 + 0.812067X_3 \quad (36)$$

The model illustrates that every 1% decrease in the average temperature percentage will increase the cure rate for *COVID-19* patients by 0.129. For a 1% decrease in air humidity, there will be an increase of 0.740 *COVID-19* recovery rates. Based on the model interpretation results, it was found that there was a mismatch in the direction of the relationship between several predictor variables and a large number of *COVID-19* cure rates. This is possible due to a reasonably high correlation between the variable predictor.

G. Generalized Poisson Regression Partial Test

Generalized Poisson Regression was carried out to overcome overdispersion in the Poisson model. Generalized Poisson regression can overcome overdispersion because the probability distribution function contains the dispersion parameter. The following is a summary of the modeling possibilities carried out in the five best models that have been selected based on the smallest AIC and the most significant parameters.

TABLE 10
TOP 5 MODEL ASSUMPTION IN GPR

Variable	Significant Parameters	AIC
X_3	β_0	512.7356
X_1, X_2	$\beta_0, \beta_1, \beta_2$	508.1198
X_2, X_3, X_5	$\beta_0, \beta_2, \beta_3$	501.4445
X_1, X_2, X_3, X_5	$\beta_0, \beta_1, \beta_2, \beta_3, \beta_5$	491.6475
X_1, X_2, X_3, X_4, X_5	$\beta_0, \beta_1, \beta_2, \beta_3$	492.0908

Table 10 shows that model 4 has the smallest AIC, followed by the most significant parameters. The following are parameter estimates with variables X1, X2, X3, and X5.

TABLE 11
GENERALIZED POISSON REGRESSION PARAMETER ESTIMATION

Parameter	Estimation	Standard Error	Z-Value	Pr(>|z)	Decision	
β_0	-92.7663	61.16558	-1.517	0.129357		
β_0	-4.67494	0.15741	-29.461	<2e-16		
β_1	5.13501	1.46178	3.513	0.000443	*Reject H_0*	Significant
β_2	-1.05664	0.26646	-3.965	7.33E-05	*Reject H_0*	Significant
β_3	11.90205	2.62115	5.546	2.92E-08	*Reject H_0*	Significant
β_5	0.21532	0.21535	3.762	0.000168	*Reject H_0*	Significant

Based on the table above, four variables reject H_0 because the P-value is less than 0.05.

$$\hat{\mu} = \exp(-92.7663 - 4.67494 + 5.13501X_1 - 1.05664X_2 + 11.90205X_3 + 0.21535X_5) \tag{37}$$

Based on the model above, it can be concluded that the number of recovered *COVID-19* patients will increase by 169.86 if the temperature is increased by one unit, provided that the other variables are constant. This number also increases by 147,568.83 people if the wind speed increases by one unit. This number will also increase by 1.24 if the duration of solar radiation increases by one unit. However, the total cure rate will decrease by 2.88 if the humidity is increased by one unit.

H. Best Model Selection

The best model selection is made by choosing the smallest AIC value. The following summarizes the selected models of poisson regression, negative binomial, and generalized poisson regression.

TABLE 12
TOP 5 MODEL ASSUMPTION

Method	AIC	Significant Parameter	Overdispersion
Poisson	860150.3	5	Yes
Negative Binomial	502.5942	3	No
GPR	492.0908	4	No

It can be concluded that the best model is generalized Poisson regression with three significant parameters with the model:

$$\hat{\mu} = \exp(-92.7663 - 4.67494 + 5.13501X_1 - 1.05664X_2 + 11.90205X_3 + 0.21535X_5)$$

V. CONCLUSION

Based on the analysis and result, the following conclusions are obtained:

1. Based on the results of the analysis using Poisson regression, negative binomial, and Generalized Poisson regression, it can be concluded that the best model in analyzing the factors that influence the recovery rate of COVID-19 patients in DKI Jakarta based on the AIC value is Generalized Poisson regression.

2. After analyzing using generalized Poisson regression, it was found that the factors that influence the Covid 19 recovery rate in DKI Jakarta are temperature, air humidity, wind velocity, and duration of sun exposure with the following equation:
$$\hat{\mu} = \exp(-92.7663 - 4.67494 + 5.13501X_1 - 1.05664X_2 + 11.90205X_3 + 0.21535X_5)$$
3. Based on the model above, it can be concluded that the number of recovered *COVID-19* patients will increase by 169.86 if the temperature is increased by one unit, provided that the other variables are constant. This number also increases by 147,568.83 people if the wind speed increases by one unit. This number will also increase by 1.24 if the duration of solar radiation increases by one unit. However, the total cure rate will decrease by 2.88 if the humidity is increased by one unit
4. Sugiyono, as reported by CNN News, said that the decreased resistance of the virus due to high temperature and humidity automatically made the spread of the virus predicted to slow down even more. The duration of solar radiation is also useful for keeping the earth's temperature warm (Hamdi, 2014) so this is in line with the results of the interpretation of this research model. For now, the correlation of wind speed to the recovery rate still needs to be researched.

REFERENCE

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, and W. Tan, “A Novel Coronavirus from Patients with Pneumonia In China,” *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020, https://doi.org/10.1056/nejmoa2001017

[2] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, Z. Feng, et al, “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia”, *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199-1207, 2020, doi:10.1056/nejmoa2001316

[3] Y. Wang, Y. Chen, and Q. Qin, “Unique Epidemiological and Clinical Features of The Emerging 2019 Novel Coronavirus Pneumonia (COVID-19) Implicate Special Control Measures,” *Journal of Medical Virology*, vol. 92, no. 6, pp. 568-576, 2020, doi:10.1002/jmv.25748

[4] Ministry of Health Republic of Indonesia, 2020b, “Guidelines for the Prevention and Control of Corona Virus Disease (COVID-19) July 2020,” 13 July 20, *Jakarta: Ministry of Health Republic of Indonesia*.

[5] World Health Organization, 2020, “Timeline of WHO’s response to COVID-19”, *Newsroom*. Available: https://www.who.int/news-room/detail/29- 06-2020-covidtimeline [Accessed at 8 April 2022].

[6] World Health Organization, “Climate change and human health: risks and responses: summary,” 2003, *World Health Organization*. Available: https://apps.who.int/iris/handle/10665/42749

[7] P. Shi, Y. Dong, H. Yan, C. Zhao, X. Li, et al, “Impact Of Temperature On The Dynamics Of The Covid-19 Outbreak In China,” *Science Of The Total Environment*, vol. 728, no. 138890, 2020, doi: 10.1016/j.scitotenv.2020.138890

[8] R. Tosepu, J. Gunawan, D.S. Effendy, L.O Ahmad, H. Lestari, H. Bahar, and P. Asfian, “Correlation between weather and COVID-19 pandemic in Jakarta, Indonesia,” *Science of The Total Environment*, vol. 725, no. 138436, 2020, doi: 10.1016/j.scitotenv. 2020.138436

[9] R.H. Myers, *Classical and Modern Regression with Application*, PWS-KENT Publishing Company, 1990.

[10] W. W. Hines, “Probability And Statistics In Engineering and Management Science,” Hoboken, New Jersey: John Wiley and Sons, 2003.

[11] L. Amaliana, “Geographically Weighted Zero Inflated Poisson Regression Model, Case Study: The Number of Filariasis Cases in East Java 2012,” S,Si Thesis, Institut Teknologi Sepuluh Nopember, Surabaya, 2014.

[12] J, M. Hilbe, *Negatif Binomial Regression*, 2nd ed. Cambridge: Cambridge University Press, 2011, doi: 10.1017/cbo9780511973420

[13] Mc P. Cullagh and J.A. Nelder, *Generalized Linier Models*, 2nd ed. Chapman and Hall, London, 1989, doi: 10.1007/978-1-4899-3242-6

[14] H.A. Park, “An Introduction To Logistic Regression: From Basic Concepts To Interpretation With Particular Attention To Nursing Domain,” *Journal of Korean Academy of Nursing*, vol. 43, no. 2, pp. 154, 2013, https://doi.org/10.4040/jkan.2013.43.2.154

[15] Marlene Müller, *Generalized Linear Model: Fraunhofer Institute for Industrial Mathematics (ITWM)*, Kaiserlauten, 2014.

[16] F. Fadhillah, “Aplikasi Regresi Binomial Negatif Dan Generalized Poisson Dalam Mengatasi Overdispersion Pada Regresi Poisson (Studi Kasus Data Kemiskinan Provinsi di Indonesia Tahun 2009),” *Institutional repository Uin Syarif Hidayatullah Jakarta*, 2019, Available: https://repository.uinjkt.ac.id/dspace/

[17] S. Abdullah, S. Susilo, S. Ula, A. Aswata, N. Valentika, and S. I. Chasanah, “Algoritma Membangkitkan Proses Poisson Majemuk Dengan Komponen Proses Poisson Nonhomogen Fungsi Linear dan Komponen Berdistribusi Eksponensial,” *Statmat: Jurnal Statistika Dan Matematika*, vol. 2, no. 1, pp. 81, https://doi.org/10.32493/sm.v2i1.4224

[18] H. Ihsan. W. Sanusi, and R. Ulfadwiyanti, “Model Generalized Poisson Regression (GPR) dan Penerapannya pada Angka Pengangguran bagi Penduduk Usia Kerja di Provinsi Sulawesi Selatan,” *Journal of Mathematics, Computations, and Statistics*, 2020, pp. 109 – 117.

[19] Jakarta Tanggap COVID-19, (n.d), Retrieved April 8, 2022, Available: https://corona.jakarta.go.id/id

[20] BPS provinsi DKI Jakarta, (n.d.), Retrieved April 8, 2022, Available: https://jakarta.bps.go.id/publication/2021/02/26/bb7fa6dd5e90b534e3fa6984/provinsi-dki-jakarta-dalam-angka-2021.html

[21] BPS provinsi DKI Jakarta, (n.d.), Retrieved April 8, 2022, Available: https://jakarta.bps.go.id/publication/2022/02/25/5979600247867d861a1f334c/provinsi-dki-jakarta-dalam-angka-2022.html