Volume. 7 No 2 December (2024)

e-ISSN: **2615-3270**
p-ISSN: **2615-3599**

Eigen Mathematics Journal

Homepage jurnal: https://eigen.unram.ac.id/ index.php/eigen

Nowcasting of Indonesia's Gross Domestic Product Using Mixed Sampling Data Regression and Google Trends Data

Niken Yuliana[a], Umu Arifatul Syifa[b], Shashella Zelicha Ramadhanty[c], Fitri Kartiasih[d*]

[a.] DIV Statistical Computing Study Program, Politeknik Statistika STIS, Jakarta, Indonesia. Email: 222011685@stis.ac.id
[b.] DIV Statistical Computing Study Program, Politeknik Statistika STIS, Jakarta, Indonesia. Email: 222011527@stis.ac.id
[c.] DIV Statistical Computing Study Program, Politeknik Statistika STIS, Jakarta, Indonesia. Email: 222011818@stis.ac.id
[d.] DIV Statistics Study Program, Politeknik Statistika STIS, Jakarta, Indonesia. Email: fkartiasih@stis.ac.id

ABSTRACT

This study aims to compare the results of the GDP nowcasting of the accommodation and food service activities sector without and with the pandemic time using the MIDAS method. The MIDAS method is an econometric approach used to predict economic development using real-time available high-level and low-frequency data. In this study, Google Trend acts as a predictor variable consisting of 16 search categories which are then reduced by Principal Component Analysis, resulting in several principal components. For GDP data, the data period collected is Quarter I 2010 to Quarter I 2023. This period will later be partitioned into the period before the COVID-19 pandemic, namely Quarter I 2010 to Quarter IV 2019 and a combined period, namely Quarter I 2010 to Quarter I of 2023. This partition was carried out to see the performance and sensitivity of the model before and after the shock due to the COVID-19 pandemic. From the models that have been made, nowcasting is carried out and it is found that the RMSE and MAE values for the pre-pandemic model are smaller than the combined model. The RMSE values for each of the pre-pandemic and combined models were 0.005753 and 0.056032 and the MAE values were 0.00359 and 0.048976 for the pre-pandemic and combined models. However, from this study it is not advisable to make predictions on the nominal GDP of the accommodation and food service activities sector because the results of the nowcasting predictions are still far from the actual value, but can be a reference if you want to predict the growth direction of the accommodation and food service activities sector.

Keywords: COVID-19, GDP, Google Trends, nowcasting, MIDAS

Submitted: 11-09- 2023; Accepted: 29-07- 2024

Doi: https://doi.org/10.29303/emj.v7i2.212

1. Introduction

The COVID-19 pandemic is a respiratory disease caused by the SARS-CoV-2 virus. In 2020, Indonesia faces a big challenge in stopping the spread of COVID-19. COVID-19 has had a negative impact on all economic sectors and all fields, so that it has also affected economic activities and people's daily lives (Hanoatubun, 2020). Indonesia's Gross Domestic Product (GDP) is one of the main indicators that illustrates this impact.

COVID-19 causes a decline in economic growth and threatens national economic stability. The Indonesian government has made significant efforts in dealing with the pandemic to mitigate its impact on the country's economy, including on GDP. The economic conditions in early 2020 were full of uncertainties and could not be predicted by economic actors. The old economic model can no longer accurately predict economic indicators independently (Muditomo, 2020). One of the time series data is GDP growth. Time series is a sequence of observations made sequentially in time (Box, Jenkins, and Reinsel, 2008). To find out the current economic condition, forecasting is needed because of the importance of accurate GDP data.

* Corresponding author.
e-mail address: fkartiasih@stis.ac.id

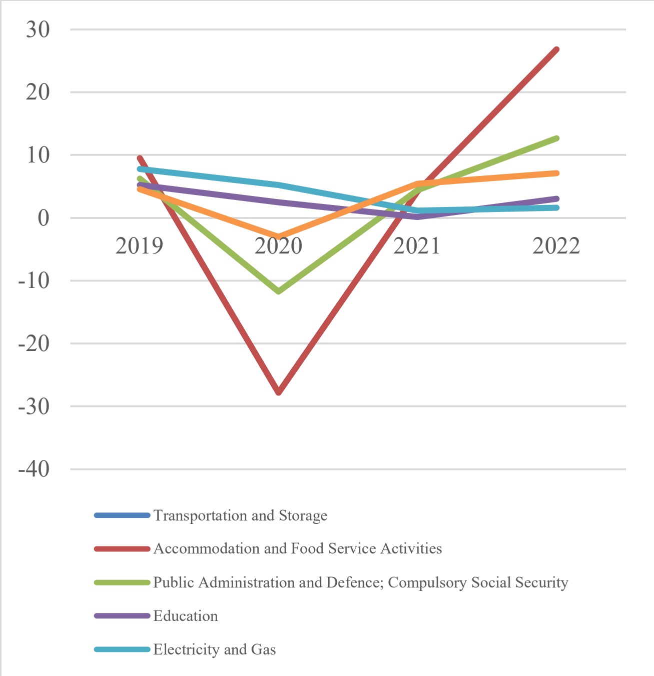


Figure. 1. Percentage Changes in GDP in Five Sectors.

The survey results from the Statistics Indonesia (BPS) noted that 82.85% of companies were affected by the Covid-19 corona virus pandemic. Based on the sector, the accommodation and food service activities experienced the most decline in income, at 92.47%. The sector has seen significant revenue declines in many countries around the world due to travel restrictions, restaurant closures and changes in consumer behavior. The tourism and hospitality industry has experienced a significant decline marked by a decrease in the number of foreign tourist visits and a ban on tourism activities in Indonesia (Paludi, 2022).

In this respect, the accommodation and food service activities are an important part of the economy, with tourism, hospitality, restaurants, cafes and similar businesses. This industry provides many jobs for the people. Thousands of people work directly or indirectly in restaurants, hotels, cafes and similar businesses. The accommodation and food service activities sectors are often closely related to the tourism industry, so the growth of this industry helps reduce unemployment rates and increase people's economic welfare. Money will be spent by tourists on accommodation, food and drinks. This industrial growth increases the country's foreign exchange earnings due to the arrival of foreign tourists, which can be used for infrastructure development and other economic projects. Therefore, governments, industry players and economic analysts can gain important insights from accurate forecasts for the sector.

The main focus of this research is nowcasting, a real-time prediction method that aims to estimate economic data using available information in real-time or near real-time. This method benefits decision makers by providing a quick and accurate understanding of current economic conditions. According to Ma'arif (2019) for the last ten years, the term "nowcasting" has been widely used. The nowcasting method is growing quite rapidly. The methods or approaches that are widely used include: (1) Bridge Equation developed by Parigi and Schlitzer (1995), (2) Mixed Data Sampling (MIDAS) developed by Ghysels, Clara, and Valkanov (2004), (3) Mixed -Frequency Vector Autoregressive (MF-VAR) was developed by Mariano and Murasawa (2010), (4) Dynamic Factor Model (DFM) was developed by Mariano and Murasawa (2003) and Giannone, Reichlin, and Small (2008).

One of the data sources used in nowcasting is search data on search engines such as Google. A popular tool in this regard is Google Trends, which allows analysis of trends and consumer behavior based on keywords searched on Google. In the context of the accommodation and dining sector, Google Trends can provide valuable insights into people's interests and demand for services and products in this sector. By looking at search trends in hospitality, restaurants, cafes and similar industries, we can gain a better understanding of consumer interests and dynamics within these sectors.

Current research focuses on using the Mixed Data Sampling (MIDAS) method to improve GDP predictions in accommodation and food service activities sectors. In nowcasting the GDP of the accommodation and food service activities sector, data from various economic indicators that reflect the sector's activity may have non-uniform time intervals, but the MIDAS model allows integration of data with different time intervals. The use of the MIDAS model to predict quarterly GDP data with a monthly predictor variable produces forecasts that are more significant than the use of the Autoregressive Distributed Lag Model (Claudio, et al., 2020). In addition, MIDAS allows the use of non-economic data for GDP predictions in the accommodation and consumption industries. For example, Google Trend data can show consumer interest and demand in a particular industry. By incorporating these non-economic data into the MIDAS model, we can obtain more accurate information about GDP predictions. Based on this background, this study aims to evaluate the performance of the MIDAS model in nowcasting the GDP of the accommodation and food service activities sector using Google Trend data.

2. Literature Review

2.1. Gross Domestic Product (GDP)

The economic condition of a country is one of the crucial things. An important indicator that can measure these conditions in a certain period of time is GDP (Hawari & Kartiasih, 2017; Kartiasih, 2019). GDP is defined as the total added value that is the result of business units in a country, or is the total value of final goods and services produced by all economic units.

2.2. Google Trends

Volume data or certain keywords from search results on the web, images, news, YouTube and Google Shopping in a geographic area and at a certain time range (Evita & Ihdina, 2019). According to Femmy et al (2021), many companies use Google Trends to get social patterns to make predictions because they are able to make significant use of big data.

The data obtained from these tools is time series data which will then be normalized on a scale of 0 to 100 to obtain an index described by Woloszko (2020) in the following equation

$$\rho = \frac{\vec{E}}{J_C(T = const)\left(P\left(\frac{\vec{E}}{E_C}\right)^m + (1 - P)\right)} \quad (1)$$

Where:

SVI_{ct} = search volume index
SV = search volume
SVT = total search volume
c = indicating subscript category search certain
t = indicating subscript time search certain
C_c = a constant that can afford normalize index in the interval 0-100 on the category search certain

The data generated by Google *Trends* is *real time* and available on tiers national and regional with frequency daily, weekly and monthly (Alfaris, 2022). The frequency of data provided by Google Trends depends on the request user. Data is available on the frequency daily if user enter range time not enough from or the same with three months. Weekly data available for user with data requirements on the range time more from three months and less from or the same with five years. While the annual data available for users who need more data of five years.

2.3. Nowcasting

Nowcasting is a short-term forecasting technique using the static forecasting method (Ahmed et al., 2020). Alfaris (2022) states that this method is called the "bridge equation" because the prediction model used is able to accommodate differences in frequency between data. In addition, Gill et al. (2019) stated that nowcasting is a forecasting method that is suitable for forecasting data in a very short term (very short-term forecast).

The main challenge in nowcasting is dealing with differences in data release dates which cause the available information set to differ from time to time within a quarter (Knut et al, 2016). Wallis refers to this problem as "ragged edge" data. Giannone et al. (2008) highlighted the importance of release dates and concluded that forecasting models formed from these data may experience a decrease in accuracy.

2.4. Mixed Data Sampling (MIDAS) Regression

MIDAS regression is a time series regression that combines data with different frequencies, namely combining data that has a low frequency with data with a higher frequency. In time series regression, the dependent and independent variables usually involve data samples at the same frequency, for example daily, monthly, quarterly, yearly, etc. While the available data does not always have the same frequency or the frequency varies, therefore researchers cannot use higher frequency data directly. However, data available at a higher frequency can potentially provide valuable information.

Some researchers overcome the frequency differences in time series regression through data interpolation. Data that has a higher frequency will be converted to a lower frequency, for example daily data is interpolated into monthly data. The usual method for such interpolation is the average method. This method may lose valuable information in interpolated data, making it more difficult to detect relationships between variables. Therefore, Ghysels (2002), Santa-Clara (2004), and Valkanov (2005) proposed a regression that can accommodate variables at different frequencies directly, this regression is usually called Mixed Data Sampling (MIDAS) Regression. Ringo and Monika (2021) conducted research on nowcasting economic growth by comparing various methods and found that the MIDAS model provides forecasting results with the best accuracy. The following is the general equation of the MIDAS regression.

$$Y_t^{(Q)} = \beta_0 + \beta_1 B\left(L^{\frac{1}{h}};\, \theta\right) X_t^{(M)} + \varepsilon_t \quad (2)$$

Where:

$B\left(L^{\frac{1}{h}};\, \theta\right)$ = weight function for polynomial *lag* from variable with Xhigher frequency
β = coefficient of the regression model
Q = information for variables $Y$with quarterly frequency
M = information for variables $X$with monthly frequency
ε = residual of the regression model

Parameters β_0, β_1, and θ need to be parameterized using Almon's Polynomial Distributed Lag (PDL) function. Weighing this PDL or Almon Lag suitable used for difference data frequency is not too away (Vollaro et al., 2021). Following equality Almon lag counter:

$$B = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \alpha_3 i^3 + \ldots + \alpha_t i^t \quad (3)$$

Where:

B = weight
t = time

3. Methodology

3.1. Scope of Study

The locus of this research is the national scope using data on Indonesia's GDP in the accommodation and food service activities sector and Google Trend. For GDP data, the data period collected is the first quarter of 2010 to the first quarter of 2023. This period is partitioned into the period before the COVID-19 pandemic, namely the first quarter of 2010 to the fourth quarter of 2019 and the combined period, namely the first quarter of 2010 to the first quarter 2023. This partition was carried out to see the performance and sensitivity of the model before and after the shock due to the COVID-19 pandemic. The GDP used in calculating GDP growth is GDP at Current Market Prices by Industry because this method reflects the actual market value of all goods and services produced in a country in a certain period, so it can provide a more accurate picture of the actual economic value of a country. In addition, this method helps in comparing the economies of countries over time, because it takes into account changes in prices. To nowcast GDP growth, variables from Google Trend are used. The Google Trend variable consists of many indicators adapted to the accommodation and food service activities sector. This period for Google Trend is January 2010 to May 2023 with monthly frequency.

3.2. Data and Data Sources

Data collection in this study was carried out by collecting secondary data sources obtained from two sources main that is Publication of the Central Bureau of Statistics and Google. For sector GDP data accommodation and meals drink taken from the website of the Central Statistics Agency and for Google Trend data taken from the Google Trends website. For Google Trend variable, used category relevant search with the target variable is sector GDP accommodation and meals drink. Following Table 1 which lists the Google Trend variables used.

Table 1 – Details of Google Trends Variable.

No	Variables / Queries	No	Variables / Queries
1	Hotel	9	Roadside stall Eat
2	Cottage Tour	10	Shop Food
3	Lodging Teenager	11	Catering
4	Earth camp	12	Bar
5	Villas	13	Discotheque
6	Apartment	14	Café
7	Restaurant	15	Shop Drink
8	Canteen	16	Shop Drug

3.3. Method Analysis

3.3.1. Principal Component Analysis

Description of data in general without any purpose to draw conclusions. In this study, the descriptive analysis used was Principal Component Analysis to reduce variables. Principal component analysis is a method for reducing variables that explains the diversity of a group of variables through several linear combinations of the actual data. In this study, principal component analysis was used to reduce search category variables on Google Trend, so that the number of variables becomes smaller so that the model is simpler and can overcome multicollinearity problems. This is because Big Data, which in this study is Google Trend, is difficult to fulfill the classical assumptions of parametric statistical models (Bzdok and Yeo, 2017). To produce a good model, the model must meet all the classical assumptions, resulting in a model that is the Best Linear Unbiased Estimator (Hyndman and Athanasopoulus, 2018).

Principal Component Analysis (PCA) is a multivariate analysis technique used to see the structure or pattern of covariance of the data. Identification of the pattern of variance can be done by creating a new variable which is the result of reducing the previous variable, so that the dimensions become smaller without losing important information in the data. The new variable is usually referred to as the main component or principal component (PC).

The formation of the principal components can be obtained from the covariance matrix Σ or the correlation matrix R. Both of these matrices can be used to calculate the eigenvalues λ_i contained in the eigenvectors γ_i. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ are obtained through the equation

$$|\Sigma - \lambda I| = 0 \tag{4}$$

Whereas eigenvectors can be obtained from the equation $\gamma_1, \gamma_2, \ldots, \gamma_p$

$$(\Sigma - \lambda_i I)\gamma_i = 0; i = 1,2, \ldots, p \tag{4}$$

For example there is a random vector $X = [X_1, X_2, \ldots, X_p]$that follows a certain distribution with a *mean value* $\mu$and has mutually orthonormal pairs of eigenvalues and eigenvectors $(\lambda_1, \gamma_1), (\lambda_2, \gamma_2), \ldots, (\lambda_p, \gamma_p)$, then a linear combination of Xcan form a component main with equality

$$PC_i = \gamma_{i1}'X_1 + \gamma_{i2}'X_2 + \ldots + \gamma_{ip}'X_p \tag{5}$$

PC compresses data dimensions of p variable free become k component main For $k \leq p$. Components are selected $k < m$after determining the score of each PC that is formed. One of PC's selection criteria is:k component main own minimum eigenvalue of one or capable explain a minimum of 70% proportion cumulative variance of the initial data, as well proportion diversity own mark Enough big.

3.3.2. Analysis Inference

Analysis inference in research this there on the formation of a GDP growth model nowcasting using method Mixed Data Sampling (MIDAS) regression. At stages modeling MIDAS regression, some testing inferences made is as following.

Time series data used for form a model must characteristic stationary. Time series data is said stationary if the mean, *variance*, and *covariance* No influenced by time or can said constant. More details, properties stationary data sequence time happen if

1. Average streak data time worth constant for every period observation
$$E(Y_t) = \mu \text{ for every } t$$
2. Sequential data time worth constant for every period observation
$$var(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \text{ for every } t$$
3. variances two data sequences time worth constant for every period observation
$$cov(Y_t, Y_{t-k}) = E\big((Y_t - \mu)(Y_{t-k} - \mu)\big) = \gamma_k \text{ for every } t$$

Can also be serial data time said as stationary data if no own element trends.

Stationarity of time series data can be tested using several methods such as graphical analysis, autocorrelation function, and unit root test. However, in general, researchers often use the unit root test to test data stationarity because they already use a significance level and a definite statistical value. Suppose a random walk equation is

$$Y_t = \rho Y_{t-1} + \mu_t \tag{6}$$

If $\rho = 1$ then the model is said as random walks without intercepts. Something Y_t for example No own constant variance so Y_t contain unit root or data not stationary. If equation (8) is subtracted with Y_{t-1} on second side right and left equality, then will become

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \mu_t$$

$$\Delta Y_t = (\rho - 1)Y_{t-1} + \mu_t$$

$$\Delta Y_t = \delta Y_{t-1} + \mu_t \quad (7)$$

Equation (7) is a random walk model without intercept and elements trends.

Serial correlation between residuals and ΔY_tcan be expressed in the general form of an autoregressive process:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \ldots + \alpha_k \Delta Y_{t-k} + \mu_t$$

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta Y_{t-i} + \mu_t \quad (8)$$

Testing with use above equation known as Augmented Dickey Fuller (ADF) test. Testing This done with count test statistic:

$$\tau = \frac{\hat{\rho}}{SE(\hat{\rho})}$$

Hypothesis:

$H_0: \delta = 0$ or $\rho = 1$ (Data has no unit root or Y_t is not stationary)

$H_1: \delta < 0$ or $\rho < 1$

Criteria rejection hypothesis for the ADF *test* is moment $\tau >$ τ-McKinnon *Critical Values* or $p - value > alpha$.

Due to the data used is a time series data, then especially formerly must checked stationarity from the data for avoid spurious regression occurs. Test statistics used for testing stationarity is the Augmented Dickey-Fuller test with hypothesis zero data no stationary and hypothetical stationary data alternative.

3.3.3. Mixed Data Sampling Model

The Mixed Data Sampling Model (MIDAS) allows integration of data with different time intervals. In addition, MIDAS allows the use of non-economic data for predictions. This is in line with the condition of the data that will be included in the model which has two-time frames, namely monthly and quarterly. The MIDAS weight used in this study is Polynomial Distributed Lag (PDL). In general, the specifications of the MIDAS model to be applied in this study are as follows.

For the set of Google Trend variables, the variables that will be included in the model are those that have gone through a reduction process using principal component analysis. Here are the equations.

Pre-pandemic model:

$$\begin{aligned} GROWTH_t^{(Q)} = \delta_0^Q &+ \sum_{l=1}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND1_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND2_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND3_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND4_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND5_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND6_t^{(M)} + \varepsilon_{3_t} \end{aligned}$$

Combined model:

$$\begin{aligned} GROWTH_t^{(Q)} = \delta_0^Q &+ \sum_{l=1}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND1_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND2_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND3_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND4_t^{(M)} \\ &+ \sum_{l=a}^{j} \delta_l B(L^{\frac{1}{3}};\theta_l) GTREND5_t^{(M)} + \varepsilon_{3_t} \end{aligned}$$

3.3.4. Model Accuracy

In addition to the data used in model estimation must meet the assumptions, the results of the selected model must also be measured for its accuracy with several indices. Generally, time series data uses the BIC value to determine the best model, but other than that there are other values used to measure model accuracy, for example RMSE. RMSE can be calculated using the equation

$$RMSE = \sqrt{\frac{(y_i - \hat{y}_i)^2}{n}} \quad (9)$$

The level of model accuracy can also be measured using MAE. The range of MAE values is smaller than RMSE so that it can be interpreted more clearly. MAE also has several advantages as described in a study by Willmott & Mastsuura (2005), which is the most obvious and natural error measure of the mean. The MAE equation can be formulated as follows

$$MAE = \frac{|y_i - \hat{y}_i|}{n} \quad (10)$$

Where, y is true value, $\hat{y}$ is estimated value, and n is many observation

3.4. Stages Analysis

In form a GDP growth nowcasting model, the stages through which is as following.

1. Reduction Google Trends variable.
2. Every variable done testing stationarity using the ADF test.
3. If assumptions stationarity is not yet fulfilled, then done transformation differentiation and or natural logarithm.
4. If all variable has stationary, next two models were carried out, namely the pre-pandemic model and the combined model.
5. Furthermore, is inspection assumption classic.
6. Stage furthermore is GDP growth nowcasting simulation.
7. For each model nowcasting is counted accuracy.
8. Stage end is do model comparison.

In form a GDP growth nowcasting model, the stages through which is as following.

Figure. 2 - Research Flowchart.

4. Results and Analysis

4.1. GDP Trends in the Accommodation and Dining Sector Drink

The accommodation and food service activities sector are one of the GDP-forming sectors with a percentage distribution of GDP in 2022 of 2.41. This figure has decreased by 0.02 compared to the percentage distribution in 2021 and a decrease of 0.14 compared to the percentage distribution in 2020. This means that from 2020 to 2022, the accommodation and food service activities sector's contribution to GDP will continue to decline.

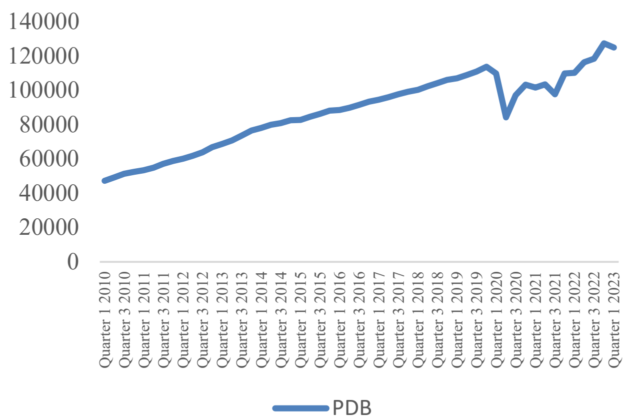


Figure. 3 - GDP of the Accommodation and Dining Sector Drink.

From the figure above it can be seen that in general the GDP of the accommodation and food service activities sector has a positive trend, which is increasing every year. However, in 2020 there was a significant decrease as a result of the COVID-19 that hit Indonesia at that time. The existence of mobility restrictions and work from home policies resulted in Indonesia's economic activity weakening (Nasution et al., 2020). These up and down fluctuations continue until the fourth quarter of 2021 and return to a positive trend in 2022, but decline again in the first quarter of 2023.

This significant decrease arose due to people's concerns about carrying out activities outside the home, which also had an impact on the accommodation and food service activities sectors. This can be seen, among others, from the decrease in room occupancy rates in both star and non-star hotels, which were originally at 54.81 and 31.48 respectively in 2021 to 33.79 and 18.31 in 2020. In addition, COVID-19 also had an impact on culinary business operations, starting from limiting the number of customers, only serving takeaway or delivery orders, to the requirement to attach a COVID-19 vaccination certificate, thereby reducing public interest in enjoying the culinary business. However, there are also groups of people who choose not to eat non-homemade food at all with the aim of avoiding the COVID-19 virus, which was currently being discussed. Based on Statistics Indonesia data, the accommodation and food service activities sector are the sector most affected by COVID-19, with a decrease in income of 92.47%.

4.2. Principal Component Analysis of Google Trend Data

4.2.1. Pre-Pandemic Model

The search categories on Google Trend that are used to nowcast the pre-pandemic GDP of the accommodation and food service activities sector consist of 16 categories with a time span from 2010 to 2019. The following is a visualization of the correlation matrix of the 16 categories of search results with Google Trend with that time range.

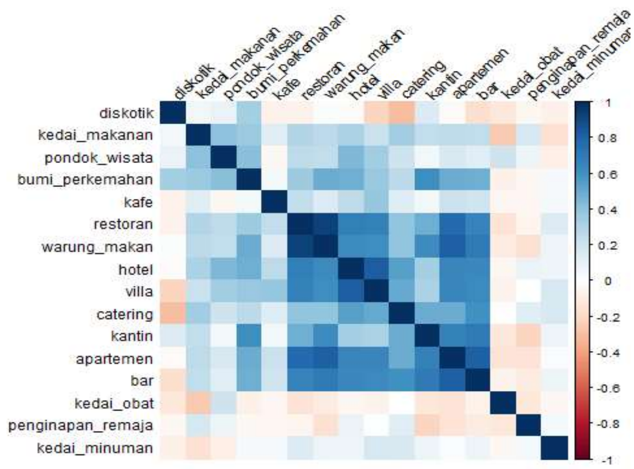


Figure. 4 - Matrix Correlation of Pre-Pandemic Model.

From Figure 4 it can be seen that in some categories there are boxes with a fairly dense blue color. This means that between these categories have a fairly high and positive correlation, this means that there is multicollinearity between variables in the data.

Table 2 – Analysis Main Pre-Pandemic Components.

Principal Component (PC)	Standard Deviation	Proportion Variance	Proportion Cumulative
PC 1	2.4675	0.3805	0.3805
PC 2	1.2923	0.1044	0.4849
PC 3	1.2406	0.0962	0.5811
PC 4	1.1269	0.0794	0.6605
PC 5	1.0406	0.0677	0.7281
PC 6	1.0110	0.0639	0.7920

Apart from multicollinearity, the number of categories used is also the reason for principal component analysis. Principal component analysis is carried out in order to reduce the categories used, so that the number of categories becomes fewer and the assumptions of independence can be fulfilled. The following is a summary table of the results of the principal component analysis that has been carried out.

From Table 2 it is known that the 6 principal components are able to produce a cumulative proportion of 79%. That is, these 6 principal components are able to explain 79% of the variation from the 16 categories used. However, prior to forming the main data components, they were standardized first, because of the 16 categories used, several Google Trend categories had a range of values that were quite different from the other categories.

4.2.1.1. Principal Component 1

Principal component 1 can modeled through equality

$$PC_1 = -0.34X_1 - 0.15X_2 + 0.024X_3 - 0.257X_4 - 0.33X_5 - 0.36X_6 - 0.34X_7 - 0.28X_8 - 0.35X_9 - 0.17X_{10} - 0.26X_{11} - 0.35X_{12} + 0.0184X_{13} - 0.113X_{14} - 0.046X_{15} + 0.053X_{16}$$

This component can describe categories of hotels, villas, apartments, restaurants, stalls dining, and bars, as well capable explains 38% of the diversity of Google Trend data.

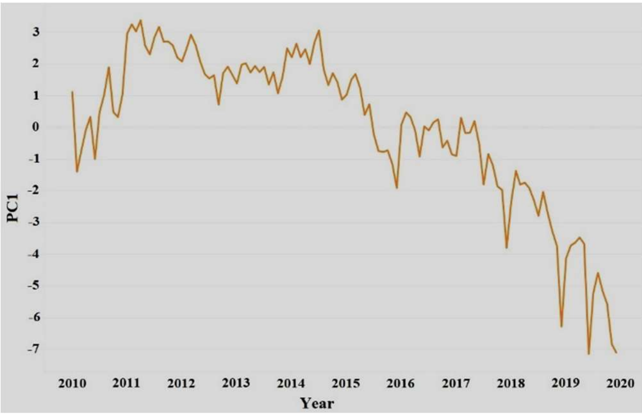


Figure. 5 - Principal Component 1 of Pre-Pandemic Model.

The image above is a visualization of the principal component 1. From the figure it can be seen that this component has a downward trend. This means that internet searches using queries for hotels, villas, apartments, restaurants, food stalls and bars are decreasing.

4.2.1.2. Principal Component 2

Principal component 2 can modeled through equality

$$PC_2 = -0.020X_1 + 0.32X_2 + 0.043X_3 + 0.42X_4 - 0.21X_5 - 0.039X_6 - 0.089X_7 + 0.14X_8 + 0.006X_9 + 0.34X_{10} - 0.19X_{11} - 0.12X_{12} + 0.58X_{13} - 0.22X_{14} - 0.29X_{15} - 0.12X_{16}$$

This component capable to describe category earth campgrounds and discotheques, as well explain about 10% of diversity of Google Trend data.

Based on the line chart in Figure 6, it can be seen that the line patterns formed have increased and also decreased. In 2010, the line pattern looks more volatile than in other years. Then, it increased in mid-2013 and gradually decreased until the end of 2019. However, in general the line patterns that are formed tend to be around 0. This means that during the period 2010 to 2019 searches for campground and discotheque queries tended to be average or in other words not experiencing a significant increase or decrease.

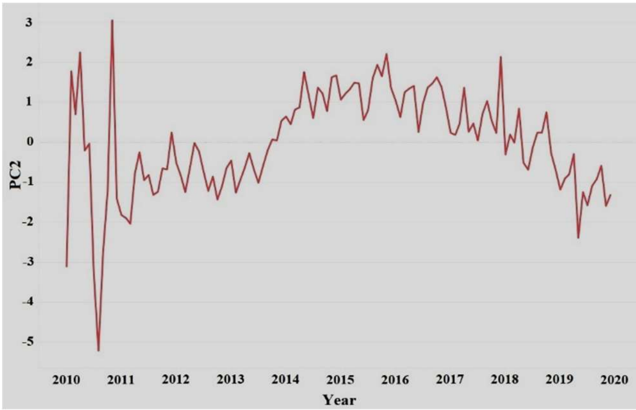


Figure. 6 - Principal Component 2 of Pre-Pandemic Model.

4.2.1.3. Principal Component 3

Principal component 3 can modeled through equality

$$PC_3 = 0.23X_1 + 0.445X_2 + 0.52X_3 - 0.069X_4 + 0.18X_5 - 0.18X_6 - 0.052X_7 - 0.36X_8 - 0.17X_9 + 0.34X_{10} + 0.21X_{11} - 0.13X_{12} - 0.23X_{13} + 0.082X_{14} - 0.062X_{15} + 0.098X_{16}$$

This component able to describe category cottage travel, accommodation youth, canteen, and shop food. This component able to explain about 9% the diversity of Google Trend data, so proportion cumulative possible variance described by components main 1 to 3 is by 58%.

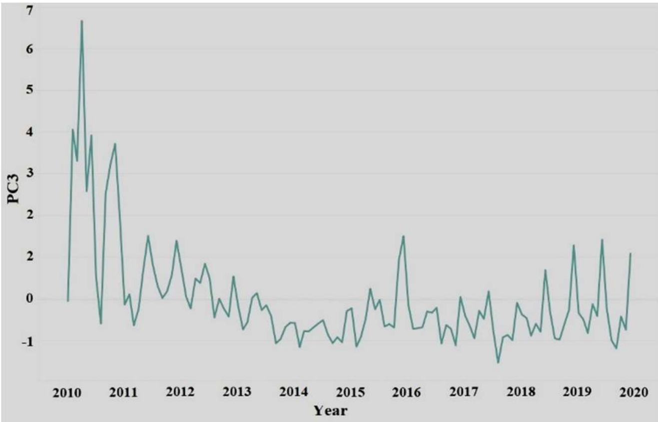


Figure. 7 - Principal Component 3 of Pre-Pandemic Model.

4.2.1.4. Principal Component 4

Principal component 4 can modeled through equality

$$PC_4 = -0.079X_1 - 0.39X_2 + 0.39X_3 - 0.075X_4 - 0.048X_5 + 0.015X_6 + 0.036X_7 + 0.039X_8 - 0.038X_9 + 0.24X_{10} + 0.047X_{11} - 0.006X_{12} + 0.053X_{13} + 0.14X_{14} + 0.13X_{15} - 0.76X_{16}$$

Different with principal component before, principal component 4 's only able to describe 1 category that is shop medicine. This component able to explain about 7.9% diversity from overall Google Trends.

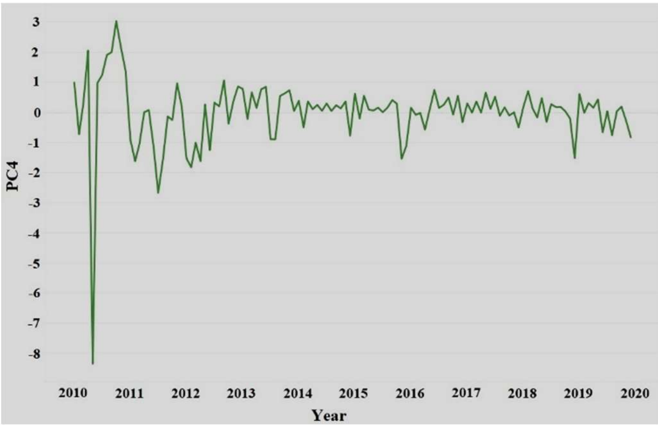


Figure. 8 - Principal Component 4 of Pre-Pandemic Model.

4.2.1.5. Principal Component 5

Principal component 5 can modeled through equality

$$PC_5 = 0.12X_1 + 0.009X_2 - 0.24X_3 - 0.13X_4 + 0.19X_5 + 0.040X_6 + 0.15X_7 - 0.28X_8 + 0.058X_9 + 0.016X_{10} - 0.39X_{11} - 0.13X_{12} + 0.16X_{13} + 0.66X_{14} - 0.36X_{15} - 0.096X_{16}$$

Principal component 5 able explain about 6.8% diversity from the entirety of Google Trend and consists from category catering and cafe search.

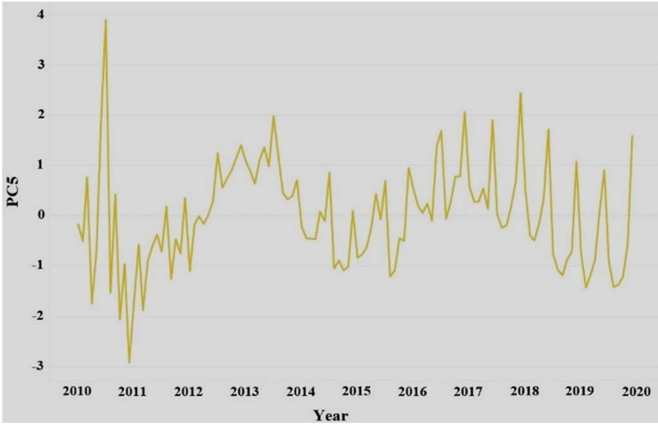


Figure. 9 - Principal Component 5 of Pre-Pandemic Model.

Based on the figure above, seen that line pattern formed really fluctuating and reaching point highs and lows in 2010.

4.2.1.6. Principal Component 6

Principal component 6 can modeled through equality

$$PC_6 = -0.15X_1 - 0.12X_2 - 0.14X_3 - 0.18X_4 - 0.17X_5 + 0.099X_6 - 0.068X_7 + 0.13X_8 - 0.014X_9 + 0.29X_{10} + 0.20X_{11} + 0.17X_{12} - 0.38X_{13} - 0.11X_{14} - 0.73X_{15} - 0.030X_{16}$$

Component main 6 only able to describes 1 category that is shop drinks and can explain about 6.4% diversity whole Google Trend variable with proportion cumulative variance by 79%.

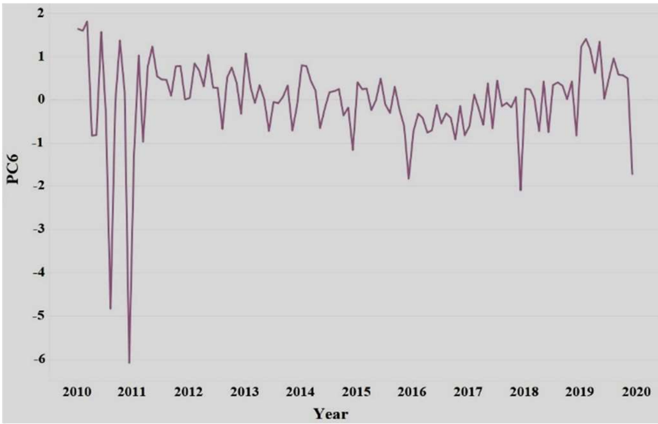


Figure. 10 - Principal Component 6 of Pre-Pandemic Model.

Same thing with component line chart major 4, in general component line chart it is also around 0 and exists drastic decline in 2010. The difference is component line chart tend a little more fluctuating compared to line graph on component 4.

4.2.2. Combined Model

For modeling *nowcasting* combined used range time from 2010 to May 2023.

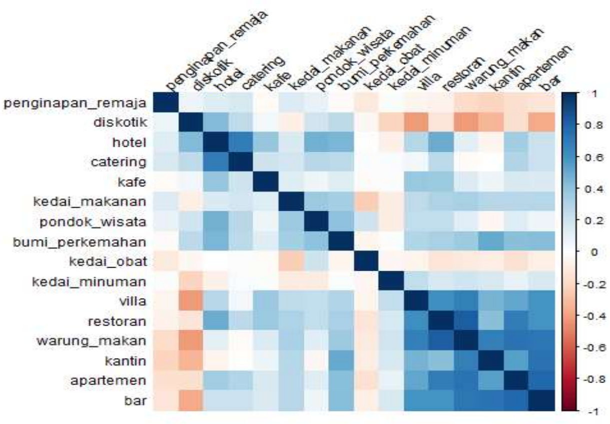


Figure. 11 - Matrix Correlation of Combined Model.

From Figure 11 it can be seen that in some categories there are boxes with a fairly dense blue color which indicates a fairly high and positive correlation. Next, principal component analysis is carried out to reduce the categories.

Table 3 – Analysis Combined Model Components.

Principal Component (PC)	Standard Deviation	Proportion Variance	Proportion Cumulative
PC 1	2.2700	0.3220	0.3220
PC 2	1.6306	0.1662	0.4882
PC 3	1.1300	0.07981	0.56799
PC 4	1.1104	0.0771	0.6451
PC 5	1.0555	0.0696	0.7147

From the principal component analysis (Table 3), 5 principal components were formed with a cumulative proportion of 71%. That is, these 5 principal components are able to explain 71% of the variation from the entire category used.

4.2.2.1. *Principal Component 1*

Principal component 1 can modeled through equality

$$PC_1 = -0.34X_1 - 0.15X_2 + 0.024X_3 - 0.26X_4 - 0.33X_5 - 0.36X_6 - 0.34X_7 - 0.28X_8 - 0.35X_9 - 0.17X_{10} - 0.26X_{11} - 0.35X_{12} + 0.018X_{13} - 0.11X_{14} - 0.046X_{15} + 0.053X_{16}$$

This principal component can describe 6 categories, namely villas, apartments, food stalls, bars, restaurants and canteens. This first component has been able to explain 32% of the diversity of the entire category. Similar to the principal component 1 in pre-pandemic, the variables in this component also have a fairly strong and positive correlation.

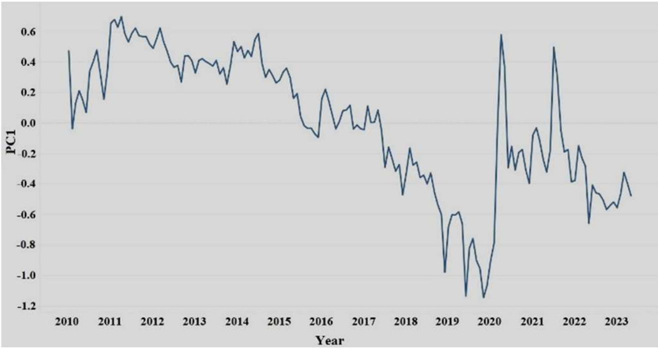


Figure. 12 - Principal Component 1 of Combined Model.

Based on the figure above, the line chart formed up to 2019 shows the same pattern as the pre-pandemic principal component 1 chart, namely a declining pattern. However, in this component it can be seen that the graph rose again at the time when Covid-19 hit. This means that internet searches for villas, apartments, food stalls, bars, restaurants and canteens increased during the Covid-19 outbreak and have decreased until now.

4.2.2.2. *Principal Component 2*

Principal component 2 can modeled through equality

$$PC_2 = -0.020X_1 + 0.32X_2 + 0.043X_3 + 0.42X_4 - 0.21X_5 - 0.039X_6 - 0.089X_7 + 0.14X_8 + 0.006X_9 + 0.34X_{10} - 0.196X_{11} - 0.12X_{12} + 0.58X_{13} - 0.22X_{14} - 0.29X_{15} - 0.12X_{16}$$

This principal component is able to describe the categories of hotels, cottages, catering, and discotheques. This component is able to explain 16% of the variance from Google Trend data, so that the cumulative proportion of the variance that can be explained by principal components 1 to 2 is 48%.

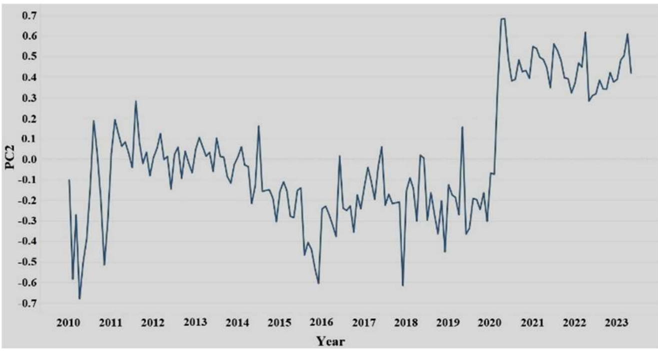


Figure. 13 - Principal Component 2 of Combined Model.

Based on the line chart in Figure 13, it can be seen that the line patterns formed have increased and also decreased. From 2010 to 2019, line patterns tended to be around or less than 0. Then it increased starting in early 2020. This means that when Covid-19 occurred, internet searches for hotel, catering and discotheque queries experienced an increase.

4.2.2.3. Principal Component 3

Principal component 3 can modeled through equality

$$PC_3 = 0.23X_1 + 0.44X_2 + 0.52X_3 - 0.069X_4 + 0.18X_5 - 0.18X_6 - 0.052X_7 - 0.36X_8 - 0.17X_9 + 0.34X_{10} + 0.21X_{11} - 0.13X_{12} - 0.23X_{13} + 0.082X_{14} - 0.062X_{15} + 0.098X_{16}$$

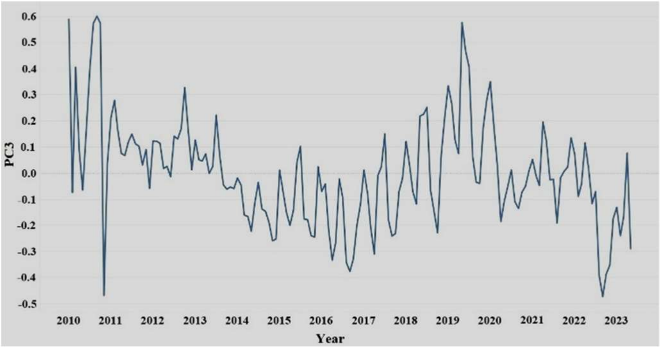


Figure. 14 - Principal Component 3 of Combined Model.

From these equations it can be said that this principal component is able to describe the categories of campsites, food stalls, and drugstores. As can be seen from the figure, the search for these queries is very fluctuating. However, in general, from 2010 to 2016 there was a decline and it started to increase again at the end of 2016. Search queries decreased again in 2019 or since the start of the pandemic.

4.2.2.4. Principal Component 4

Principal component 4 can modeled through equality

$$PC_4 = -0.079X_1 - 0.39X_2 + 0.39X_3 - 0.075X_4 - 0.048X_5 + 0.015X_6 + 0.036X_7 + 0.039X_8 - 0.038X_9 + 0.24X_{10} + 0.047X_{11} - 0.006X_{12} + 0.053X_{13} + 0.14X_{14} + 0.13X_{15} - 0.76X_{16}$$

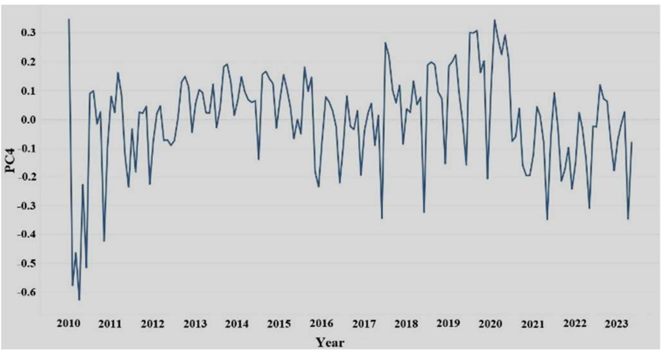


Figure. 15 - Principal Component 4 of Combined Model.

This principal component can describe category shop food with fluctuating patterns however the trend is around 0.

4.2.2.5. Principal Component 5

Principal component 5 can modeled through equality

$$PC_5 = 0.12X_1 + 0.009X_2 - 0.24X_3 - 0.13X_4 + 0.19X_5 + 0.040X_6 + 0.15X_7 - 0.28X_8 + 0.058X_9 + 0.016X_{10} - 0.39X_{11} - 0.13X_{12} + 0.16X_{13} + 0.66X_{14} - 0.36X_{15} - 0.096X_{16}$$

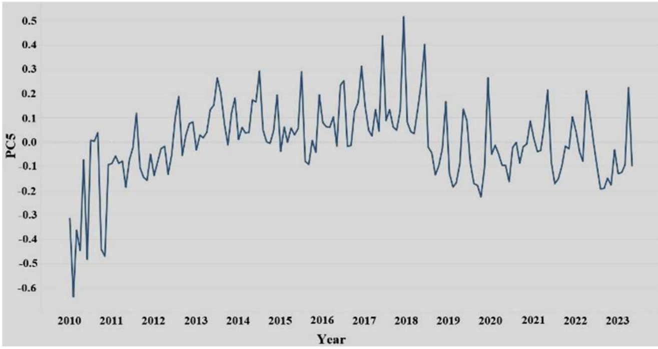


Figure. 16 - Principal Component 5 of Combined Model.

This principal component is able to describe the categories of youth lodging and cafes, and is able to explain about 7% of the diversity of Google Trend data. Based on Figure 16, it can be seen that in 2010 the line was below 0, then gradually increased until it tended to be around 0.

4.3. Stationarity Test

4.3.1. Pre-Pandemic Model

Because the data used is time series data, it is necessary to test the stationarity of the independent variables, namely the 6 principal components that have been formed and the dependent variable, namely the GDP of the accommodation and food service activities sector. Stationarity testing aims to obtain a model that is more precise and accurate, and does not produce spurious regression. The stationarity testing method used is Augmented Dickey-Fuller (ADF).

Based on the table 4, it is known that the dependent variable, namely GDP, is not stationary at the level, so a logarithmic transformation is performed and produces stationary test results. The only independent variables that are not stationary at the level are PC1 and PC5, namely the principal components 1 and 5. For this reason, a first difference transformation is performed to produce stationary test results.

Table 4 – Stationarity Test of Pre-Pandemic Model.

Group Variable	Variable	ADF p-values	Conclusion
Variable Dependent	GDP	0.8994	Not Stationary
	GDPLOG	0.0003	Stationary
Variable Independent	PC1	0.9866	Not Stationary
	D1PC1	0.0000	Stationary
	PC2	0.0046	Stationary
	PC3	0.0000	Stationary
	PC4	0.0000	Stationary
	PC5	0.0926	Not Stationary
	D1PC5	0.0002	Stationary
	PC6	0.0000	Stationary

4.3.2. Combined Model

For the combined nowcasting model, the independent variables used are the 5 principal components that have been formed and the dependent variable is the GDP of the accommodation and food service activities sector with the time span from 2010 to May 2023.

Table 5 – Stationarity Test of Combined Model.

Group Variable	Variable	ADF p-values	Conclusion
Variable	GDP	0.7536	Not Stationary
Dependent	GDPLOG	0.4099	Not Stationary
	D1GDPLOG	0.0000	Stationary
Variable	PC1	0.0631	Not Stationary
Independent	D1PC1	0.0000	Stationary
	PC2	0.0264	Stationary
	PC3	0.0000	Stationary
	PC4	0.0000	Stationary
	PC5	0.035	Stationary

Based on the Table 5, it is known that the dependent variable, namely GDP, is neither stationary at the level nor after the logarithmic transformation is performed, so a first difference is performed to produce stationary test results. The independent variables that are not stationary at the level are only the variable PC1 or principal component 1, but are stationary at the first difference.

4.4. Formation of the Nowcasting Model

4.4.1. Pre-Pandemic Model

The establishment of the pre-pandemic COVID-19 nowcasting model uses data ranging from 2010 to 2018 for the training model and 2019 for nowcasting. By using the MIDAS method with Polynomial Distributed Lag (PDL) weights, the following results are obtained.

Table 6 – Nowcasting of Pre-Pandemic Model.

Variable	Coef-ficient	Std. Error	t-Statistics	Prob.
C	0.3668	0.0612	5,992	0.000
GDPLOG(-1)	0.9693	0.0055	177.6	0.000
	Variable: D1PC1(-3) Lags: 4			
PDL01	-0.0042	0.0014	-2.9059	0.0074
	Variable: D1PC2(-3) Lags: 2			
PDL01	-0.0029	0.0017	-1.6556	0.1098
	Variable: D1PC3(-3) Lags: 2			
PDL01	-0.0025	0.0012	-2.1525	0.0408
	Variable: D1PC4(-3) Lags: 2			
PDL01	0.0028	0.0015	1.9078	0.0675
	Variable: D1PC5(-3) Lags: 2			
PDL01	-0.0016	0.0012	-1.3476	0.1894
	Variable: D1PC6(-3) Lags: 3			
PDL01	0.0023	0.0008	3.0573	0.0051

Established MIDAS model is a model with PDL weighing with degrees polynomial of 1 and a maximum lag of 4 which is substantially automatic will selected by Eviews. Election amount maximum lag used is experimental results, meanwhile election degrees polynomial based on the characteristics of the data. The following is the model equation formed:

$$\begin{aligned} GROWTH_t^{(Q)} &= 0.3668 + 0.9693\, GROWTH_{t-1}^{Q} \\ &\quad - 0.0042 D1PC11_t^{(M)} - 0.0029 D1PC2 \\ &\quad - 0.0025 D1PC3_t^{(M)} + 0.0028 D1PC4_t^{(M)} \\ &\quad - 0.0016 D1PC5_t^{(M)} + 0.0023 D1PC6_t^{(M)} \\ &\quad + \varepsilon_{3_t} \end{aligned}$$

4.4.2. Combined Model

The pre-pandemic COVID-19 nowcasting model is built by using data with range time 2010 to 2021 for *training* model and 2022 to first quarter of 2023 for *nowcasting* with use MIDAS method with Almon lag PDL weighing.

Table 7 – Nowcasting of Combined Model.

Variable	Coef-ficient	Std. Error	t-Statistics	Prob.
C	0.0028	0.0053	0.5249	0.6027
D1 GDPLOG(-1)	0.2909	0.1115	2.6098	0.0128
	Variable: D1PC1(-3) Lags: 4			
PDL01	-0.0489	0.0060	-8.1305	0.0000
	Variable: D1PC2(-3) Lags: 2			
PDL01	-0.0195	0.0063	-3.0842	0.0037
	Variable: D1PC3(-3) Lags: 2			
PDL01	-0.0059	0.0030	1.9870	0.0540
	Variable: D1PC4(-3) Lags: 2			
PDL01	0.0172	0.0063	2.7088	0.0100
	Variable: D1PC5(-3) Lags: 2			
PDL01	-0.0128	0.0045	-2.8455	0.0070

The MIDAS model parameters used to form the combined model are the same as those used to form the pre-pandemic model, namely the model with a PDL weight with a polynomial degree of 1 and a maximum lag of 4 which will be automatically selected by Eviews.

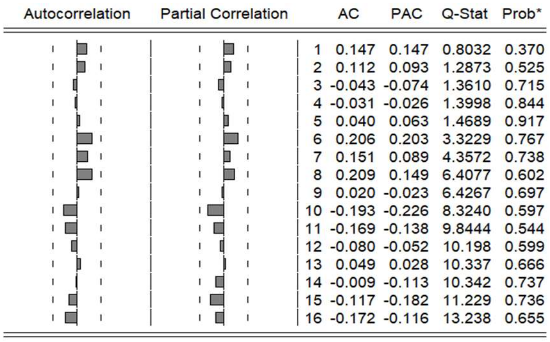
$$\begin{aligned} GROWTH_t^{(Q)} &= 0.0028 + 0.2909 GROWTH_{t-1}^{Q} \\ &\quad - 0.04892 D1PC11_t^{(M)} - 0.0195 D1PC2 \\ &\quad - 0.0059 D1PC3_t^{(M)} + 0.0172 D1PC4_t^{(M)} \\ &\quad - 0.0128 D1PC5_t^{(M)} + \varepsilon_{3_t} \end{aligned}$$

4.5. Classical Assumption Test

The classical assumption test and the test method performed were normality with Jarque Berra, homoscedasticity with the quadratic correlogram graph, non-autocorrelation with Durbin Watson, and non-multicollinearity with the VIF value using alpha = 5%.

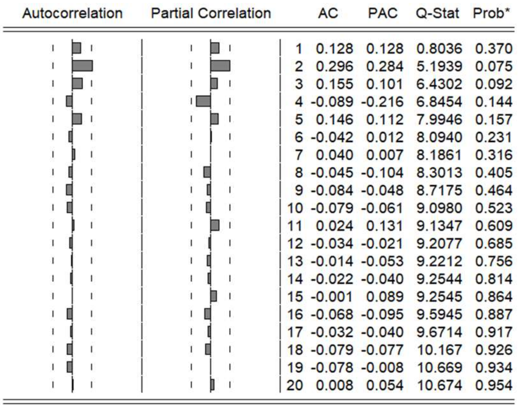
Table 8 – Results of Normality Assumption Test.

Model	Test Statistics	p-values	Decision
Pre-Pandemic	1.4167	0.4925	Fulfilled
Combined	8.0169	0.0135	Not Fulfilled



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.147	0.147	0.8032	0.370
		2	0.112	0.093	1.2873	0.525
		3	-0.043	-0.074	1.3610	0.715
		4	-0.031	-0.026	1.3998	0.844
		5	0.040	0.063	1.4689	0.917
		6	0.206	0.203	3.3229	0.767
		7	0.151	0.089	4.3572	0.738
		8	0.209	0.149	6.4077	0.602
		9	0.020	-0.023	6.4267	0.697
		10	-0.193	-0.226	8.3240	0.597
		11	-0.169	-0.138	9.8444	0.544
		12	-0.080	-0.052	10.198	0.599
		13	0.049	0.028	10.337	0.666
		14	-0.009	-0.113	10.342	0.737
		15	-0.117	-0.182	11.229	0.736
		16	-0.172	-0.116	13.238	0.655

Figure. 17 - Quadratic Residual Correlogram of Pre-Pandemic Model.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.128	0.128	0.8036	0.370
		2	0.296	0.284	5.1939	0.075
		3	0.155	0.101	6.4302	0.092
		4	-0.089	-0.216	6.8454	0.144
		5	0.146	0.112	7.9946	0.157
		6	-0.042	0.012	8.0940	0.231
		7	0.040	0.007	8.1861	0.316
		8	-0.045	-0.104	8.3013	0.405
		9	-0.084	-0.048	8.7175	0.464
		10	-0.079	-0.061	9.0980	0.523
		11	0.024	0.131	9.1347	0.609
		12	-0.034	-0.021	9.2077	0.685
		13	-0.014	-0.053	9.2212	0.756
		14	-0.022	-0.040	9.2544	0.814
		15	-0.001	0.089	9.2545	0.864
		16	-0.068	-0.095	9.5945	0.887
		17	-0.032	-0.040	9.6714	0.917
		18	-0.079	-0.077	10.167	0.926
		19	-0.078	-0.008	10.669	0.934
		20	0.008	0.054	10.674	0.954

Figure. 18 - Quadratic Residual Correlogram of Combined Model.

Table 9 – Checking Results of Non-Multicollinearity Assumptions.

Variable	VIF Value	
	Pre-Pandemic	Combined
GDPLOG(-1)	1.2803	1.4324
PC1	1.9083	2.2648
PC2	1.6340	1.8087
PC3	1.1303	1.2118
PC4	1.2181	1.2529
PC5	1.9706	1.2541
PC6	1.3343	-
Decision	Fulfilled	Fulfilled
GDPLOG(-1)	1.2803	1.4324

4.6. Comparison of Pre-Pandemic and Combined Model Nowcasting Result

After forming the MIDAS model with PDL weighing with a polynomial degree of 1 and a maximum lag of 4, pre-pandemic and combined model nowcasting was carried out. For the pre-pandemic model, nowcasting is carried out on the gdplog variable for the first to fourth quarters of 2019, while the combined model nowcasting is carried out on the d1pdblog variable for the first quarter of 2022 to the first quarter of 2023.

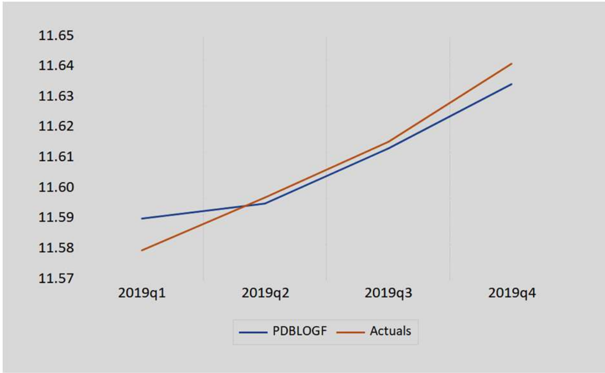


Figure. 19 - Nowcasting Result of Pre-Pandemic Model.

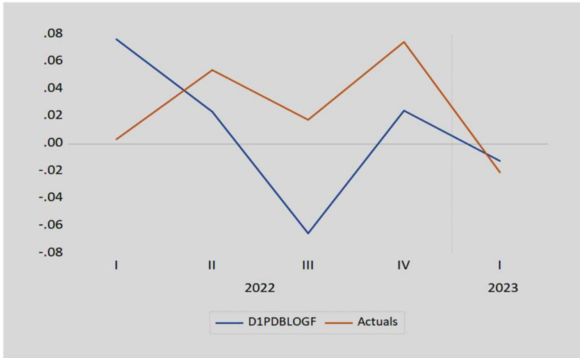


Figure. 20 - Nowcasting Result of Combined Model.

Table 10 – Evaluation Results.

Model	RMSE	MAE
Pre-Pandemic	0.005753	0.003359
Combined	0.056032	0.048976

Based on table 12, it can be seen that both RMSE and MAE pre-pandemic nowcasting results show a smaller value than the combined model nowcasting results. In addition, if you look at the graph of the nowcasting results, both the pre-pandemic and combined models tend to show the direction of movement of the nowcasting results correctly.

5. Conclusion

This study aims to compare the GDP nowcasting model for the accommodation and food service activities sector before the pandemic and combined with the MIDAS method. The parameters used in the MIDAS model are models with PDL weights with a polynomial degree of 1 and a maximum lag of 4 which will be automatically selected by Eviews (automatic lag selection). From the models that have been made, nowcasting is carried out and it is found that the RMSE and MAE values for the pre-pandemic model are smaller than the combined model. The RMSE values for each of the pre-pandemic and combined models were 0.005753 and 0.056032 and the MAE values were 0.00359 and 0.048976 for the pre-pandemic and combined models. If viewed graphically or the direction of the nowcasting results, the direction of the nowcasting results produces a direction that tends to be the same as the direction of the GDP of the accommodation and food service activities sector for both the pre-pandemic and combined models. Even though the results of nowcasting are always below the expected value, the resulting direction of GDP growth tends to be in accordance with the actual value. For this reason, the use of the MIDAS model with Google Trend as a predictor variable for the GDP of the accommodation and food service activities sector is considered quite appropriate in predicting its growth direction.

However, the model produced in this study is difficult to interpret, because there are several transformations made to make the variable stationary, moreover the principal component analysis of the Google Trend variable is carried out to reduce the variable. For this reason, it is recommended that further research be used with other variables in order to produce a more interpretable model. In addition, this research is not recommended for predicting the nominal GDP of the accommodation and dining and drinking sector, but can be used as a reference if you want to predict the growth direction of the accommodation and dining and drinking sector.

REFERENCES

Aastveit, KA, Gerdrup, KR, Jore, AS, & Thorsrud, LA (2014). Nowcasting GDP in real time: A density combination approach. Journal of Business & Economic Statistics, 32(1), 48-68. doi: https://doi.org/10.1080/07350015.2013.844155

Ahmed, WS, Sheikh, J., Ur-Rehman, K., Shad, SA, & Butt, FS (2020). New continuum of stochastic static forecasting model for mutual funds at investment policy level. *Chaos, Solitons & Fractals*, *132*, 109562. doi: https://doi.org/10.1016/j.chaos.2019.109562

Banbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting. https://ideas.repec.org/p/ecb/ecbwps/20101275.html

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow (1564). European Central Bank. doi: https://doi.org/10.1016/B978-0-444-53683-9.00004-9

Box, GE, Jenkins, GM, Reinsel, GC, & Ljung, GM (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Bzdok, D., & Yeo, BTT (2017). Inference in the age of big data: Future perspectives on neuroscience. doi: https://doi.org/10.1016/j.neuroimage.2017.04.061

Claudio, JC, Heinisch, K., & Holtemöller, O. (2020). Nowcasting East German GDP growth: a MIDAS approach. Empirical Economics, 58(1), 29–54. doi: https://doi.org/10.1007/s00181-019-01810-5

Efendy, F., Huriyati , R., Disman , D., & Sultan, MA (2021). Using Google Trends In Content Marketing Strategy Planning for Increasing Competitiveness _ Perpetrator Business in the Internet World. National Seminar: Innovation & Adoption Technology, 2(2), 192-200.

Ghysels, E., Clara, PS, & Valkanov, R. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models (2nd ed.). University of North Carolina and CIRANO. https://ideas.repec.org/p/cir/cirwor/2004s-20.html

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: the real-time informational content of ` macroeconomic data. Monetary Journal Economics, 55, p. 665–676. doi: https://doi.org/10.1016/j.jmoneco.2008.05.010

Gil, M., Leiva -Leon, D., Pérez, JJ, & Urtasun, A. (2019). An application of dynamic factor models to nowcast regional economic activity in Spain. *Banco de Espana Occasional Pape*, (1904). doi: https://dx.doi.org/10.2139/ssrn.3349124

Ghysels, E., Sinko, A., & Valkanov , R. (2007). MIDAS regressions: Further results and new directions. Econometric reviews, 26(1), 53-90. doi: https://doi.org/10.1080/07474930600972467

Hanoatubun, S. (2020). Impact of Covid-19 on Indonesian Economy. *EduPsyCouns: Journal of Education, Psychology and Counseling*, *2* (1), 146-153. Retrieved from https://ummaspul.e-journal.id/Edupsycouns/article/view/423

Havranek, T., & Zeynalov , A. (2021). Forecasting tourist arrivals: Google Trends meets mixed-frequency data. Tourism Economics, 27(1), 129-148. doi: https://doi.org/10.1177/1354816619879584

Hawari, R., & Kartiasih, F. (2017). Kajian Aktivitas Ekonomi Luar Negeri Indonesia Terhadap Pertumbuhan Ekonomi Indonesia Periode 1998-2014. *Media Statistika*, *9*(2), 119. doi: https://doi.org/10.14710/medstat.9.2.119-132

Human, N. (n.d.). Principal component analysis (PCA). RPubs https://rpubs.com/nadhifanhf/principal-component-analysis

Hyndman, RJ, & Athanasopoulos , G. (2018). Forecasting Principles and Practice (2nd ed.). OTexts. https://otexts.com/fpp2/index.html

Indonesia, S. Hotel and Accommodation Statistics Others in Indonesia 2019. Statistics Indonesia.

Indonesia, S. Hotel and Accommodation Statistics Others in Indonesia 2020. Statistics Indonesia.

Indonesia, S. Hotel and Accommodation Statistics Others in Indonesia 2021. Statistics Indonesia.

Kartiasih, F. (2019). Dampak Infrastruktur Transportasi Terhadap Pertumbuhan Ekonomi Di Indonesia Menggunakan Regresi Data Panel. *Jurnal Ilmiah Ekonomi Dan Bisnis*, *16*(1), 67–77. https://doi.org/10.31849/jieb.v16i1.2306

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. International Journal of Forecasting, 27(2), 529-542. doi: https://doi.org/10.1016/j.ijforecast.2010.02.006

Laine, OM, & Lindblad, A. (2021). Nowcasting Finnish GDP growth using financial variables: A MIDAS approach. Journal of the Finnish Economic Association, 2(1), 74-108. https://ssrn.com/abstract=3973878

Ma'Arif, S. (2019). Growth Nowcasting Product Domestic Indonesian Gross Using Dynamic Factor Model [Master's thesis].

Manalu, EP, Muditomo , A., Adriana, D., & Trisnowati , Y. (2020, August). Role of information technology for successful responses to Covid-19 pandemic. In 2020 international conference on information management and technology (ICIMTech) (pp. 415-420). IEEE. doi: https://doi.org/10.1109/ICIMTech50083.2020.9211290

Mariana. (2013). Analysis Main Components. Journal Mathematics and Learning, 2(2), 99-114.

Mariano, RS, & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly. Journal of Applied Econometrics, 4(18), p. 427-443. doi: https://doi.org/10.1002/jae.695

Mariano, RS, & Murasawa Y. (2010). A Coincident Index, Common Factors, and Monthly Real GDP. Oxford Bulletin of Economics and Statistics, 1(72), pp 27-46. doi: https://doi.org/10.1111/j.1468-0084.2009.00567.x

Meiryan. (2021, August 6). Understanding the autocorrelation test in the regression model. Accounting. https://accounting.binus.ac.id/2021/08/06/memahami-uji-autocorelasi-dalam-model-regresi/

Nasution, DAD, Erlina , E., & Muda, I. (2020). Impact the Covid-19 pandemic against Indonesian economy. *Journal Benefita*, *5* (2), 212-224. doi: http://doi.org/10.22216/jbe.v5i2.5313

Parigi, G., & Schlitzer G. (1995). Quarterly Forecast of The Italian Business Cycle by Means of Monthly Economic. Journal of Forecasting, 14, p. 117-141. doi: https://doi.org/10.1002/for.3980140205

Paludi, S. (2022). a year The Covid-19 Pandemic and Its Impact To Industry Indonesian Tourism. *Equilibrium: Journal Education and Economics Research*, *19* (01), 49-60. doi: https://doi.org/10.25134/equi.v19i01.4337

Purnaningrum1, E., & Ariqoh, I. (2019). Google Trends Analytics in Field Tourism. Economic Magazine, 24(2), 232-243. doi: https://doi.org/10.36456/majeko.vol24.no2.a2069

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. International Journal of Forecasting, 32(2), 257-270. doi: https://doi.org/10.1016/j.ijforecast.2015.07.004

Rahayu, A. (2021, November 30). Heteroscedasticity and Gamma Regression. Binus University Malang. https://binus.ac.id/malang/2021/11/heteroskedastisitas-dan-regresi-gamma/#:~:text=Heteroskedastisitas%20merupakan%20false%20one%20factor,(coefficient)%20regression%20will%20disturb

Ringo, J., & Monika, A. (2021). Dynamic Factor Model Application for Nowcasting Regional Economic Growth Using Google Trends Data in Indonesia. *Official Statistics National Seminar*, *2021* (1), 157-165. doi: https://doi.org/10.34123/semnasoffstat.v2021i1.806

Rosita, R. (2020). Influence Covid-19 pandemic against MSMEs in Indonesia. *Journal Lantern Business*, *9* (2), 109-120. doi: https://doi.org/10.34127/jrlab.v9i2.380

Viceira, LM (2012). Bond risk, bond return volatility, and the term structure of interest rates. International Journal of Forecasting, 28(1), 97–117. doi: https://doi.org/10.1016/j.ijforecast.2011.02.018

Vollaro, M., Raggi, M., & Viaggi , D. (2021). Public R&D; and European agriculture: impact on productivity and returns on R&D; expenditures. *Bio-based and Applied Economics Journal*, *10* (1050-2021-1257), 73-86. doi: http://dx.doi.org/10.22004/ag.econ.312984

Willmott, CJ, & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 30(1), 79-82. doi: https://doi.org/10.3354/cr030079

Woloszko, N. (2020). Activity tracking in real time with Google Trends.

Xu, Q., Bo, Z., Jiang, C., & Liu, Y. (2019). Does Google search index really help predicting stock market volatility? Evidence from a modified mixed data sampling model on volatility. Knowledge-Based Systems, 166, 170-185. doi: https://doi.org/10.1016/j.knosys.2018.12.025