Predicting the Number of Forest and Land Fire Hotspot Occurrences Using the ARIMA and SARIMA Methods

Angga Bayu Santoso [1] \*, Tri Widodo [2]

Faculty of Engineering and Computer Science [1], [2]

Universitas Teknokrat Indonesia

Bandar Lampung, Lampung, Indonesia

anggabayu@teknokrat.ac.id [1], tri\_widodo@teknokrat.ac.id [2]

***Abstract*—** **Forests are an area and part of the environmental cycle that is very important for survival because forests are areas on Earth that regulate the balance of the ecosystem. Forest fires rank second to illegal logging in Indonesia's list of forest destruction causes. Forest fires can occur due to two factors, namely natural and human factors. Therefore, the hotspot factor that can cause forest fires is an independent variable. The population of hotspots in the West Kalimantan region in 2020 amounted to 1,416 hotspots. This study aims to predict the number of hotspots on land and forests that cause fires before the fires spread and are challenging to overcome or extinguish. The method to indicate the number of hotspot occurrences uses the Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) methods. Modeling ARIMA (0,1,1) and SARIMA (0,1,1) (2,2,1)12 obtained Root Mean Square Error (RMSE) evaluation results for ARIMA of 6.61 while SARIMA of 7.61. The Mean Squared Error (MSE) evaluation value for ARIMA is 43.70 and SARIMA is 58.05. Based on these results, it can be concluded that the ARIMA model provides excellent and accurate performance in describing the trend of hotspot events that will occur in the future with a smaller RMSE value compared to SARIMA.**

***Keywords—*** **Forestry, Land and Forest Fires, Hotspot, ARIMA, SARIMA**

***Abstrak*— Hutan merupakan wilayah dan bagian dari siklus lingkungan hidup yang sangat penting untuk kelangsungan hidup, sebab hutan merupakan wilayah di bumi yang mengatur keseimbangan ekosistem. Kebakaran hutan berada di peringkat kedua setelah *ilegal logging* dalam daftar penyebab kerusakan hutan di Indonesia. Bencana kebakaran hutan dapat terjadi karena dua faktor, yakni faktor alam dan manusia. Oleh karena itu, faktor titik panas yang dapat menimbulkan kejadian kebakaran hutan berperan sebagai variabel independen. Populasi titik api wilayah Kalimantan Barat tahun 2020 berjumlah 1.416 titik. Tujuan dari penelitian ini untuk memprediksi jumlah kemunculan titik api pada lahan dan hutan yang menyebabkan kebakaran, sebelum kebakaran tersebut menyebar dan sulit untuk diatasi atau dipadamkan. Metode yang digunakan untuk melakukan prediksi jumlah kemunculan *hotspot* menggunakan metode *Autoregresif Integrated Moving Average* (ARIMA) dan *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Pemodelan ARIMA (0,1,1) dan SARIMA (0,1,1) (2,2,1)12 didapatkan hasil evaluasi *Root Mean Square Error* (RMSE) untuk ARIMA sebesar 6,61 sedangkan SARIMA sebesar 7,61. Nilai evaluasi *Mean Squared Error* (MSE) untuk ARIMA sebesar 43,70 dan SARIMA sebesar 58,05. Berdasarkan hasil tersebut dapat disimpulkan bahwa model ARIMA memberikan kinerja yang baik dan akurat dalam menggambarkan trend kejadian titik api (*hotspot*) yang akan terjadi di masa mendatang dengan nilai RMSE lebih kecil di bandingkan SARIMA.**

***Kata Kunci—*** Kehutanan, Kebakaran Hutan dan Lahan, Hotspot, ARIMA, SARIMA

# Introduction

Forests are areas and parts of the environmental cycle that are very important for survival because forests are areas on Earth that regulate the balance of the ecosystem [1]. According to the Basic Forestry Law No.41 of 1999 concerning Forestry, a forest is an ecosystem unit in the form of an expanse of land containing biological natural resources dominated by trees in their natural environment, one and the other inseparable [2]. Indonesia is a country where almost all provinces have forest areas. Indonesia's forest and water conservation areas 2019 were around 123.8 million hectares [3].

But today, Indonesia's deforestation rate is still relatively high, with around 115.46 thousand hectares per year in 2020 [4]. The high rate of deforestation can undoubtedly threaten the existence and sustainability of forests in Indonesia. One of the most significant contributors to the high rate of deforestation is the forest and land fires carried out illegally by local communities and large private companies to clear land [5].

Forest and land fires rank second to illegal logging in Indonesia's list of forest destruction causes [6]. Every year, millions of hectares of forest are burned. Forest and land fires in Indonesia generally occur on the islands of Sumatra and Kalimantan, where both islands are areas dominated by wetlands, especially flammable peatlands [7]. High organic matter content, porous nature, and low vertical conductivity are the causes of peatlands being highly volatile. Fires on peatlands are difficult to extinguish because they spread below the ground surface

A hotspot is a forest fire indicator that detects an area with a relatively higher temperature than the surrounding temperature [10]. The site is represented in a point with specific coordinates. Hotspots are usually used as an early indicator to detect the possibility of forest fires in an area; the greater the number of hotspots that appear, the greater the potential for forest and land fires that occur in a room [9]. Hotspots can be used as a reference for early identification of forest and land fires [11].

West Kalimantan is a tropical region with high air temperatures, with an average air temperature of 27.6°. The population of hotspots in the West Kalimantan region in 2019 was recorded at 4,005 points [12]. By this situation, the government needs effective and efficient forest and land fire prevention measures in the West Kalimantan region. One of the effective preventive measures the government can take is observing the appearance of hotspots through satellites. Hotspots that appear in forests and land can be observed through the National Aeronautics and Space Administration (NASA) satellite

Several related studies based on similar problems and solutions in preventing forest and land fires, one of which is a study aimed at forecasting the number of hotspots on Kalimantan peatlands using a zero-inflated Poisson regression model based on climatic factors, namely sun exposure and rainfall. This model can map the Kalimantan region into four good areas, with the smallest RMSE value of 10.05 in area 2

Different from the previous research, another study aims to predict the occurrence of hotspots in peatlands in Riau province using long- and short-term memory. The research made predictions for the next six months, from August 2019 to January 2020. LSTM can predict time series with an RMSE value of 363.38 [13]. The research aimed at predicting the number of occurrences of hotspots was also carried out in the Rokan Hilir district using the Elman recurrent neural network, with the results of the research producing a constant actual value with the prediction results, namely the RMSE value of 437.603 [14]. Furthermore, a similar study was conducted to predict the number of hotspot events through a time series approach using the seasonal arima model using the NASA MODIS satellite data. The research produced an excellent prediction model with an RMSE value of 5.85 [15].

Another method in prediction techniques (time series) is the Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model is a model that completely ignores independent variables in forecasting [16]. ARIMA aims to determine an excellent statistical relationship between the predicted variable and the historical value of the variable so that forecasting can be done with the model [17]. Furthermore, the ARIMA model was developed into the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. SARIMA is the development of the ARIMA model, which can analyze recurring or seasonal data patterns when time is fixed, such as quarterly, semi-annually and annually. This prediction model aims to determine a relationship between predicted variables through data on the historical value of these variables so that predictions can be made [18].

Based on the previous problems and some related research that has been mentioned, this research aims to build a prediction model for the number of forest and land fire hotspot occurrences in the West Kalimantan region using the ARIMA and SARIMA methods using data from the early detection information system for forest and land fire control (Sipongi) from 2013 to 2022.

# Literature Review

## Forest and Land Fires

Forest fires are one of the environmental problems that threaten forests, causing adverse environmental impacts, creating economic problems, and ecological damage [19]. Forest fires are burning events that spread freely in unplanned areas and use natural fuels from the forest, such as litter, grass, twigs or dead tree branches [20].

## Hotspot.

Hotspots are indicators of forest fires that detect an area with a relatively higher temperature than the surrounding temperature [10]. Hotspots are points on the image (pixels or sub-pixels) that have very high temperatures and are associated with active Earth surface fires [21].

## Data Mining

Data mining combines several computer science disciplines defined as discovering new patterns from massive data sets, including methods related to artificial intelligence [22].

## Prediction

Prediction is systematically estimating something that will happen in the future based on past and present information to get forecast results close to the actual results. There are two techniques for predicting, namely qualitative prediction techniques and quantitative prediction techniques [23].

## Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) is a forecasting model that produces forecasts based on synthesizing historical data patterns. The ARIMA method will work well if the data in the time series are dependent or statistically related [24]. The advantage of the ARIMA method is that it is suitable for forecasting data with a simple and relatively easy application in analyzing data containing seasonal and trend patterns [24].

## Seasonal Autoregressive Integrated Moving Average (SARIMA)

Seasonal Autoregressive Integrated Moving Average (SARIMA) is developing the Autoregressive Integrated Moving Average (ARIMA) model on time series with seasonal patterns. The SARIMA model is denoted by ARIMA (*p,d,q*) (𝑃, 𝐷, 𝑄)𝑠. In the SARIMA model, two parts are designated by lowercase and capital letters, and (*p,d,q*) is the notation of the part of the model that is not seasonal. While (*P, D, Q*) is the notation of the seasonal component of the model. The power s is a notation for calculating the number of periods per season [25].

# Research Method

The methodology in this study adopts the process in the ARIMA and SARIMA model framework [26]. The central core of the framework consists of five stages: data collection or acquisition, preprocessing, model identification, model parameter estimation, diagnostic checking, prediction and result evaluation.

## Akuisisi Data

This study uses hotspot data in the West Kalimantan region. The hotspot data is obtained from the early detection information system for forest and land fire control (Sipongi)—the dataset used from 2013 to 2022. Features or variables in the dataset include date, district, sub-district, village, time, satellite and confidence (hotspot level). The total data obtained was 166,980 records.

## Preprocessing

After collecting the data, the next step is to preprocess the forest and land fire hotspot data in West Kalimantan. This process aims to prepare raw data that will later be used as data for modelling input so that the resulting model will be of higher quality. Thus, data preprocessing is crucial to identifying and removing irrelevant and redundant features. In addition, the implication of these conditions will reduce the dimensionality of the data and allow the model to operate more quickly and effectively.

## Model Identification

In this research, model identification is carried out with several steps, namely as follows:

1. They plotted the time series data to see the stationarity of the data, whether it is stationary either in mean or variance. After that, Decomposing time series data into several sub-components is carried out to determine the effect of each component on the data series. Generally, two models are used, namely Additive Decomposition and Multiplicative Decomposition.
2. In addition to data plotting, the data stationary test can use the Augmented Dickey-Fuller Test to see whether the statistical test value is lower than the critical value. Also, the time series data is fixed by looking at the p-value, whether the value is smaller than 0.05. If not, then the data is non-stationary. If the time series data is not stationary to the mean or variance, it can be fixed with the Differencing process. Differencing is calculating the change or difference in observation values [27]. The difference value obtained is rechecked whether it is stationary or not. Stationarity means that there is no growth or decline in the data. So, data fluctuations are around a constant average value, independent of time and variance of these fluctuations or remain stable at all times [28]. The differencing equation [29], is expressed in equation (1).

(1)

Where:

: First differencing

: *X* value at order *t*

: *X* value at order *t-1*

1. Once the data is stationary, then use the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify the model. This step is done by placing seasonally patterned data by specifying the ACF plot and PACF plot to determine the order of the model used.

## Model Parameter Estimation

At this stage, the parameters of Moving Average (MA), Autoregressive (AR), seasonal, and non-seasonal are estimated, and the significance of each parameter is tested. If there is a model whose parameters are not significant, it is declared not to pass the test. Each insignificant parameter will be eliminated to obtain a model with substantial parameters.

## Diagnostic Checking

Diagnostic Checking testing in this study determines whether the model used is suitable and feasible for forecasting. The appropriate conjecture model has properties like the original data. This test is carried out by conducting a white noise diagnostic test and a standard distribution test using model diagnostics. The best model is that the residuals obtained are expected to have white noise properties, namely residuals that follow a normal distribution.

## Prediction

In this research, the prediction stage is carried out with two models, namely ARIMA and SARIMA. The ARIMA model is a time series analysis approach using autocorrelation and time series residual variation. The ARIMA model arrangement consists of Autoregressive (AR), Moving Average (MA) and Integrated (I) models. Integrated shows the value of the differencing order of data from non-stationary to stationary. The general form of the ARIMA model [30] is expressed in equation (2).

(2)

Dimana:

: Autoregressive parameters

: Backward sliding operator

*d* : Differencing Parameter

: Observation value at time *t*

: Constant parameters

: Moving average parameters

: Residual value (*error*)

While the SARIMA model on time series data to identify patterns in past historical data to estimate future variables. The SARIMA (*p,d,q*) (*P,D,Q*)*s* model equation with mean where *p,d,q, P, D,* and *Q* are integers and (*s*) is periodicity. The general form of the ARIMA model is expressed in equation (3).

(3)

Where:

: Non-seasonal autoregressive level

: Seasonal autoregressive rate

: Non-seasonal differencing level

: Seasonal differencing level

: *Non-seasonal moving average*

: *Seasonal moving average*

: Actual data t-th

: t-period error

## Evaluation of Results

Evaluate the results of forecasting that has been done. The forecasting results will be a graph predicting the number of future hotspot occurrences in the West Kalimantan region. Use Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for system testing. This test is carried out to analyse whether the system is running well or still needs updating so that the system can be by the needs. As for MSE can be calculated by equation (4).

(4)

Where:

: Original data value

: Forecasting data value

*n* : Amount of data

Meanwhile, the RMSE can be calculated with equation (5).

*R* (5)

Where:

: Nilai data asli

: Forecasting data value

*n* : Amount of data

Mean Squared Error (MSE) is the average squared error between actual and forecasting values. The Mean Squared Error method is generally used to check the estimation of how much the error value is in forecasting. Meanwhile, the Root Mean Squared Error (RMSE) obtains the error rate of the forecasting results. MSE or RMSE, if the results obtained are smaller or closer to zero, the prediction results will be more accurate. If the results obtained are equal to zero, then the prediction results are perfect.

# Results and Discussion

This section discusses the results of each stage described in the previous section, starting from the data collection stage, preprocessing, transformation (differencing), model identification, model parameter estimation, diagnostic checking, and prediction and evaluation of results.

## Data Acquisition

This study uses hotspot data in the West Kalimantan region. The hotspot data is obtained from the early detection information system for forest and land fire control (Sipongi). The dataset used is time series data from 2013 to 2022. Features or variables in the dataset include date, district, sub-district, village, time, satellite and confidence (hotspot level). The total data obtained was 166,980 records. The details of the hotspot dataset can be seen in Fig 1.



Fig 1. West Kalimantan hotspot dataset 2013-2022

## Preprocessing

### Missing Value Handling

After collecting the data, the next step is to preprocess the forest and land fire hotspot occurrence dataset in West Kalimantan. Preprocessing is done by checking and removing missing values. The results of checking missing values can be seen in Fig 2.

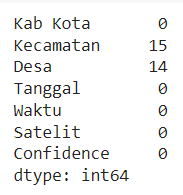


Fig 2. Result of checking missing values dataset

The check results obtained missing values in the sub-district variable as much as 15 data and in the village variable as much as 14. Therefore, the missing values can be handled by deleting the data rows containing missing values, and the results can be seen in Fig 3.

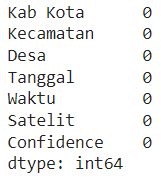


Fig 3. Results of handling missing values dataset

Missing values handling is done by deleting data in empty rows or columns (null) so that the amount of data decreases after handling missing values to 166,965 records. As for seeing the details of the data frame data after holding missing values, it can be seen in Fig 4.



Fig 4. Dataset of missing value handling results

### Data Integration

Furthermore, data integration is carried out by checking and changing the data type, sorting data based on district variables, date and confidence (hotspot level), grouping the number of hotspot events, and grouping hotspot data based on date, month and year. Based on this process, 120 records were generated with two variables, namely date (index variable) and count (hotspot count variable). The results of data integration can be seen in Fig 5.

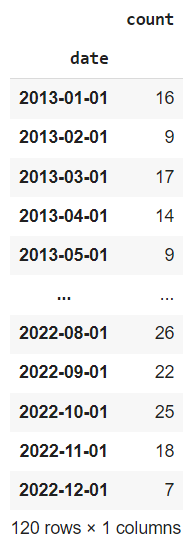


Fig 5. Datasets resulting from the data integration stage

## Model Identification

Several processes will be carried out in the model identification stage, such as making time series plots, performing the Augmented Dickey-Fuller Test stationary test, and making ACF and PACF plots. This identification aims to get the best model that matches the available data, namely West Kalimantan Forest and land fire hotspot data.

### Create a time series plot.

Make a time series plot to see the data's stationarity in mean or variance. Then, the data is used to decompose the components in the time series data. The results of the time series plot can be seen in Fig 6.

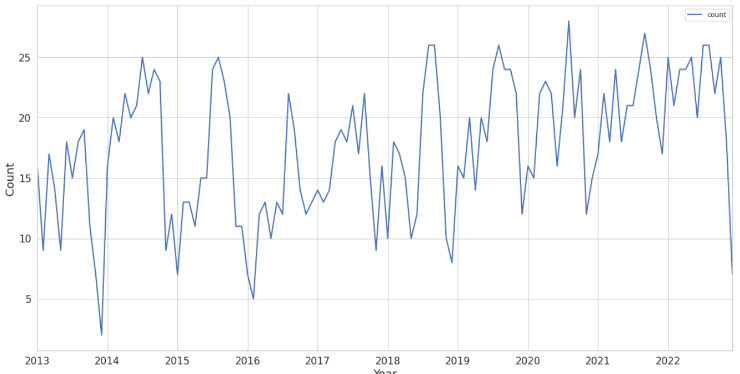


Fig 6. Time series plot of the hotspot dataset

The resulting time series plot results show that the mean and variance of hotspots are not constant throughout 2013-2022, so the hotspot data is not stationary in mean and variance.

After seeing the trend of the data from the time series plot, decompose the data, which aims to decompose the time series data into several components or identify seasonality and trends from a series of data. The results of the hotspot dataset decompose can be seen in Fig 7 and Fig 8.

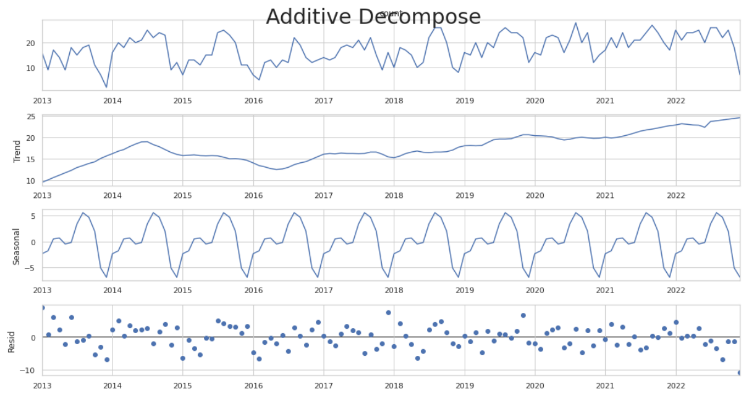


Fig 7. Additive decomposition plot of data hotspots

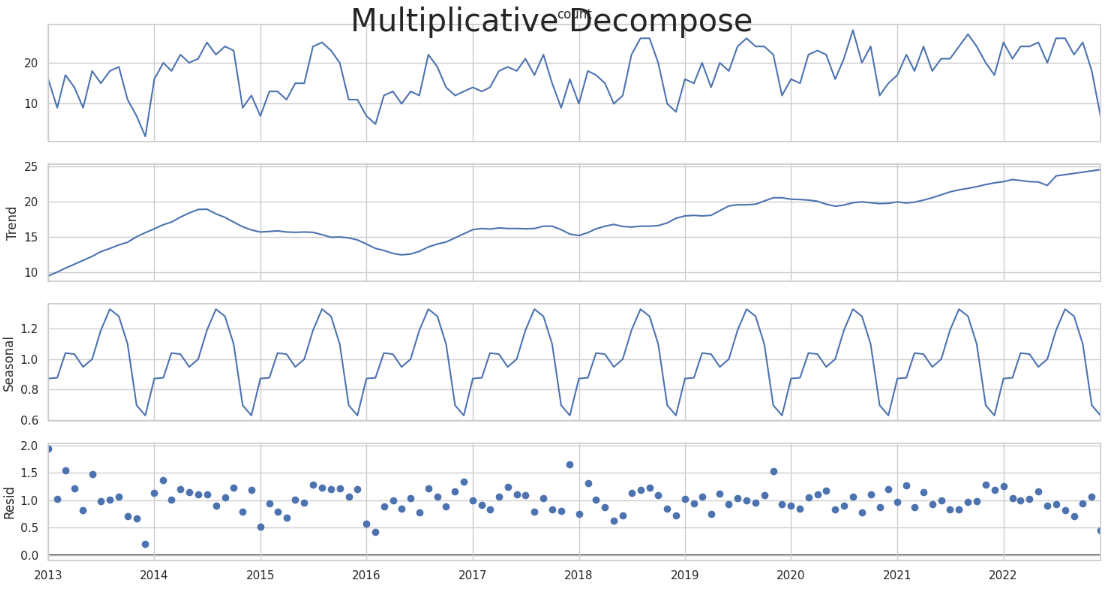


Fig 8. Multiplicative decompose plot of hotspot data

The results of additive decomposition and multiplicative decomposition explain the trend component related to the fluctuation of the data; the seasonal part shows the recurring pattern of variation that occurs in the hotspot data, and the residual component shows the remaining noise from the hotspot data. The three members explain the fluctuations of non-stationary time series data.

### Augmented Dickey-Fuller Test

They are using the Augmented Dickey-Fuller Test to see the value of the test statistic, whether it is lower than the critical value or not. Also, by looking at the p-value, whether the value is smaller than 0.05 so that the time series data can be said to be stationary; otherwise, the data is non-stationary. Based on the time series plot and decomposed data, it is concluded that the data is non-stationary. This is evidenced by the ADF test value, namely the p-value, with a value of 0.4567. The results' data plot and ADF value before differencing can be seen in Fig 9 and Fig 10.

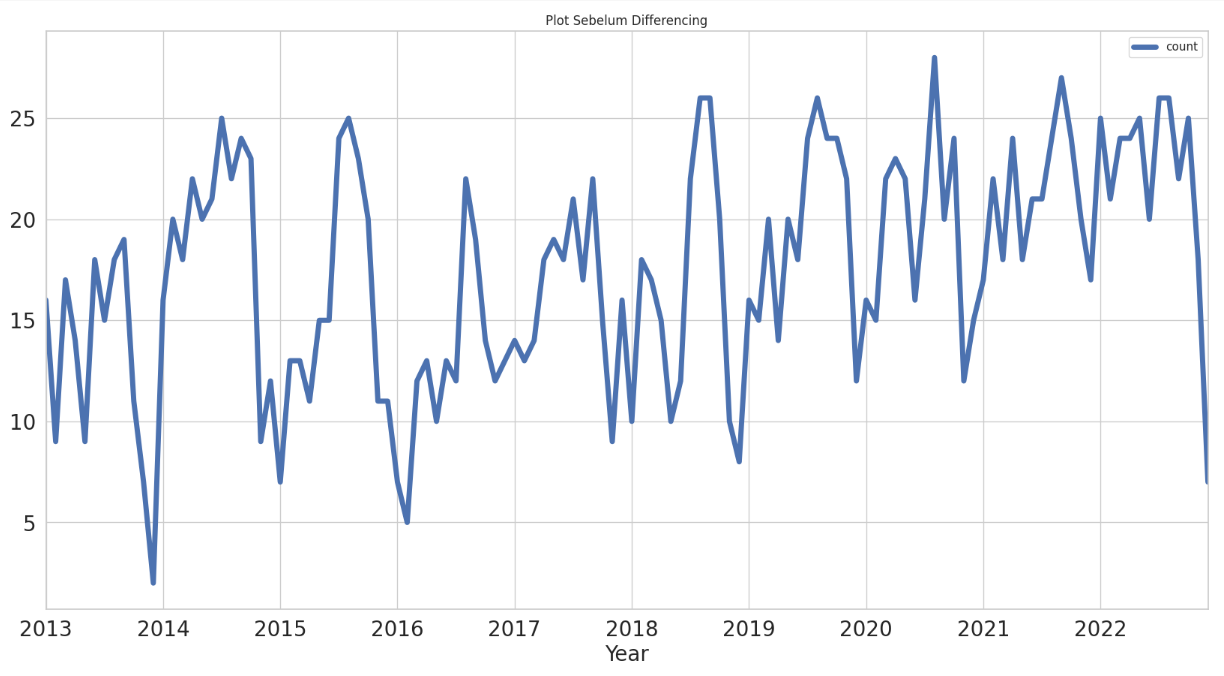


Fig 9. Plot of data before differencing

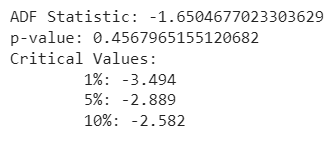


Fig 10. ADF test results before differencing

The next stage of differencing is calculating the change or difference in observation values. In this study, differencing was carried out two times to make the data stationary. The plot of differencing results can be seen in Fig 11 and Fig 12.

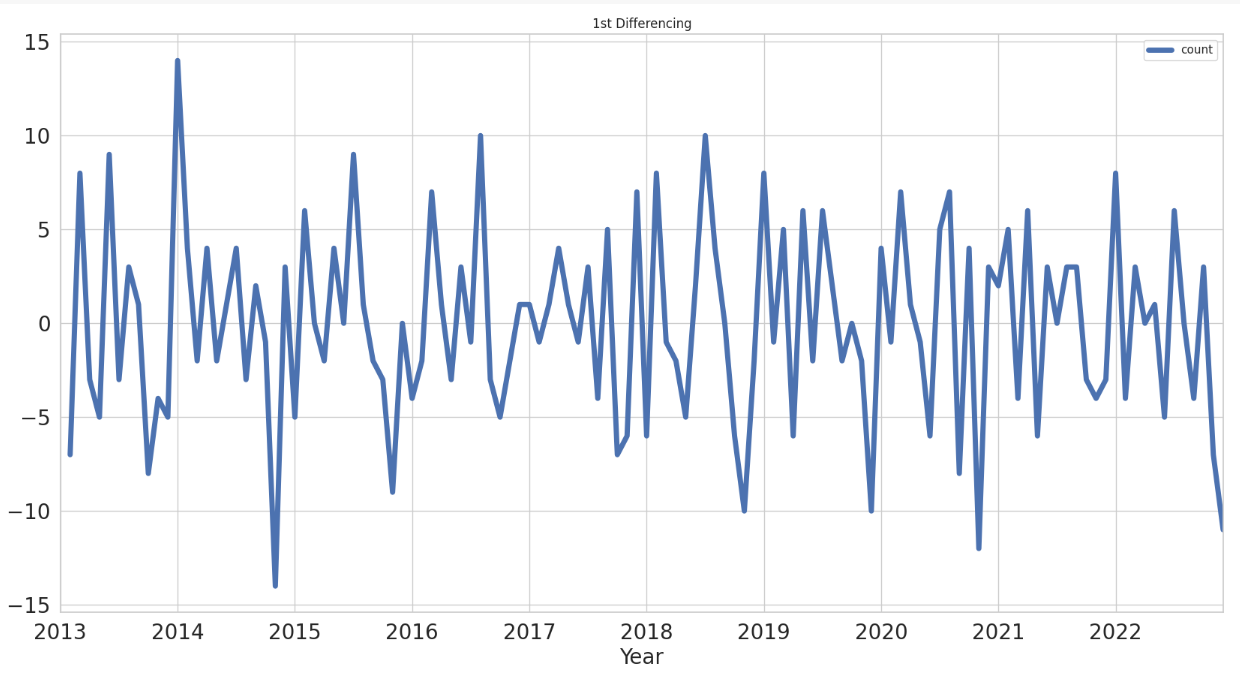


Fig 11. Data plot after first differencing

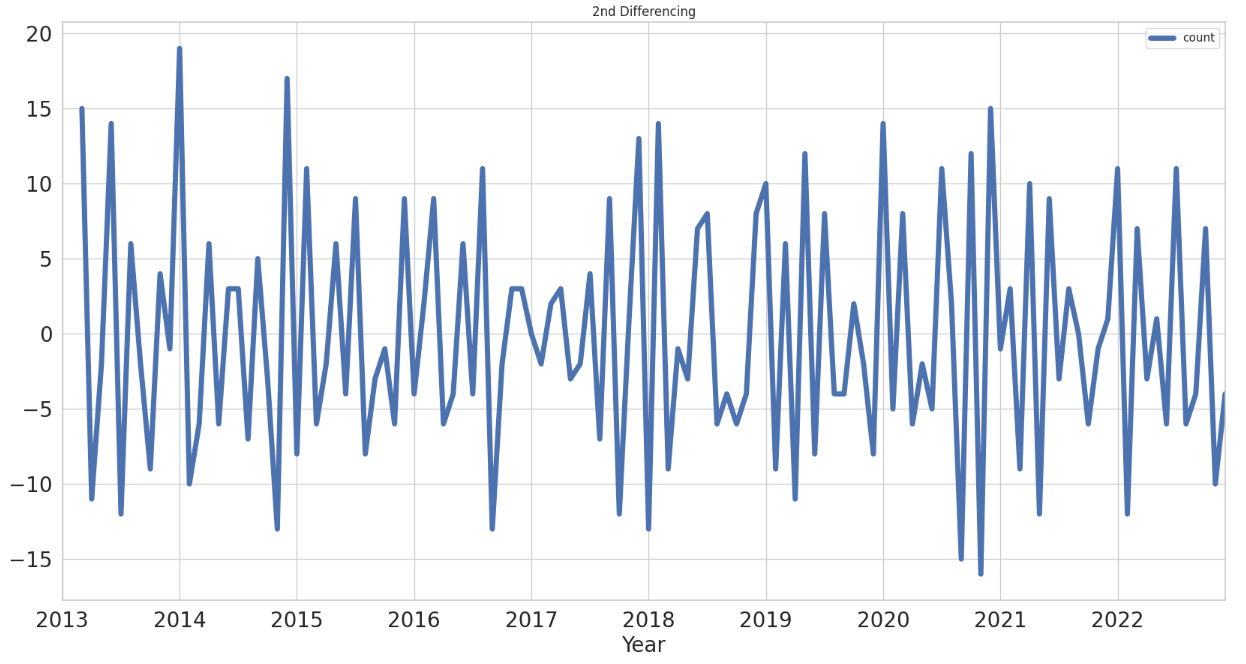


Fig 12. Data plot after second differencing

The ADF test value after differencing is the p-value with a value of 0.046. Based on these results, it can be concluded that the hotspot data shows stationary data.

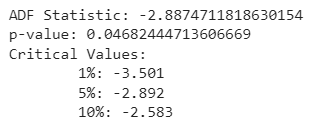


Fig 13. ADF test results after differencing

### Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

ACF and PACF plots after the data is stationary, and the next step is to identify seasonally patterned data by specifying the ACF plot and PACF plot to determine the order of the model used.

1. *Autocorrelation Function* (ACF)

The ACF plot is used to identify the order of the Moving Average (MA) model. This study aims to see the stationarity of the data in the mean. In Fig 14, it can be seen that lags 1, 2, 3, 11, 12 and 13 pass through the confidence interval. These results explain that there is still autocorrelation, and the data is not stationary in the mean. In Fig 15, it can be seen that the lags are already relatively in the confidence interval, so the data is more stationary compared to the previous ACF plot.

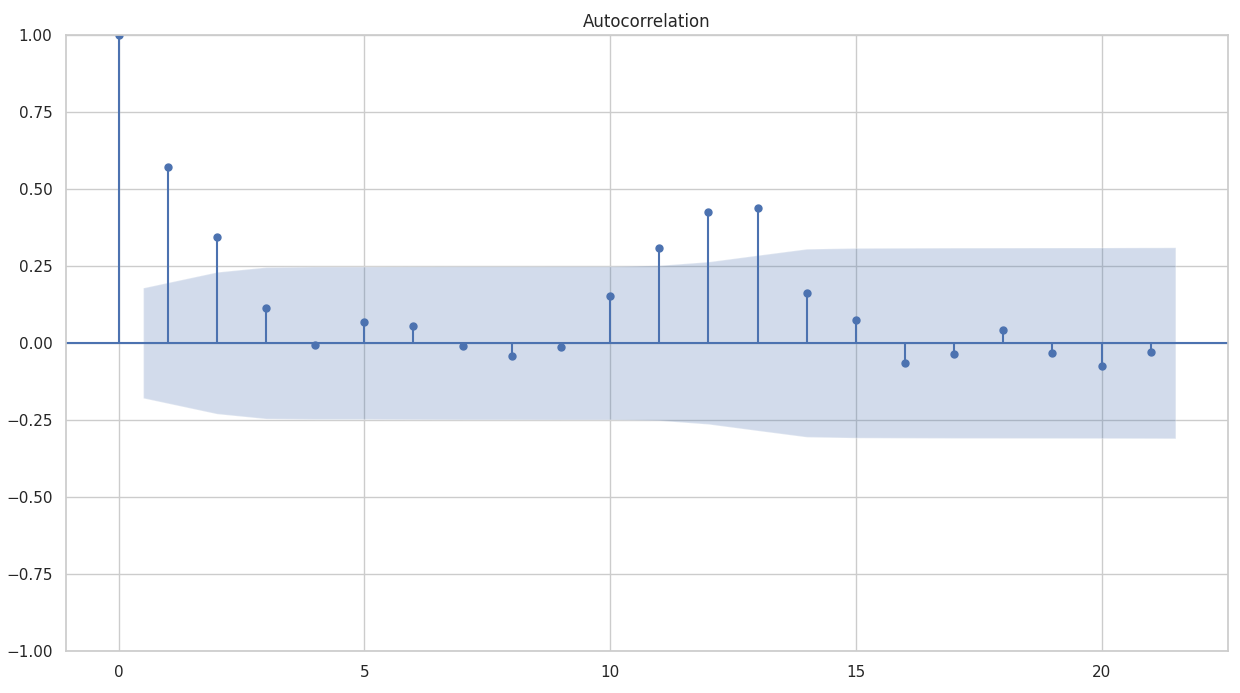


Fig 14. ACF plot results before differencing

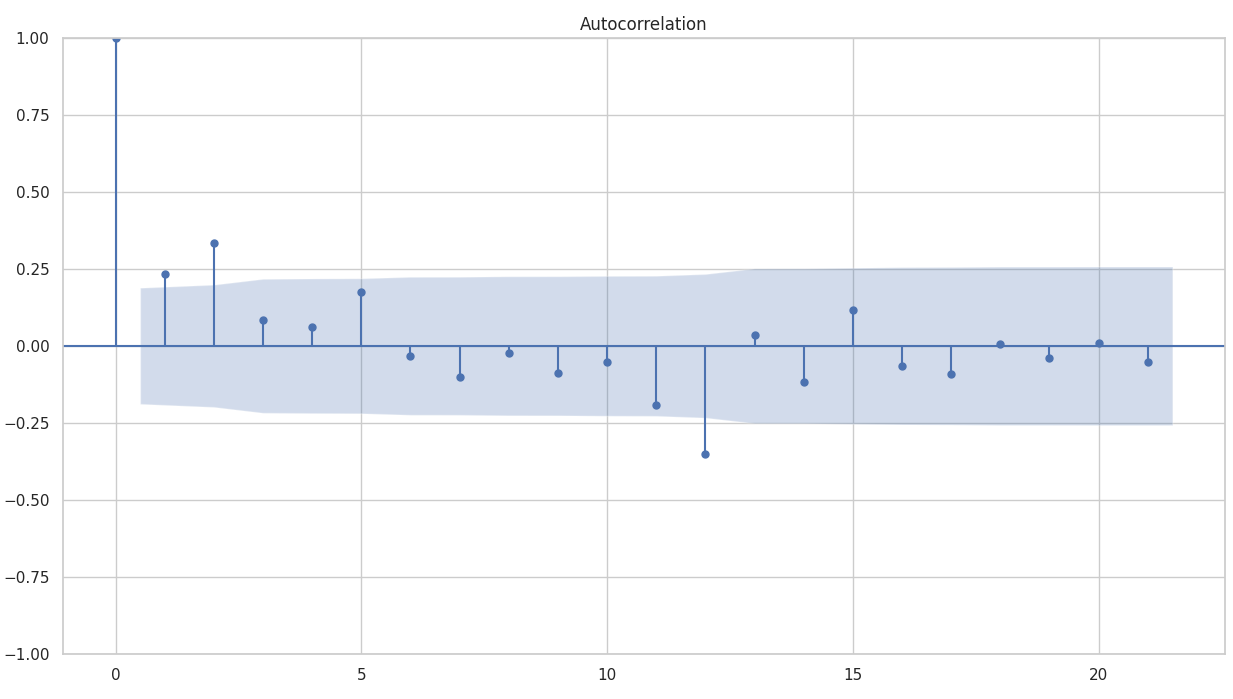


Fig 15. ACF plot results after differencing

1. *Partial Autocorrelation Function* (PACF)

The PACF plot is used to identify the Autoregressive (AR) model. In Fig 16, it can be seen that lags 1, 2 and 14 cross the confidence interval. These results explain that there is still autocorrelation, and the data is not stationary in the mean. In Fig. 17, it can be seen that the lags are already relatively in the confidence interval, so the data is more stationary compared to the previous PACF plot.

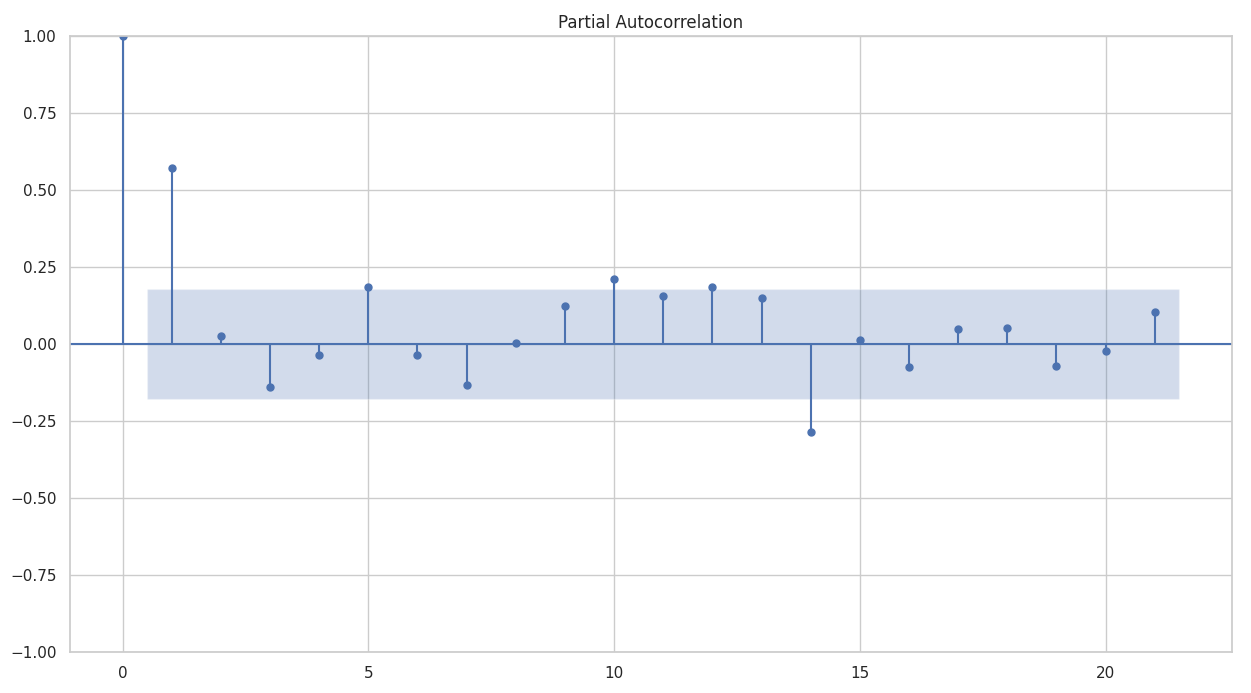


Fig 16. PACF plot results before differencing

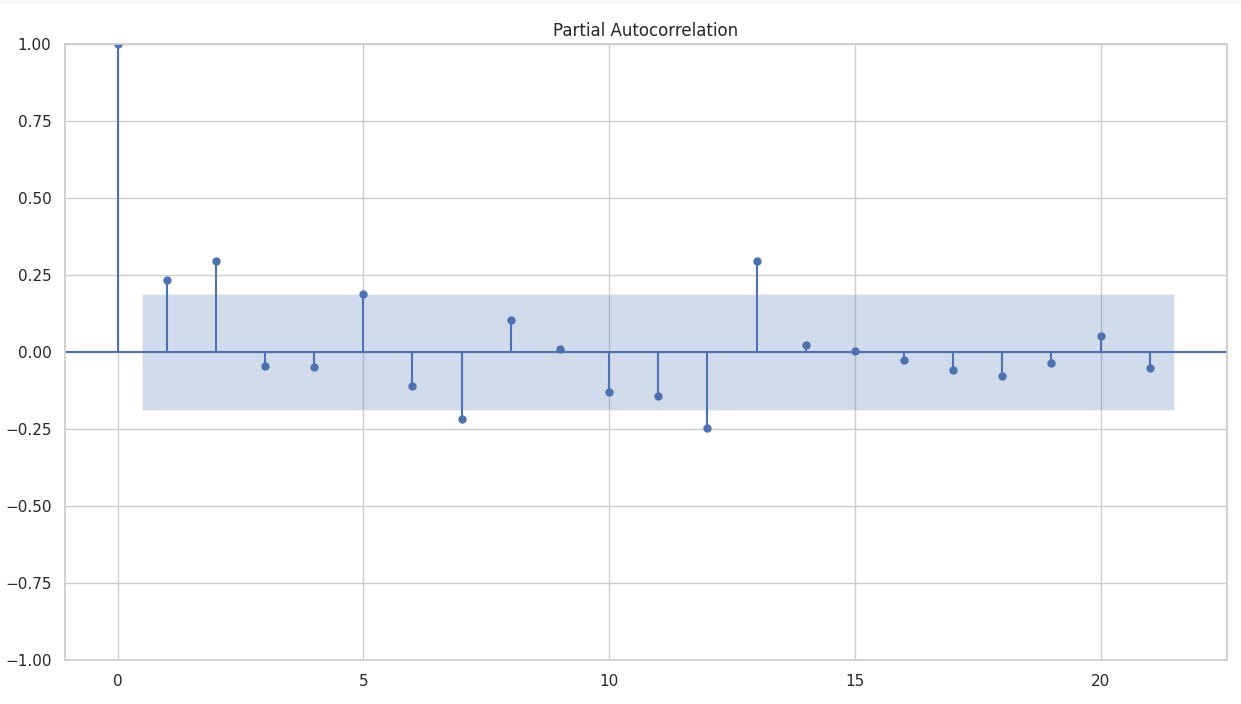


Fig 17. PACF plot results after differencing

## Estimasi Parameter Model

At this stage, MA, AR, seasonal, and non-seasonal parameters are estimated, and the significance of each parameter is tested. If there is a model whose parameters are not significant, it is declared not to pass the test. For determining model parameters for either ARIMA or SARIMA, the information criteria commonly used in selecting ARIMA models are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Parameter estimates for ARIMA and SARIMA are as follows.

The ARIMA model significance test results can be seen in Fig. 18 shows that there is only one significant ARIMA model, namely the ARIMA (0, 1, 1) model.

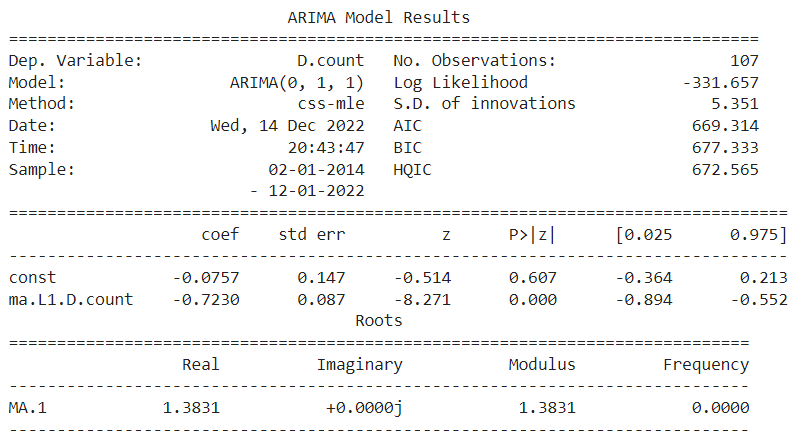


Fig 18. Parameter Estimation of ARIMA Model

The significance test results in Figure 19 show only one significant SARIMA model, namely the SARIMA (0, 1, 1) (2, 2, 1, 12) model. Based on these results, it can be concluded that the best possible model of the ARIMA method is the ARIMA (0, 1, 1) model, while the best model of the SARIMA method is the SARIMA (0, 1, 1) model (2, 2, 1, 12).

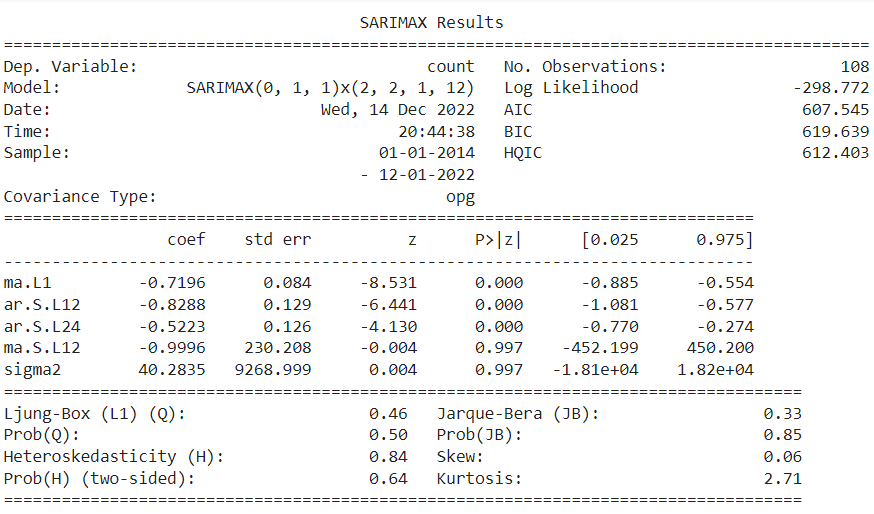


Fig 19. Parameter Estimation of SARIMA Model

## Diagnostic Checking

After the ARIMA and SARIMA models have their parameters estimated, the next stage is diagnostic checking to determine whether the model used is suitable and feasible for forecasting.

Fig 20 is the result of diagnostic checking on the ARIMA model by looking at the residuals and density plots. A good ARIMA model is when there is no accumulation of residual plots. A good ARIMA model is also when the density graph is bell-shaped around the value of 0 on the X-axis.

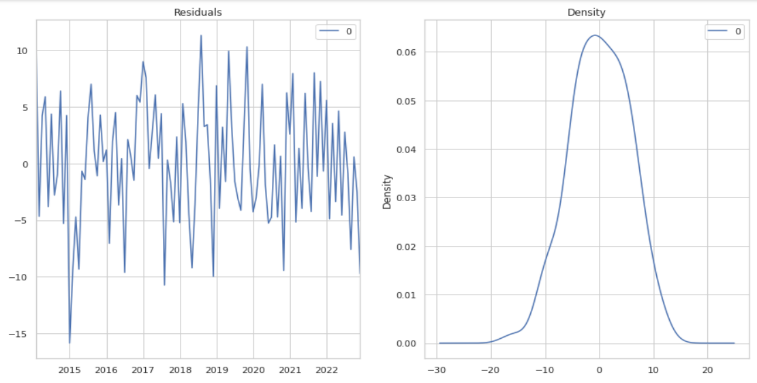


Fig 20. Diagnostic check results of the ARIMA model

Diagnostic checking is also carried out on the SARIMA model. In Fig 21, the SARIMA model is checked by looking at four graphs. The standardized residual plot in this study explains the fluctuation of hotspot data around the average value of 0 (zero). The second plot is the histogram density, which shows the normal distribution of the hotspot data. Then, the correlation plot shows a normal linear distribution that follows a linear line. The last plot is the ACF plot, which shows that no lag passes through the confidence interval, so the data is stationary in the mean*.*

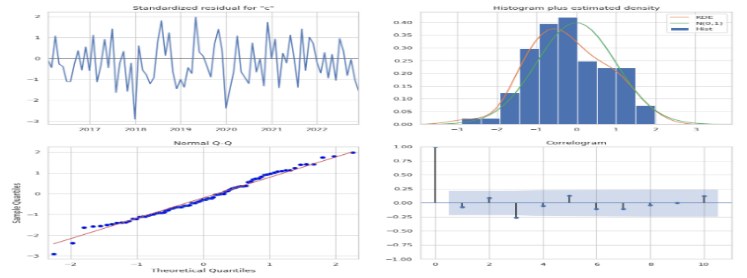


Fig 21. SARIMA model diagnostic checking results

## Prediction

After diagnostic checking, the next step is to perform forecasting on each model obtained by comparing the predicted value with the real (actual) value of the time series data, which can help the accuracy of the forecast.

Predictions are made using test data on the ARIMA model and SARIMA model. Predictions are made to determine the alignment between actual data and predicted data. Fig 22 (ARIMA model prediction) and Fig 23 (SARIMA model prediction) show that the expected data fluctuations are similar to the actual data fluctuations. Although there are several various hotspot occurrences, they are not too significant. So, the ARIMA (0, 1, 1) and SARIMA (0, 1, 1) (2, 2, 1, 12) models can be maximally used for predicting hotspot occurrence.

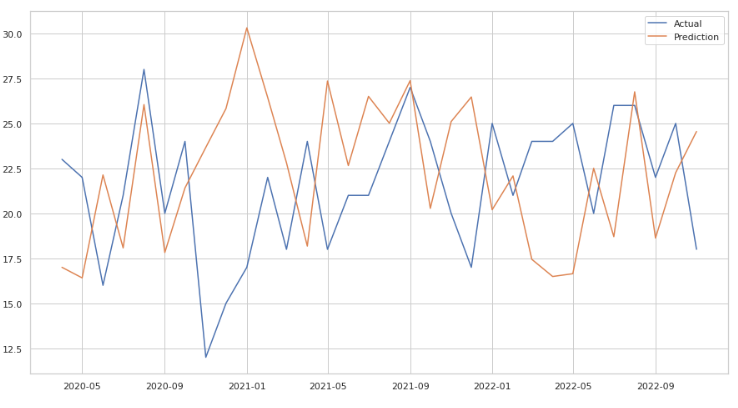


Fig 22. Prediction results of ARIMA model test data

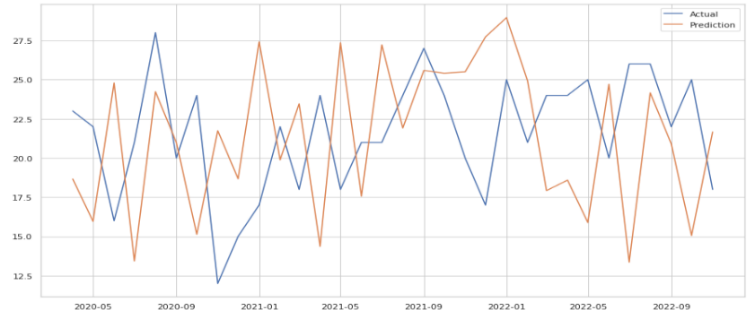


Fig 23. SARIMA model test data prediction results

Furthermore, forecasting is carried out using training data on the ARIMA and SARIMA models. Forecasting in this study was carried out to see the trend of forecasting data results for the next year, namely from 2022 to 2023, based on training data. Fig 22 (ARIMA model forecasting) and Fig 23 (SARIMA model prediction) show a downward trend at the beginning, then a decrease in the middle and a decrease again at the end of 2023. Based on these results, it is concluded that the highest occurrence of hotspots is in the middle of 2023, so that forest and land fires can be prevented.

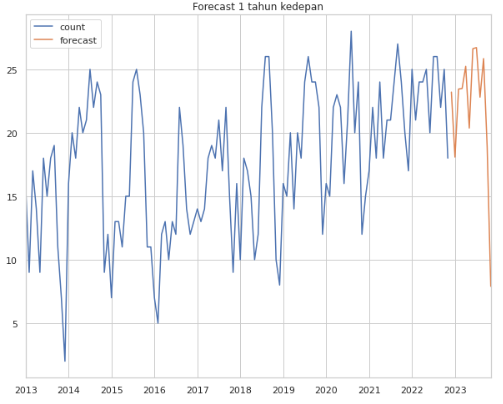


Fig 24. ARIMA model training data forecasting results

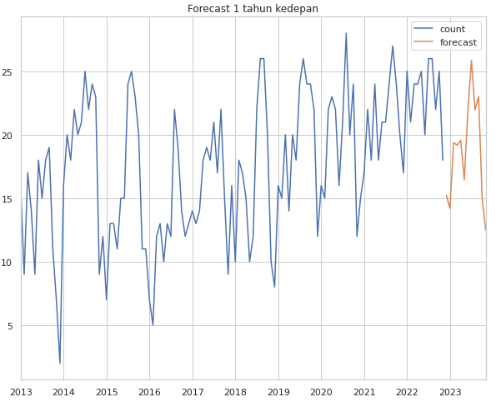


Fig 25. SARIMA model training data forecasting results

The graphs in Fig 24 and Fig 25 are visualizations of forecasting for the next year, namely 2023. In this research, forecasting is carried out monthly because the data used is daily but must be combined into months.

Furthermore, Fig 26 shows the value of the ARIMA model forecasting results. The value of the ARIMA model forecasting results is seen from 2020 to 2021. The lowest occurrence of hotspots was seen in the five months of 2020, then experienced the highest increase in month 1 of 2021 and dropped again in month 11 of 2021. Meanwhile, Fig 27 is the value of the SARIMA model forecasting results. The value of the ARIMA model forecasting results is seen from 2020 to 2021. The lowest occurrence of hotspots was seen in month 1 of 2020, then experienced the highest increase in month 1 of 2021 and dropped again in month 10 of 2021. Based on the value of the forecasting results, it can be used as a reference in preventing forest and land fires in that period.

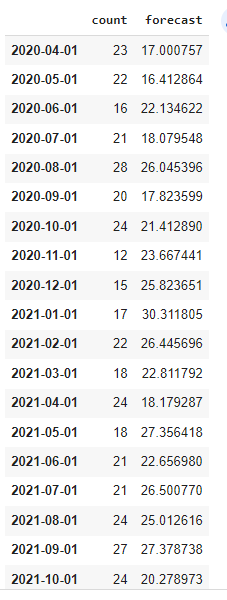


Fig 26. ARIMA model forecasting result value

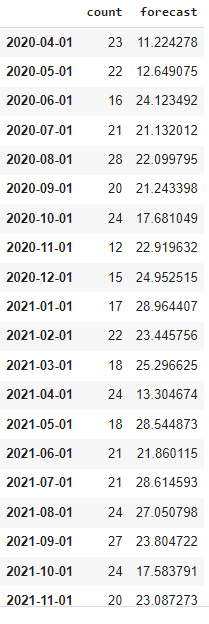


Fig 27. SARIMA model forecasting result value

## Prediction Model Evaluation

The last stage after prediction is to evaluate the results of forecasting that has been done. Model testing is done using the calculation of Mean Squared Error (MSE) and Root Mean Square Error (RMSE). In this study, the ARIMA model obtained an MSE value of 43.70 and an RMSE of 6.61. At the same time, the SARIMA model received an MSE value of 58.05 and an RMSE of 7.61. Based on the results of this comparison, it can be concluded that the ARIMA (0, 1, 1) model is best for predicting the occurrence of hotspots of forest and land fires in the West Kalimantan region. The details of the model evaluation results can be seen in Table 1.

Table 1. Evaluation of ARIMA and SARIMA prediction models

|  |  |  |
| --- | --- | --- |
| Metode | ARIMA | SARIMA |
| MAP | 43,70 | 58,05 |
| RMSE | 6,61 | 7,61 |

# Conclusions

Based on the modelling that has been done to predict the number of forest and land fire hotspot occurrences in the West Kalimantan region, the best models are ARIMA (0,1,1) and SARIMA (0,1,1) (2,2,1)12 models. The ARIMA's Root Mean Square Error (RMSE) evaluation result is 6.61, while SARIMA is 7.61. Meanwhile, ARIMA's Mean Squared Error (MSE) evaluation value is 43.70, and SARIMA is 58.05. Therefore, the ARIMA model performs well in predicting the occurrence of hotspots with a smaller RMSE value than SARIMA. Then, the highest average event of hotspots in the West Kalimantan region occurs in the mid-year period so that it can be used as a reference in preventing forest fires that are more effective and efficient in that period.

##### References

[1] D. P. Melaponty, F. Fahrizal, and T. F. Manurung, “Species diversity of forest vegetation in city forest Bukit Senja area middle Singkawang district Singkawang city,” *J. Hutan Lestari*, vol. 7, no. 2, pp. 893–904, 2019.

[2] A. Suhaibah, “Tinjauan Yuridis Terhadap Peranan Pemerintah Dalam Pelaksanaan Rehabilitasi Hutan Manggrove,” *J. Sos. Hum. Sigli*, vol. 2, no. 2, pp. 17–34, 2019, doi: 10.47647/jsh.v2i2.168.

[3] S. Nurbaya, Efransyah, S. Murniningtyas, Erwinsyah, and E. Damayanti, *Status Hutan dan Kehutanan Indonesia 2020*, vol. 17, no. 2. 2021. [Online]. Available: http://journal.ummgl.ac.id/index.php/urecol/article/view/719/804%0Ahttp://www.forestprogramme.com/files/2011/05/FOREST-Standard-Guide\_V04\_UK.pdf

[4] H. Wahyuni and S. Suranto, “Dampak Deforestasi Hutan Skala Besar terhadap Pemanasan Global di Indonesia,” *JIIP J. Ilm. Ilmu Pemerintah.*, vol. 6, no. 1, pp. 148–162, 2021, doi: 10.14710/jiip.v6i1.10083.

[5] A. H. Putra, F. Oktari, and A. M. Putriana, “Deforestasi dan Pengaruhnya Terhadap Tingkat Bahaya Kebakaran Hutan di Kabupaten Agam Provinsi Sumatera Barat,” *J. Dialog Penanggulangan Bencana*, vol. 10, no. 2, pp. 191–200, 2019.

[6] A. D. Nurhayati and L. Arhami, “Gangguan Hutan di KPH Kuningan Divisi Regional Jawa Barat dan Banten,” *J. Trop. Silvic.*, vol. 10, no. 3, pp. 159–165, 2019, doi: 10.29244/j-siltrop.10.3.159-165.

[7] A. Muzaki, R. Pratiwi, and S. R. Az Zahro, “Pengendalian Kebakaran Hutan Melalui Penguatan Peran Polisi Kehutanan Untuk Mewujudkan Sustainable Development Goals,” *LITRA J. Huk. Lingkungan, Tata Ruang, dan Agrar.*, vol. 1, no. 1, pp. 22–44, 2021, doi: 10.23920/litra.v1i1.579.

[8] M. Alkaff, A. F. Zulkarnain,

[9] S. Santriwati, H. Halide, and H. Hasanuddin, “Faktor Osean – Atmosfer untuk Memprediksi Titik Panas (Hostspot) di Wilayah Asia Tenggara Bagian Selatan,” *J. Geocelebes*, vol. 5, no. 2, pp. 116–130, 2021, doi: 10.20956/geocelebes.v5i2.13454.

[10] T. A. Pratiwi, M. Irsyad, and R. Kurniawan, “Klasifikasi Kebakaran Hutan dan Lahan Menggunakan Algoritma Naïve Bayes (Studi Kasus: Provinsi Riau),” *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 101, 2021, doi: 10.26418/Justin.v9i2.42823.

[11] Endrawati, J. Purwanto, S. Nugroho, and R. A. S, “Identifikasi Areal Bekas Kebakaran Hutan Dan Lahan Menggunakan Analisis Semi Otomatis Citra Satelit,” *Semin. Nas. Geomatika 2017 Inov. Teknol. Penyediaan Inf. Geospasial untuk Pembang. Berkelanjutan*, pp. 273–282, 2016.

[12] B. H. Saharjo and D. A. Nugraha, “Pengaruh Curah Hujan terhadap Penurunan Titik Panas (Hotspot) ti Indonesia pada Tahun 2019-2020,” *J. Trop. Silvic.*, vol. 13, no. 03, pp. 184–190, 2022, doi: 10.29244/j-siltrop.13.03.184-190.

[13] U. Khaira, M. Alfalah, P. C. S. Gulo, and R. Purnomo, “Prediksi Kemunculan Titik Panas Di Lahan Gambut Provinsi Riau Menggunakan Long Short Term Memory,” *J. Inform. J. Pengemb. IT*, vol. 5, no. 3, pp. 77–82, 2020, [Online]. Available: http://ejournal.poltektegal.ac.id/index.php/informatika/article/view/1931

[14] W. Maulida, “Model Prediksi Jumlah Kemunculan Titik Panas di Kabupaten Rokan Hilir Menggunakan Elman Recurrent Neural Network,” 2018, [Online]. Available: https://repository.ipb.ac.id/handle/123456789/93678

[15] M. Alkaff and N. E. Yulianto, “Prediksi Jumlah Kejadian Titik Api Melalui Pendekatan Deret Waktu Menggunakan Model Seasonal Arima,” *J. ELTIKOM*, vol. 3, no. 2, pp. 54–63, 2019, doi: 10.31961/eltikom.v3i2.122.

[16] I. B. B. Mahayana, I. Mulyadi, and S. Soraya, “Peramalan Penjualan Helm dengan Metode ARIMA (Studi Kasus Bagus Store),” *Inferensi*, vol. 5, no. 1, p. 45, 2022, doi: 10.12962/j27213862.v5i1.12469.

[17] Zulhamidi and R. Hardianto, “Peramalan Penjualan Teh Hijau Dengan Metode Arima (Studi Kasus Pada Pt. Mk),” *J. PASTI*, vol. XI, no. 3, pp. 231–244, 2017.

[18] M. I. Rizki and T. A. Taqiyyuddin, “Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia,” *J. Sains Mat. dan Stat.*, vol. 7, no. 2, pp. 62–72, 2021, doi: 10.24014/jsms.v7i2.13168.

[19] H. Harafani and R. S. Wahono, “Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan,” *J. Intell. Syst.*, vol. 1, no. 2, pp. 82–90, 2015.

[20] B. H. Saharjo and M. R. A. Nasution, “Pola Sebaran Titik Panas (Hotspot) Sebagai Indikator Terjadinya Kebakaran Hutan Dan Lahan Di Kabupaten Aceh Barat,” *J. Silvikultur Trop.*, vol. 12, no. 2, pp. 60–66, 2021, [Online]. Available: http://tanahair.indonesia.go.id/

[21] Putra, E. Heryanto, A. Sopaheluwakan, R. P. Pradana, and U. Haryoko, “Sebaran Spasial Dan Temporal Titik Panas (Hotspot) Di Indonesia Dari Satelit Modis Dengan Metode Gridding,” *Semin. Nas. Geomatika*, no. 3, pp. 1123–1128, 2019, [Online]. Available: https://scholar.google.co.id/scholar?oi=bibs&cluster=3216153241389431135&btnI=1&hl=id

[22] H. Sastypratiwi and R. D. Nyoto, “Analisis Data Artikel Sistem Pakar Menggunakan Metode Systematic Review,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 250, 2020, doi: 10.26418/jp.v6i2.40914.

[23] W. Pangestu, A. Widodo, and B. Rahayudi, “Prediksi Jumlah Kendaraan Bermotor di Indonesia Menggunakan Metode Average-Based Fuzzy Time Series Models,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, p. 964, 2018.

[24] T. Safitri, N. Dwidayati, and K. Kunci, “Perbandingan Peramalan Menggunakan Metode Exponential Smoothing Holt-Winters dan Arima,” *Unnes J. Math.*, vol. 6, no. 1, pp. 48–58, 2017, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/ujm

[25] G. Christie, D. Hatidja, and R. Tumilaar, “Penerapan Metode SARIMA dalam Model Intervensi Fungsi Step untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa (Application of the SARIMA Method in the Step Function Intervention to Predict the Number of Visitors at Londa Tourism Object),” *J. Ilm. Sains*, vol. 22, no. 2, p. 96, 2022, doi: 10.35799/jis.v22i2.40961.

[26] O. A. Amalia and A. H. M. Putri, “Comparison of backpropagation artificial neural network and SARIMA in predicting the number of railway passengers,” *J. Phys. Conf. Ser.*, vol. 1663, no. 1, 2020, doi: 10.1088/1742-6596/1663/1/012033.

[27] W. Anbiya and F. C. Garini, “Application of GARCH Forecasting Method in Predicting The Number of Rail Passengers ( Thousands of People ) in Jabodetabek Region,” vol. 18, no. 2, pp. 198–223, 2022, doi: 10.20956/j.v18i2.18382.

[28] S. Faradilla, “Peramalan Penjualan Produk Baja dan Besi di PT MSU dengan Pendekatan Metode ARIMA dan Single Moving Average,” vol. 12, no. 1, 2023.

[29] A. Sinardi and R. D. E. Putra, “Prediksi Awal Musim Hujan Di Waingapu Menggunakan Metode ARIMA,” vol. 01, pp. 1–6, 2020.

[30] H. Djoni, “Penerapan Model ARIMA untuk Memprediksi Harga Saham PT. Telkom Tbk.,” *J. Ilm. Sains*, vol. 11, no. 1, pp. 116–123, 2011.