

Forecasting Indonesia's Unemployment Rate with Official Statistics and Big Data: A U-MIDAS Approach

Raisa Meidy Mustapa¹, Ribut Nurul Tri Wahyuni^{2*}

¹BPS-Statistics Cilegon Municipality, Indonesia

²Politeknik Statistika STIS, Indonesia

E-mail: ¹raisameidy@bps.go.id, ²rnurult@stis.ac.id

^{*}Corresponding Author

JEL Classification:

C55

E24

J64

Received: 10 October 2024

Revised: 25 May 2025

Accepted: 02 June 2025

Available online: September 2025

Published regularly: September 2025

Abstract

Research Originality: The current model is unable to forecast the unemployment rate utilizing varying periods of predictor variables. Furthermore, the use of official statistics and big data in previous studies to forecast Indonesia's unemployment rate has been limited.

Research Objectives: This study forecasts Indonesia's biannual unemployment rate (UR) by utilizing monthly Google Trends Index (GTI) data, quarterly Gross Domestic Product (GDP) data, and monthly inflation data.

Research Methods: The unrestricted mixed data sampling (U-MIDAS) model is applied to forecast Indonesia's UR using data from the second semester of 2006 to the first semester of 2024.

Empirical Results: This study finds that the best model for predicting UR is one that utilizes a combination of big data and official statistics. Using 34 GTI keywords relevant to job seekers' cultural and behavioral patterns in Indonesia, Indonesia's UR in February 2024 was 4.7%.

Implications: This study demonstrates that employing GTI and macroeconomic variables for forecasting unemployment enhances predictive accuracy compared to utilizing either variable independently.

Keywords:

U-MIDAS; unemployment rate; google trends index; big data

How to Cite:

Mustapa, R. M., & Wahyuni, R. N. T. (2025). Forecasting Indonesia's Unemployment Rate with Official Statistics & Big Data: A U-MIDAS Approach. *Etikonomi*, 24(2), 587 – 604. <https://doi.org/10.15408/etk.v24i2.41760>.

INTRODUCTION

The labor market has undergone significant changes since the introduction of the internet, transitioning from a print-centric economy to a more global digital landscape. The concurrent rise in labor demand within the digital domain is suggested by the transition of the labor supply system to digital platforms. LinkedIn and Indeed have emerged as prominent digital marketplaces that facilitate connections between employers and the majority of job seekers (Dillahunt et al., 2021). Interaction on these digital platforms produces a big data source, characterized by rapid accessibility, high granularity, and low cost (Ghasemaghaei, 2021).

In contrast to conventional data sources, such as censuses, surveys, and administrative records, big data can reveal user behavior and interests that fluctuate over time and vary across geographical regions. The Google Trends Index (GTI), a significant data source, can be used to assess user interest (Afkham et al., 2021; D'Amuri & Marcucci, 2017). The value of the GTI will increase as a result of an increase in the number of job searches conducted by users, which will indicate a comparable increase in its prominence on Google. GTI has significantly altered the prognosis of various socio-economic indicators, including unemployment rates. Numerous studies have demonstrated its unique ability to predict unemployment. The results demonstrate that GTI's forecasting capabilities surpass those of conventional benchmark models. GTI has significantly improved the accuracy of forecasting by serving as a proxy for labor supply, quickly and accurately illustrating unemployment (Costa et al., 2024; Fondeur & Karamé, 2013; Mulero & Garcia-Hiernaux, 2022; Simionescu & Cifuentes-Faura, 2022).

Previous research in Indonesia has focused more on using a few GTI keywords to forecast unemployment. For example, Fajar et al. (2020) precisely forecasted the monthly unemployment rate utilizing the "layoffs" search trend on GTI. Research by Naccarato et al. (2018) also demonstrates that utilizing Google Trends has enhanced the accuracy of forecasting youth unemployment. Few studies still use more keywords to represent the broad patterns of the labor market. Furthermore, Widyarsi & Usman (2021) employed gross domestic product (GDP), inflation, and GTI. In addition to big data, official statistics, including GDP and inflation, can be used to forecast the unemployment rate (Davidescu et al., 2021). Okun's law asserts that economic contraction is indicative of high unemployment (Ball et al., 2017), as evidenced by the escalating unemployment rate during the COVID-19 pandemic, which was exacerbated by the lethargic global economy.

Additionally, the Phillips curve illustrates inflation as a macroeconomic indicator by analyzing trends in unemployment. This curve describes a potential trade-off between inflation and unemployment, where lower unemployment rates tend to be associated with higher inflation, and vice versa (Crump et al., 2024). Because official statistics are reported at different times (GDP growth every three months and inflation and GTI every month), data interpolation was used to combine the three variables (Widyarsi & Usman, 2021). While interpolation may provide a potential remedy, this approach neglects essential data patterns during brief economic intervals.

This study aims to forecast Indonesia's unemployment rate by combining official statistics and GTI while employing the Unrestricted-Mixed Data Sampling (U-MIDAS) model, which effectively handles predictors with differing data frequencies and eliminates the need for interpolation. High unemployment rates can have an impact on numerous aspects. Unemployed individuals relate to mental health issues and social exclusion (Pohlan, 2019). In addition, unemployment has a negative impact on food security (Haini et al., 2023), increases poverty (Ngubane et al., 2023), and increases criminality (Jawadi et al., 2021). For mitigating the negative impacts of increasing unemployment, the government needs to undertake unemployment forecasts in order to implement appropriate policies.

The MIDAS model has been utilized in prior studies to contrast it with alternative models, including the autoregressive (AR) model and the autoregressive integrated moving average with exogenous variables (ARIMAX) model (Smith, 2016; Tungtrakul et al., 2016). This model, however, has constraints in addressing variations in quantity and complexity due to its reliance on a nonlinear estimation approach (Tarsidin et al., 2018). This study introduces an innovative strategy that utilizes a more flexible MIDAS regression technique, known as the U-MIDAS model, to address these issues. The U-MIDAS model demonstrates greater flexibility than the MIDAS model in managing varying quantities of variables and complex dimensions, and it provides estimates more easily due to its use of Ordinary Least Squares (OLS) (Ghysels et al., 2016).

This study will present three new contributions to the field of unemployment forecasting. First, BPS-Statistics Indonesia publishes data on Indonesia's unemployment rate biannually. The unemployment rate, derived from surveys, is published three months after survey completion. The unemployment rate for the first semester is typically released in May, while the rate for the second semester is typically released in November. The prolonged publication delay is attributed not only to the duration of the survey implementation series and constrained expenses but also to significant changes required prior to data dissemination (Borup & Schütte, 2022). The utilization of GTI may address this issue, as it provides more current unemployment rate forecasts, given that GTI data is accessible monthly. Second, we utilize the U-MIDAS model to predict low-frequency data (unemployment data) released every six months by BPS-Statistics Indonesia, using high-frequency data (quarterly GDP, monthly inflation, and monthly GTI) without requiring data aggregation. This model preserves important high-frequency information that could be lost in standard aggregation methods. Third, this study uses more keywords (34) to build GTI compared to previous studies, such as Adu et al. (2023), Fajar et al. (2020), Naccarato et al. (2018), Simionescu and Cifuentes-Faura (2022), Smith (2016), Tarsidin et al. (2018), Tungtrakul et al. (2016), and Widyarsi and Usman (2021). Job seekers in Indonesia frequently employ these keywords on digital platforms such as Jobstreet, LinkedIn, and Fiverr.

METHODS

This study employs official statistical data, such as the biannual unemployment rate ($UR_{t,(SA)}$), quarterly GDP at constant prices ($GDP_{t,(Q)}$), and monthly consumer inflation ($inflation_{t,(M)}$) from BPS-Statistics Indonesia, as well as monthly GTI data from Google. In this study, $GTI_{t,(M)}$ denotes the principal component variable of the monthly GTI. Data from the second semester of 2006 to the first semester of 2024 is used because GTI data for Indonesia has been accessible since mid-2006. This period includes several significant macroeconomic phases—namely the 2008 global financial crisis, the COVID-19 pandemic, and the post-pandemic recovery—thus offering a thorough temporal framework to analyze fluctuations in labor market behavior and assess the predictive efficacy of job-search intensity on unemployment dynamics.

This paper's primary innovation is incorporating keywords that reflect the culture and behavior of job seekers in Indonesia, alongside the application of the U-MIDAS model to mitigate the constraints of utilizing variables with varying frequencies, thus improving the model's predictive efficacy relative to prior research. The efficacy of utilizing GTI data necessitates the careful selection of suitable keywords. The selection of GTI keywords will prioritize terms most commonly searched by job seekers due to their role as a labor supply (Vicente et al., 2015). We present an innovative approach, unprecedented in Indonesia, utilizing 34 keywords corresponding to job seekers' cultural and behavioral patterns: Pendaftaran CPNS, Lowongan BUMN, Interview Kerja, Fiverr, LinkedIn, Jobstreet, Curriculum Vitae, Resume, Cover Letter, Motivational Letter, CV, Pekerjaan Baru, Pekerjaan Sementara, Interview User, Kesempatan Kerja, Wawancara Kerja, Karir, Pekerjaan Paruh Waktu, Pekerjaan Gaji Tinggi, Outsourcing, Bursa Kerja, Rekrutmen, Lamaran Pekerjaan, Cara Mencari Kerja, Cari Uang, Cari Kerja, Pengusaha, Wirausaha, Wiraswasta, Pekerjaan Online, Bisnis, Lowongan Kerja, Peluang Usaha, dan Lowongan Pekerjaan. The determination of the number of principal components is based on the requirement that the cumulative variance explained by the principal components must reach at least 75 percent.

The U-MIDAS model is a regression technique that addresses frequency discrepancies between response and predictor variables without requiring aggregating predictor variables. In this study, the response variable is a low-frequency variable (UR), and high-frequency predictor variables (GDP , $inflation$, GTI) and their lags elucidate the response variable. The U-MIDAS model employs a straightforward polynomial lag computed by OLS.

Prior to modeling, the data series in time series analysis must demonstrate stationarity. A stationarity evaluation on all variables utilizes the Augmented Dickey-Fuller (ADF) test. It includes both trend and intercept components in Equation 1. We can infer that the variable (Y) is stationary when δ is less than 0.

$$\Delta Y_t = c_0 + \beta_0 T + \delta Y_{t-1} + \sum_{k=1}^h \beta_k \Delta Y_{t-k} + \varepsilon_t \quad (1)$$

ΔY_t is first difference ($Y_t - Y_{t-1}$), T indicates the trend, and k is the number of lag terms. Subsequent to performing the stationarity test, the following phase constructs three U-MIDAS models. This study employs Eviews data processing software to construct the

three proposed research models. Equation 2 delineates a formula that employs official statistical data (official statistics model). Equation 3 delineates a formula that incorporates GTI data (big data model). Finally, Equation 4 employs a combination of official statistical data and GTI data (combination model). The developed models are as follows:

$$UR_{t,(SA)} = a_1 + \sum_{p=1}^s \gamma_p GDP_{t-(p-1),(Q)} + \sum_{q=1}^v \omega_{inflasi_q} inflation_{t-(q-1),(M)} + \varepsilon_{1t} \quad (2)$$

$$UR_{t,(SA)} = a_2 + \sum_{q=1}^v \omega_{1q} GTI_{1t-(q-1),(M)} + \sum_{q=1}^v \omega_{2q} GTI_{2t-(q-1),(M)} + \sum_{q=1}^v \omega_{3q} GTI_{3t-(q-1),(M)} + \dots + \sum_{q=1}^v \omega_{zq} GTI_{zt-(q-1),(M)} + \varepsilon_{2t} \quad (3)$$

$$UR_{t,(SA)} = a_3 + \sum_{p=1}^s \gamma_p GDP_{t-(p-1),(Q)} + \sum_{q=1}^v \omega_{inflasi_q} inflation_{t-(q-1),(M)} + \sum_{q=1}^v \omega_{1q} GTI_{1t-(q-1),(M)} + \sum_{q=1}^v \omega_{2q} GTI_{2t-(q-1),(M)} + \sum_{q=1}^v \omega_{3q} GTI_{3t-(q-1),(M)} + \dots + \sum_{q=1}^v \omega_{zq} GTI_{zt-(q-1),(M)} + \varepsilon_{3t} \quad (4)$$

where z is the number of the principal component variables of the monthly GTI.

The optimal model selection is conducted in phases to guarantee that the selected model is the most parsimonious while exhibiting superior unemployment prediction capabilities. The initial phase is choosing the model with the most optimal lag for Equation 2. The methodology entails parameter estimation with a minimum lag value of one and a maximum lag value of eleven. The maximum lag for each method will vary among model situations, determined by the predictor variables and the number of observations linked to each variable. The model with the best lag is determined by assessing the highest Adjusted R-square in conjunction with the lowest values of the Akaike Information Criterion (AIC), Schwarz Criterion (SC), and Hannan-Quinn (HQ) Criterion. The model with the selected optimal lag must satisfy the conditions of normality, homoscedasticity, and non-autocorrelation, as it is estimated using OLS. A model with the optimal lag for Equation 2 will be derived from the initial stage. The identical procedure is employed to construct a model with the optimal lag for Equations 3 and 4.

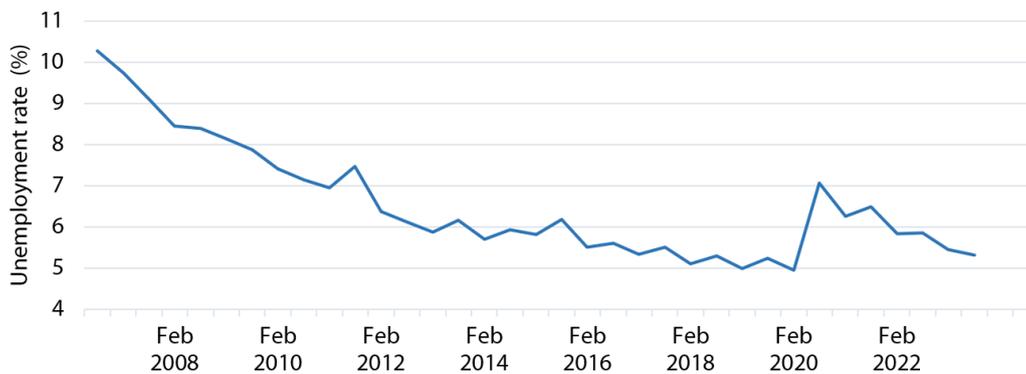
RESULTS AND DISCUSSION

This section elucidates the correlation between unemployment conditions and the economy in Indonesia over time, as well as how GTI data exemplifies unemployment patterns. Figure 1 demonstrates a decline in the unemployment trend in Indonesia from 2006 to 2023. The consistent decline in unemployment occurred from 2006 to 2012. This decline suggests the ongoing stability of the labor market, despite global economic uncertainty stemming from the 2008 financial crisis and rising commodity prices. Variations in the unemployment rate occurred between the first semester of 2012 and the second semester of 2020, due to the prolonged business cycle and the ongoing surge in commodity prices. Consequently, the decline in the unemployment rate was less significant than in previous years. The increase in the unemployment rate during the COVID-19 pandemic was most pronounced, as shown in Figure 1, reflecting the

economic downturn in Indonesia resulting from measures such as reduced working hours, extensive social restrictions, and layoffs. The unemployment rate declined to 1.84 percent in the second semester of 2022 and remained at that level until the end of 2023.

Figure 2 depicts an upward trend in GDP from 2006 to 2023. The persistent upward trend of GDP signifies annual economic growth. Nonetheless, economic contraction also transpired. The COVID-19 pandemic, which impacted economic conditions, led to a deceleration of Indonesia's economy after a period of sustained expansion. Following the most severe contraction in 2020, the economy exhibited improvement, as demonstrated by the rise in GDP from 2021 to 2023.

Figure 1. Indonesia's Unemployment Rate in the First Semester of 2006 – the Second Semester of 2023

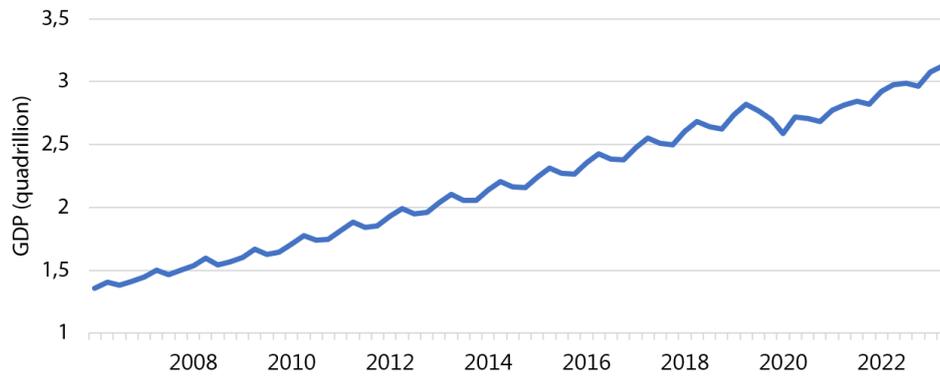


Source: BPS-Statistics Indonesia

The month-to-month inflation trend in Indonesia from 2006 to 2023 remains stable at 0.30 percent. Figure 3, however, indicates that elevated inflation transpired in June 2008. The rising inflation that year was a result of the global financial crisis. Kang et al. (2020) elucidate that escalating commodity and energy prices triggered the inflation surge in 2008, resulting from the global financial crisis. The peak inflation rate during the study period was recorded in September 2013, at 3.29 percent. The primary factor driving this inflation was the escalation of food prices, which was additionally influenced by the surge in fuel prices. Conversely, Indonesia encountered deflation of 0.09 percent in February 2015, primarily attributed to a decline in the food expenditure category, which typically sees an increase.

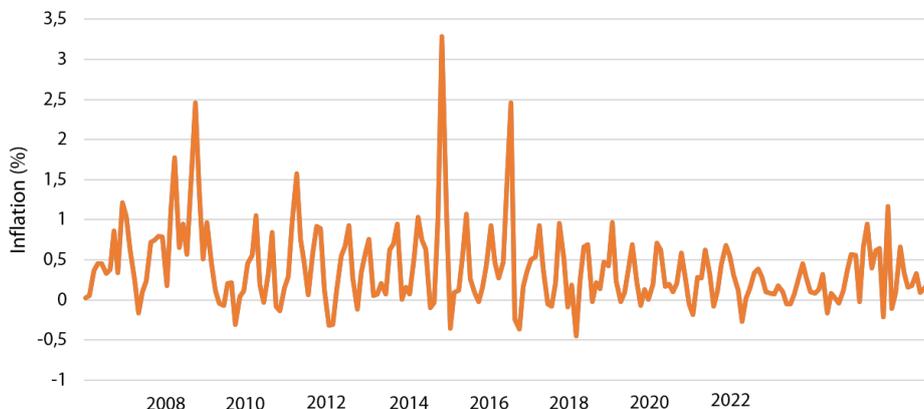
The primary issue in forecasting with GTI data is the appropriate selection of keywords. This study presents originality by employing 34 distinct keywords to reflect the dynamic patterns of labor behavior. Figure 4 illustrates a significant correlation among various keywords. LinkedIn exhibits a strong correlation with job vacancies, as it is one of the most widely used platforms for job searching. The principal component analysis reduced the 34 GTI datasets to satisfy the assumption of independence. With a criterion of a minimum cumulative variance of 75% (Mathai et al., 2022), we identified three principal components of GTI.

Figure 2. Indonesia's GDP in the Third Quarter of 2006 – the Fourth Quarter of 2023



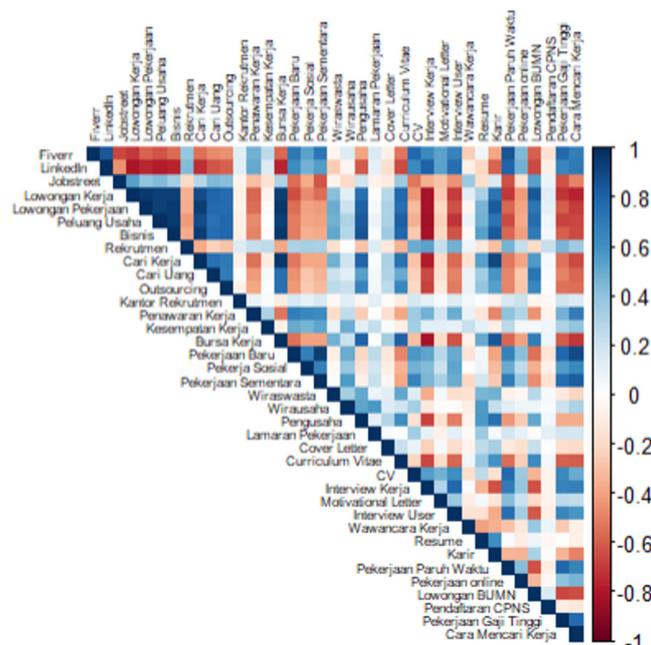
Source: BPS-Statistics Indonesia

Figure 3. Monthly Inflation in Indonesia from 2006-2023



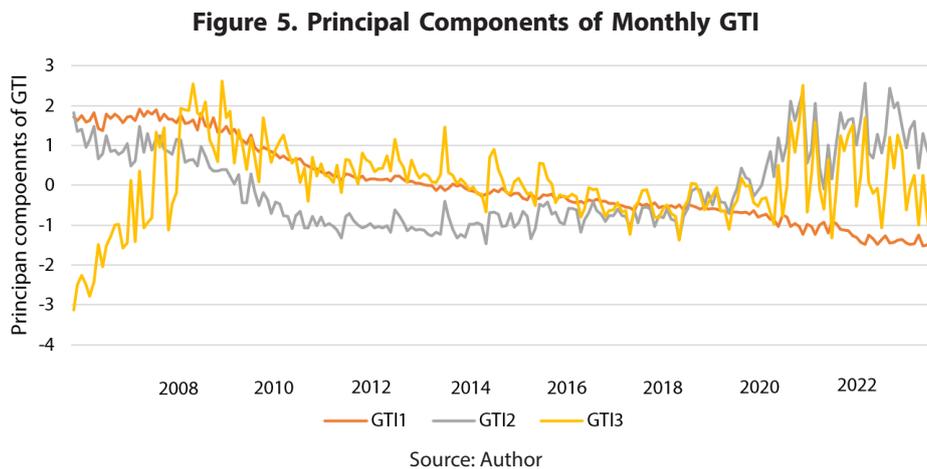
Source: BPS-Statistics Indonesia

Figure 4. Correlation Matrix between GTI Keywords



Source: Author

The patterns of the three principal components of GTI illustrated in Figure 5 demonstrate notable variations. It indicates a transformation in worker behavior about job searches on online platforms. The varying patterns evident in the three main GTI component graphs indicate that the keywords included in the GTI predictor have consistently featured prominent concepts over time, thereby preserving their usefulness for modeling purposes.



The next step involves preliminarily identifying relationships among the variables to ensure the accuracy of variable selection. We implemented adjustments for all variables because the initial detection method, which employs Pearson correlation, fails to consider the frequency differences among the variables. The adjustment entailed utilizing partial data for the correlation calculation: utilization of GDP data from the first and third quarters; utilization of inflation data from February and August; utilization of the average data of the principal GTI components over six months in the first and second semesters; and utilization of unemployment rate data from the first and second semesters. The use of GDP and inflation data during that period is attributable to the collection of unemployment rate data from the same timeframe. The application of a six-month average for GTI data is supported by a study examining the duration of job searches in Indonesia over six months (Zatzaha et al., 2020).

Table 1. Pearson Correlation between Variables

Variable	Correlation	Statistics	<i>p-value</i>
GDP	-0.904***	-7.426	0.000
inflation	0.417*	2.639	0.013
GTI1	0.845***	9.079	0.000
GTI2	0.396**	2.478	0.018
GTI3	-0.136	-0.788	0.436

Note: (***) significant at 1%, **) significant at 5%, *) significant at 10%
Source: Author

Table 1 shows the Pearson correlation between variables. A higher value of Pearson correlation indicates a stronger relationship between two variables (Johnson & Bhattacharyya, 2019). Table 1 reveals a negative correlation between the GDP variable, which acts as an indicator of economic growth, and the unemployment rate, highlighting a significant relationship between these two factors. The results indicate the existence of Okun's law in Indonesia, supported by recent literature. This finding indicates a correlation between economic growth and the expansion of job opportunities in the formal sector. As a result, the decrease in the number of informal workers who are vulnerable to unemployment is likely to lead to a reduction in the unemployment rate.

In Indonesia, there is a significant correlation between inflation and the unemployment rate, which diverges from the expectations set by the Phillips curve theory. This situation can occur when a country demonstrates relatively low inflation due to policies designed to manage prices, which typically support job creation more effectively. The inflation rate suggests that maintaining price stability may lead to increased job opportunities. The evidence indicates that maintaining low and stable inflation is crucial for creating more job opportunities.

However, the three principal components of the GTI demonstrate differing correlations. This outcome is defensible, as the differences in terminology used can elucidate the relationships among the variables in accordance with the existing literature (Widyarsi & Usman, 2021). The modeling process can proceed efficiently even when predictor variables show a weak correlation with the unemployment rate. The exclusion of unnecessary variables is not obligatory, as these variables could signal economic shifts.

U-MIDAS modeling begins by performing a stationarity test on all variables by applying the ADF test. After analyzing the pre- and post-pandemic periods, it is clear that the GDP variance exhibits fluctuations over time, which requires the application of a natural logarithm transformation.

Table 2. Stationarity Test

Variable	At Level	First Difference	Stationarity Level
UR	-2.666	-8.919***	First difference
Ln GDP	-1.626	-3.498**	First difference
inflation	-10.824***	-	Level
GTI1	-1.224	-19.512***	First difference
GTI2	-1.625	-3.572**	First difference
GTI3	-3.558**	-	Level

Note: (***) significant at 1%, (**) significant at 5%, (*) significant at 10%
Source: Author

Table 2 indicates that the stationary variables exist at varying levels. The subsequent step is to determine the best lag for each model. According to the testing criteria, the ideal lags for the official statistics, big data, and combination model are three, five, and five, respectively. The subsequent stage involves identifying the optimal prediction model. Table 3 presents a detailed comparison of the three models and their corresponding selection criteria.

Table 3. Selection Criteria for The Best Prediction Model

Model	Adjusted R-Square	AIC	SC	HQ
Official statistics	0.761	0.381	0.695	0.488
Big data	0.284	1.604	2.322	1.848
Combination	0.873	-0.390	0.790	0.007

Source: Author

The adjusted R-squared value for the official statistics model (Equation 2) stands at 0.761. The analysis indicates that the predictor variables in the combined model account for 76.13 percent of the variation in unemployment rate changes. In contrast, the remaining variation is attributed to factors not included in the model. In the big data model (Equation 3), the adjusted R-squared value is 0.284, indicating that the predictor variables included in the model can account for 28.40% of the variation in unemployment rate changes. In contrast, the remaining variation is attributed to other factors not encompassed by the model. The adjusted R-squared value for the combined model (Equation 4) is 0.873, indicating that 87.30% of the variation in unemployment rate changes can be accounted for by the predictor variables in this model. The subsequent step involves evaluating the assumptions of the three best models.

Table 4. Normality Assumption Test

Model	Statistical Value	Critical Value	p-value	Decision
Official statistics	0,549	5,991	0,760	Failed to Reject H_0
Big data	0,760	5,991	0,684	Failed to Reject H_0
Combination	0,114	5,991	0,945	Failed to Reject H_0

Source: Author

Table 4 indicates that the three models satisfy the normality assumption. It also demonstrates that statistical values for the official statistics model, big data, and their combination are less than the crucial value, with a p-value exceeding 0.05. Consequently, it may be concluded that the model's error is normally distributed with a mean of zero and constant variance.

Table 5 indicates that the non-autocorrelation assumption for the three models is satisfied. Table 5 illustrates that the computed statistic values for each latency are inferior to their critical thresholds and exhibit p-values over 0.05. Upon examining the computed statistics, all models exhibit values below the critical threshold and p-values exceeding 0.05. Consequently, it can be inferred that the errors are independent or not serially associated.

Table 5. Non-autocorrelation Assumption Test

Model	Statistical Value	Critical Value	p-value	Decision
Official statistics	5,944	18,307	0,989	Failed to Reject H_0
Big data	10,767	18,307	0,824	Failed to Reject H_0
Combination	17,429	18,307	0,358	Failed to Reject H_0

Source: Author

Table 6 indicates that the three models satisfy the homoscedasticity assumption. Upon examining the computed statistics, all models exhibit values below the crucial threshold and p-values exceeding 0.05. Consequently, it can be concluded that the error variance remains constant, thereby satisfying the assumption of homoscedasticity.

Table 6. Homoscedasticity Assumption Test

Model	Statistical Value	Critical Value	p-value	Decision
Official statistics	14,024	26,296	0,597	Failed to Reject H_0
Big data	8,628	26,296	0,928	Failed to Reject H_0
Combination	14,330	26,296	0,574	Failed to Reject H_0

Source: Author

Table 3 indicates that the optimal method is the combined model, as evidenced by its superior Adjusted R-squared value and the lowest AIC, SC, and HQ metrics. In the current analysis, the official statistics model is more effective than the big data model. The constraint of GTI, functioning solely as a proxy for labor supply, remains inadequate for generating a reasonably accurate forecast. The findings align with the study by Ormerod et al. (2014), indicating that incorporating GTI did not significantly enhance the model's accuracy. These findings could be attributed to a constraint of GTI, which remains affected by the social conditions prevalent in the region and the temporal context of the study. When extensive searches are conducted during a period of heightened interest in a particular topic, the resulting forecasts may demonstrate low accuracy, as evidenced by the findings of Google Flu Trends (Ormerod et al., 2014).

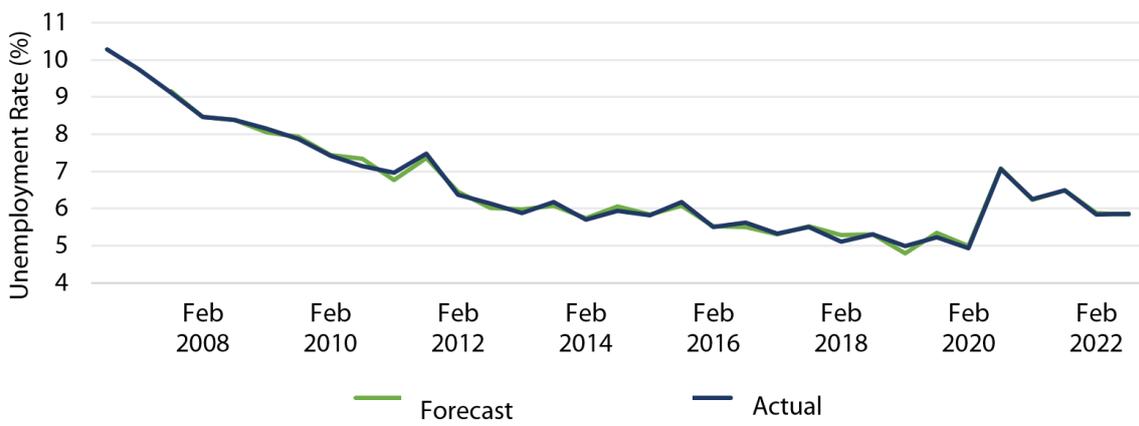
Table 7. Best Prediction Model

Variable	Coefficient	Std. Error	T-statistics	Prob.
Variable: D(ln GDP) Lags: 5				
Lag 1	-7.695	12.877	-0.598	0.569
Lag 2	-22.362***	4.397	-5.085	0.001
Lag 3	-33.464**	10.524	-3.180	0.016
Lag 4	-16.469**	5.676	-2.902	0.023
Lag 5	7.782	9.342	0.833	0.432
Variable: Inflation Lags: 5				
Lag 1	0.536	0.301	1.778	0.119
Lag 2	-0.070	0.142	-0.495	0.636
Lag 3	0.505	0.311	1.624	0.148
Lag 4	-0.301	0.438	-0.688	0.513
Lag 5	0.272	0.248	1.094	0.310
Variable: D(GTI1) Lags: 5				
Lag 1	-4.368*	2.249	-1.942	0.093
Lag 2	-4.435**	1.375	-3.226	0.014
Lag 3	-5.354**	1.597	-3.353	0.012
Lag 4	-3.824**	1.449	-2.640	0.033
Lag 5	-0.358	1.279	-0.280	0.788
Variable: D(GTI2) Lags: 5				
Lag 1	-0.768	0.407	-1.888	0.101
Lag 2	-0.691	0.574	-1.204	0.268
Lag 3	-1.365**	0.441	-3.095	0.017
Lag 4	-1.200**	0.465	-2.583	0.036
Lag 5	-0.412	0.327	-1.261	0.248
Variable: GTI3 Lags: 5				
Lag 1	0.343	0.229	1.496	0.178
Lag 2	-0.457	0.269	-1.696	0.134
Lag 3	0.084	0.227	0.369	0.723
Lag 4	-0.446*	0.216	-2.061	0.078
Lag 5	-0.462	0.248	-1.857	0.106

Note: (***) significant at 1%, (**) significant at 5%, (*) significant at 10%
 Source: Author

The integration of GTI variables with macroeconomic indicators is considered to yield more accurate forecasts. The findings align with the research conducted by Smith (2016), which demonstrated that the MIDAS model, utilizing survey data predictor variables alongside GTI, was more effective in forecasting the unemployment rate in the UK. The involvement of the GTI, characterized by its real-time and high-frequency nature, effectively captures patterns of unemployment fluctuations that survey indicators, which exhibit a lag, fail to detect. Table 7 presents the outcomes of the optimal prediction model.

Figure 6. Line Graphs of Forecasted and Actual Unemployment Rate for Best Model



Source: Author

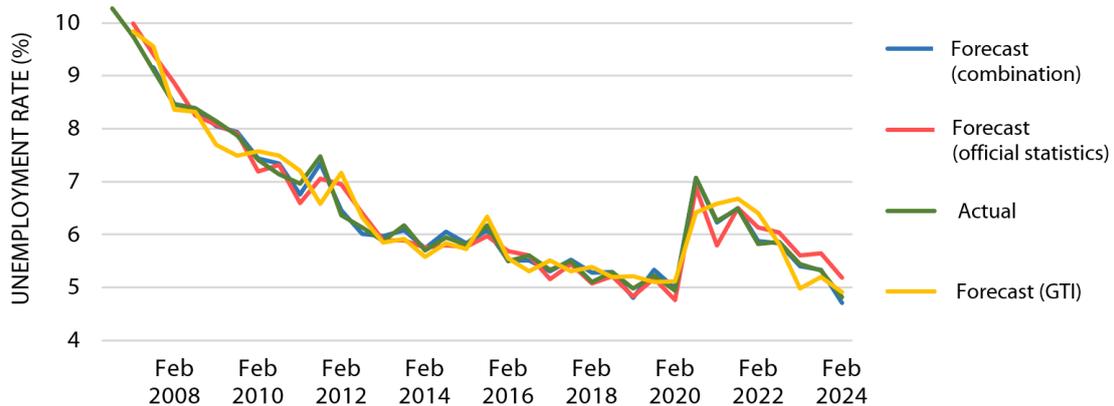
The forecast results for the unemployment rate in February 2024 are illustrated in Figure 6, utilizing the best model. The image analysis indicates that the projected unemployment rate for Indonesia in February 2024, utilizing the combined model, is 4.7 percent. At this point, the precise value stands at 4.82 percent. Both values have shown a decline compared to the previous period (August 2023) and the same month in the prior year (February 2023), indicating a positive trend in Indonesia's unemployment situation. The value has approached the Sustainable Development Goals (SDGs) target for the unemployment rate of 4.6 percent for 2024. The forecasting results demonstrate a high level of accuracy, evidenced by the negligible discrepancies between actual data and forecast outcomes over the observed period. The best model successfully identified the rise in the unemployment rate in 2020.

Figure 6 indicates that unemployment forecasting using a mix of macroeconomic indicators and GTI is most suitable for implementation in Indonesia. This outcome aligns with the findings of Gayaker & Türe (2025), which demonstrated that integrating macroeconomic indicator variables with GTI enhances forecasting accuracy relative to utilizing a single indicator alone. The amalgamation of the two variables integrates the adaptability of labor market behavior signals with the consistency of economic indicators, yielding more dependable forecasts. Incorporating varied

keywords has improved the model's prediction efficacy relative to conventional models alone.

Based on Figure 7, forecasting using macroeconomic indicators tends to underestimate the actual values during various intervals, particularly from August 2016 to February 2021. The extended duration of underestimating is likely due to the inconsistencies in economic swings that impacted the precision of the unemployment rate trend throughout that era. Economic shocks can diminish the precision of projections for unemployment derived from economic data. The GTI technique produced excessively optimistic expectations after COVID-19, particularly from February 2021 to February 2022, whereas it underestimated projections from February 2013 to February 2015. The overestimate during this period is consistent with research conducted by Dilmaghani (2018), which suggests that social incentives can influence the stability of predicting GTI in specific groups and time frames. For instance, atypical variations in GTI occurred during the COVID-19 era, characterized by a surge in layoff narratives, which led to the keyword "unemployment" gaining significant traction, even among individuals not actively seeking employment. Consequently, misleading popularity emerged due to insufficient control within the model.

Figure 7. Line Graphs of Forecasted and Actual Unemployment Rate for All Model



Source: Author

The forecasting results from the combination model closely align with the unemployment rate. The absence of control variables that accurately represent the economic conditions in the GTI data has been tackled by integrating economic indicators derived from surveys with the GTI data. The U-MIDAS model, incorporating macroeconomic indicator predictor variables and GTI, demonstrates superior performance in forecasting the unemployment rate for February 2024 compared to its benchmark model.

CONCLUSION

The results of this study demonstrate that the U-MIDAS model, which integrates official statistics and big data, more precisely forecasts Indonesia's unemployment rate than either using big data or official statistics alone. The optimal model predicted an unemployment rate of 4.7 percent in February 2024, which was 0.12 percentage points lower than the actual data released by BPS-Statistics Indonesia. The use of a broader spectrum of keywords and alignment with Indonesian culture has demonstrated enhanced model performance. Therefore, to enhance the accuracy of the prediction results, additional studies can incorporate keywords that job seekers in Indonesia frequently use to create GTI.

This study presents a strategy for the government to implement more rapid and precise policies regarding unemployment in Indonesia. BPS-Statistics Indonesia currently releases official unemployment statistics biannually, with a dissemination delay of three months following the survey. The U-MIDAS model enables policymakers to swiftly estimate Indonesia's unemployment rate, even on a monthly basis, by utilizing a combination of GTI data and official statistics. When the estimation value suggests a rise in the unemployment rate, the government may enact relevant policies to mitigate unemployment, such as instituting tax incentives to promote job creation, investing in education and vocational training to align worker skills with market demands, and initiating labor-intensive public works programs.

REFERENCES

- Adu, W. K., Appiahene, P., & Afrifa, S. (2023). VAR, ARIMAX and ARIMA Models For Nowcasting Unemployment Rate in Ghana Using Google Trends. *Journal of Electrical Systems and Information Technology*, 10(1), 1–16. <https://doi.org/10.1186/s43067-023-00078-1>.
- Afkham, M., Ghoddsi, H., & Rafizadeh, N. (2021). Google Search Explains Your Gasoline Consumption! *Energy Economics*, 99, 105305. <https://doi.org/10.1016/j.eneco.2021.105305>.
- Ball, L., Leigh, D., & Loungani, P. (2017). Okun's Law: Fit at 50? *Journal of Money, Credit and Banking*, 49(7), 1413–1441. <https://doi.org/10.1111/jmcb.12420>.
- Borup, D., & Schütte, E. C. M. (2022). In Search of a Job: Forecasting Employment Growth Using Google Trends. *Journal of Business & Economic Statistics*, 40(1), 186–200. <https://doi.org/https://doi.org/10.1080/07350015.2020.1791133>.
- Costa, E. A., Silva, M. E., & Galvão, A. B. (2024). Real-Time Nowcasting the Monthly Unemployment Rates with Daily Google Trends Data. *Socio-Economic Planning Sciences*, 95, 101963. <https://doi.org/10.1016/j.seps.2024.101963>.
- Crump, R. K., Eusepi, S., Giannoni, M., & Şahin, A. (2024). The Unemployment–Inflation Trade-Off Revisited: The Phillips Curve in COVID Times. *Journal of Monetary Economics*, 145, 103580. <https://doi.org/10.1016/j.jmoneco.2024.103580>.

- D'Amuri, F., & Marcucci, J. (2017). The Predictive Power of Google Searches in Forecasting US Unemployment. *International Journal of Forecasting*, 33(4), 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>.
- Davidescu, A. A., Apostu, S. A., & Marin, A. (2021). Forecasting The Romanian Unemployment Rate in Time of Health Crisis—A Univariate Vs. Multivariate Time Series Approach. *International Journal of Environmental Research and Public Health*, 18(21), 1–31. <https://doi.org/10.3390/ijerph182111165>.
- Dillahunt, T. R., Israni, A., & Lu, A. J. (2021). Examining The Use of Online Platforms for Employment: A Survey of U.S. Job Seekers. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411764.3445350>.
- Dilmaghani, M. (2018). The Racial Digital Divide in The Predictive Power of Google Trends Data for Forecasting the Unemployment Rate. *Journal of Economic and Social Measurement*, 43(3–4), 119–142. <https://doi.org/10.3233/JEM-180458>.
- Fajar, M., Rizky Prasetyo, O., Fajar, M., Prasetyo, O. R., & Nonalisa, S. (2020). Forecasting Unemployment Rate in the Time of COVID-19 Pandemic Using Google Trends Data (Case of Indonesia). *International Journal of Scientific Research in Multidisciplinary Studies*, 6(11), 29–33.
- Fondeur, Y., & Karamé, F. (2013). Can Google Data Help Predict French Youth Unemployment? *Economic Modelling*, 30(1), 117–125. <https://doi.org/10.1016/j.econmod.2012.07.017>.
- Gayaker, S., & Türe, H. (2025). National Unemployment Rate Forecast with Google Trends. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12(1), 107–123.
- Ghasemaghaei, M. (2021). Understanding The Impact of Big Data on Firm Performance: The Necessity of Conceptually Differentiating Among Big Data Characteristics. *International Journal of Information Management*, 57, 102055. <https://doi.org/10.1016/j.ijinfomgt.2019.102055>.
- Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed Frequency Data Sampling Regression Models: The R Package midasr. *Journal of Statistical Software*, 72(4), 1–35.
- Haini, H., Musa, S. F., Loon, P. W., & Basir, K. H. (2023). Does Unemployment Affect the Relationship Between Income Inequality and Food Security? *International Journal of Sociology and Social Policy*, 43(1-2), 48–66. <https://doi.org/10.1108/ijssp-12-2021-0303>.
- Jawadi, F., Mallick, S. K., Idi Cheffou, A., & Augustine, A. (2021). Does Higher Unemployment Lead to Greater Criminality? Revisiting The Debate Over the Business Cycle. *Journal of Economic Behavior and Organization*, 182, 448–471. <https://doi.org/10.1016/j.jebo.2019.03.025>.
- Johnson, R. A., & Bhattacharyya, G. K. (2019). *Statistics: Principles and Methods* (8th edition). New Jersey: Wiley.

- Kang, W., Ratti, R. A., & Vespignani, J. (2020). Global Commodity Prices and Global Stock Market Volatility Shocks: Effects Across Countries. *Journal of Asian Economics*, 71, 101249. <https://doi.org/10.1016/j.asieco.2020.101249>.
- Mathai, A., Provost, S., & Haubold, H. (2022). *Multivariate Statistical Analysis in the Real and Complex Domains*. Berlin: Springer International Publishing.
- Mulero, R., & Garcia-Hiernaux, A. (2022). Forecasting Unemployment with Google Trends: Age, Gender and Digital Divide. *Empirical Economics*, 65(2), 587–605. <https://doi.org/10.1007/s00181-022-02347-w>.
- Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining Official and Google Trends Data to Forecast the Italian Youth Unemployment Rate. *Technological Forecasting and Social Change*, 130, 114–122. <https://doi.org/10.1016/j.techfore.2017.11.022>.
- Ngubane, M. Z., Mndebele, S., & Kaseeram, I. (2023). Economic Growth, Unemployment and Poverty: Linear and Non-Linear Evidence from South Africa. *Heliyon*, 9(10), e20267. <https://doi.org/10.1016/j.heliyon.2023.e20267>.
- Ormerod, P., Nyman, R., & Bentley, R. A. (2014). Nowcasting Economic and Social Data: When and Why Search Engine Data Fails, An Illustration Using Google Flu Trends. *ArXiv Preprint*. <https://doi.org/https://doi.org/10.48550/arXiv.1408.0699>.
- Pohlan, L. (2019). Unemployment and Social Exclusion. *Journal of Economic Behavior and Organization*, 164, 273–299. <https://doi.org/10.1016/j.jebo.2019.06.006>.
- Simionescu, M., & Cifuentes-Faura, J. (2022). Can Unemployment Forecasts Based on Google Trends Help Government Design Better Policies? An Investigation Based on Spain and Portugal. *Journal of Policy Modeling*, 44(1), 1–21. <https://doi.org/10.1016/j.jpolmod.2021.09.011>.
- Smith, P. (2016). Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data. *Journal of Forecasting*, 35(3), 263–284. <https://doi.org/10.1002/for.2391>.
- Tarsidin, Idham, & Rakhman, R. N. (2018). Nowcasting Household Consumption and Investment in Indonesia. *Bulletin of Monetary Economics and Banking*, 20(3), 375–404. <https://doi.org/https://doi.org/10.21098/bemp.v20i3.858>.
- Tungtrakul, T., Kingnetr, N., & Sriboonchitta, S. (2016). An Empirical Confirmation of the Superior Performance of MIDAS over ARIMAX. In: Huynh, V. N., Inuiguchi, M., Le, B., Le, B., Denoeux, T. (Eds). *Integrated Uncertainty in Knowledge Modelling and Decision Making. IUKM 2016. Lecture Notes in Computer Science*, 9978. Cham: Springer.
- Vicente, M. R., López-Menéndez, A. J., & Pérez, R. (2015). Forecasting Unemployment with Internet Search Data: Does It Help to Improve Predictions When Job Destruction Is Skyrocketing? *Technological Forecasting and Social Change*, 92, 132–139. <https://doi.org/10.1016/j.techfore.2014.12.005>.

- Widyarsi, L., & Usman, H. (2021). The Use of Google Trends Data for Forecasting Open Unemployment Rates at National and Regional Levels in West Java Province. *Seminar Nasional Official Statistics, 2021(1)*, 980–990.
- Zatzaha, N. M., Suprihatinb, B., & Zunaidahc. (2020). The Influence of Socio Demographic Factors On Job Search Duration For Educated Workers in South Sumatera Province. *Jurnal Dialektika, 6*, 55–64.