Utilizing Google Trends Data to Examine the Impact of Open Unemployment Rates on Indonesia's Gross Domestic Product

[1]Giani Jovita Jane, [2]Rafif Hasabi, [3]Sinatrya Dwi Purnatadya, [4]Fitri Kartiasih[*]
[1,2,3]Program Studi DIV Komputasi Statistik, Politeknik Statistika STIS
[4]Program Studi DIV Statistika, Politeknik Statistika STIS
[1,2,3,4]Jl. Otto Iskandardinata No.64C Jakarta 13330, Indonesia
*e-mail: fkartiasih@stis.ac.id

(***received*****:** 15 October 2024, ***revised*****:** 28 October 2024, ***accepted*****:** 5 November 2024)

Abstract

Data related to the economy have varying frequencies and have delays in publication time. Such as data on the Open Unemployment Rate (OUR) with a semi-annual frequency and Gross Domestic Product at Constant Prices (riil GDP) according to expenditure with a quarterly frequency. So, frequency conversion is required to conduct simple regression modelling using these data. On the other hand, big data such as Google Trends is an additional predictor to estimate OUR and GDP data to overcome delays in publication time. Then the estimated data is modelled to investigate the effect of OUR on GDP. Data conversion uses the Chow-Lin method, while estimation with Google Trends data uses robust regression. The study shows that the estimation results using Google Trends as an additional predictor provide more accurate results than without Google Trends data for OUR and GDP data. Based on the robust regression results, it can be concluded that the OUR has a negative and significant effect on GDP. The findings provide valuable insights for supporting sustainable economic policy and further research on economic analysis.

Keywords: OUR, GDP, random forest, robust regression, chow-lin interpolation

1 Introduction

In today's modern economic era, the progress of a country is always associated with economic growth. Economic growth can be used to see a country's economic success [1], [2], [3]. Economic growth is closely related to development to achieve prosperity. Development is achieved by overcoming various community economic and social problems, such as unemployment. Thus, economic growth will lead to prosperity and be a sign that development has been successful [4], [5], [6].

The success of development activities in a country can be seen through several macroeconomic indicators such as Gross Domestic Product (GDP) and Open Unemployment Rate (OUR) [7]–[10]. In Indonesia, GDP and OUR data are released every three months and once every six months by Statistics Indonesia (BPS) as the National Statistical Office (NSO). The publication of these indicators shows a time lag. For example, the first quarter GDP was released in the second quarter due to the time-consuming evaluation process. This lag causes a reduction in important information that can be used for policymaking. The success of development activities in a country can be seen through several macroeconomic indicators such as GDP and OUR [7], [8], [9]. In Indonesia, GDP and OUR data are released every three months and once every six months by BPS as the National Statistical Office (NSO). The publication of these indicators shows a time lag. For example, the first quarter GDP was released in the second quarter due to the time-consuming evaluation process. This lag causes a reduction in important information that can be used for policymaking. The success of development activities in a country can be seen through several macroeconomic indicators such as GDP and OUR [7], [8], [9]. In Indonesia, GDP and OUR data are released every three months and once every six months by BPS as the National Statistical Office (NSO) [11]. The publication of these indicators shows a time lag. For example, the first quarter GDP was released in the second quarter due to the time-consuming evaluation process. This lag causes a reduction in important information that can be used for policymaking.

 ISSN:2302-8149
e-ISSN:2540-9719

OUR and GDP macroeconomic indicators have a close relationship with each other, as is Okun's view, which states that an increase in output by 2% will reduce the unemployment rate by 1% [12]. This view means that the increase in output (approached by GDP) and OUR have a negative relationship. Many studies have proven this relationship, such as research [13], which uses the MIDAS regression model to model the relationship between the two indicators. Besides that, research [14] used data on economic growth and the annual OUR to see the causal relationship between the two.

One of the simplest methods to see the effect of a variable on other variables is simple linear regression. However, this method requires a data frequency similarity between the response and predictor variables [15]. For example, if the response variable is annual data, the predictor variable must also be annual data. So it will be challenging to apply this regression model to analyze the relationship between GDP and OUR indicators, considering that both have different data frequencies. In previous related studies, the modeling of macroeconomic indicators differed in frequency, so MIDAS regression was used. The MIDAS regression model can only see the relationship between the two macroeconomic indicators but cannot see the magnitude of the effect.

The problem with simple linear regression can be solved by converting low-frequency data into high-frequency data or vice versa. For example, the conversion of quarterly data to monthly is done by desegregating or interpolating using a specific method. However, this frequency conversion method cannot be used for future periods due to the lag in publishing economic indicators. For example, suppose you want to do a frequency conversion by desegregating GDP data for April and May. In that case, we must have GDP data for the second and third quarters. Therefore, a suitable forecasting method is needed to obtain this value.

On the other hand, big data is a data source that has great potential in Official Statistics. With its great potential, Google Trends data, which is publicly available, accessible, and real-time (daily frequency data), can be an asset to improve the quality of Official Statistics data such as GDP and OUR. Google Trends captures a person's interests or behavior through the search keywords entered. Research [16] concluded that Google Trends data could be a good predictor for improving forecasting results. So, it is hoped that using Google Trends data can improve GDP and OUR data forecasting results.

Based on some of the problem descriptions above, to investigate the effect of the OUR on Indonesia's GDP, this study will use OUR data and estimated GDP using Google Trends data predictors with monthly frequency. This study aims to evaluate the estimation results of OUR data and monthly GDP using Google Trends data predictors and to analyze the OUR effect on GDP. The evaluation uses OUR interpolation and GDP desegregation results as reference values. The findings from this research are expected to provide significant contributions, such as supporting the achievement of the United Nations' Sustainable Development Goals (SDGs), particularly Goal 8, which promotes sustained, inclusive, and sustainable economic growth and employment [17]. Moreover, this study offers a valuable framework for policymakers and researchers by demonstrating how real-time data sources like Google Trends can be leveraged to produce timely economic insights, which can facilitate proactive decision-making and stimulate further research in the field of economic forecasting and data-driven policymaking.

2 Literature Review

2.1. Open Unemployment Rate

According to International Labour Organization (ILO), the standard definition underlies the unemployment rate concept. This definition was established at the *19th International Conference of Labour Statisticians* in 2013, consisting of three indicators. They are employment, unemployment, and the labor force. Employment means working-age people bound to any activity producing goods or benefits. Unemployment are people of working age who are not working or looking for a job during a specific period and are available to accept job opportunities. At the same time, the labor force is the supply of labor to produce goods and services in return for wages or profits. Thus, these three indicators define the unemployment rate formula in Equations (1) and (2).

$$OUR = \frac{Unemployment}{Labor\ force} \times 100 \quad (1)$$

 ISSN:2302-8149
e-ISSN:2540-9719

$$OUR = \frac{Unemployment}{Unemployment + Employment} \times 100 \quad (2)$$

The Open Unemployment Rate (OUR) is a labor indicator generally used to explain the labor market and economic ability to create jobs. The indicator can support economic decisions, especially in the labor market [4], [18], [19]. Another usual implementation using the unemployment rate is to monitor the economic cycle, such as the unemployment rate trend, which is also related to economic performance and its variation. A low OUR does not necessarily mean the labor market is good and vice versa. Complementary data is needed to describe the quality of employment, other measures of underutilization of labor, and unemployment conditions to provide a complete picture of the labor market. An example of a case that can be taken is the condition of older people in the labor market. Older workers tend to have lower unemployment rates than other working-age populations. However, on closer inspection, it turns out that this low number does not directly reflect that the elderly labor market is in a good situation. Far from this, the elderly experience difficulties in the labor market. Low labor utilization means experiencing hopelessness in finding work. Also, seasonal unemployment is not considered unemployment (ILO Department of Statistics' Data Production and Analysis Unit, 2019).

2.2. Gross Domestic Product

Gross Domestic Product (GDP) is one of the macroeconomic indicators used to measure the economic condition of a country at a particular time [11], [20], [21], [22]. GDP can also be used for static and dynamic analysis—especially concerning macroeconomics—and for international comparisons. In its evolution, this indicator offers the possibility of an appreciation of income per inhabitant. This aspect leads to an assessment of living standards, to quality of life [23]. Attempts to measure the aggregate income of a country started in the 17th century. They became one of the main contributions to economic knowledge during the late 20th century [24].

According to BPS (2022), calculating GDP using the constant price approach is helpful for the economic growth rate of a country. The economic growth rate is usually calculated by comparing the amount of GDP at constant prices in one year to another. In other words, changes in the GDP value at constant prices each year can determine a country's economic growth. Thus it is clear that GDP is an important indicator to determine a country's economic growth, which in turn becomes a measure of the success of development carried out by the government in a country [25], [26], [27], [28]. According to BPS (2022), calculating GDP using the constant price approach is helpful for the economic growth rate of a country. The economic growth rate is usually calculated by comparing the amount of GDP at constant prices in one year to another. In other words, changes in the GDP value at constant prices each year can determine a country's economic growth. Thus it is clear that GDP is an important indicator to determine a country's economic growth, which in turn becomes a measure of the success of development carried out by the government in a country [25], [26], [27]. Calculating GDP using the constant price approach is helpful for the economic growth rate of a country [11]. The economic growth rate is usually calculated by comparing the amount of GDP at constant prices in one year to another. In other words, changes in the GDP value at constant prices each year can determine a country's economic growth. Thus it is clear that GDP is an important indicator to determine a country's economic growth, which in turn becomes a measure of the success of development carried out by the government in a country [27], [28].

Based on the information on the google.com page, Google Trends is a site that Google makes to analyze the popularity of top search queries on Google Search in various locations and languages. Google Trends allows the user to access big-size sample from the search on Google that is anonymous, categorized, and aggregated [29]. There are two types of samples available on this site. They are real-time data from the last seven days and non-real-time data (a separate sample from real-time data with a search period from 2004 to 72 hours before the current search). Research [29] state that Google Trends offer a volume index of Google queries based on category and geography of search frequency. The index is obtained by normalizing the total query volume of a search term of a geographic location divided by the total number of queries of a particular region in a given period. A zero index refers to a query search on January 1, 2004. The bigger the index, the closer the query search to the recent period. It is because the obtained index explains the deviation percentage from the query share on January 1, 2004. The index data is available at the country and state levels (for the United States and several other countries).

 ISSN:2302-8149
e-ISSN:2540-9719

The data can be downloaded using several procedures. The data is downloaded using Application Programming Interface (API) in this research. Pytrend is an API that provides many methods to download information from Google Trends in Python [30]. This information is interest over time (interest based on a period, either from a single period or a different range of periods in the same time), historical hourly interest, interest by region, related topics, related queries, trending searches, top charts, and suggestion information. The interest over time method returns the index of a historical search when a specific keyword is the most searched according to interest over time.

A bunch of previous research has experimented doing nowcasting, the method to predict the present using real-time data. Google Trends is one of the most favorable open-source sites and offers real-time data. Research [31] has forecasts macroeconomic indicators using Google Trends. The research chose Google Trends because it is built on the daily public interest through behavioral search records to explain economic activity at a real-time level. This research proved that the Google Trends model succeeded in capturing the change in GDP better than the model without Google Trends.

2.3. Open Unemployment Rate Data Interpolation

Open Unemployment Rate (OUR) data has a semi-annual frequency every February and August. In this study, the analysis was carried out with monthly data frequency, so it is necessary to convert low-frequency data, such as semi-annual, into high-frequency data, such as monthly or weekly. Research [32] interpolated OUR data using the Chow-Lin method approach. This research converts data from annual frequency data to monthly data so that the estimation results are in the form of monthly data. The results obtained, the value of the OUR and the value of the OUR resulting from the interpolation, is almost the same on average and standard deviation. Based on this research, this study uses the Chow-Lin method approach to interpolate semi-annual OUR data into monthly.

2.4. Gross Domestic Product Data Disaggregation

The availability of short-term (monthly) economic data, such as Gross Domestic Product (GDP), can reflect actual economic developments that are currently taking place [33]. In addition, in macroeconomic analysis, differences in the time-frequency of the data used are often found [13], as in this study which has data with different frequencies. So desegregating GDP data is one way that can be done to overcome this. Annual GDP data is the sum of quarterly GDP values. So the quarterly GDP value is also the sum of the monthly GDP values [34]. Instead, annual GDP values can be distributed over a quarterly observation period. Thus, quarterly GDP makes it possible to be distributed (distributed) into monthly GDP. The order used is $m = 3$ if the desegregation is carried out from quarterly GDP data to monthly GDP data. Equation (3) expresses vector desegregation.

$$c' = \left(\frac{1}{m}, \frac{1}{m}, \frac{1}{m}\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \quad (3)$$

Suppose the elements of the vector above are added up. In that case, the result can be expressed as a monthly average from one quarter, referred to as the aggregate value (Luthfiana & Nasrudin, 2018). The GDP value measured monthly from quarterly GDP data is called the GDP desegregation value. Several methods that can be used to desegregate GDP data are Cubic Spline, Denton, Denton-Cholette, Chow-Lin Max Log, Chow-Lin Min RSS, Fernandez, and Litterman. Research [35] used the Chow-Lin Max-Log method to desegregate GDP data and produced quite good results. In addition, research [36] shows that using the Chow-Lin Max Log method is better than using the Chow-Lin Min RSS method. Based on previous research results, researchers used the Chow-Lin Max Log method to desegregate GDP data.

2.5. Relationship Between Open Unemployment Rate and Gross Domestic Product

Akmal [37] stated that the high Open Unemployment Rate (OUR) value contributes to low GDP value because of the high amount of unemployment. High employment indicates slow economic growth affected by too much unemployment, leading to an unproductive economy. The output would grow when labor is utilized fully. A supporting publication mentions that OUR and GDP have a negative relationship over an extended period [38]. If the GDP growth rate went down below the labor growth rate, the new available jobs would not be enough to accommodate all new job seekers. Then the proportion of the employed labor force falls, and the unemployment rate rises. A similar idea was sparked in 1962 called Okun's Law. Specifically, Okun's Law initiates the one percent decline in the unemployment rate. Real GDP should grow about two percentage points faster than the potential GDP for a certain period [38].

 ISSN:2302-8149
e-ISSN:2540-9719

As stated before, Okun's Law concludes a negative relationship between unemployment and output gaps, as explained by the equation (4) (Lee & Huruta, 2019).

$$U - U^n = -0.5(Y - Y^n) \quad (4)$$

Where:
$U - U^n$: Unemployment rate gap
$Y - Y^n$: Output gap

In actual practice, the conditions in each country vary. Not all countries experience the Okun's Law economic phenomenon. Some research has been conducted to study the existence of the Law in Indonesia. Lee and Huruta successfully found the Okun's Law phenomenon in Indonesia using Structural Vector Autoregression between OUR and GDP from 1987 to 2017 [39]. Remarkably, there is a causality relationship between OUR on GDP. Another research found a similar thing to this research. Okun's law exists in Indonesia when OUR is significant to GDP, as explained by the first differentiation model between OUR and GDP from 1986 to 2018 [40].

3 Research Method

This chapter outlines the methodological approach employed to address the research questions and objectives in a systematic manner. It begins by providing a comprehensive account of the data utilized in this research, along with a detailed explanation of its sources. Subsequently, the research workflow is presented, delineating the procedures and analytical techniques employed to attain the research objectives.

3.1 Research Data

This study uses secondary data from the Statistics Indonesia (BPS) website and Google as shown in Table 1. The data from the BPS used is the GDP data according to spending at a constant price basis (ADHK) for the 2010 base year from 2005 to 2022. In that period, there are differences in the base year in the released constant price GDP data. By BPS, it is necessary to carry out the process of equalizing the base year for GDP data based on constant prices for 2005 – 2009 using the 2000 base year. The GDP data is then converted to data frequency to change the data frequency from quarterly to monthly.

After the GDP data, the data taken from BPS is the OUR data from 2005 to 2022. This OUR data frequency is semi-annual and is released twice a year. BPS routinely releases OUR data in February and August. This OUR data will then be converted to data frequency to change the data frequency from semi-annual to monthly. In addition, this study also used big data generated by the Google search platform in the form of the Google Trends search index. The search index is conversion data with a value of 0 – 100. This data is publicly available, and all features can be accessed via the link https://trends.google.com/trends/?geo=ID. Specifically, the Google Trends data retrieved is data of interest over time by region (search index based on region). The specified region is Indonesia, with monthly data from 2004 to 2022. This data from Google Trends will then be used as a predictor in estimating Indonesia's ADHK OUR and GDP data from 2005 to 2022.

Table 1. Research data periods

Data	Periods	Frequency
Open Unemployment Rate	2005 – 2022	*Semi-annual*
Gross Domestic Bruto	2004 – 2022	Quarterly
Google Trends Search Index	2004 – 2022	Monthly

Due to the difference in the data periods obtained (see Table 1), the data analysis period is determined from 2005 to 2022. Then, as previously mentioned, because the frequencies of the three data are different, they will be equated from low to high frequencies. High-frequency data such as monthly or weekly data. While low-frequency data, such as annual data.

Data retrieval from Google Trends is done by using several search keywords. Selecting search keywords is essential because it will affect the results of predictions or estimates. The selection of keywords can refer to previous research with the same focus. Research [41] estimated OUR data using Google Trends data with the keywords "Indonesian business", "online business", "job market", "find money", "find work", "career", and "job vacancies". Another study on reference [42] estimated OUR

 ISSN:2302-8149
e-ISSN:2540-9719

data using Google Trends data with the keywords "job vacancies", "business", "job", and "job market". Furthermore, research [16] uses search keywords "car", "phone", "vacation", "hotel", "food", and many more to estimate data GDP using Google Trends data. The selection of search keywords in this study refers to previous research with some additional search keywords. For clarity, the search keywords used are shown in Table 2. Category filtering was not done in collecting data because this study wanted to capture all Google Trends search index values in all categories.

Table 2. Google trends search keyword

Search Keyword for Unemployment Rate Data	Search Keyword for GDP Data
business	crisis
Job fair	economic crisis
carrier	financial crisis
job vacancy	recession
job	loan
business opportunity	bankrupt
lay off	out of business
idle, unemployment	debt

3.2 Research Flow

Data estimation of the OUR and GDP Based on Constant Prices (GDP ADHK), as well as analysis of the effect of the OUR on GDP using Google Trends data, are carried out in several stages. The earliest stage is data collection, including OUR and GDP data from the BPS website and Google Trends data. Then the OUR and GDP data are converted monthly for further estimation using the random forest model. The final stage is to analyze the effect of OUR on GDP using simple linear regression and robust regression models. The research flow is shown in Figure 1.

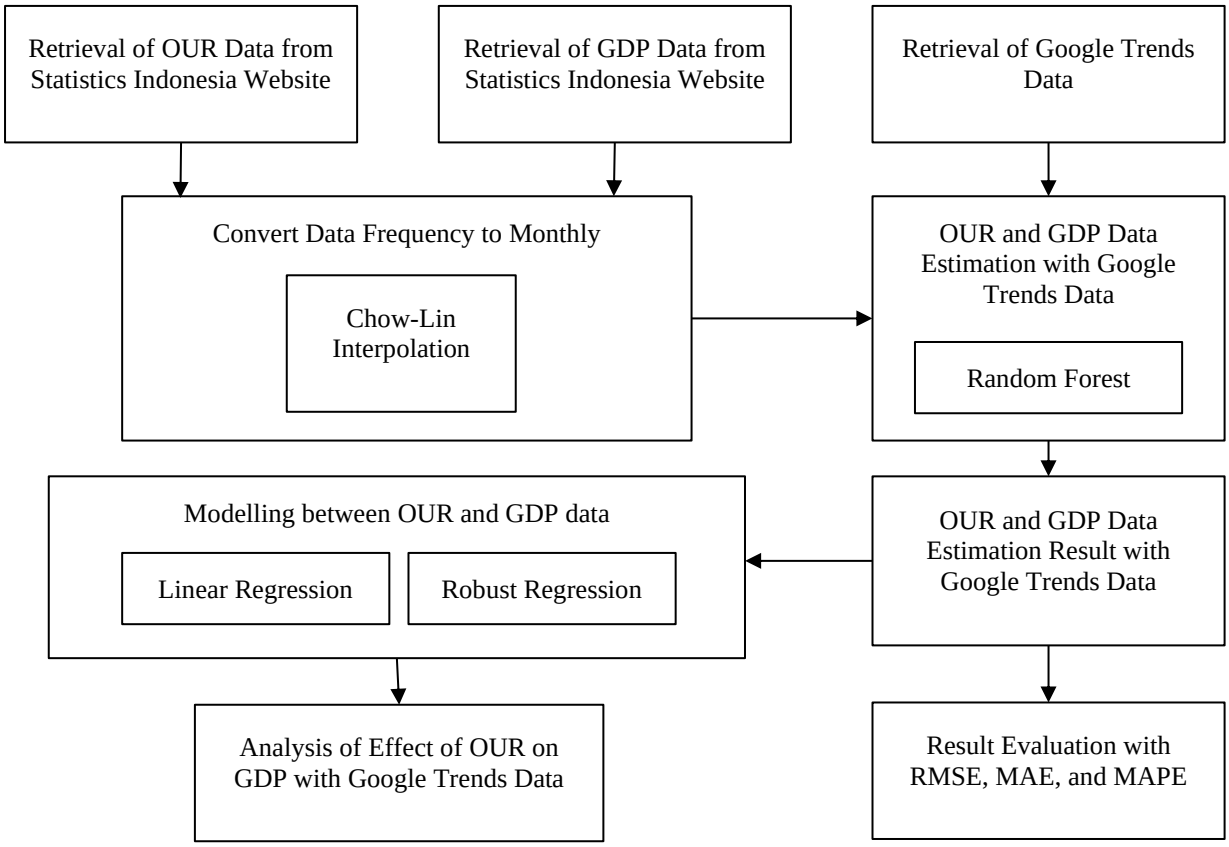

Figure 1. Research flow

3.3 Method

The methods used in this study include the method of equalizing the base year of GDP, Chow-Lin interpolation for GDP data desegregation and frequency conversion of OUR data, the Random

 ISSN:2302-8149
e-ISSN:2540-9719

Forest model for estimation of OUR and GDP data along with the evaluation size used, as well as the Pearson correlation coefficient, linear regression. Simple and robust regression is to investigate the effect of OUR on GDP.

3.3.1 Equates the Based Year of Gross Domestic Product

The value of GDP at constant prices according to expenditure generated by Statistics Indonesia (BPS) in specific years experiences a difference in the base year where variations of the base year are used, such as the 1993, 2000 and 2010 base years. The value of GDP is used at constant prices for the 2010 base year. The process of equalizing the base year must be carried out to get the best results [43]. The steps taken to equalize the base year for GDP data are as follows [44].

1. Determine one data that is calculated using two base years. The data in this study were calculated using two base years in 2010, where GDP in that year was calculated using the 2000 base year and 2010 base year.
2. Find the multiplier by dividing the GDP value for the 2010 base year by the GDP for the 2000 base year, then multiply it by the GDP value for the year you want to equate to the base year.

The following shows GDP data at constant prices with different base years, namely the 2000 and 2010 base years (see Table 3). Suppose you want to equate the base year for the 2009 GDP value from the 2000 base year to the 2010 base year. Then the GDP value is calculated using two; the base year is the GDP value 2010. Then the value for that year is used to find the multiplier. Based on the previous stages, the GDP value for the first quarter of 2009 is obtained using the 2010 base year (see equation (5) and (6)).

Table 3. Equate the based year of GDP

Periods	GDP (2000 = 100)	GDP (2010 = 100)
January 2009	528056.5	…
April 2009	540677.8	…
Juli 2009	561637.0	…
Oktober 2009	548479.1	…
Januari 2010	559683.4	1642356.30
April 2010	574712.8	1709132.00
Juli 2010	594250.6	1775109.90
Oktober 2010	585812	1737534.90
Januari 2011		1748731.20
April 2011		1816268.20
Juli 2011		1881849.70
Oktober 2011		1840786.20

$$GDP_{2009Q1(2010=100)} = \frac{GDP_{2010Q1(2010=100)}}{GDP_{2010Q1(2000=100)}} \times GDP_{2009Q1(2000=100)} \quad (5)$$

$$GDP_{2009Q1(2010=100)} = \frac{1642356.30}{559683.4} \times 528056.5 = 1549549.12 \quad (6)$$

3.3.2 Chow-Lin Interpolation

In extracting high-frequency data (such as monthly or weekly) from low-frequency data (such as annual data), the Chow-Lin approach can be used which uses the desegregation method [35]. Chow-Lin interpolation is used in this study to convert OUR data from a semi-annual frequency to become monthly. It desegregates GDP data with a quarterly frequency to become monthly. Converting the OUR data using the Chow-Lin interpolation method is done with the E-Views data processing application. Meanwhile, desegregating GDP data is done using the RStudio data processing application. The Chow-Lin method ensures that the average value, the first and last data from the high-frequency data series, matches the low-frequency data sets [32]. Equation (7) represents interpolation using the Chow-Lin approach.

$$v_j = \overline{v_j} + \sum_{i=1}^{n} F_{ji}\left(y_i - \sum_{q=1}^{m} H_{iq}\overline{v_q}\right) \quad (7)$$

 ISSN:2302-8149
e-ISSN:2540-9719

The following steps are generally carried out to apply the Chow-Lin interpolation method [32].

1. Create an initial high-frequency (v_j) data set using additional data such as past information. Regression with the least squares method is used to combine these data.
2. Analyze the residual differences between the observed low-frequency series and the high-frequency series that have been aggregated to a low-frequency scale (through the $H \in f^{n \times m}$) matrix).
3. Creates temporally consistent high-frequency (y_i) data by distributing those differences between high-frequency periods using the $F \in R^{n \times m}$ Distribution matrix.

OUR data frequency conversion from semi-annual to monthly with the Eviews data processing application is carried out by selecting the Frequency Conversion Options, namely the Chow-Lin method. For example, the input data entered are the February and August 2005 OUR values. Then with Chow-Lin interpolation, we will fill in the OUR values between those time ranges, namely from March to July. From BPS data, the OUR value in February 2005 was 10.26%, and in August 2005 it was 11.24%. With Chow-Lin interpolation, the results are shown in Table 4.

Table 4. Chow-lin interpolation on OUR data

Period	Value of OUR
February 2005	10.26
March 2005	10.42
April 2005	10.59
May 2005	10.75
June 2005	10.91
July 2005	11.08
August 2005	11.24

As in the research [32], with this interpolation, the actual OUR data remains the same and only fills in the values between February and August 2005. The OUR values will then be used for analysis in the next section. GDP data desegregation using the RStudio data processing application using the tempdisagg package. The specification of the method used is Chow Lin – Max Log. This method ensures that the aggregated GDP's aggregate (sum) value will be the same or close to the quarterly GDP value [34]. The results are shown in Table 5 with Chow-Lin Max Log interpolation.

Table 5. Chow-lin interpolation on GDP data

Periods	GDP Value
April 2005	426187.9
May 2005	432078.3
June 2005	438710.1
July 2005	446083.2
August 2005	448370.1
September 2005	445571

The interpolation results in Table 5 show the value of monthly GDP from April 2005 to September 2005. In other words, these are the months in the second and third quarters. Data from the Central Statistics Agency show that the value of GDP at Constant Prices (2010) in the second and third quarters was Rp. 1296976.28 billion rupiah and Rp. 1340024.25 billion, respectively. Meanwhile, compared with the desegregated data in Table 5, the total value of monthly GDP from April 2005 to June 2005 was IDR 1296976.28 billion. From July 2005 to September 2005, it was IDR 1340024.25 billion—the results obtained following actual quarterly GDP data.

3.3.3 Random Forest

Random forest is a machine-learning method that is quite popular today. Breiman revealed this method in the early 2000s. Random Forest is an algorithm that consists of a collection of tree-structured

 ISSN:2302-8149
e-ISSN:2540-9719

classifiers (decision trees) [45]. Each tree votes for units for the most popular class of inputs [46]. This method has become very popular because it can be applied to various prediction problems, including forecasting on time series data. Besides being easy to use, this method is generally recognized for its accuracy and ability to handle small sample sizes and high-dimensional feature spaces [47]. Research [48] even shows that, in their case, forecasting with Random Forest is better than classical methods for time series data such as ARIMA.

This study will apply Random Forest to predict the monthly value of OUR and GDP. Forecasting will use predictors in the form of lag or data at a previous time. Furthermore, Google Trends data will be included as an additional predictor. The model building will be carried out using the Sklearn package, which is available in the Python programming language.

Forecasting OUR data will use the original value of the data, while forecasting GDP will use the growth value in the relevant month because GDP data has a relatively vital trend component. So, that data transformation is carried out to refine the data and hope the results can be better. GDP growth is defined as the ratio of the increase in GDP to the previous month, as presented in equation (8).

$$Growth\ GDP_t = \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}} \tag{8}$$

The forecasting results will be returned to the GDP value following the equation (9).

$$GDP_t = Growth\ GDP_t \times GDP_{t-1} + GDP_{t-1} \tag{9}$$

3.3.4 Model Evaluation

Several evaluation methods are used to evaluate the estimation results on the OUR and GDP data [49].

1. Mean Absolute Error (MAE) is the average value of the absolute difference between the actual value and the value predicted by the model. A small MAE value indicates that the model is more accurate in predicting the actual value. The formula for calculating MAE is shown in the equation (10). Where Y_t is the actual value and $\hat{Y}_t$ is the predicted value.
$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t| \tag{10}$$
2. Mean Absolute Percentage Error (MAPE) is the average value of the absolute percentage difference from the actual value with the model's predicted value. MAPE value can be calculated using equation (11).
$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{Y_t} \tag{11}$$
3. Root Mean Squared Error (RMSE) is the root of Mean Squared Error (MSE). A small RMSE value indicates that the model is more accurate in predicting the actual value. RMSE can be calculated using equation (12).
$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(Y_t - \hat{Y}_t)^2}{N}} \tag{12}$$

In addition to the evaluation measures above, to measure the goodness of a robust regression model between OUR and GDP data, one more evaluation method is added, namely the coefficient of determination (R^2). The coefficient of determination helps know the model's ability to explain the independent variable; the value of the coefficient of determination is between zero and one. If the value of the coefficient of determination is close to one, the stronger the ability of the independent variables to explain the dependent variable [49]. The value of the coefficient of determination can be calculated using equation (13).

 ISSN:2302-8149
e-ISSN:2540-9719

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (13)$$

3.3.5 Pearson Correlation

Pearson correlation coefficient (ρ) is a linear relationship measurement between two variables [50]. The value of this coefficient varies between -1 to 1. The positive value (more than zero) would mean there is a tendency if one variable increases, the other will also increase or vice versa. A negative value (less than zero) refers to the tendency that if one variable increases, the other will decrease or vice versa. Meanwhile, a zero coefficient means no relationship between the two variables. The closer the absolute value to 1, the stronger the relationship. Some assumptions that must be fulfilled to count the coefficient are the data needs to be in interval or ratio level, the proposed two variables have a linear relationship, and the data must be in normal bivariate distribution ("Pearson's Correlation Coefficient," 2008). Table 6 interprets the absolute value of the Pearson correlation coefficient [51].

Table 6. Correlation coefficient interpretation

Absolute Coefficient Interval	Relationship Level
0,8 – 1	Very Strong
0,6 – 0,7999	Strong
0,4 – 0,5999	Strong Enough
0,2 – 0, 3999	Weak
0,0 – 0,1999	Very Weak

The equation (14) is the formula to calculate the Pearson correlation.

$$r_{xy} = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{\{n\sum X^2 - (\sum X)^2\}\{n\sum Y^2 - (\sum Y)^2\}}} \quad (14)$$

3.3.6 Simple Linear Regression

A simple linear regression model is a model that expresses a probabilistic linear relationship between a predictor variable that is considered to influence a response variable [52]. In general, simple linear regression is written in a mathematical equation as in equation (15).

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (15)$$

Where:
Y = Response Variable
X = Predictor Variabele
ϵ = Residual

The values of β_0 and β_1 are estimated through several methods that zero out the error (residue) from the entire data set. There are two methods commonly used in estimating these parameters, namely Ordinary Least Square (OLS) and Maximum Likelihood. In this study OLS will be used to estimate the value of these parameters. The values of β_0 and β_1 are estimated in equations (16) and (17), respectively.

$$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \quad (16)$$

$$b_0 = \frac{n\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i}{n} \quad (17)$$

Simple Linear Regression requires fulfillment of the assumptions so that the resulting regression equation becomes valid [53]. Several classic assumptions must be met in a simple linear regression model, including:

 ISSN:2302-8149
e-ISSN:2540-9719

1. Homoskedasticity of residuals is a state where every random error has the same variance (homogeneous). If the error variance is not homogeneous, it will make it difficult to measure the correct standard deviation of the error prediction, usually resulting in confidence intervals that are too wide or too narrow. This will later affect the difficulty of obtaining the correct conclusion. Homoscedasticity can be calculated through the Breusch-pagan-godfrey test on the residual values of simple linear regression modeling results.
2. Residual non-autocorrelation means that the residuals must be independent. If this assumption is not fulfilled, it will complicate subsequent analyses, such as difficulty in estimating the variance of errors and difficulty in testing hypotheses in forming confidence intervals. This assumption can be detected with the Breusch-Godfrey Autocorrelation test.
3. Normality of the residual means that the random errors have (or at least approach) a normal distribution. This assumption can be measured through the Jarque-Berra test applied to the residual values generated by the model.

3.3.7 Robust Regression

Outlier is an extreme observation where the absolute residual is more significant than the other observations [54]. Outlier influences model learning or insignificant t-test statistics [55]. Another problem with outliers among the observations is the violation of classic assumptions in regression analysis [56]. To cope with these issues, eliminating outliers is a standard method. However, the consequence of removing outliers is the diminishing number of observations since outliers also explain the actual situations that happened to exist in research. To prevent this consequence in modeling a regression, the solution is to find a regression method that can 'weaken' the effect of outliers while making the regression. The most suitable model that has this characteristic is the robust regression model. The robust regression reduces the effect of outliers with the OLS (*Ordinary Least Square*) method [57]. This model was designed to analyze data with outliers to generate a sturdy and stable model. There are five types of robust regression. These types consist of M-estimators, Least Median Square-estimator, Least Trimmed Square-estimator, S-estimator, and MM-estimator. The M-estimator is the most common robust model when making a regression model. The model derives from the generalized Maximum Likelihood estimator [58]. By minimizing the following objective function, a robust M-estimator model is obtained as in equation (18).

$$\sum_{i=1}^{n} x_{ij} w_i \left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right) = 0 \quad (18)$$

Where:
x_{ij} = observation of the *i*-th predictor variable with the *j*-th parameter
w_i = weight
y_i = observation of the *i*-th dependent variable
β_j = regression coefficient of *j*-th parameter

The regression coefficient is obtained by calculating the *Maximum Likelihood* (see equation (19)) estimator that minimizes the sum of residual function.

$$min \sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - \sum_{j=0}^{k} x_{ij}\beta_i) \quad (19)$$

The matrix notation of the function is written as equation (20).

$$X^T W_0 X b = X^T W_0 Y \quad (20)$$

where:
W_0 = diagonal matrix of weight with $n \times n$ size
X = matrix of the dependent variable with dimensions size $(n \times (p+1))$
b = estimator of outlier (β)

 ISSN:2302-8149
e-ISSN:2540-9719

Y = matrix of the dependent variable with $(n \times n)$ size

Equation (21) below calculates the estimator of β in M-estimator robust regression (IRLS).

$$b_{l+1} = (X^T W_l X^{-1})^{-1}(X^T W_l Y) \quad (21)$$

While the weight function of the M-estimator could be determined by the Huber or the Tukey Bisquare function (see Table 7) [54]. The m-estimator is a robust estimator with a mathematical intrinsic structure where the estimator is not too sensitive to spurious deviation [58]. The estimator evaluates most of the observations around the mean and ignores the false value (observations usually located far from the mean) simultaneously. These characteristics allow the accurate regression to be made even though a priori of outlier or data structure error is unknown. Also, this method returns more information about the nature of data statistics after estimating the residue.

Table 7. Weight function m-estimator

Method	Weight Function	Range
Huber	$w(e_i^*)\{1\ \frac{c}{\lvert e_i^* \rvert}$	$\lvert e_i^* \rvert \le c$ $\lvert e_i^* \rvert > c$ $c = 1.345$
Tukey Bisquare	$w(e_i^*)\{\left[1 - \left(\frac{e_i^*}{c}\right)^2\right]^2$ 0	$\lvert e_i^* \rvert \le c$ $\lvert e_i^* \rvert > c$ $c = 4.685$

4 Result and Analysis

This section presents the results and analyses related to the estimation of the Open Unemployment Rate (OUR) and the Gross Domestic Product (GDP) in Indonesia using Google Trends data. The utilization of Google Trends as a proxy for economic indicators is intended to ascertain the relationship between individuals' online search patterns and macroeconomic variables, particularly the OUR and GDP. The results and analysis are organized into three sections: firstly, the estimation of the OUR and GDP with Google Trends data; secondly, the exploration of the relationship between the OUR and GDP; and finally, the analysis of the impact of the OUR on GDP.

4.1. Unemployment Rate Estimation with Google Trends Data

OUR data converted to frequency will then be used for estimation using Google Trends data as an additional predictor. The model used in this estimation is a random forest with 100 trees. To make a comparison, we also estimate OUR data without additional predictors in the form of Google Trends data. The list of predictor and response variables used in estimating the OUR data is shown in Table 8.

Table 8. Response and predictors variables of the model

Model	Response Variables	Predictor Variables
Random Forest with Google Trends	Monthly OUR	Lag, Google Trends search index each search keyword
Random Forest without Google Trends	Monthly OUR	Lag (OUR(-1))

The trend of monthly OUR values obtained from the previous interpolation results is shown in Figure 2. The OUR from 2005 to 2022 shows a downward trend. There were several increases, such as in 2020 which increased to 7% from 2019, which was 5%. This may have happened due to the Covid-19 pandemic, which has resulted in many economic sectors going out of business and severing work relations [59].

 ISSN:2302-8149
e-ISSN:2540-9719

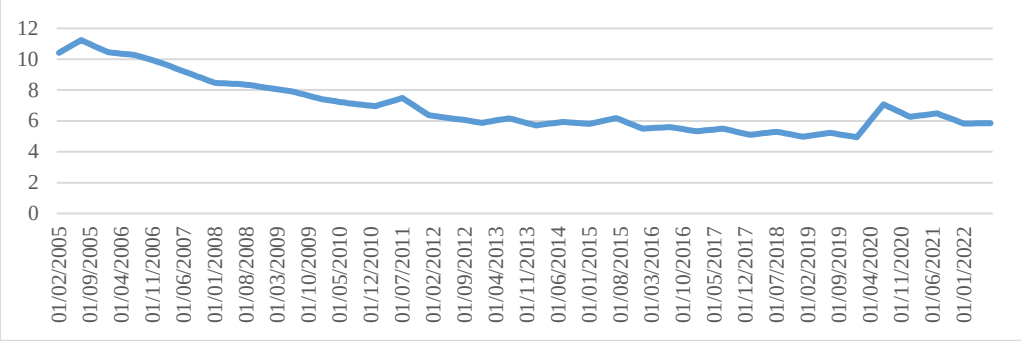

Figure 2. Open unemployment rate trends 2005 - 2022

From Figure 3 below, the search trend related to the OUR tends to decrease from 2004 to 2022. This result aligns with the OUR data trend in Figure 2. From these results, it is also known that the keywords "unemployed" and "unemployed" adequately reflect the condition of open unemployment in Indonesia. The predictor variable in the form of lag is a predictor variable that refers to the use of variable values at the previous time.

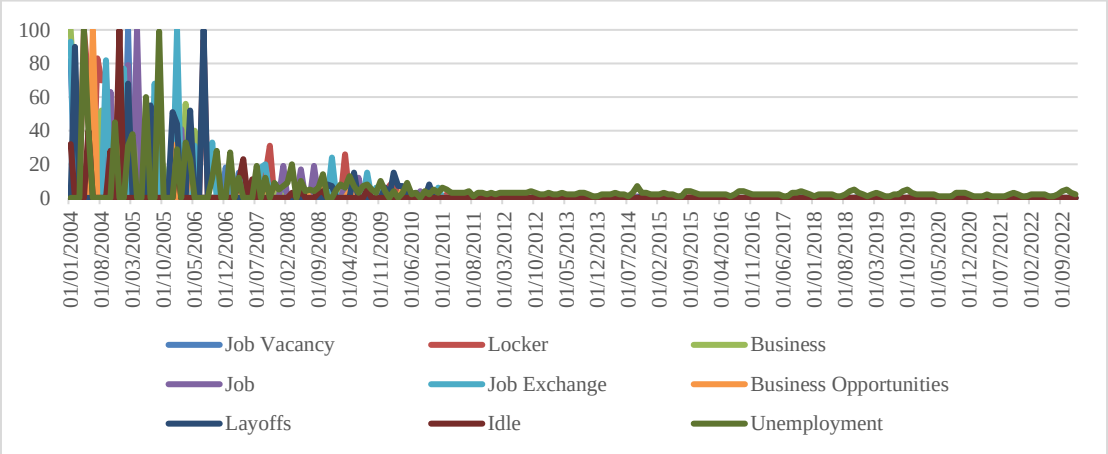

Figure 3. Google trends keyword trends related to OUR 2004 – 2022

The comparison between the estimated OUR data and the original OUR data with additional predictors in the form of Google Trends data is shown in Figure 4. The graph shows the OUR values and their estimates tend to have the same value. Figure 4 shows that modeling using Random Forest with additional predictors in the form of Google Trends search index data is quite good at estimating OUR data.

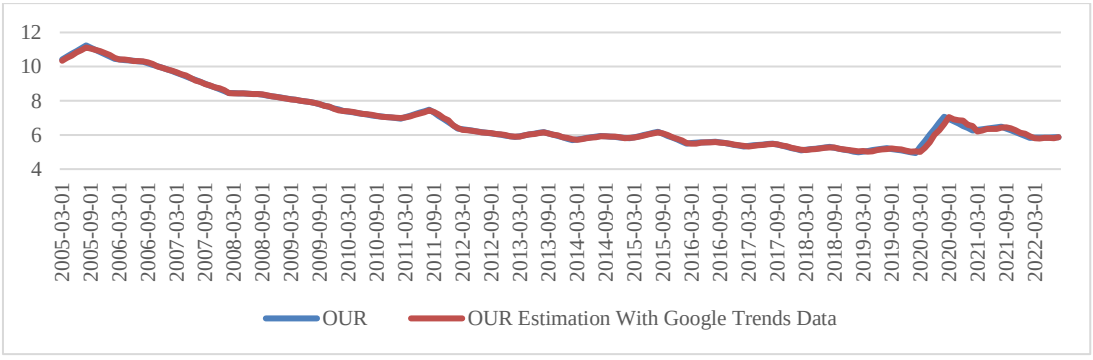

Figure 4. Unemployment rate value and estimated OUR with google trends data

Then to compare, the following uses the same model, namely random forest. Still, it does not add Google Trends data as an additional predictor. The comparison between the estimated OUR data and the original OUR data without additional predictors in the form of Google Trends data is shown in Figure 5. It shows that estimating the OUR value without Google Trends data (using only the lag

 ISSN:2302-8149
e-ISSN:2540-9719

variable) gives good results. Compared to the estimation results using Google Trends data, the estimation results are visually better using Google Trends data.

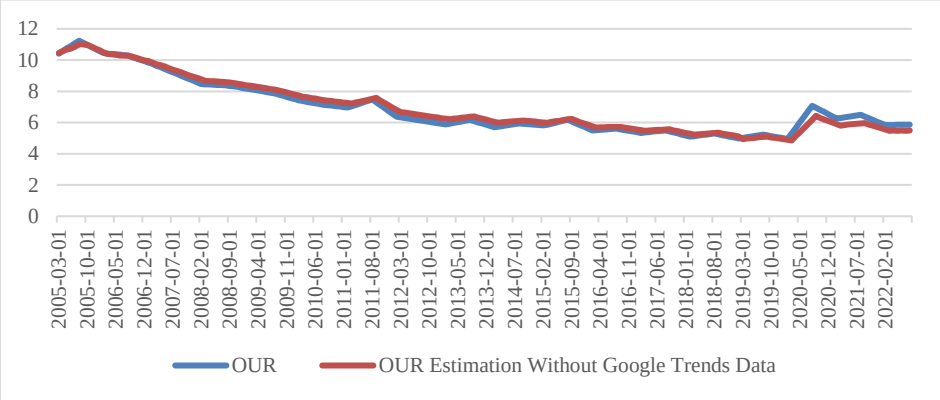

Figure 5. Unemployment rate value and estimated unemployment rate without google trends data

To know for sure the goodness of the two models, evaluation measures are used in the form of RMSE (Root Mean Square Error), MAE (Mean Squared Error), and MAPE (Mean Absolute Percentage Error) values. The results of calculating the evaluation size for each model are shown in Table 9. From the overall evaluation size shown in Table 9, the random forest model with Google Trends has a smaller value than the random forest model without Google Trends. These results strengthen the results of the visual interpretation of the visualization of Figure 4 and Figure 5, which shows that adding additional predictors in the form of Google Trends search index data can improve the estimation results of the OUR. These results also show that Google Trends data can be used to estimate OUR with a relatively small error. OUR data calculated using Google Trends will be used in the following analysis.

Table 9. Unemployment rate estimation model evaluation

Model	RMSE	MAE	MAPE (%)
Random Forest with Google Trends	0.0867	0.0470	0.7211
Random Forest without Google Trends	0.2682	0.2237	3.4471

4.2. GDP Estimation with Google Trends Data

Estimated GDP data uses frequency conversion and Google Trends data as additional predictors. The model used in this estimation is a random forest with 100 trees. For comparison, GDP data was also estimated without different predictors in the form of Google Trends data. The list of predictor and response variables used in this GDP data estimation is shown in Table 10. GDP estimation is carried out at its growth rate. The estimation results are then calculated to obtain the GDP value for the month in question.

Table 10. Response and predictor variable

Model	Response Variable	Predictor Variable
Random Forest with Google Trends	Monthly GDP Growth Rate	Lag, Google Trends search index of each search keyword
Random Forest without Google Trends	Monthly GDP Growth Rate	Lag (GDP Growth(-1))

Trends in monthly GDP values obtained from previous interpolation results are shown in Figure 6. GDP value from 2005 to 2022 shows an upward trend with a visible seasonal pattern. However, in 2020, the GDP value decreased due to the Covid-19 pandemic. The decline occurred by up to 6 percent compared to the previous year, 2019. Towards 2022, the trend of GDP values will start to increase in line with the improving economic conditions in Indonesia.

 ISSN:2302-8149
e-ISSN:2540-9719

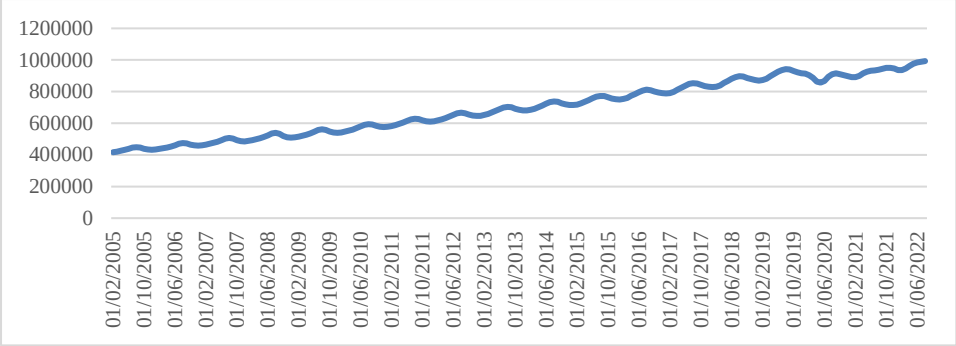

Figure 6. GDP trends in 2005 – 2022

The search keywords used to get the search index from Google Trends as mentioned in the Methodology section. The search index trend of the keywords used from 2004 to 2022 is shown in Figure 7. The search trend for keywords related to GDP tends to fluctuate from 2004 to 2007. Then searches tend to stagnate from 2008 to 2022. However, it is different for the keyword "debt," which tends to increase in that period. This result aligns with the GDP data trend in Figure 6. From these results, it is also known that the keyword "debt" adequately reflects the condition of Indonesia's GDP.

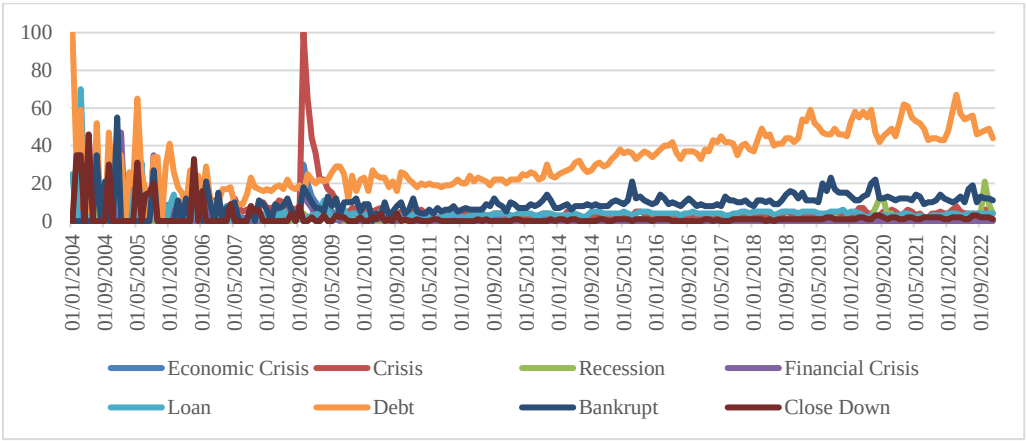

Figure 7. Google trends keyword trends related to GDP in 2004 – 2022

The predictor variable in the form of lag is a predictor variable that refers to the use of variable values at the previous time. A comparison between the original GDP data and the estimated GDP using additional predictors in the form of Google Trends data is shown in Figure 8. The GDP values and estimates tend to have similar values. This indicates that modelling using Random Forest with additional predictors in the form of Google Trends search index data is quite good at estimating GDP data.

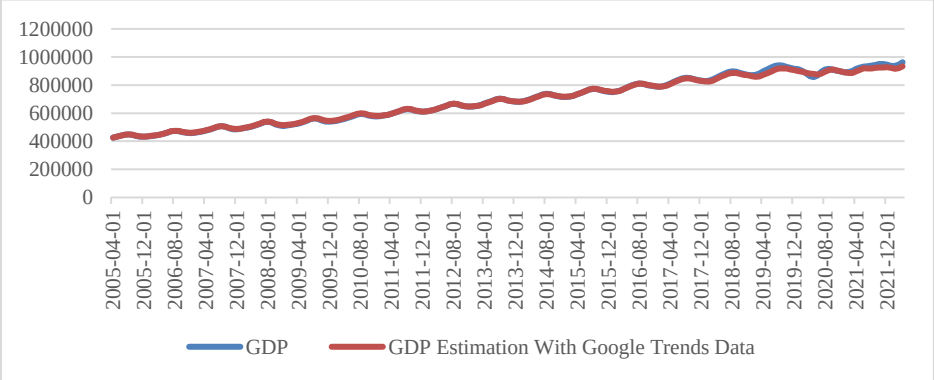

Figure 8. GDP value and estimation with google trends data

As a comparison, the following uses the same model, namely a random forest. Still, it does not add Google Trends data as an additional predictor. The comparison between the original GDP data and the estimated GDP results without different predictors is shown in Figure 9. It shows that the estimated

 ISSN:2302-8149
e-ISSN:2540-9719

GDP value without Google Trends data (only with the lag variable) gives good results. If we compare these results with those before (using Google Trends data), visuals are not better.

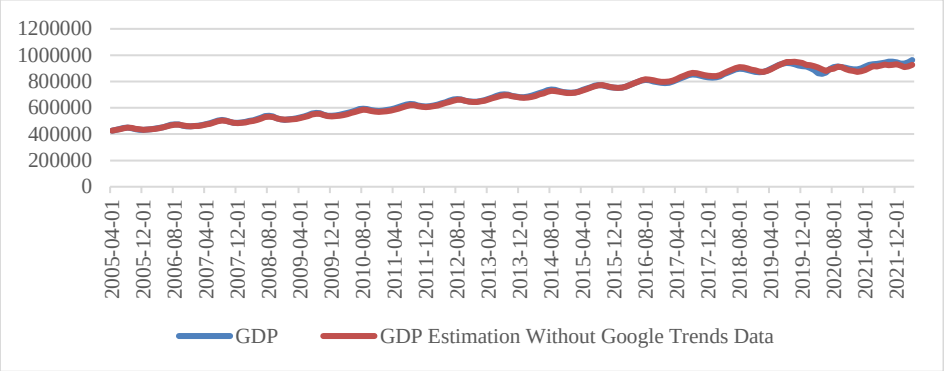

Figure 9. GDP value and estimation without google trends data

To be confident for sure the goodness of the two models, evaluation measures are used in the form of RMSE, MAE, and MAPE values. The results of calculating the evaluation size for each model are shown in Table 11. From the overall evaluation measures, the random forest model with Google Trends Data has a smaller value than the random forest model without Google Trends Data. These results reinforce the results of the previous visualization that the additional predictors in the form of Google Trends search index data can increase the estimated goodness of GDP data. From these results, we have also learned that Google Trends data can be used to calculate GDP data with errors that tend to be small. GDP data estimated using Google Trends will be used in the following analysis.

Table 11. Evaluation measures for each GDP estimation model

Model	RMSE	MAE	MAPE (%)
Random Forest with Google Trends	8298.96	4936.41	0.63
Random Forest without Google Trends	10332.47	7461.64	1.01

4.3. Analyzing the Relationship Between OUR and GDP

Calculating the correlation Pearson coefficient between the two variables is needed to measure the relationship between OUR and GDP initially. Suppose the result shows that there is a strong enough relationship (see Table 12). In that case, the advanced analysis will be conducted using the regression method.

Table 12. Pearson correlation coefficient between OUR and GDP

Variable	Correlation Coefficient (ρ)
Interpolated OUR with Interpolated GDP	-0.906
OUR estimated by *Google Trends* with GDP estimated by *Google Trends*	-0.851

The line chart in Figure 10 below compares GDP and OUR trends from 2005 to 2022. The chart shows a negative relationship between the variables in line with the Pearson coefficient. Table 12 explains the power of the general relationship between the OUR and output. The correlation between OUR and GDP obtained from the interpolation is -0.906. Meanwhile, the relationship between OUR and GDP from Google Trends results is -0.851. These values are a strong negative relationship between both variable pairs. The conclusion that can be taken is the relationship is expressed in the opposite direction. Lower OUR is the consequence of higher GDP or vice versa. These results align with the chart in Figure 10, which explicitly pictures the relationship between the two variables. However, regression modeling will be performed to find out if this is a genuine causality relationship.

 ISSN:2302-8149
e-ISSN:2540-9719

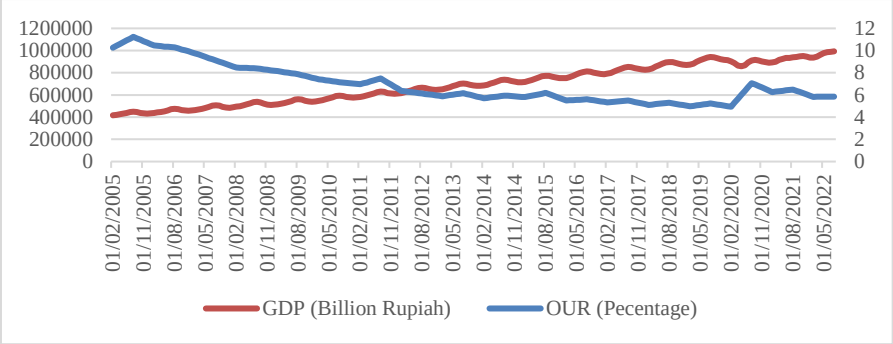

Figure 10. GDP and OUR in indonesia 2005 – 2022

4.4. The Impact of OUR on GDP

Regression analysis is used to investigate the relationship. The target variable of research would be OUR, and the predictor target of an investigation would be GDP. Equation (22) is the proposed model of regression.

$$\ln GDP_i = \beta_0 + \beta_1 OUR_i \quad (22)$$

where:
$\ln GDP_i$ = natural logarithm of GDP i-month
β_0 = intercept
β_1 = regression coefficient for OUR_i
OUR_i = open unemployment rate of the particular month

Natural logarithm transformation of GDP is used instead of the real GDP because the research intends to obtain the percentage of GDP change monthly from the employment side. Also, in this linear modeling, linear variables are needed. This transformation would fulfill the additive assumption (linear) and transform unnormal distribution observations into normally distributed ones [60]. Table 13 and Table 14 show the equation model generated in Eviews.

Table 13. Simple linear regression of interpolated OUR with interpolated GDP

Variable	Coefficient	Standard Error	t-Statistics Values	Probability
β_0	15,423	0,043	355,445	0,000
β_1	-1,046	0,023	-46,48	0,000

Table 14. Simple linear regression of OUR estimated by google trends with GDP estimated by google trends

Variable	Coefficient	Standard Error	t-Statistics Values	Probability
β_0	15,233	0,062	246,259	0,000
β_1	-0,953	0,032	-29,585	0,000

All the regression coefficients are significant in model-making. However, both models don't meet all classical assumption tests. Only homoscedastic and non-multicollinearity assumptions were fulfilled from both models. The researchers used M-estimator robust regression with the same proposed model as simple linear regression to continue performing the analysis. Checking classical assumptions is unnecessary in a robust model because this model assumes that robust standard error is resistant to heteroscedasticity, normality, and autocorrelation assumptions [61]. Table 15 show the robust model that was generated in Eviews.

Table 15. Robust m-estimator regression of interpolated OUR with interpolated GDP

Variable	Coefficient	Standard Error	Z-Statistics Values	Probability
β_0	15,376	0,041	370,286	0,000
β_1	-1,026	0,021	-47,632	0,000

 ISSN:2302-8149
e-ISSN:2540-9719

Table 15 shows that for every one-unit increase in the value of OUR in a month, it can reduce 1.026% of GDP in the same month. These results also indicate a significant influence of the OUR variable on GDP. This is characterized by a probability value that is less than the alpha significance level, which is 5%. Not only that, in line with the results of the correlation coefficient, OUR harms GDP.

Meanwhile, the model generated from Google Trends estimates in Table 16. is not much different from the previous model. The resulting model has the same negative OUR regression coefficient, and both variables significantly influence the natural logarithm of GDP. This model states that for every decrease in the value of OUR in a month, a GDP increase of 0.933%.

Table 16. Robust m-estimator regression of OUR estimated by google trends with GDP estimated by google trends

Variable	Coefficient	Standard Error	Z-Statistics Values	Probability
β_0	15,168	0,042	363,683	0,000
β_1	-0,933	0,022	-42,925	0,000

Evaluation measures are calculated to see how close the prediction results of the two models are in describing the relationship between OUR to GDP and the actual value (GDP as a result of interpolation), namely MAE, MAPE, MSE, and RMSE. The smaller the resulting value, the closer the prediction results of a model are to the actual value. Then, R^2-adj is also calculated to measure how well the predictor variables can describe the existing target variable. From Table 17, it can be concluded that the model resulting from the estimation results of the interpolation results is overall better than the other models.

Table 17. Robust model evaluation

Model	MAE	MAPE (%)	RMSE	$R^2 - adj$
M-Estimator Regression with Interpolated Variables	0,054	0,403	0,074	0,902
M-Estimator Regression with Estimated Variables using Google Trends	0,075	0,559	0,112	0,789

5 Conclusion

Providing an alternative to model economic data with different data frequencies to overcome delays in the publication of financial data, this study successfully used a machine learning approach, namely random forest, by utilizing Google Trends as an additional predictor in estimating data on the Open Unemployment Rate (OUR) and Gross Domestic Product (GDP) based on constant prices according to expenditure. OUR and GDP estimation using Google Trends data as an additional predictor gives quite good results compared to those without using Google Trends as another predictor. The estimation of the Google Trends data approach and machine learning can be used to predict economic data that has delays. These estimated data can be used to pre-indicate economic conditions at a particular time before the original data is published. There is a negative and significant effect between the OUR on GDP as indicated by a negative regression coefficient (1), a negative Pearson correlation coefficient (R) value, and the results of the parameter significance test found that the probability value is less than the significance level of 0.05. For every one-unit decrease in the value of the Open Unemployment Rate in a month, there is an increase in GDP at constant prices according to expenditure by 1.046%.

References

[1] P. Romhadhoni, D. Z. Faizah, and N. Afifah, "Pengaruh Produk Domestik Regional Bruto (PDRB) daerah terhadap pertumbuhan ekonomi dan tingkat pengangguran terbuka di Provinsi DKI Jakarta," *J. Mat. Integr.*, vol. 14, no. 2, p. 113, 2019, doi: https://doi.org/10.24198/jmi.v14i2.19262.

[2] R. Hawari and F. Kartiasih, "Kajian Aktivitas Ekonomi Luar Negeri Indonesia Terhadap Pertumbuhan Ekonomi Indonesia Periode 1998-2014," *Media Stat.*, vol. 9, no. 2, p. 119, 2017, doi: 10.14710/medstat.9.2.119-132.

[3] A. R. Subian, D. A. Mulkan, H. H. Ahmady, and F. Kartiasih, "Comparison Methods of Machine Learning and Deep Learning to Forecast The GDP of Indonesia," *Sist. J. Sist. Inf.*, vol. 13, no. 1, pp. 149–166, 2024.

[4] A. I. Widiyasari, I. A'mal, N. F. Putri Yunardi, and F. Kartiasih, "Analisis Variabel Ketenagakerjaan Terhadap Produktivitas Pekerja Indonesia Tahun 2022," *Semin. Nas. Off. Stat.*, vol. 2023, no. 1, pp. 117–126, 2023, doi: 10.34123/semnasoffstat.v2023i1.1855.

[5] F. A. Sibagariang, L. M. Mauboy, R. Erviana, and F. Kartiasih, "Gambaran Pekerja Informal dan Faktor-Faktor yang Memengaruhinya di Indonesia Tahun 2022," in *Prosiding Seminar Nasional Official Statistics 2023: Peran Official Statistics dan Sains Data dalam Mewujudkan Ketahanan Pangan Nasional*, 2023, pp. 151–161.

[6] F. Kartiasih, N. D. Nachrowi, I. D. G. K. Wisana, and D. Handayani, *Potret Ketimpangan Digital dan Distribusi Pendapatan di Indonesia: Pendekatan Regional Digital Development Index*, 1st ed. Depok, Indonesia: UI Publishing, 2023.

[7] Y. P. Ningsih and F. Kartiasih, "Dampak Guncangan Pertumbuhan Ekonomi Mitra Dagang Utama terhadap Indikator Makroekonomi Indonesia," *J. Ilm. Ekon. dan Bisnis*, vol. 16, no. 1, pp. 78–92, 2019, doi: https://doi.org/10.31849/jieb.v16i1.2307.

[8] S. J. Adwendi and F. Kartiasih, "Penggunaan Error Correction Mechanism dalam Analisis Pengaruh Investasi Langsung Luar Negeri Terhadap Pertumbuhan Ekonomi Indonesia," *Stat. J. Theor. Stat. Its Appl.*, vol. 16, no. 1, pp. 17–27, 2016, doi: 10.29313/jstat.v16i1.1767.

[9] N. Hartati, "Pengaruh Inflasi Dan Tingkat Pengangguran Terhadap Pertumbuhan Ekonomi Di Indonesia Periode 2010 – 2016," *J. Ekon. Syariah Pelita Bangsa*, vol. 5, no. 01, pp. 92–119, 2020, doi: 10.37366/jespb.v5i01.86.

[10] I. K. P. Widyarta, C. P. D. Samosir, P. Aysyah, and F. Kartiasih, "Revisiting Unemployment in Indonesia: Error Correction Model (ECM) Analysis," *J. Akuntansi, Manajemen, Bisnis dan Teknol.*, vol. 4, no. 2, pp. 222–241, 2024.

[11] Badan Pusat Statistik, *Produk domestik bruto indonesia menurut pengeluaran 2018-2022*. Jakarta: Badan Pusat Statistik, 2022.

[12] A. M. Okun, *Potential GNP: its measurement and significance*. Cowles Foundation for Research in Economics at Yale University, 1963.

[13] M. Fajar, "Aplikasi Model MIDAS pada Pemodelan Hubungan antara Tingkat Pengangguran Terbuka dan Pertumbuhan Ekonomi," *Semin. Nas. Off. Stat. 2021*, no. June, 2021, doi: 10.13140/RG.2.2.31228.59529.

[14] R. Efrianti, A. Irawan, and A. Akbar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Pengangguran Di Provinsi Sumatera Selatan Tahun 2002 – 2019," *Http://Journal.Unbara.Ac.Id/Index.Php/Klassen*, vol. 1, no. 1, pp. 36–49, 2021.

[15] C. Foroni, M. Marcellino, and C. Schumacher, "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 178, no. 1, pp. 57–82, Jan. 2015, doi: 10.1111/rssa.12043.

 ISSN:2302-8149
e-ISSN:2540-9719

[16] I. Bouayad, J. Zahir, and A. Ez-Zetouni, "Nowcasting and Forecasting Morocco GDP growth using Google Trends data," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 3280–3285, 2022, doi: 10.1016/j.ifacol.2022.10.129.

[17] United Nation, *The Sustainable Development Goals Report 2022*. New York: United Nation, 2022.

[18] A. M. Suhaib and F. Kartiasih, "The Impact of Digital Technology on Women's Participation in the Labor Force in Papua Province," *J. Ketenagakerjaan*, vol. 19, no. 2, pp. 218–232, 2024, doi: 10.47198/jnaker.v19i2.342.

[19] K. Fadillah and F. Kartiasih, "The effect of COVID-19 and population mobility on the underemployment rate in Indonesia," *J. Kependud. Indones.*, vol. 18, no. 2, pp. 217–236, 2023, doi: 10.55981/jki.2023.2037.

[20] F. Y. Kamal, M. I. Sari, M. F. G. U. Utami, and F. Kartiasih, "Penggunaan Remote Sensing dan Google Trends untuk Estimasi Produk Domestik Bruto Indonesia," *Equilib. J. Penelit. Pendidik. dan Ekon.*, vol. 21, no. 02, pp. 37–59, 2024.

[21] W. E. Pramesthy, P. Muthi, M. A. Budiman, Z. Ahmad, and F. Kartiasih, "The Effect of E-Commerce on Gross Regional Domestic Product and Clustering of Its Characteristics by Utilizing Official Statistics and Big Data," *J. Econ. Business, Account. Ventur.*, vol. 27, no. 1, pp. 14–32, 2024, doi: 10.14414/jebav.v27i1.4136.

[22] N. Yuliana, U. A. Syifa, S. Z. Ramadhanty, and F. Kartiasih, "Nowcasting of Indonesia's Gross Domestic Product Using Mixed Sampling Data Regression and Google Trends Data," *Eig. Math. J.*, vol. 7, no. 2, pp. 67–80, 2024.

[23] C. Anghelache, ÿtefan V. Iacob, and D. L. Grigorescu, "Analysis of the quarterly evolution of the Gross Domestic Product," *Theor. Appl. Econ.*, vol. XXVII, no. 3, pp. 243–260, 2020.

[24] R. England W., "Measurement of social well-being: alternatives to gross domestic product," *Ecol. Econ.*, vol. 25, no. 1, pp. 89–103, 1998.

[25] A. Kusumasari and F. Kartiasih, "Aglomerasi Industri Dan Pengaruhnya Terhadap Pertumbuhan Ekonomi Jawa Barat 2010-2014," *J. Apl. Stat. Komputasi Stat.*, vol. 9, no. 2, pp. 28–41, 2017, doi: https://doi.org/10.34123/jurnalasks.v9i2.143.

[26] F. Kartiasih, "Dampak Infrastruktur Transportasi Terhadap Pertumbuhan Ekonomi Di Indonesia Menggunakan Regresi Data Panel," *J. Ilm. Ekon. Dan Bisnis*, vol. 16, no. 1, pp. 67–77, 2019, doi: 10.31849/jieb.v16i1.2306.

[27] F. Kartiasih, "Transformasi Struktural dan Ketimpangan Antardaerah di Provinsi Kalimantan Timur," *Inov. J. Ekon. Keuang. dan Manaj.*, vol. 15, no. 1, pp. 105–113, 2019, doi: https://doi.org/10.30872/jinv.v15i1.5201.

[28] A. A. G. R. B. D. Pemayun, M. Z. Azizi, N. A. Daulay, N. H. Apriliani, and F. Kartiasih, "Estimation of Java GRDP in Regency/City Level: Satellite Imagery and Machine Learning Approaches," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. X, no. 2, pp. 379–386, 2024.

[29] H. Choi and H. Varian, "Predicting the Present with Google Trends," *Econ. Rec.*, vol. 88, no. SUPPL.1, pp. 2–9, 2012, doi: 10.1111/j.1475-4932.2012.00809.x.

[30] L. Samaras, E. García-Barriocanal, and M.-A. Sicilia, "Comparing Social media and Google to detect and predict severe epidemics," *Sci. Rep.*, vol. 10, no. 1, p. 4747, 2020, doi: 10.1038/s41598-020-61686-9.

 ISSN:2302-8149
e-ISSN:2540-9719

[31] E. A. Costa, M. E. Silva, and A. B. Galvão, "Real-time nowcasting the monthly unemployment rates with daily Google Trends data," *Socioecon. Plann. Sci.*, vol. 95, no. May, p. 101963, 2024, doi: 10.1016/j.seps.2024.101963.

[32] W. K. Adu, P. Appiahene, and S. Afrifa, "VAR, ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google trends," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, pp. 1–16, 2023, doi: 10.1186/s43067-023-00078-1.

[33] G. Bruno, T. Di Fonzo, R. Golinelli, and G. Parigi, "Short-run GDP forecasting in G7 countries : temporal disaggregation techniques and bridge models," *Eur. Comm. (Eurostat, EuroIndicators) Work. Pap. Stud.*, 2005.

[34] P. S. Luthfiana and N. Nasrudin, "Disaggregation and Forecasting of the Monthly Indonesian Gross Domestic Product (GDP)," *Bul. Ekon. Monet. dan Perbank.*, vol. 20, no. 4, pp. 497–526, 2018, doi: 10.21098/bemp.v20i4.905.

[35] L. Mosley, I. A. Eckley, and A. Gibberd, "Sparse temporal disaggregation," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 185, no. 4, pp. 2203–2233, 2022, doi: 10.1111/rssa.12952.

[36] C. Sax and P. Steiner, "Temporal Disaggregation of Time Series," *R J.*, vol. 5, no. 2, pp. 80–87, 2013, doi: 10.32614/RJ-2013-028.

[37] M. S. Akmal, "An Empirical Analysis of the Relationship Between GDP and Unemployment," *Humanomics*, vol. 19, no. 3, pp. 1–6, 2003, doi: 10.1108/eb018884.

[38] L. Levine, "Economic growth and the unemployment rate," *Labor Mark. Recover. After Recess. 2007-2009 Sel. Anal.*, pp. 55–64, 2014.

[39] C.-W. Lee and A. D. Huruta, "Okun's law in an emerging country: An empirical analysis in Indonesia," *Int. Entrep. Rev.*, vol. 5, no. 4, pp. 141–161, 2019, doi: 10.15678/IER.2019.0504.09.

[40] D. W. Suryono, A. Burda, and R. Chandra, "Does Indonesia's Economic Growth Reduce Unemployment?," vol. 132, no. AICMaR 2019, pp. 169–171, 2020, doi: 10.2991/aebmr.k.200331.037.

[41] R. Nooraeni, N. S. Purba, and N. P. Yudho, "Using Google Trend Data as an Initial Signal Indonesia Unemployment Rate," *Contrib. Pap. Sesssion*, vol. 3, no. 1, pp. 265–273, 2020.

[42] L. Widyarsi and H. Usman, "Penggunaan Data Google Trends untuk Peramalan Tingkat Pengangguran Terbuka di Tingkat Nasional dan Regional di Provinsi Jawa Barat," *Semin. Nas. Off. Stat.*, vol. 2021, no. 1, pp. 980–990, 2021, doi: 10.34123/semnasoffstat.v2021i1.842.

[43] C. A. Medel, "A Comparison Between Direct and Indirect Seasonal Adjustment of the Chilean GDP 1986–2009 with X-12-ARIMA," *J. Bus. Cycle Res.*, vol. 14, no. 1, pp. 47–87, 2018, doi: 10.1007/s41549-018-0023-3.

[44] V. Krisdiantoro, "Pengaruh Tenaga Kerja, Investasi Dan Kurs Tukar Terhadap Pertumbuhan Ekonomi Di Indonesia," Mar. 2020.

[45] R. I. Arumnisaa and A. W. Wijayanto, "Comparison of Ensemble Learning Method: Random Forest, Support Vector Machine, AdaBoost for Classification Human Development Index (HDI)," *Sist. J. Sist. Inf.*, vol. 12, no. 1, p. 206, 2023, doi: 10.32520/stmsi.v12i1.2501.

[46] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016, doi: 10.1007/s11749-016-0481-7.

[47] G. Biau and E. Scornet, "A Random Forest Guided Tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.

[48] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of ARIMA and Random Forest Time Series Models for Prediction of Avian Influenza H5N1 Outbreaks," *BMC Bioinformatics*, vol. 15, no. 1, 2014, doi: 10.1186/1471-2105-15-276.

[49] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[50] P. Sedgwick, "Pearson's correlation coefficient," *Bmj*, vol. 345, 2012.

[51] S. Sugiyono, "Metode penelitian kuantitatif dan kualitatif dan R&D," *Alf. Bandung*, pp. 170–182, 2010.

[52] J. Neter, *Applied Linear Statistical Models*. in Irwin series in statistics. Irwin, 1996. [Online]. Available: https://books.google.co.id/books?id=q2sPAQAAMAAJ

[53] R. Kurniawan and B. Yuniarto, *Analisis Regresi: Dasar dan Penerapannya dengan R*, 1st ed. Jakarta: Kencana, 2016.

[54] J. Nahar and S. Purwani, "Application of Robust M-Estimator Regression in Handling Data Outliers," *4th ICRIEMS Proc.*, pp. 53–60, 2017.

[55] T. W. Gress, J. Denvir, and J. I. Shapiro, "Effect of Removing Outliers on Statistical Inference: Implications to Interpretation of Experimental Data in Medical Research," *Marshall J. Med.*, vol. 4, no. 2, 2018, doi: 10.18590/mjm.2018.vol4.iss2.9.

[56] A. Fitrianto and S. H. Xin, "Comparisons Between Robust Regression Approaches in the Presence of Outliers and High Leverage Points," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 1, pp. 243–252, Mar. 2022, doi: 10.30598/barekengvol16iss1pp241-250.

[57] Y. W. Wen, Y. W. Tsai, D. B. C. Wu, and P. F. Chen, "The Impact of Outliers on Net-Benefit Regression Model in Cost-Effectiveness Analysis," *PLoS One*, vol. 8, no. 6, pp. 1–9, 2013, doi: 10.1371/journal.pone.0065930.

[58] D. Q. F. de Menezes, D. M. Prata, A. R. Secchi, and J. C. Pinto, "A review on robust M-estimators for regression analysis," *Comput. Chem. Eng.*, vol. 147, 2021.

[59] N. Donthu and A. Gustafsson, "Effects of COVID-19 on business and research," *J. Bus. Res.*, vol. 117, pp. 284–289, 2020, doi: https://doi.org/10.1016/j.jbusres.2020.06.008.

[60] D. N. Gujarati and D. C. Porter, *Basic Econometric*, 5th ed. New York: New York: McGraw Hill, 2015.

[61] A. P. Utomo, N. P. Sumartini, A. P. G. Siregar, N. Nagari, and S. Triyaningsih, "Regresi Robust untuk Memodelkan Pendapatan Usaha Industri Makanan Non-Makloon Berskala Mikro dan Kecil di Jawa Barat Tahun 2013," *J. Mat. Sains, dan Teknol.*, vol. 15, no. 2, pp. 63–74, 2014.