

# Indonesian GDP movement detection using online news classification

Dinda Pusparahmi Sholawatunnisa<sup>a</sup>, Lya Hulliyatus Suadaa<sup>a,b,\*</sup>, Usep Nugraha<sup>b</sup> and Setia Pramana<sup>a,b</sup>

<sup>a</sup>*Politeknik Statistika STIS, Jakarta, Indonesia*

<sup>b</sup>*BPS Statistics Indonesia, Jakarta, Indonesia*

**Abstract.** Gross Domestic Product (GDP) stands as a pivotal indicator, offering strategic insights into economic dynamics. Recent technological advancements, particularly in real-time information dissemination through online economic news platforms, provide an accessible and alternative data source for analyzing GDP movements. This study employs online news classification to identify patterns in the movement and growth rate of Indonesia's GDP. Utilizing a web scraping technique, we collected data for analysis. The classification models employed include transfer learning from pre-trained language model transformers, with classical machine learning methods serving as baseline models. The results indicate superior performance by the pre-trained language model transformers, achieving the highest accuracy of 0.8880 and 0.7899. In comparison, hyperparameter-tuned classical machine learning models also demonstrated commendable results, with the best accuracy reaching 0.845 and 0.7811. This research underscores the efficacy of leveraging online news classification, particularly through advanced language models. The findings contribute to a nuanced understanding of economic dynamics, aligning with the contemporary landscape of information accessibility and technological progress.

**Keywords:** Pre-trained language model transformers, machine learning, transfer learning, natural language processing, economic growth

## 1. Introduction

Since 2020, there have been a number of crises affecting countries all over the world, ranging from the social restrictions and health-related problems caused by the COVID-19 pandemic to armed conflicts and trade wars among nations that have disrupted global supply chains. The profound impact of the coronavirus outbreak has led to the widespread destruction of economies. The decline in GDP is particularly pronounced in nations less resilient to the global downturn, with financial markets experiencing unprecedented collapses. This pervasive economic downturn has significantly affected major industries, with the exception of healthcare and Fast-Moving Consumer Goods (FMCG), which represent essential necessities [1]. As a result of the crisis,

several countries including Indonesia, experienced the worst economic downturn and officially declared a recession, which put a great deal of pressure on the global economy. According to the widely accepted convention, a recession is defined as a contracting economy for at least two consecutive quarters in terms of real Gross Domestic Product (GDP) [2]. Commencing in the second quarter of 2020, where economic growth plummeted by 5.32%, Indonesia's recession persisted for four consecutive quarters. Subsequent quarters experienced further economic contractions, with declines of 3.49% and 2.17%, respectively. However, signs of economic recovery emerged, with a notable positive upswing of 2.97% in GDP during the first quarter of 2021, followed by a substantial rebound with a GDP growth rate of 7.07% in the second quarter of 2021 [3], indicating the Indonesian economy's emergence from the recession.

The fluctuations in GDP not only signify economic recessions or downturns but also offer vital insights into national economic growth, serving as a crucial

---

\*Corresponding author: Lya Hulliyatus Suadaa, Politeknik Statistika STIS, Otto Iskandardinata 64C, Jakarta 13330, Indonesia. E-mail: [lya@stis.ac.id](mailto:lya@stis.ac.id).

metric for various stakeholders. The GDP is a well-known economic indicator because it is considered to be the best single indicator of people's well-being and it can simultaneously measure the entire income of all individuals participating in the economy and the total state spending on goods and services produced by the economy [4]. Consequently, GDP is regularly employed as a benchmark for assessing and comparing economic development among nations. The official institution in charge of computing and disseminating GDP is BPS-Statistics Indonesia.

The calculation of official GDP statistics involves the intricate and time-intensive task of aggregating and analyzing data from various sources. The current quarterly GDP release period, taking approximately 35 days from the commencement of a new quarter, reflects an improvement by the national statistics office compared to prior performance, despite the fact that there is still a fairly significant time gap. Given the dynamic nature of the economy, there is a growing need for real-time access to information on economic growth. This is particularly crucial for making timely and informed decisions in areas such as investment and economic policy formulation that aligns with the forward-looking principle. Consequently, relying on outdated GDP growth data poses limitations to its optimal utilization. To address this challenge, research has been conducted to shorten the time gap between official GDP releases and the current period by detecting GDP movements and growth rates using online news classification. This study is pioneering in Indonesia, utilizing non-conventional data sources as supplementary information for official GDP statistics.

Aside from the urgency to shorten the time lag in the GDP release period, according to Smith and Lorenc [5], the ongoing COVID-19 crisis might encourage adaptations in the statistical process to become more robust to unforeseen economic shocks. To assess the necessity for adaptation in statistical production, Delden et al. [6] argue that they need to differentiate between three categories of changes. The first category consists of those indicators and monitoring instruments designed to detect economic changes that can affect the presumptions made in the methodology used to produce the short-term statistics (STS). Those indicators and monitoring instruments can be permanently integrated into production systems and used as needed; however, it is not necessary to implement all indicators and monitoring instruments within a stable economic setting. The second category focuses on adaptations that can be made to the STS methodology in order to expand its appli-

cability to a wider variety of diverse economic conditions. The methods employed rely on assumptions that are not always explicitly stated. There is a chance that some of the assumptions will turn out to be invalid due to shifting economic conditions. The modifications in the second category offer alternative techniques that are insensitive or less sensitive to the consequences of an economic downturn. The third category consists of interventions that are crisis-specific. For instance, to enhance the imputation they used an auxiliary data source. Hence, this conducted study is in line with these findings and strengthens the need to analyze new approaches to detect economic changes because economic shocks such as the COVID-19-related crisis won't be the last to have an impact on the estimations for statistical production.

The rapid detection of GDP movements through online news classification provides the public with timely and strategic information for informed decision-making in a dynamically changing economy. The government, corporations, and the general public as financial stakeholders can optimize profit potential by leveraging projections of positive GDP movements. On the contrary, if negative GDP movements are detected the detection can be utilized as an early warning system for a downturn in economic growth or even an economic crisis. As a result, the risk of a recession can be detected earlier. However, the aggregate movement of GDP, encompassing various macroeconomic components, remains too complex for effective utilization. GDP, as a single quantification of a country's domestic economy. Recognizing this complexity, it becomes crucial to categorize GDP movement detection based on industrial sectors, allowing stakeholders in each sector to tailor their responses accordingly. This study applies the GDP categorization using the production approach, which consists of 17 industry sectors that support the domestic economy that correspond with Indonesia's official statistics guidelines for computing GDP. The GDP production approach was chosen by considering the practicalities of classifying and collecting online news data. The data collection process employs scraping techniques, utilizing identified keywords associated with the 17 industry sectors. This methodology ensures a targeted and sector-specific analysis of GDP movement, enhancing the relevance and applicability of the information for stakeholders in each segment of the economy.

In contrast to traditional news media, online news stands out as a more contemporary journalistic product, benefiting from recent technological advancements that enable real-time reporting, which presents both oppor-

tunities and challenges for statistical analysis to draw the proper conclusions from an overwhelming amount of data and even almost leads to information overload. Besides the huge amount of data, the ambiguity that frequently occurs in the context of language semantics also presents an issue for natural language analysis. In the context of a sentence, a single word may carry diverse interpretations depending on its surroundings. To address this issue, a transformers-based Language Model (LM) with an “attention” mechanism has been employed. This model comprehends the context of the entire sentence, providing nuanced meanings for related words and phrases based on their positions within the sentence. The introduction of this transformers-based language model has significantly advanced benchmarks for natural language processing [5]. This study predicts that the implementation of a transformers-based language model will yield superior results compared to prior studies.

Few studies have explored the detection of GDP fluctuations through online news. In a recent study, Khairani et al. [8] used online news collected during the first quarter of 2021 to identify changes in GDP in 14 industry sectors. News data scraping was conducted from the five online sources with the highest traffic levels. Among the eight classification algorithms tested, the Random Forest algorithm demonstrated the highest accuracy, achieving a remarkable 96.51%.

Wongso et al. [9] conducted a study focused on online news classification, aiming to identify an effective algorithm for categorizing Indonesian news articles into five distinct categories: economics, health, technology, sports, and politics. The research involved comparisons of various classifiers and feature selection techniques. The classifiers assessed included Multinomial Naïve Bayes, Multivariate Bernoulli Naïve Bayes, and Support Vector Machines (SVM), while the feature selection methods tested were TF-IDF and the Singular Value Decomposition (SVD) algorithm. The study’s findings reveal that the combination of TF-IDF and Multinomial Naïve Bayes emerges as superior, achieving an accuracy of 85% in categorizing Indonesian news articles.

While other related studies may focus on different research objects, they still encompass research objectives and methods that bear relevance to the ongoing study. For instance, Duarte et al. [10] conducted research forecasting price drops in the Brazilian stock market utilizing 11 machine learning algorithms. The study identified Gaussian Naïve Bayes (NB-G), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) as the most effective methods for predict-

ing price changes. The study concluded that NB-G and SVM were highly effective, exhibiting quick training times and ease of setup. Although Multilayer Perceptron (MLP) and XGBoost (XGB) produced commendable results, they introduced complexity to the research process, necessitating a more extended training time for parameter optimization.

Aligned with the findings of relevant studies, classical machine learning algorithms like random forest, SVM, and Naïve Bayes have consistently demonstrated high performance in research related to the detection of economic movements. Consequently, these three traditional machine learning algorithms are employed as a baseline model for comparison with the performance results of the classification achieved by pre-trained language model transformers.

Other researchers have delved into price prediction by incorporating transfer learning from pre-trained models. Chen’s study [11] employs the Fine-Tuned Contextualized-Embedding Recurrent Neural Network (FT-CE-RNN) model to forecast short-term stock price movements. The model utilizes news headlines represented by contextualized vectors generated from BERT. Results from the FT-CE-RNN research demonstrate a 4.1% increase in accuracy, surpassing all baseline models and showcasing the effectiveness of the RNN implementation. In another study, Wei and Nguyen [12] leverage the BERT model to analyze capital and financial market news for forecasting changes in stock prices. Employing sentiment analysis and utilizing the pre-trained transformers finBERT on news and headlines, Sonkiya et al.’s research [13] utilizes a Generative Adversarial Network (GAN) to predict nominal stock values. These studies collectively illustrate the diverse applications of transfer learning and pre-trained models in enhancing the accuracy and effectiveness of price prediction.

Prior studies have shown that the detection of GDP fluctuations based on online news is currently limited, particularly in the context of Indonesia. While numerous studies focus on detecting price changes in volatile financial investment instruments, research on this specific topic in Indonesia remains relatively sparse. Although recent studies have begun incorporating pre-trained model transformers, conventional machine learning methodologies still prevail in academia. Several research findings using transfer learning indicate that pre-trained transformers exhibit superior performance in detecting price movements from news text inputs, surpassing conventional machine learning methods. However, these studies primarily focus on the re-

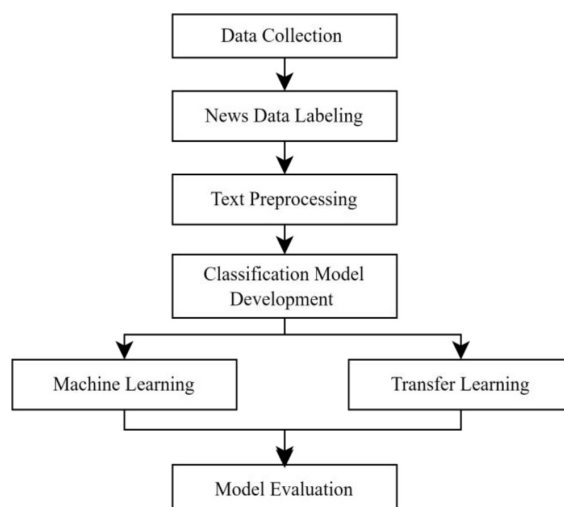


Fig. 1. Research workflow.

search object of stocks, known for their volatile price movements. Hence, adjustments are necessary to apply these findings to the detection of relatively non-volatile GDP movements, a context significantly different from the investment fluctuation discussed in prior research. Therefore, this study aims to:

1. Develop a dataset for classifying GDP movements and GDP growth rates in 17 industry sectors using online news data.
2. Detect GDP movement and GDP growth rates in 17 industry sectors using online news, employing transfer learning from pre-trained model transformers, and comparing its performance with classical machine learning methods used as a baseline model.

## 2. Methodology

### 2.1. Research limitation

This study will employ the real GDP growth approach to detect the movement and growth rate of GDP during the third quarter of 2022. This choice is made to mitigate potential biases arising from price volatility. The labeling of 'up' and 'down' categories will be based on news narratives reflecting changes in the production of goods and services rather than current price movements. This approach avoids the need for standardizing base-year prices associated with the conventional assessment of price movements. In the context of our study, we specifically analyze online news articles, and the journalistic characteristic in reporting

economic news often involves presenting information in a way that is engaging and likely to capture public attention. As a result, news articles tend to focus on highlighting either positive (up) or negative (down) aspects of economic activities. The occurrence of a truly "neutral" sentiment in economic news, where there is no clear indication of positive or negative movement, is relatively rare based on our observations and their interpretation can be ambiguous. Practically, terms denoting neither positive nor negative sentiment might be challenging to discern in economic contexts. Furthermore, our choice of using "up" and "down" labels is informed by feedback from practitioners at BPS-Statistics Indonesia, who play a pivotal role in GDP analysis. They emphasized that terms indicating "up" generally correspond to positive growth in GDP, while terms indicating "down" are associated with negative growth that may lead to economic challenges or crises. In the context of GDP movement, where terms inherently convey a sense of direction, either indicating positive growth, negative growth, or conditions of acceleration or deceleration, the consideration for a "neutral" label becomes less relevant.

### 2.2. Research workflow

The study will follow the workflow illustrated in Fig. 1, progressing through its stages in a systematic manner.

#### 2.2.1. Data collection

Web scraping was employed to gather data from the online news portal Detik.com using identified key-

Table 1  
Industrial sectors defined within the standard classification of Indonesian business fields  
(Klasifikasi Baku Lapangan Usaha Indonesia or KBLI)

No.	Category	Industrial sectors
1	A	Agriculture, forestry, and fisheries
2	B	Mining and quarrying
3	C	Manufacturing industry
4	D	Electricity and gas supply
5	E	Water supply, waste management, and recycling
6	F	Construction
7	G	Wholesale and retail trade; repair of motor vehicles and motorcycles
8	H	Transportation and warehousing
9	I	Accommodation and food and beverage service activities
10	J	Information and communication
11	K	Financial and insurance activities
12	L	Real estate
13	M, N	Business services
14	O	Government administration, defense, and mandatory social security
15	P	Education services
16	Q	Health and social activities
17	RSTU	Other services

**"Menhub: Mobilitas Kereta Api Dunia Anjlok 50%"**

Jakarta - Menteri Perhubungan Budi Karya Sumadi membeberkan ngerinya dampak pandemi COVID-19 kepada mobilitas kereta api di seluruh dunia. Mengutip Badan Pusat Statistik (BPS), **sektor transportasi dan pergudangan mengalami kontraksi paling dalam hingga minus 30,84%**. Penurunan ini menyebabkan kontribusi sektor transportasi dan pergudangan terhadap produk domestik bruto (PDB) turun 3,57% di kuartal II-2020, dari 5,57% pada kuartal II-2019.

"Hal tersebut berimplikasi pada **penurunan pendapatan di mana sektor transportasi dan pergudangan termasuk tiga sektor tertinggi yang berdampak lebih dari 90%** pelaku usaha mengalami penurunan pendapat dan sektor transportasi dan pergudangan mengalami suatu masalah menurut BPS 2020," paparnya.

Fig. 2. News indicating a downturn in GDP in transportation and storage industry sector.

words. Detik.com, a media company under PT. Trans Corporation, is exclusively accessible online. The portal includes various channels, such as DetikFinance focusing on the economy, and specific channels for local news covering Indonesian provinces (e.g., Detik-Sumut for North Sumatra, DetikJabar for West Java). This channel feature facilitates the differentiation of news categories and geographical coverage. According to SimilarWeb, Detik.com had 516.7 million visitors from July to September 2022, securing its top rank in the News and Media Publisher sites category [14]. The choice of Detik.com for article collection aligns with a study conducted by Brajawidagda et al. [15], which also utilized online news from the same source. Detik.com stands out as one of Indonesia's two most popular online news sites according to alexa.com. It's noteworthy that the freedom of the press is protected under Law 40/1999 in Indonesia, a country known for being one of the most populous democracies globally, fostering an environment without official censorship and provid-

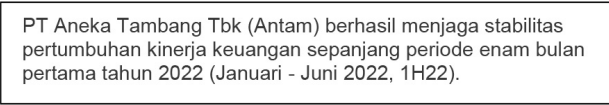
ing the press with a relatively objective voice in this populous democracy.

The primary language of the scraped data is Indonesian. However, we acknowledge the possibility that there may be instances where certain terms or phrases in other languages, such as English, could be present in the data.

To ensure comprehensive representation of each industry sector in the dataset, the keywords selected are terms that characterize the 17 different industrial sectors, as outlined in Table 1. Consequently, the collected data has been categorized for each industry sector. An example of the online news used in the study is illustrated in Fig. 2.

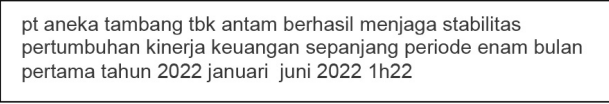
### 2.2.2. News data preparation

The dataset preparation comprises two stages of text preparation: filtering and labelling. The data filtering process involves a meticulous examination of each news article to determine its validity. However, it's important to note that this method is resource and time-intensive.



PT Aneka Tambang Tbk (Antam) berhasil menjaga stabilitas pertumbuhan kinerja keuangan sepanjang periode enam bulan pertama tahun 2022 (Januari - Juni 2022, 1H22).

Fig. 3. Raw text data before undergoing preprocessing.



pt aneka tambang tbk antam berhasil menjaga stabilitas pertumbuhan kinerja keuangan sepanjang periode enam bulan pertama tahun 2022 januari juni 2022 1h22

Fig. 4. Text data after case folding and data cleaning.

The criteria for filtering include the exclusion of articles containing advertisements, biography stories, entertainment content, and any irrelevant information not related to economic activities measured in the GDP. The primary goal of this data filtering is to distinguish between valid and invalid news data.

During the labeling stage, seven annotators, well-versed in System National Account and GDP through coursework, were engaged. Their foundational knowledge qualified them to annotate the news data for the research focusing on GDP. To address the labor-intensive nature of annotation due to the thorough manual reading required for each news release, three different annotators labeled each news release independently. The final label was determined through a majority vote as a consensus measure. The inter-rater reliability was assessed using the Krippendorff alpha metric instead of distribution.

Given the concern about annotator bias, a set of labeling guidelines was established by the researchers and strictly followed by the annotators. This guideline serves as a mitigation strategy against potential bias. To further ensure the validity of each news article before processing it for detecting GDP movements and growth rates, the labeled dataset underwent validation by a subject matter expert from the BPS-Statistics Indonesia. This expert validation enhances the credibility and reliability of the dataset.

### 2.2.3. Krippendorff's Alpha

Krippendorff's Alpha is a widely adopted statistical measure for inter-rater reliability. It gauges the variability resulting from the choice of analysis units rather than the characteristics of annotators [16]. Known as a nonparametric measurement for annotator agreement, it assesses the consistency of scoring among two or more annotators on the same analysis unit [17]. Krippendorff's Alpha offers high flexibility regarding measurement scales, the number of annotators, and can handle missing data effectively [18].

### 2.2.4. Text pre-processing

Pre-processing procedures conducted in this study include the following:

#### 1) Case folding and data cleaning

During this procedure step, the letters are transformed into lowercase, while duplicate news, punctuation, and excessive spaces are removed. To provide a visual representation of this process, Fig. 3 displays the raw data, while Fig. 4 illustrates the data after the application of cleaning and case folding.

#### 2) Stemming and stopwords removal

This processing stage is exclusively applied to classical machine learning algorithms. It encompasses two essential tasks: stemming, where affixes are removed to obtain root words, and removal of frequently occurring words with limited semantic value, such as conjunctions and prepositions [19]. The PorterStemmer library is employed to execute this step. Figure 5 visually represents this stage, illustrating the impact of stemming and stopwords removal on the processed text.

#### 3) Tokenization

In this pre-processing stage, the text document undergoes tokenization, breaking it down into smaller units known as tokens, which serve as the fundamental elements for classical machine learning algorithms [20]. In the context of pre-trained transformers, these tokens may represent subwords. The NLTK library is employed for efficient tokenization. Figure 6 visually depicts the impact of tokenization on the processed text.

### 2.2.5. Features selection

In the feature selection process, the representation of news data is transformed into vectors. The vector representation is calculated based on the Term Frequency-Inverse Document Frequency (TF-IDF) weights, ob-

pt aneka tambang tbk antam hasil jaga stabil tumbuh kerja uang panjang periode enam bulan pertama tahun 2022 januari juni 2022 1h22

Fig. 5. Text data after stemming and stopwords removal.

['pt', 'aneka', 'tambang', 'tbk', 'antam', 'hasil', 'jaga', 'stabil', 'tumbuh', 'kerja', 'uang', 'panjang', 'periode', 'enam', 'bulan', 'pertama', 'tahun', '2022', 'januari', 'juni', '2022', '1h22']

Fig. 6. Text data after tokenization.

tained using the sci-kit learn library. The term frequency formula (TF) used in `sklearn` is expressed as follows [21]:

$$TF_{(t,d)} = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}, \quad (1)$$

where,  $f_{t',d}$  = frequency of the term  $t$  in the document  $d$ ,  $\sum_{t' \in d} f_{t',d}$  = total term  $t$  that appears in each document  $d$ .

The Inverse Document Frequency (IDF) formula in `sklearn` adds a constant equal to 1 in the numerator and denominator to prevent division by zero, and it is expressed as follows:

$$IDF_{(t)} = \log \frac{1 + n}{1 + df(t)} + 1, \quad (2)$$

where,  $n$  = the number of training documents used,  $df(t)$  = the number of training documents containing term  $t$ .

#### 2.2.6. Classification model development

The study involved building both classical machine learning classification models and pre-trained language model transformers to detect GDP movements and GDP growth rates using online news. The BERT models employed in this study were adapted for classification tasks through the augmentation of a custom classification layer atop the base BERT architecture. This supplementary layer, comprising a pre-classifier dense layer, a classifier dense layer, and a dropout layer, facilitated the training of the model for the specific supervised task. Throughout the training process, model parameters, including those within the appended classification layer, were optimized using backpropagation. Two transformer-based language models, namely `IndoBERT` and `MultilingualBERT`, were utilized to facilitate natural language processing in Indonesian. Brief descriptions of these models are provided below:

Table 2

Hyperparameter grid for random forest classifier

Hyperparameter	Values
n_estimators	340, 380, 400, 425, 1000, 2000
max_features	'log2', 'sqrt'

#### 1) Random Forest

Random forest is an algorithm that generates a set of decision trees to categorize class labels based on a majority vote for classification task [22]. In this approach, multiple trees, trained independently, are combined to formulate predictions. The ensemble of numerous trees contributes to enhanced accuracy in classification results. The large number of trees will improve the accuracy of the classification results [23]. Random Forest is adept at handling large input variables and can effectively balance errors in imbalanced datasets [24]. The hyperparameters specified for this classifier are outlined in Table 2. These hyperparameters dictate the number of trees (n\_estimators) and the maximum number of features considered for each split (max\_features) during the random forest model construction.

#### 2) Support vector machine (SVM)

SVM works on the premise of determining a hyperplane that can divide distinct classes. One of the advantages of SVM is the ability to check the optimal combination of features, making it highly robust for high-dimensional data [25]. SVM excels in handling text data for classification tasks because text data is typically sparse and high-dimensional, meeting the condition where certain features may be irrelevant but likely to correlate with each other, often organized into linearly separable categories [26]. SVM is widely employed in text classification and is considered one of the most effective non-deep learning classifiers for handling large datasets with numerous features [24]. In the context of growth detec-

Table 3  
Hyperparameter grid for SVM classifier

Hyperparameter	Values
C	1, 10, 100, 1000
Gamma	0.1, 0.01, 0.001, 0.0001, 'scale', 'auto'
Kernel	'rbf', 'poly', 'linear', 'sigmoid'

tion, hyperparameter tuning revealed that the sigmoid kernel yielded the best performance among the kernels tested. However, for the task of GDP movement detection, the results from hyperparameter tuning indicated that the linear kernel demonstrated the optimal performance. Table 3 outlines the range of values explored during the grid search for each hyperparameter.

### 3) Naïve Bayes Classifier

Naïve Bayes is a probabilistic classifier that relies on strong independence assumptions in feature space. It applies probability concepts, with parameters learned from training data using Bayesian probability rules [27]. The classifier assumes, naively, that the appearance of a word in a document is independent of other words, contrary to the interrelated nature of words in human language. Despite this simplification, Naïve Bayes often performs well in text classification tasks [26]. In this study, two Naïve Bayes models are employed: Bernoulli Naïve Bayes and Multinomial Naïve Bayes. The hyperparameters for these classifiers are specified in Table 4, with particular emphasis on the alpha values, which play a crucial role in shaping the behavior of the Naïve Bayes models during the training process.

### 4) IndoBERT

IndoBERT is a pre-trained language model based on BERT transformers that has been specifically trained with Indonesian vocabulary using a masked language model within the Hugging Face framework [28]. Using IndoLEM, a sentiment analysis dataset in Indonesian, IndoBERT surpasses conventional machine learning models, including Naïve Bayes (+ 13.2 points), Multilingual-BERT (+ 7.5 points), and MalayBERT (+ 2.1 points), achieving an F1-score of 84.13%. This highlights IndoBERT's superior performance in sentiment analysis tasks in Indonesian [29]. Hyperparameter tuning was not conducted for this particular BERT model. Instead, a set of predefined parameters was employed, and model selection was performed by identifying the epoch with the optimal performance evaluation metrics after a specified num-

ber of iterations. This approach allowed for effective model training and selection while avoiding the complexity of hyperparameter tuning. The parameters used for training the IndoBERT model include:

- Learning Rate: 5e-6
- Number of Epochs: 10
- Batch Size: 4

### 5) Multilingual BERT

Multilingual BERT (mBERT) is a pre-trained version of the BERT model that has been trained in 102 languages (uncased) and 104 languages (cased) on Wikipedia, including Indonesian. So that it can be used for NLP tasks in Indonesian. However, mBERT faces challenges in effectively representing all 104 languages, especially low-resource ones [7]. In our study, we compare mBERT with Monolingual-BERT, specifically IndoBERT, tailored for Bahasa Indonesia. The hyperparameters for mBERT were selected as follows:

- Learning Rate: 5e-6
- Number of Epochs: 10
- Batch Size: 4

These parameters were chosen to strike a balance between effective model training and computational efficiency. The specific language-agnostic nature of mBERT allows it to capture linguistic nuances across different languages, making it a versatile choice for multilingual NLP applications.

The BERT models utilized in this study underwent fine-tuning for the specific task of GDP movement and growth rate classification. The architecture of the fine-tuned BERT model is illustrated in Fig. 7, providing an overview of the tokenization and embedding process. Each sentence token undergoes token embedding, resulting in subword token embeddings within the BERT model. The token embeddings include special tokens, such as [CLS] at the beginning and [SEP] at the end, signifying sentence separation. The segment embedding designates the sentence segment, represented by  $E_A$  in this illustration for a single sentence.

Position embeddings are assigned to each token, denoted as  $E_1, E_2, \dots, E_{m+1}$ , where  $m+1$  is the position of the [SEP] token. These position embeddings serve as input to the transformer layer. The transformer layer processes the embeddings, generating output representations such as  $C_{[CLS]}$ ,  $C_{pt}$ ,  $C_{ane}$ ,  $C_{\#ka}$ , and so on. The output representations are then fed into the classification layer, where the model learns to map these repre-



Table 4  
Hyperparameter grid for naïve bayes classifier

Hyperparameter	Values
alphas	[1.0e-10, 0.0001, 0.001, 0.01, 0.1, 0.5, 1.0, 2.0, 10.0]

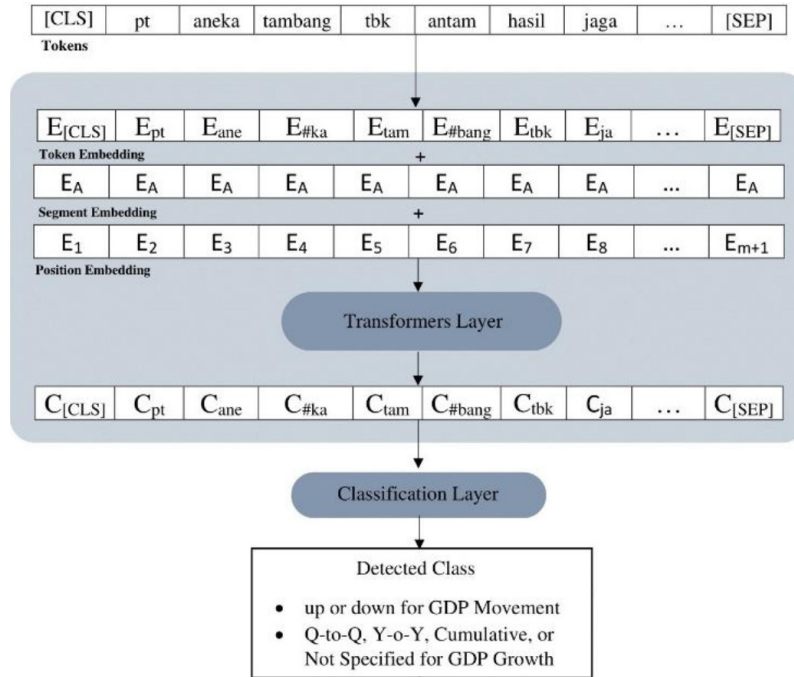


Fig. 7. Fine-tuned BERT architecture for GDP movement and growth rate detection.

sentations to the detected class, facilitating the classification of GDP movement and growth rates. This fine-tuned architecture enables the BERT model to adapt its pre-trained knowledge to the specific task at hand, enhancing its ability to discern patterns and features relevant to economic indicators in online news data.

2.2.7. Model evaluation

To evaluate the performance of the generated classifier, the dataset is divided into two main parts: training and testing. According to Pramana, et.al [28], the training set is used to build the model, while the test set is used to determine how good the final classifier is. There are several techniques that can be employed, including the hold-out method, cross-validation, and bootstrap, to split the dataset into training sets and test sets. In this study, the dataset division technique employed is stratified cross-validation with five data folds. The validation set is also utilized for hyperparameter tuning in classical machine learning classifiers and fine-tuning in pre-trained language model transformers. The dataset is split in the following proportions: 80% for the training

set, 10% for the validation set, and 10% for the testing set. The partition mechanism is illustrated in Fig. 8.

The evaluation of the GDP movement and growth rate detection model encompasses both data testing and 5-fold cross-validation. The k-fold Cross Validation method is employed, which divides the data into training and testing sets, distributing them according to the specified number of k-folds [33]. The metrics used to measure model performance in text classification are accuracy, precision, recall, and F1-score.

- a. Accuracy measures the overall correctness of the model by dividing the number of correctly classified items by the total number of items classified. It is determined by dividing the percentage of correctly classified items by the total number of items classified.
- b. Precision, reflecting the magnitude of the positive classification by computing the proportion of true positive classifications to total positive classifications.
- c. Recall assesses the ability of the model to identify all relevant instances. It calculates the proportion

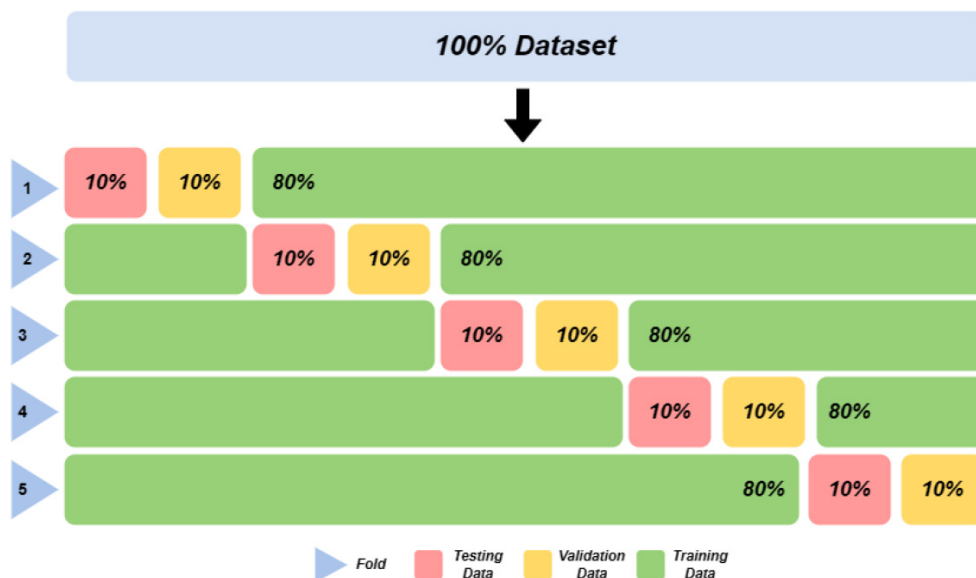


Fig. 8. Cross-validation mechanism for classifier evaluation.

of positive classifications that are correct among all classifications that should be positive.

- d. F1-score is the harmonic average of recall and precision. It provides a balanced measure that considers both false positives and false negatives.
- e. Macro averaging of F1-scores calculates F1 for each class and then takes the average, treating all classes equally.

### 3. Result and discussion

#### 3.1. Dataset construction

The research collected a total of 13,286 data points through web scraping from Detik.com. Following the strategic filtering process, 1,586 news items were retained for further analysis. The data filtering aimed to address imbalances in our machine learning models by targeting 100 news entries per industrial sector. However, some sectors, including Mining and Quarrying, Water Supply, Waste Management, and Recycling, Real Estate, Business Services, and Government Administration, Defense, and Mandatory Social Security, did not meet the target. Despite this, we proceeded with the analysis, imposing a minimum threshold of over 60 news articles for sectors falling below the target. The initial filtered news totaled 1,540. Recognizing the limited number of data points with the q-to-q growth label, 46 additional data points labeled q-to-q were added

from random industrial sectors, resulting in a total of 1,586 processed news data.

To ensure a comprehensive representation of each industry in our datasets, a stratified cross-validation framework was implemented. This approach played a crucial role in maintaining a balanced distribution of industries across the training, validation, and test sets. By stratifying the data based on industry, our objective was to capture the distinct nuances and dynamics of each sector, mitigating the impact of sparse observations from certain industries. Figure 9 below visually depict the distribution of both raw and filtered data for each industrial sector.

The news data used for GDP movement detection was labeled as either “up” or “down.” Meanwhile, GDP growth rates were classified into categories such as “Not specified,” “Year on Year,” “Cumulative to Cumulative,” and “Quarter to Quarter.” In the context of GDP growth rates, the “Cumulative to Cumulative” label encompasses periods such as half a year or a semester, capturing year-to-date temporal information. It specifically refers to the cumulative growth rate over a specific time interval, offering a comprehensive measure of economic performance during that period. The annotator’s labeling results led to imbalanced data, as depicted in Fig. 10. However, this study does not address the issue of imbalanced data, maintaining it to ensure that the classification result reflects the actual condition of the economy. The low alpha value in the GDP growth rate data indicates a relatively low level of reliability in the labeling of GDP growth rates by annotators. Upon

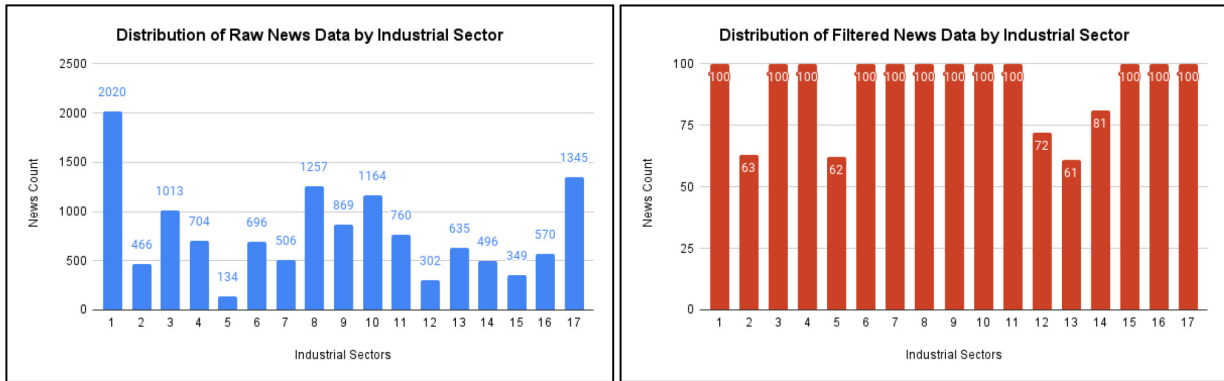


Fig. 9. Distribution of raw and filtered news data by industrial sector.

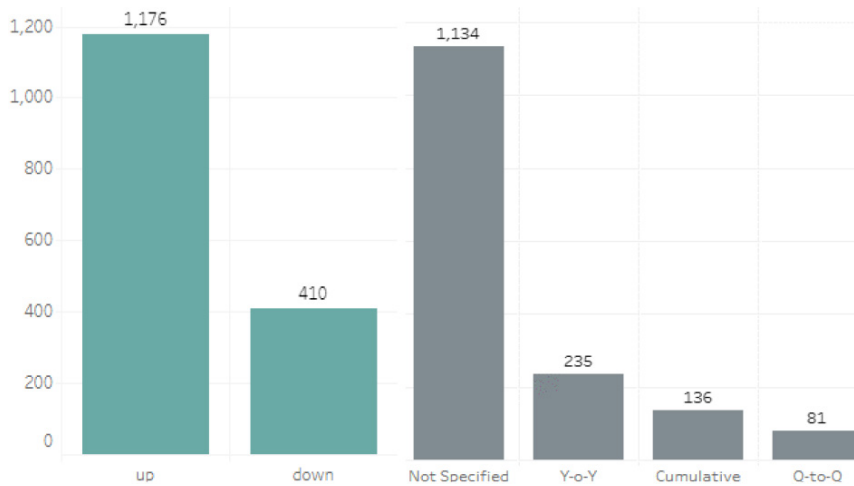


Fig. 10. Comparison of the label data count.

further analysis, it was found that several news data contained more than one type of growth rate period, and some had inadequate time descriptions in the news narrative.

### 3.2. GDP movement detection using news classification

The classification was conducted using three classical machine learning models, and the best parameters were obtained through hyperparameter tuning using the grid search technique with a five-fold data cross-validation. SVM emerged as the model with the most optimal performance in the classification of GDP movement data.

In Tables 5 and 6, we present the evaluation metrics for both the original class assignment and performance metric comparison of the three classical machine learning models and pre-trained transformers models. In our binary classification task, where 'up' represents posi-

tive instances indicative of positive GDP movements or accelerations, and 'down' represents negative instances indicative of negative GDP movements or slowdowns, we report precision, recall, and F-1 scores for both classes. This approach offers a comprehensive analysis, allowing us to assess the model's performance in detecting both positive and negative movements in GDP, acknowledging the equal importance of both directions in our study.

In Table 5, the SVM model exhibits outstanding performance in recall, F1-score, and accuracy metrics, based on the four evaluation metrics. However, the precision metric of SVM shows a slightly lower score compared to Random Forest, which has the lowest performance among the other three evaluation metrics.

According to Table 6, the performance of classification with transfer learning outperforms classification with classical machine learning for the IndoBERT model. The classification results using the IndoBERT-

Table 5  
Performance evaluation of classical machine learning models for GDP movement classification

Model	Label	Precision	Recall	F1-Score	Accuracy
SVM	down	0.7257	0.6484	0.6811	<b>0.8440</b>
	up	0.8802	0.9134	0.8960	
	macro avg	<b>0.8030</b>	<b>0.7809</b>	<b>0.6848</b>	
Random Forest	down	0.8389	0.3439	0.4859	0.8126
	up	0.8101	0.9760	0.8848	
	macro avg	0.8245	0.6599	0.6854	
Bernouli Naïve Bayes	down	0.7169	0.5970	0.6498	0.8352
	up	0.8675	0.9162	0.8910	
	macro avg	0.7922	0.7566	0.7704	

Table 6  
Classification performance of transfer learning from pre-trained language model transformers for detecting GDP movement

Model	Label	Precision	Recall	F1-score	Accuracy
IndoBERT Base	down	0.7324	0.7858	0.7556	0.8704
	up	0.9216	0.9016	0.9112	
	macro avg	0.8270	0.8437	0.8334	
IndoBERT Large	down	0.7544	0.8551	0.7957	<b>0.8881</b>
	up	0.9470	0.9006	0.9220	
	macro avg	<b>0.8507</b>	<b>0.8779</b>	<b>0.8589</b>	
Multilingual BERT uncased	down	0.6593	0.7085	0.6775	0.8314
	up	0.8959	0.8772	0.8855	
	macro avg	0.7776	0.7929	0.7815	

Large model show the best performance, achieving an accuracy of 0.8880. The accuracy obtained from IndoBERT-Base is not significantly different, standing at 0.8704. Considering resource efficiency and computing time, IndoBERT-Base could be chosen as a viable alternative. On the other hand, the classification results using MultilingualBERT (mBERT) did not outperform the baseline classical machine learning model with hyperparameter tuning. In comparison to SVM and Bernoulli Naïve Bayes, mBERT yielded a lower accuracy score, which is 0.8314. The performance metrics of the GDP movement classification model with transfer learning from pre-trained language model transformers are listed in Table 3.

Following the recommendations provided by practitioners and subject matter experts from BPS-Statistics Indonesia, a more precise interpretation of the results of GDP movement detection using online news classification is to identify accelerated GDP growth for labels indicating an “up” movement and a slowdown in GDP growth for labels indicating a “down” movement. The analysis takes into account real GDP growth trends over the past 12 years on a quarterly basis. The observed pattern of GDP movement suggests a seasonal effect due to the use of quarterly periods, but when ignoring the seasonal effect, the pattern of GDP movement tends to exhibit stability, as depicted in the graph in Fig. 11.

Negative GDP growth is a rare occurrence in the Indonesian economy. From 2010 to 2022, negative GDP growth occurred for three quarters in 2020 due to the COVID-19 pandemic outbreak. Therefore, detecting declining GDP movements obtained using online news is more relevant for use as an early warning system for a potential economic crisis.

### 3.3. GDP growth rate detection using news classification

Hyperparameter tuning using the grid search technique was performed on all machine-learning models to enhance model performance. The evaluation results of GDP growth rate classification with classical machine learning are presented in Table 7. Among the three compared models, SVM demonstrates the best performance with an accuracy of 0.7811. However, the classification performance of GDP growth rate is lower compared to the classification of GDP movements. One contributing factor is the presence of numerous news articles with unspecified growth rate periods, causing imbalanced data. To address this issue, a discussion with the subject matter expert from BPS-Statistics Indonesia’s National Production Account Consolidation Sub-Directorate is conducted, leading to the conclusion that news data with an unspecified growth rate period can be assumed as the annual GDP growth rate.

Table 7  
Performance evaluation of classical machine learning models classification for detecting GDP growth rates

Model	Label	Precision	Recall	F1-Score	Accuracy
SVM	Not specified	0.8182	0.9777	0.8907	<b>0.7811</b>
	Cumulative	0.5000	0.0719	0.1183	
	Q-to-Q	0.6600	0.2733	0.3585	
	Y-o-Y	0.4911	0.3076	0.3752	
Random Forest	macro avg	<b>0.6173</b>	<b>0.4076</b>	<b>0.4357</b>	0.7610
	Not specified	0.7899	0.9896	0.8783	
	Cumulative	0.3667	0.0530	0.0901	
	Q-to-Q	0.4667	0.0983	0.1612	
Multinomial Naïve Bayes	Y-o-Y	0.4442	0.1727	0.2438	0.7597
	macro avg	0.5169	0.3284	0.3434	
	Not specified	0.7959	0.9687	0.8737	
	Cumulative	0.2400	0.0515	0.0821	
	Q-to-Q	0.4600	0.1217	0.1774	
	Y-o-Y	0.4680	0.2633	0.3325	
	macro avg	0.4910	0.3513	0.3665	

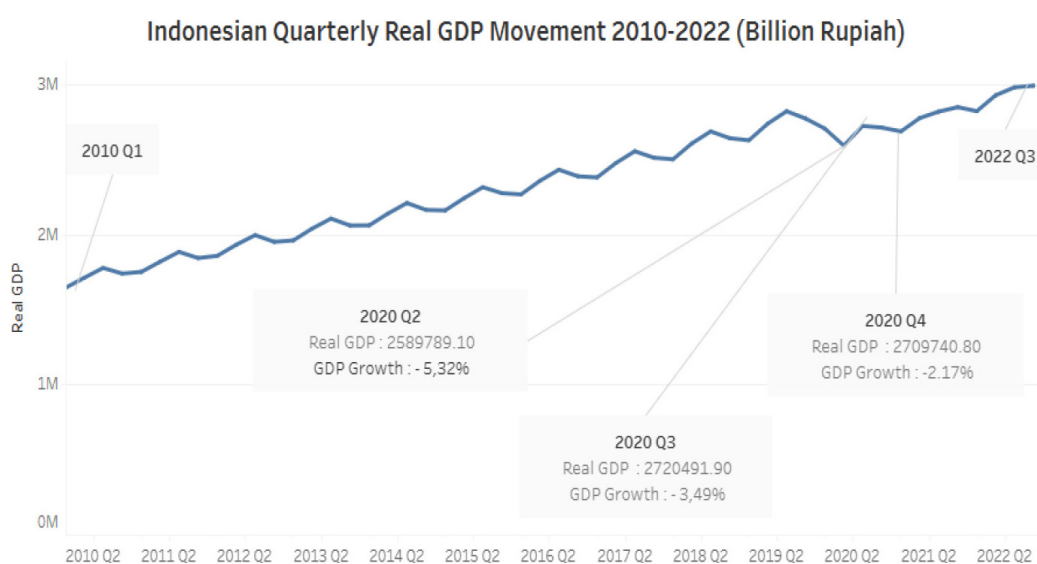


Fig. 11. Trends in quarterly real GDP growth from 2010 to 2022.

The application of classification using transfer learning from the pre-trained language model transformer is then applied. The evaluation results of the GDP growth rate classification with pre-trained language model transformers are shown in Table 8. The pre-trained language model transformers that provide the best performance is IndoBERT-Base with an accuracy of 0.7899. Based on recall, f1-score, and accuracy, IndoBERT-Base outperforms SVM as the best baseline model. In our investigation, we observed that the performance of IndoBERT-Large on the classification task at hand was suboptimal. From Table 8, it can be seen that IndoBERT-Large, which gives the best performance in the classification of GDP movements, actually gives poor performance in the classification of

the GDP growth rate. Several factors may contribute to its diminished effectiveness for this specific application. Notably, imbalanced class distribution could be a contributing factor, as the dataset contains significantly fewer examples for certain classes, potentially hindering the model's ability to generalize across all classes uniformly. Further research and analysis are warranted to gain deeper insights into the model's behavior on this particular task and dataset.

### 3.4. Consideration of class imbalance remedies

In response to valuable feedback from reviewers, we revisited the literature, exploring potential remedies for class imbalance as suggested by Kuhn and Johnson

Table 8  
Classification Performance of pre-trained language model transformers for detecting GDP growth rates

Model	Label	Precision	Recall	F1-score	Accuracy
IndoBERT Base	Not specified	0.8896	0.9242	0.9055	<b>0.7899</b>
	Cumulative	0.3809	0.4055	0.3832	
	Q-to-Q	0.6155	0.6783	0.6133	
	Y-o-Y	0.4972	0.3264	0.3759	
	macro avg	<b>0.5958</b>	<b>0.5836</b>	<b>0.5694</b>	
IndoBERT Large	Not specified	0.7895	0.9793	0.8716	0.7597
	Cumulative	0.1000	0.0167	0.0286	
	Q-to-Q	0.2000	0.1700	0.1833	
	Y-o-Y	0.1910	0.2182	0.2036	
	macro avg	0.3201	0.3460	0.3218	
Multilingual BERT uncased	Not specified	0.8768	0.9395	0.9069	0.7887
	Cumulative	0.4212	0.3131	0.3487	
	Q-to-Q	0.5667	0.3483	0.4214	
	Y-o-Y	0.4702	0.4067	0.4333	
	macro avg	0.5837	0.5019	0.5276	

(2016) [34]. While these remedies are recognized as effective in certain contexts, it is crucial to contextualize their application within the specific nature of our research on GDP prediction.

#### 3.4.1. Model tuning

The approach involves tuning the model to maximize the accuracy of the minority class. The effectiveness may vary across models and datasets. In this case, sensitivity increased with more complex models, but the trade-off with specificity was not favorable. In our research, we have indeed performed extensive hyperparameter tuning for each classical machine learning classifier. However, our optimization strategy did not specifically focus on maximizing the accuracy of the minority class to avoid introducing bias. Our objective was to strike a balance between sensitivity and specificity without disproportionately favoring one class over the other. We recognize the importance of a fair evaluation across all classes, and our hyperparameter tuning process was designed to achieve a robust and unbiased performance across the entire dataset.

#### 3.4.2. Alternate cutoffs

This method explores adjusting cutoffs for predicted probabilities using the ROC curve. It allows for a balance between sensitivity and specificity. However, the choice of cutoff involves trade-offs and should be carefully considered based on the context. It's worth noting that the method is typically applied when dealing with two possible outcome categories. In our research, we encountered the challenge of class imbalance in a multiclass classification setting, where the balancing of sensitivity and specificity across multiple classes is a nuanced task. While the ROC curve and alternate

cutoffs are valuable for binary classification, our study involves a broader spectrum of outcome categories. We have considered various strategies, including hyperparameter tuning and cross-validation, to ensure a comprehensive and fair evaluation across all classes. It's worth noting that the method mentioned is typically applied when dealing with two possible outcome categories. In our research, we encountered the challenge of class imbalance in a multiclass classification setting, where the balancing of sensitivity and specificity across multiple classes is a nuanced task. While the ROC curve and alternate cutoffs are valuable for binary classification, our study involves a broader spectrum of outcome categories.

#### 3.4.3. Adjusting prior probabilities

This method involves adjusting prior probabilities to counteract class imbalance. Balanced priors or a training set can be used to address the bias toward the majority class. In our research, we specifically, employ classifier model like Naive Bayes that consider prior probabilities. Rather than adjusting prior probabilities using a balanced training set, we opted for hyperparameter tuning, focusing on the alpha parameter of Naive Bayes. Additionally, to ensure a robust evaluation, we adopted a 5-fold cross-validation approach, enhancing the generalization performance of the model. It's important to note that our study involves a comparative analysis with other classifiers, such as SVM and Random Forest, which do not rely on prior probabilities. This intentional diversity in classifier selection aims to provide a comprehensive evaluation of different methodologies. Adjusting prior probabilities might introduce bias, impacting the fairness of the comparison.

#### 3.4.4. Unequal case weights

Models can assign different weights to individual data points during training. This method emphasizes certain points, potentially aiding in mitigating class imbalance. However, determining weights for individual data points often requires prior knowledge or information from previous research. In the context of research on natural language processing (NLP) and official statistics, especially GDP, obtaining such prior knowledge might be challenging due to the limited availability of relevant information specifically that aligns with Indonesian economy, unlike certain domains where pre-existing knowledge may readily guide the determination of prior probabilities, our study operates within the context of official statistics and NLP, where such a level of detailed prior knowledge may not be readily accessible.

#### 3.4.5. Sampling methods

Sampling methods, such as up-sampling and down-sampling, aim to balance class frequencies in the training set. Post hoc sampling approaches can help address imbalance during model training. We did consider a form of sampling approach in our study. We strategically determined the target sample to ensure a representative representation of all 17 industrial sectors in our dataset. However, the challenge we faced was the inherent trade-off between achieving balance in the representation of industrial sectors and maintaining balance in the classification of GDP growth rate categories. The nature of our dataset poses a complex problem where balancing one aspect inevitably leads to an imbalance in the other. We prioritized the comprehensive representation of the 17 industrial sectors, considering them as fundamental components that contribute to the construction of GDP. To address the imbalance in the minority class of the quarter-to-quarter (q-to-q) GDP growth rate, we introduced 46 additional q-to-q data points with a random industrial sector. We also validated our approach through consultation with subject matter experts at BPS-Statistics Indonesia.

## 4. Conclusion

Based on the comprehensive analysis and findings discussed in this study, several key conclusions can be drawn related to the proposed research objectives:

- 1) Based on the constructed online news dataset, online news emerges as a reliable alternative data source for the early detection of GDP move-

ments and growth rates. Achieving an accuracy of 88.8% with IndoBERT-Large and 78.9% with IndoBERT-Base pre-trained language model transformers, this study underscores the effectiveness of leveraging online news for economic analysis. Despite encountering imbalanced labels during dataset construction, the study deliberately chose to retain the actual label distribution, avoiding extensive data manipulation such as augmentation or resampling. This decision aims to preserve the dataset's authenticity, ensuring it accurately reflects real-world economic conditions. The study primarily employs classification techniques to detect GDP movement, acknowledging the challenges posed by imbalanced labels and emphasizing a realistic portrayal of economic scenarios.

- 2) The classification of online news for detecting GDP movement and growth rates has been successfully conducted. In comparison to classical machine learning models, the transfer learning approach employing IndoBERT-Large and IndoBERT-Base pre-trained language model transformers exhibits superior performance based on accuracy, precision, recall, and F1-score metrics. This highlights the efficacy of leveraging pre-trained language models for the task of GDP detection, showcasing improved capabilities over traditional machine learning methods in capturing nuanced patterns in online news data related to economic indicators.

Further research plans to be proposed is improving the classification performance of the GDP growth rate using sentence selection based on keywords that indicate the important features with several adjustments [35].

## References

- [1] Manna S, Mukherjee S, Das D, Saha A. Impact of COVID-19 on the world economy. *Journal of Climate Change*. 2023 Jan 1; 9(1): 67-72.
- [2] Eggers AC, Ellison M, Lee SS. The economic impact of recession announcements. *Journal of Monetary Economics*. 2021 May 1; 120: 40-52.
- [3] BPS. Produk Domestik Bruto Indonesia Triwulanan 2017–2021. Badan Pusat Statistik. 2021.
- [4] Mankiw G, Quah E, Wilson P. *Pengantar Ekonomi Makro, Edisi Ketiga, Salemba Empat*. Jakarta. ISBN: 9789790613560.
- [5] Smith PA, Lorenc B. Robust official business statistics methodology during COVID-19-related and other economic downturns. *Statistical Journal of the IAOS*. 2021; 37: 1079-1084.

- [6] van Delden A, de Winter M, van Bommel K. Identifying and evaluating COVID-19 effects on short-term statistics. *Statistical Journal of the IAOS*. (Preprint): 1-24.
- [7] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv1810.04805*. 2018 Oct 11.
- [8] Khairani F, Kurnia A, Aidi MN, Pramana S. Predictions of Indonesia economic phenomena based on online news using Random Forest. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*. 2022 Apr 27; 7(2): 532-40.
- [9] Wongso R, Luwinda FA, Trisnajaya BC, Rusli O. News article text classification in Indonesian language. *Procedia Computer Science*. 2017 Jan 1; 116: 137-43.
- [10] Duarte JJ, Montenegro González S, Cruz JC. Predicting stock price falls using news data: Evidence from the Brazilian market. *Computational Economics*. 2021 Jan; 57: 311-40.
- [11] Chen Q. Stock movement prediction with financial news using contextualized embedding from bert. *arXiv preprint arXiv2107.08721*. 2021 Jul 19.
- [12] Wei F, Nguyen UT. Stock trend prediction using financial market news and BERT. *Wall Street Journal*. 2018 Aug 8. doi: 10.5220/0010172103250332.
- [13] Sonkiya P, Bajpai V, Bansal A. Stock price prediction using BERT and GAN. *arXiv preprint arXiv2107.09055*. 2021 Jul 18.
- [14] SimilarWeb. Detik.com Website Traffic Rank Juli-September 2022. <https://www.similarweb.com/website/detik.com/#overview>. [Accessed 4 November 2022].
- [15] Brajawidagda U, Reddick CG, Chatfield AT. Urban resilience in extreme events: Analyzing online news and Twitter use during the 2016 Jakarta terror attack. *Information Polity*. 2017 Jan 1; 22(2-3): 159-77.
- [16] Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*. 2007 Apr 1; 1(1): 77-89.
- [17] Gwet KL. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. *Advanced Analytics, LLC*. 2014 Sep 7.
- [18] Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data—which coefficients and confidence intervals are appropriate? *BMC medical research methodology*. 2016 Dec; 16: 1-0.
- [19] Vijayarani S, Ilamathi MJ, Nithya M. Preprocessing techniques for text mining-an overview. *International Journal of Computer Science and Communication Networks*. 2015 Feb; 5(1): 7-16.
- [20] Vijayarani S, Janani R. Text mining: Open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACII)*. 2016 Jan; 3(1): 37-47.
- [21] Joachims T. A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. *Carnegie-mellon univ pittsburgh pa dept of computer science*. 1996 Mar 1.
- [22] Liparas D, HaCohen-Kerner Y, Moutzidou A, Vrochidis S, Kompatsiaris I. News articles classification using random forests and weighted multimodal features. In *Multidisciplinary Information Retrieval: 7th Information Retrieval Facility Conference, IRFC 2014, Copenhagen, Denmark, 2014 November 10–12; Proceedings 7 2014*; 63-75. Springer International Publishing. doi: 10.1007/978-3-319-12979-2\_6.
- [23] Al-Ash HS, Putri MF, Mursanto P, Bustamam A. Ensemble learning approach on Indonesian fake news classification. In *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*. 2019 Oct 29; 1-6. IEEE.
- [24] Panda M, Patra MR. Evaluating machine learning algorithms for detecting network intrusions. *International Journal of recent trends in Engineering*. 2009 May 1; 1(1): 472.
- [25] Al-Ash HS, Putri MF, Mursanto P, Bustamam A. Ensemble learning approach on Indonesian fake news classification. In *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*. 2019 Oct 29; 1-6. IEEE.
- [26] Brouwer AE, Haemers WH. *Spectra of graphs*. Springer Science and Business Media. 2011 Dec 17.
- [27] Aggarwal CC, Zhai C. A survey of text classification algorithms. In *Mining Text Data*. 2012; 163-222.
- [28] Shahi TB, Pant AK. Nepali news classification using Naive Bayes, support vector machines and neural networks. In *2018 International Conference on Communication Information and Computing Technology (ICCICT)*. 2018 Feb 2; 1-5. IEEE.
- [29] Koto F, Rahimi A, Lau JH, Baldwin T. IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *arXiv preprint arXiv2011.00677*. 2020 Nov 2.
- [30] Suleymanov U, Rustamov S, Zulfugarov M, Orujov O, Musayev N, Alizade A. Empirical study of online news classification using machine learning approaches. In *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*. 2018 Oct 17; 1-6. IEEE.
- [31] Wilie B, Vincentio K, Winata GI, Cahyawijaya S, Li X, Lim ZY, Soleman S, Mahendra R, Fung P, Bahar S, Purwarianti A. IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv preprint arXiv2009.05387*. 2020 Sep 11.
- [32] Pramana S, Yuniarto B, Mariyah S, Santoso I, Nooraeni R. *Data mining dengan R konsep serta implementasi*. Bogor: In Media. 2018; 206.
- [33] Liu L, Özsu MT, editors. *Encyclopedia of database systems*. New York, NY, USA: Springer. 2009 Sep 29.
- [34] Kuhn M, Johnson K. *Applied predictive modeling*. New York: Springer. 2013 Sep 1.
- [35] Fiok K, Karwowski W, Gutierrez-Franco E, Davahli MR, Wilamowski M, Ahram T, Al-Juaid A, Zurada J. Text guide: improving the quality of long text classification by a text selection method based on feature importance. *IEEE Access*. 2021 Jul 26; 9: 105439-50.