JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Public Opinion Sentiment Analysis of the Brain Drain Phenomenon on Social Media X Using the Naive Bayes Classifier Algorithm

Risma Hidayati*, Abdul Halim Hasugian
Faculty of Science and Technology, Computer Science, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Correspondence Author Email: rismaaja201@gmail.com
Submitted **10-10-2025;** Accepted **30-10-2025;** Published **31-10-2025**

Abstract
The brain drain phenomenon is an important issue in Indonesia due to the increasing number of young professionals choosing to work abroad, which reduces the quality of human resources within the country. This study aims to analyze public opinion toward the brain drain phenomenon through the X (Twitter) social media platform and classify public sentiment using the Naive Bayes Classifier algorithm. Data were collected through a web crawling process within the last two years, resulting in 1,170 relevant Indonesian-language tweets. The preprocessing stage included cleaning, case folding, tokenizing, normalization, stopword removal, and stemming to produce clean and structured data. Word weighting was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method to measure the significance of each term. The findings show that public opinion is divided into two main sentiments: positive and negative. Positive sentiment reflects the perception that working abroad offers career advancement and experience, while negative sentiment expresses concern about the loss of skilled human resources. The classification model achieved a high level of accuracy in categorizing sentiment data. This research contributes to understanding public perceptions and provides a foundation for developing strategic policies to address the brain drain issue in Indonesia.

Keywords: Brain Drain; Sentiment Analysis; X Social Media; Naive Bayes Classifier; TF-IDF

1. INTRODUCTION

Indonesia is experiencing a brain drain phenomenon, namely the migration of young and qualified professionals abroad which offers more promising career and economic prospects [1]. It was recorded that nearly 4,000 Indonesian citizens changed citizenship between 2019 and 2022, with the majority being in the productive age group of 25–35 years [2]. Data from the Central Statistics Agency (BPS) shows that the number of college graduates who choose to work abroad increases every year. And BPS data for 2024 shows that unemployment in Indonesia rose by 270,737 people, from 7,194,862 in February to 7,465,599 in August, marking challenges in workforce absorption [3]. This phenomenon indicates a crisis of confidence in the future within the country, which is also reflected in the viral social media hashtags such as #KaburAjaDulu, as a form of satire and anxiety of the younger generation regarding the current socio-political situation [4].

The brain drain phenomenon occurs when professionals from developing countries choose to move to developed countries, attracted by more lucrative economic opportunities, a higher quality of life, and better working and research conditions. Furthermore, political uncertainty and the security situation in their home countries are also strong reasons for migration. As a result, countries of origin experience a loss of qualified human resources who should play a vital role in development and economic growth, and face a shortage of skilled workers that can slow national progress [5].

Sentiment analysis is a branch of science that studies opinions, assessments, attitudes, and emotions towards an entity, such as a product, service, organization, individual, issue, event, or topic. Sentiment analysis focuses on opinions that express positive and negative sentiments [6]. Sentiment analysis with natural language processing algorithms in social media helps identify and classify public opinion towards the Brain drain phenomenon [7]. Most research on Brain drain focuses on economic, educational, and migration aspects. For example [8], concluded that the per capita income of the destination country has a negative effect on Indonesia's Brain drain, while the security factor is not significant. [9] found that 49.08% of educational migrants did not return to their areas of origin, influenced by economic, social, and demographic factors. However, studies on public perception and opinion regarding Brain drain on social media are still limited, even though social media is an active space in representing the voice of the community.

The use of the Naive Bayes Classifier algorithm in analyzing socio-political issues, such as the Brain drain phenomenon in Indonesia, is still relatively rare as a research focus. This provides novelty value both in terms of method and application, especially in studying public opinion through social media on national strategic issues. Data will be collected through data crawling from social media X, then processed with cleaning, normalization, and tokenization. The Naive Bayes Classifier algorithm is used to classify public sentiment about the Brain drain phenomenon, ending with model evaluation to measure its accuracy and effectiveness [10].

This study will use one of the quite popular sentiment analysis methods for Twitter data, namely the Naïve Bayes Classifier. This method is a classification technique that is known to be simple but is able to provide quite accurate results in predicting the category (sentiment) of a document, based on the frequency of word appearance in the text. The superiority of this method has been proven in a study [11] that compared its performance with the Support Vector Machine (SVM) method in classifying public opinion regarding the New Normal situation in Indonesia. The results showed that the Naïve Bayes Classifier produced higher accuracy than the SVM method with a training and test data ratio of 70:30 [12]. Therefore, by applying the Naïve Bayes Classifier method in this study, it is hoped that an

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

analysis of public sentiment on Twitter regarding the Brain Drain Phenomenon in Indonesia can be carried out, both related to political policies and personal issues concerning his role as a political figure [13].

The brain drain phenomenon has had a significant impact on Indonesia's development, particularly due to the loss of educated and skilled workers. The growing public opinion surrounding this issue has made sentiment analysis a relevant method for understanding public perception more deeply. This study uses data from social media platform X (Twitter) analyzed using the Naive Bayes Classifier algorithm to classify public sentiment towards the brain drain, thus providing a systematic overview and providing input for formulating more appropriate policies [14].

The novelty of this research lies in its specific focus on the Brain Drain phenomenon in Indonesia through sentiment analysis of social media X, which is still relatively rare as an object of national socio-political research. In terms of data, this study collects and analyzes public opinion related to Brain Drain directly from social media X using data crawling techniques, thus providing relevant and up-to-date data. In terms of methods, this study applies the Naive Bayes Classifier algorithm which is known to be simple but accurate, and has been proven effective in previous studies [15]. This approach is also combined with a complete data preprocessing process, including cleaning, normalization, and tokenization, to improve the quality of the analysis. Another novelty is the application of this model in the context of national strategic issues that have not been widely explored, thus providing new contributions in terms of methods and their application in the analysis of public opinion on socio-political phenomena in Indonesia [16].

However, most previous studies on the brain drain phenomenon in Indonesia have primarily focused on its economic, educational, or migration aspects, without exploring how the public perceives and emotionally responds to this issue through social media platforms. This creates a research gap, as public opinion on social media often reflects real-time sentiment and collective attitudes that cannot be captured through conventional surveys or statistical models. In addition, there is still limited research applying Natural Language Processing (NLP) and machine learning algorithms, particularly the Naïve Bayes Classifier, to analyze socio-political issues such as the brain drain phenomenon in Indonesia. Therefore, the novelty of this study lies in combining text mining techniques with sentiment analysis on social media platform X to identify and classify public opinion regarding the brain drain phenomenon. This approach not only provides methodological innovation but also offers new insights into understanding public perception toward strategic national issues using computational linguistics.

2. RESEARCH METHODOLOGY

2.1 Data Collection Technique

At this stage, data was successfully obtained through a crawling process from social media X which contains public opinion regarding the Brain drain phenomenon, namely the tendency of educated human resources to work or settle abroad. Data collection was carried out using a number of keywords that represent positive and negative sentiments towards the topic [17]. To detect positive opinions, keywords such as: “karier lebih baik di luar negeri”, “kesempatan besar di luar”, “lebih dihargai di luar negeri”, and “gaji layak dan lingkungan mendukung” were used. These phrases reflect views that support or understand a person's decision to work abroad. Meanwhile, negative opinions were identified through keywords such as: “kehilangan talenta terbaik”, “merugikan bangsa”, “SDM unggul kabur”, and “negara tidak mampu mempertahankan tenaga ahli”, which indicate concern or disagreement towards the phenomenon. Data collection was carried out over the past two years, with a total of more than 1000 tweets successfully collected and classified into two categories of positive and negative sentiment [18].

2.2 Data Analysis

The dataset in this study consists of public opinion regarding the Brain Drain phenomenon obtained from social media X through web crawling techniques using the Python language on the Google Colaboratory platform. Data retrieval was carried out using the Snscrape library without an official API, using keywords such as "kerja di luar negeri", "Brain Drain", and "#KaburAjaDulu". The collected data includes tweet content, dates, usernames, and other metadata, then stored in CSV format for sentiment analysis purposes. The data analysis stage in this study begins with a preprocessing process to clean the data from irrelevant elements, such as punctuation, numbers, and common words (stopwords). Next, weighting is carried out using the TF-IDF method before the data is classified based on sentiment using the Naïve Bayes algorithm [19].

2.3 Application of Naïve Bayes Classifier

Before applying the Naïve Bayes Classifier, it is important to explain briefly the theoretical basis of this algorithm. The Naïve Bayes Classifier is a probabilistic classification algorithm based on Bayes’ Theorem, which assumes that the presence of a particular feature in a class is independent of the presence of other features. Despite its simplicity, this algorithm has been proven effective in text classification and sentiment analysis tasks due to its efficiency in handling large datasets and high-dimensional data. This algorithm is widely used in sentiment analysis because it can effectively classify opinions into positive or negative categories with relatively low computational cost. Several studies have also

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

shown that Naïve Bayes often performs comparably or even better than more complex algorithms, such as Support Vector Machine (SVM), especially for short-text data like social media posts.

After weighting the words using the Term Frequency-Inverse Document Frequency (TF-IDF) method, the next stage in this research was to classify the data using the Naive Bayes Classifier algorithm After weighting the words using the Term Frequency-Inverse Document Frequency (TF-IDF) method, the next stage in this research was to classify the data using the Naive Bayes Classifier algorithm [20]. This algorithm was chosen because it is efficient in grouping text data based on the probability of word occurrence. Public opinion data was then classified into positive and negative sentiments towards the Brain Drain phenomenon. This algorithm was chosen because it is efficient in grouping text data based on the probability of word occurrence. Public opinion data was then classified into positive and negative sentiments towards the Brain Drain phenomenon.

Figure 1 below illustrates the overall research flow used in this study. It shows the sequence of processes carried out, starting from data collection through web crawling, data preprocessing, word weighting using the TF-IDF method, sentiment labeling, classification using the Naïve Bayes Classifier, and evaluation of model performance. Each stage is interconnected to ensure the accuracy and validity of the sentiment analysis results.

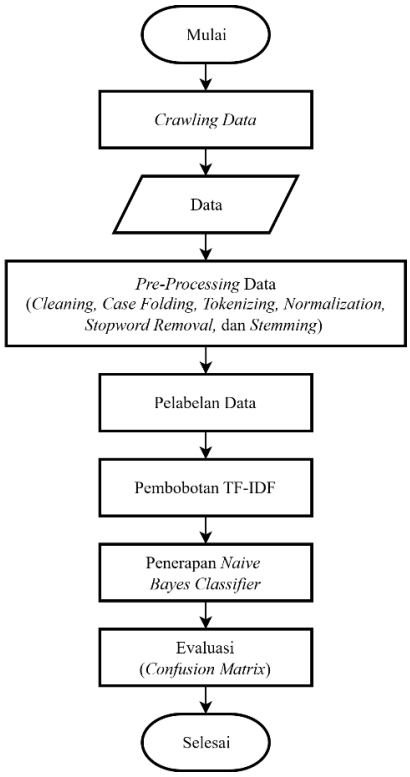

Figure 1. Research Flowchart

As seen in Figure 1, the research begins with data crawling from social media platform X to collect tweets related to the brain drain phenomenon. The data then go through several preprocessing steps to clean and structure the text before entering the TF-IDF weighting and Naïve Bayes classification stages. The final step involves evaluating model performance to determine classification accuracy and reliability.

3. RESULT AND DISCUSSION

3.1 Data Analysis

The research dataset consisted of 1,170 tweets obtained through web crawling on social media platform X using Python in Google Collaboratory. The collected data included text relevant to the brain drain phenomenon, using keywords such as "“kerja luar negeri”, “kabur aja dulu”, “brain drain”, and “SDM Indonesia”. All collected tweets were then saved in CSV format for further systematic processing. Before entering the classification stage, the data underwent a preprocessing process. This stage included cleaning and case folding to standardize the text, tokenizing to break sentences into words, normalization according to the Big Indonesian Dictionary (KBBI), stopword removal to remove meaningless words, stemming to return words to their basic form, and detokenizing to restructure the processed text. This stage is crucial to ensure the data is clean, structured, and ready for use in the weighting and classification processes.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Next, the data was sentimentally labeled using InSetLexicon, a sentiment dictionary capable of distinguishing words with positive and negative connotations. After labels are determined, word weights are calculated using the TF-IDF (Term Frequency-Inverse Document Frequency) method to measure the importance of words within a document. The weighted data is then divided into training data (80%) and testing data (20%), allowing the model to learn from a larger dataset before testing its accuracy.

3.2 Data Collection

Table 1 below presents several examples of tweet data collected through the web crawling process on social media platform X. This sample illustrates the diversity of public opinions related to the brain drain phenomenon, both positive and negative, which serve as the main dataset for this research. Each entry represents an original tweet that contains opinions or expressions about Indonesian professionals working abroad.

Table 1. Sample Comment Data

No.	Full Text
1	Pendidikan yang merata mencegah *brain drain* internal antar daerah. #SekolahRakyatkeren SekolahRakyat KerenBanget
2	Cinta Laura Soroti Fenomena *Brain drain* di Indonesia Ungkap Hal Ini https://t.co/LGrAgmI9nx
...	…
10	@sedaryut73015 @RawiNazri31634 @Dara_Cega Betul *brain drain* memang nyata di Indonesia banyak talenta ahli bahasa Inggris dan Jerman pilih tinggal di luar negeri seperti Jerman atau AS karena peluang kerja dan gaji lebih baik. Dalam analogi bus ini artinya bus perlu upgrade: perbaiki sistem pendidikan ekonomi.

As shown in Table 1, the collected data consists of short text messages from users expressing their perspectives on the brain drain issue. These examples show that the public discourse varies — from opinions supporting career opportunities abroad to concerns over the loss of skilled Indonesian talent. This table serves as an illustration of the raw data before proceeding to the text preprocessing and sentiment labeling stages.

3.3 Text Pre Processing

a. Cleaning and Case folding

The first stage of text preprocessing is the cleaning process, which aims to remove unnecessary elements such as punctuation marks, URLs, special characters, mentions, and other noise from the tweet text. This step ensures that the data used in the next process is clean and consistent. Table 2 below shows several examples of the raw tweet text before and after the cleaning process.

Table 2. Cleaning Stage

No.	Full Text	Cleaning
1	Pendidikan yang merata mencegah *brain drain* internal antar daerah. #SekolahRakyatkeren SekolahRakyat KerenBanget	pendidikan yang merata mencegah *brain drain* internal antar daerah sekolahrakyatkeren sekolahrakyat kerenbanget
2	Cinta Laura Soroti Fenomena *Brain drain* di Indonesia Ungkap Hal Ini https://t.co/LGrAgmI9nx	cinta laura soroti fenomena *brain drain* di indonesia ungkap hal ini
...	…	…
10	@sedaryut73015 @RawiNazri31634 @Dara_Cega Betul *brain drain* memang nyata di Indonesia banyak talenta ahli bahasa Inggris dan Jerman pilih tinggal di luar negeri seperti Jerman atau AS karena peluang kerja dan gaji lebih baik. Dalam analogi bus ini artinya bus perlu upgrade: perbaiki sistem pendidikan ekonomi dan	betul *brain drain* memang nyata di indonesia banyak talenta ahli bahasa inggris dan jerman pilih tinggal di luar negeri seperti jerman atau as karena peluang kerja dan gaji lebih baik dalam analogi bus ini artinya bus perlu upgrade perbaiki sistem pendidikan ekonomi

As shown in Table 2, the cleaning process successfully removed irrelevant characters and symbols that do not contribute to sentiment meaning. After this stage, the tweet text becomes more structured and standardized, making it easier for subsequent steps such as tokenizing, normalization, and stemming to process the data effectively. This cleaning process is crucial to improve the accuracy and quality of the sentiment classification model in later stages.

b. Tokenizing

After the cleaning process, the next stage in text preprocessing is *tokenizing*. Tokenizing is the process of breaking sentences or texts into smaller units called tokens, usually in the form of individual words. This stage is important for transforming unstructured sentences into word-level data that can be analyzed statistically. Table 3 below shows the results of the tokenizing process, where each cleaned text is divided into a list of words.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Table 3. Tokenizing Stage

No.	Cleaning	Tokenize
1	['pendidikan', 'yang', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'antar', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']	['pendidikan', 'yang', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'antar', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']
2	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'di', 'indonesia', 'ungkap', 'hal', 'ini']	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'di', 'indonesia', 'ungkap', 'hal', 'ini']
...	…	…
10	['betul', 'brain', 'drain', 'memang', 'nyata', 'di', 'indonesia', 'banyak', 'talenta', 'ahli', 'bahasa', 'inggris', 'dan', 'jerman', 'pilih', 'tinggal', 'di', 'luar', 'negeri', 'seperti', 'jerman', 'atau', 'as', 'karena', 'peluang', 'kerja', 'dan', 'gaji', 'lebih', 'baik', 'dalam', 'analogi', 'bus', 'ini', 'artinya', 'bus', 'perlu', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi', 'dan']	['betul', 'brain', 'drain', 'memang', 'nyata', 'di', 'indonesia', 'banyak', 'talenta', 'ahli', 'bahasa', 'inggris', 'dan', 'jerman', 'pilih', 'tinggal', 'di', 'luar', 'negeri', 'seperti', 'jerman', 'atau', 'as', 'karena', 'peluang', 'kerja', 'dan', 'gaji', 'lebih', 'baik', 'dalam', 'analogi', 'bus', 'ini', 'artinya', 'bus', 'perlu', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi', 'dan']

As seen in Table 3, each tweet that has undergone the cleaning process is successfully converted into separate tokens representing each word. This allows the system to identify the frequency of word occurrences and their relationships in the dataset. The output of this stage becomes the foundation for further preprocessing such as normalization and stopword removal, which aim to refine the text data for more accurate sentiment classification.

c. Normalization

After tokenizing, the next stage is *normalization*. The normalization process aims to standardize words so that different forms of the same word are written consistently, following proper Indonesian language conventions based on the Big Indonesian Dictionary (KBBI). This step is crucial to avoid variations that may cause errors in sentiment classification. Table 4 below presents examples of tokenized data before and after the normalization process.

Table 4. Data Normalization Process

No.	Tokenizing	Normalization
1	['pendidikan', 'yang', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'antar', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']	['pendidikan', 'yang', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'antar', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']
2	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'di', 'indonesia', 'ungkap', 'hal', 'ini']	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'di', 'indonesia', 'ungkap', 'hal', 'ini']
...	…	…
10	['betul', 'brain', 'drain', 'memang', 'nyata', 'di', 'indonesia', 'banyak', 'talenta', 'ahli', 'bahasa', 'inggris', 'dan', 'jerman', 'pilih', 'tinggal', 'di', 'luar', 'negeri', 'seperti', 'jerman', 'atau', 'as', 'karena', 'peluang', 'kerja', 'dan', 'gaji', 'lebih', 'baik', 'dalam', 'analogi', 'bus', 'ini', 'artinya', 'bus', 'perlu', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi', 'dan']	['betul', 'brain', 'drain', 'memang', 'nyata', 'di', 'indonesia', 'banyak', 'talenta', 'ahli', 'bahasa', 'inggris', 'dan', 'jerman', 'pilih', 'tinggal', 'di', 'luar', 'negeri', 'seperti', 'jerman', 'atau', 'as', 'karena', 'peluang', 'kerja', 'dan', 'gaji', 'lebih', 'baik', 'dalam', 'analogi', 'bus', 'ini', 'artinya', 'bus', 'perlu', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi', 'dan']

As shown in Table 4, the normalization stage successfully transformed inconsistent word forms into standardized ones. For instance, slang words, abbreviations, or informal spellings were converted into their formal equivalents. This ensures that all tokens represent accurate and uniform vocabulary, making the subsequent steps such as stopword removal and stemming more effective in producing high-quality sentiment data.

d. Stopword Removal

After the normalization stage, the next preprocessing step is *stopword removal*. This process aims to remove common words that do not contribute significant meaning to sentiment analysis, such as conjunctions, articles, and prepositions. Eliminating these stopwords helps the model focus only on words that carry emotional or contextual weight. Table 5 below presents examples of tokens before and after the stopword removal process.

Table 5. Stopwords Stage

No.	Normalization	Stopwords
1	['pendidikan', 'yang', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'antar', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']	['pendidikan', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']
2	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'di', 'indonesia', 'ungkap', 'hal', 'ini']	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'indonesia']

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

...	…	…
10	['betul', 'brain', 'drain', 'memang', 'nyata', 'di', 'indonesia', 'banyak', 'talenta', 'ahli', 'bahasa', 'inggris', 'dan', 'jerman', 'pilih', 'tinggal', 'di', 'luar', 'negeri', 'seperti', 'jerman', 'atau', 'as', 'karena', 'peluang', 'kerja', 'dan', 'gaji', 'lebih', 'baik', 'dalam', 'analogi', 'bus', 'ini', 'artinya', 'bus', 'perlu', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi', 'dan']	['brain', 'drain', 'nyata', 'indonesia', 'talenta', 'ahli', 'bahasa', 'inggris', 'jerman', 'pilih', 'tinggal', 'negeri', 'jerman', 'as', 'peluang', 'kerja', 'gaji', 'analogi', 'bus', 'bus', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi']

As seen in Table 5, the stopword removal stage successfully eliminates non-essential words such as “yang,” “di,” “dan,” and other filler words. The resulting text contains only the main keywords that represent the substance of each tweet. This process reduces noise in the dataset and enhances the model’s ability to accurately classify sentiment in the following stages, particularly in stemming and TF-IDF weighting.

e. Stemming

The next stage in the text preprocessing process is *stemming*. Stemming is the process of converting words into their base or root forms by removing affixes such as prefixes and suffixes. This step ensures that different variations of a word are treated as the same term, improving the model’s consistency in analyzing textual data. Table 6 below shows the results of the stemming process performed on the dataset.

Table 6. Stemming Stage

No.	Stopwords	Stemming
1	['pendidikan', 'merata', 'mencegah', 'brain', 'drain', 'internal', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']	['didik', 'rata', 'cegah', 'brain', 'drain', 'internal', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']
2	['cinta', 'laura', 'soroti', 'fenomena', 'brain', 'drain', 'indonesia']	['cinta', 'laura', 'sorot', 'fenomena', 'brain', 'drain', 'indonesia']
...	…	…
10	['brain', 'drain', 'nyata', 'indonesia', 'talenta', 'ahli', 'bahasa', 'inggris', 'jerman', 'pilih', 'tinggal', 'negeri', 'jerman', 'as', 'peluang', 'kerja', 'gaji', 'analogi', 'bus', 'bus', 'upgrade', 'perbaiki', 'sistem', 'pendidikan', 'ekonomi']	['brain', 'drain', 'nyata', 'indonesia', 'talenta', 'ahli', 'bahasa', 'inggris', 'jerman', 'pilih', 'tinggal', 'negeri', 'jerman', 'as', 'peluang', 'kerja', 'gaji', 'analogi', 'bus', 'bus', 'upgrade', 'baik', 'sistem', 'didik', 'ekonomi']

As shown in Table 6, the stemming process successfully transformed words into their root forms, such as changing “pendidikan” to “didik” and “perbaiki” to “baik.” This normalization step reduces redundancy and helps the algorithm recognize semantically similar words as one feature. The stemming results will then be used in the next process, namely detokenization, to reconstruct the cleaned and simplified text before sentiment labeling is conducted.

f. Detokenized Text

The final step of the text preprocessing stage is *detokenizing*. Detokenizing is the process of reconstructing a sequence of tokens (words) back into complete text form after all preprocessing steps such as cleaning, tokenizing, normalization, stopword removal, and stemming have been performed. This stage ensures that the processed text can be read as a coherent sentence, ready for the sentiment labeling process. Table 7 below presents examples of the text before and after detokenization.

Table 7. Detokenized Text

No.	Stemming	Detokenized Text
1	['didik', 'rata', 'cegah', 'brain', 'drain', 'internal', 'daerah', 'sekolahrakyatkeren', 'sekolahrakyat', 'kerenbanget']	didik rata cegah *brain drain* internal daerah sekolahrakyatkeren sekolahrakyat kerenbanget
2	['cinta', 'laura', 'sorot', 'fenomena', 'brain', 'drain', 'indonesia']	cinta laura sorot fenomena *brain drain* indonesia
...	…	…
10	['brain', 'drain', 'nyata', 'indonesia', 'talenta', 'ahli', 'bahasa', 'inggris', 'jerman', 'pilih', 'tinggal', 'negeri', 'jerman', 'as', 'peluang', 'kerja', 'gaji', 'analogi', 'bus', 'bus', 'upgrade', 'baik', 'sistem', 'didik', 'ekonomi']	*brain drain* nyata indonesia talenta ahli bahasa inggris jerman pilih tinggal negeri jerman as peluang kerja gaji analogi bus bus upgrade baik sistem didik ekonomi

As seen in Table 7, the detokenizing process successfully combined individual tokens back into well-structured sentences while maintaining the results of previous preprocessing steps. The resulting text is cleaner and more concise, making it suitable for the sentiment labeling stage that follows. This step marks the transition from raw text processing to sentiment analysis using the InSetLexicon dictionary and Naïve Bayes classification model.

g. Labeling Using InSetLexicon

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

The next step is the data labeling process using the InSetLexicon dictionary, an Indonesian-language sentiment dictionary containing a list of words and their polarity scores. This dictionary was obtained from the GitHub repository at https://github.com/fajri91/InSet. Each word in the tweet is matched against the word list in the dictionary. Then, the polarity scores of these words are added up. If the final score is greater than zero, the tweet is classified as positive. Conversely, if the final score is less than zero, the tweet is categorized as negative.

In this study, the labeling process uses the InSetLexicon dictionary, which contains a list of Indonesian words with their corresponding sentiment polarity scores. Words with positive sentiment values are grouped in the positive lexicon, while those with negative sentiment values are grouped in the negative lexicon. The positive and negative word lists help the system automatically determine the sentiment of each tweet. Table 8 and Table 9 below show examples of words contained in the positive and negative lexicon datasets used in this research.

Table 8. Positive Lexicon

No.	Word	Weight
	Hai	3
	Tetap	3
…	…	…
3610	Orisinal	3

Table 9. Negative Lexicon

No.	Word	Weight
	Isak	-5
	Sakit	-5
…	…	…
3610	Mencoreng	-4

As seen in Table 8 and Table 9, the InSetLexicon dictionary contains a large set of words, each assigned with a specific weight value that reflects its sentiment polarity. Positive words such as “Hai” or “Tetap” have positive weights, while negative words such as “Isak” or “Sakit” have negative weights. These lexicons play a crucial role in the labeling process, allowing tweets to be categorized as positive or negative based on the cumulative score of words they contain. The labeled data from this step then become the input for the Naïve Bayes classification model.

After identifying positive and negative words using the InSetLexicon dictionary, each tweet in the dataset was assigned a sentiment label based on the cumulative polarity score of the words it contains. Tweets with a positive total score were classified as positive, and those with a negative total score were classified as negative. Table 10 below presents examples of tweets that have been labeled automatically according to their sentiment category.

Table 10. Dataset Labeling

No.	**Detokenized_Text**	**Label**
1.	didik rata cegah *brain drain* internal daerah sekolahrakyatkeren sekolahrakyat kerenbanget	Positif
2.	cinta laura sorot fenomena *brain drain* indonesia	Positif
…	…	…
10.	*brain drain* nyata indonesia talenta ahli bahasa inggris jerman pilih tinggal negeri jerman as peluang kerja gaji analogi bus bus upgrade baik sistem didik ekonomi	Negatif

As shown in Table 10, each tweet is assigned a corresponding sentiment label — either “Positive” or “Negative” — based on the sentiment score produced by the InSetLexicon dictionary. This labeling process generates a structured and ready-to-train dataset that serves as input for the Naïve Bayes Classifier model. The resulting labeled data allow the classification algorithm to learn sentiment patterns and accurately categorize unseen tweets in subsequent stages.

Figure 2 below illustrates the distribution of sentiment categories obtained from the labeling process using the InSetLexicon dictionary. This visualization aims to show the proportion between positive and negative sentiments expressed by social media users regarding the brain drain phenomenon.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

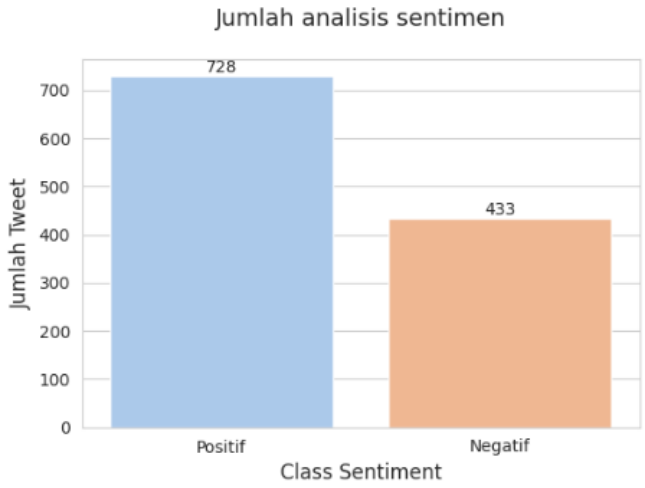

Figure 2. Visualization of the Number of Sentiment Analysis

Based on the sentiment analysis results shown in the figure, of the total 1170 tweets analyzed, there were 728 tweets (62.22%) that were included in the positive category, while 433 tweets (37.01%) were included in the negative category. These results indicate that the majority of public opinion on social media X towards the brain drain phenomenon tends to be positive. This can be interpreted as meaning that the majority of people see the brain drain phenomenon as an opportunity, for example to advance their careers, gain new experiences, or expand job opportunities abroad. Meanwhile, negative opinions are still quite significant, reflecting public anxiety about the potential loss of domestic experts, weak government policies, and risks to the nation's competitiveness in the future. Thus, public perception is polarized, but the dominant tendency is support or acceptance of the brain drain phenomenon.

3.4 Word Weighting Using TF-IDF

3.4.1 Calculating Term-Frequency (TF)

After the data labeling stage, the next step in the sentiment analysis process is word weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The first component of this process is calculating the *Term Frequency (TF)* value, which measures how often a word appears in each document or tweet. A higher TF value indicates that the word is more representative of the document's content. Table 11 below presents examples of TF calculations for several words across three sample documents.

Table 11. TF Table

Word	D1	D2	D3	DF
angka	1	0	0	1
bangsa	0	0	1	1
…	…	…	…	…
tren	2	0	0	1

As shown in Table 11, the TF values reflect the frequency of each term in different documents. For instance, words that appear more frequently, such as "tren," have higher TF values, indicating greater importance in those texts. This information becomes the foundation for calculating the *Inverse Document Frequency (IDF)* in the next stage, which will balance the influence of common versus rare words in the dataset, leading to more accurate TF-IDF weighting for sentiment classification.

Some examples of TF (rounded to 3 decimal places) or normalization:

a. D1: each word appears once $\rightarrow \frac{1}{14} = 0.071$,
   Then, the word "trend" appears twice $\rightarrow \frac{2}{14} = 0.143$
b. D2: each word appears once $\rightarrow \frac{1}{6} = 0.167$
c. D3: each word appears once $\rightarrow \frac{1}{9} = 0.111$

3.4.2 Calculating IDF

$$idf_{(t,D)} = log \frac{Total\ Number\ of\ Documents}{jNumber\ of\ Documents\ from\ TF} \quad (1)$$

After calculating the Term Frequency (TF) values, the next step in the TF-IDF process is calculating the *Inverse Document Frequency (IDF)*. This stage aims to determine the importance of each word across all documents. Words

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

that appear frequently in many documents will have lower IDF values, while rare words will have higher IDF values, indicating greater uniqueness. Table 12 below shows the results of the IDF calculations for several words in the dataset.

Table 12. Calculating IDF

No.	Word	DF	Calculating IDF	IDF
1.	angka	1	log10(3/1) = log10(3)	0.477
2.	bangsa	1	log10(3/1) = log10(3)	0.477
…	…	…	…	…
24.	tren	1	log10(3/1) = log10(3)	0.477

As seen in Table 12, each word's IDF value is obtained from the ratio of the total number of documents to the number of documents containing that word, followed by a logarithmic transformation. For example, words that occur in all documents, such as "drain," have lower IDF scores, while words that appear in only one or two documents have higher IDF scores. These IDF values are then combined with the TF values from the previous table to calculate the final TF-IDF weights, which will be used as features in the sentiment classification process.

3.4.3 Calculating TF-IDF

$$tf - idf_{t,d} = tf_{td} * idf_t \quad (2)$$

a. $tf - idf_{(tren,D1)} = 0.143 \; x \; 0.477 = 0.0682$
b. $tf - idf_{(berita,D2)} = 0.167 \; x \; 0.477 = 0.0797$
c. $tf - idf_{(indonesia,D3)} = 0.111 \; x \; 0.477 = 0.0529$

After calculating the Term Frequency (TF) and Inverse Document Frequency (IDF) values, the final stage in the weighting process is to calculate the combined *TF-IDF* score. This score reflects how important a word is to a specific document in the context of the entire dataset. The TF-IDF value increases proportionally with the frequency of a word in a document but is offset by the frequency of that word across all documents. Table 13 below presents several examples of the calculated TF-IDF values used in this study.

Table 13. Tabel TF-IDF

Word	Document	TF-IDF
tren	D1	0.0682
medsos	D1	0.0338
…	…	…
drain	D1–D3	0.0000

As shown in Table 13, words with higher TF-IDF scores—such as "tren" and "medsos"—indicate terms that are more representative and carry greater weight in determining the sentiment of the text. Meanwhile, words with very low or zero TF-IDF values, like "drain," contribute less to sentiment differentiation. These weighted features serve as the numerical input for the Naïve Bayes Classifier, allowing the model to identify sentiment patterns and make accurate predictions based on the most significant words in each tweet.

3.5 Text Classification Using Naïve Bayes and Calculating Accuracy

3.5.1 Naïve Bayes Classifier

This study uses the Naïve Bayes algorithm to classify sentiments on public opinion related to the brain drain phenomenon in Indonesia which is widely discussed on social media X. The number of data analyzed was 1161 previously 1170 because there was a removal of similar sentiments and tweets that had gone through the pre-processing stage (case folding, stopword removal, stemming, tokenizing) and manual labeling. The data was divided into two parts, namely 80% (928 data) used as training data and 20% (233 data) used as testing data.

After obtaining the TF-IDF weights, the next step is to classify the sentiment using the Naïve Bayes algorithm. In this stage, the dataset is divided into two parts: training data and testing data. The training data are used to train the model to recognize patterns of positive and negative sentiments based on word probabilities. Table 14 below presents several examples of training data that were used in this study.

Table 14. Training Data Sample

Detokenized	Label
kerja luar negeri lebih menjanjikan gaji besar daripada di indonesia	positif
brain drain bikin sdm indonesia kabur aja dulu daripada susah di negeri sendiri	negatif
sdm indonesia terus keluar negeri akhirnya negara kita kekurangan tenaga ahli	negatif

As shown in Table 14, each piece of training data consists of a processed and detokenized tweet along with its sentiment label (positive or negative). These data serve as the basis for the Naïve Bayes algorithm to learn how specific

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

words contribute to each sentiment category. The learned probability distributions from this dataset are then used to classify new, unseen tweets in the testing phase, enabling the evaluation of the model's accuracy and reliability.

a. Calculation of probability values

$$P\ (Class \mid Comentar) = \frac{Number\ of\ Class\ X}{Number\ of\ Comments} \quad (3)$$

1. P (Positif | Comentar) = $\frac{1}{3}$ = 0.333
2. P (Negatif | Comentar) = $\frac{2}{3}$ = 0.667

b. Calculation of Conditional Probability Value (Likelihood):
1. Total words in the Positive class = 10 words
2. Total words in the Negative class = 24 words
3. Number of unique words from all training data = 27 words

$$P\ (Term|\ Class) = \frac{Term\ frequency\ on\ label+1}{Total\ Words\ in\ label+Number\ of\ unique\ features} \quad (4)$$

1. Probability of the word "kerja"

P(kerja | Positif) = $\frac{1+1}{10+27} = \frac{2}{37}$ = 0.054

P(kerja | Negatif) = $\frac{0+1}{24+27} = \frac{1}{51}$ = 0.0196

The next step is to take the test data, which is to classify the test data by multiplying all the probabilities. A higher value represents a new class for the data.

Table 15. Test Data Used

Test Comments	Label (Prediction)
kerja luar negeri itu solusi terbaik saat brain drain	?
SDM indonesia kabur aja dulu cari gaji tinggi	?

As seen in Table 15, each test tweet contains text related to the brain drain phenomenon that has not been used in the training phase. These examples are used to predict sentiment labels based on the probability calculations derived from the Naïve Bayes model. The comparison between predicted and actual labels from this dataset provides the foundation for evaluating the model's accuracy, precision, recall, and F1-score in the subsequent analysis.

c. Calculation of posterior probability values

$$P\ (Comentar \mid Class) = P_{Term\ 1} x \ldots x P_{Term\ n} x\ P\ (Class|Comentar) \quad (5)$$

P (Uji | Positif) = P(positif) x P(kerja|positif) x P(luar|positif) x P(negeri|positif) x P(itu|positif)x P(solusi|positif) x P(terbaik|positif) x P(saat|positif)x P(brain|positif) x P(drain|positif)
= $0.3333 \times (0.054)^3 \times (0.027)^6$
= 0.000000000000020519

P (Uji | Negatif) = P(negatif) x P(kerja|negatif) x P(luar|negatif) x P(negeri|negatif) x P(itu|negatif)x P(solusi|negatif) x P(terbaik|negatif) x P(saat|negatif)x P(brain| negatif) x P(drain|negatif)
= $0.667 \times (0.0196)^6 \times 0.588 \times (0.0392)^2$
= 0.000000000000003427

Based on the results of the posterior probability calculation using the Naïve Bayes method, in the first test data with the sentence "working abroad is the best solution during brain drain", the probability value for the positive class was obtained at 0.000000000000020519, while for the negative class it was 0.000000000000003427. Because the probability value in the positive class is greater than the negative class, the sentence is classified as a positive sentiment. Next is the calculation for the second test data, namely:

P (Uji | Positif) = P(positif) x P(sdm|positif) x P(indonesia|positif) x P(kabur|positif) x P(aja|positif) x P(dulu|positif) x P(cari|positif) x P(gaji|positif) x P(tinggi|positif)
= $0.3333 \times (0.027)^6 \times (0.054)^2$
= 0.0000000000003796

P (Uji | Negatif) = P(negatif) x P(sdm| negatif) x P(indonesia| negatif) x P(kabur| negatif) x P(aja| negatif) x P(dulu| negatif) x P(cari|positif) x

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$P(\text{gaji}|\text{positif}) \times P(\text{tinggi}|\text{positif})$$
$$= 0.667 \times (0.0588)^2 \times (0.0392)^3 \times (0.0196)^3$$
$$= 0.0000000000010488$$

Meanwhile, in the second test data with the sentence *"sdm indonesia kabur aja dulu cari gaji tinggi"*, the probability value for the positive class is 0.0000000000003796, while for the negative class it is 0.0000000000010488. The probability value of the negative class is higher than the positive class, so the sentence is classified as a negative sentiment.

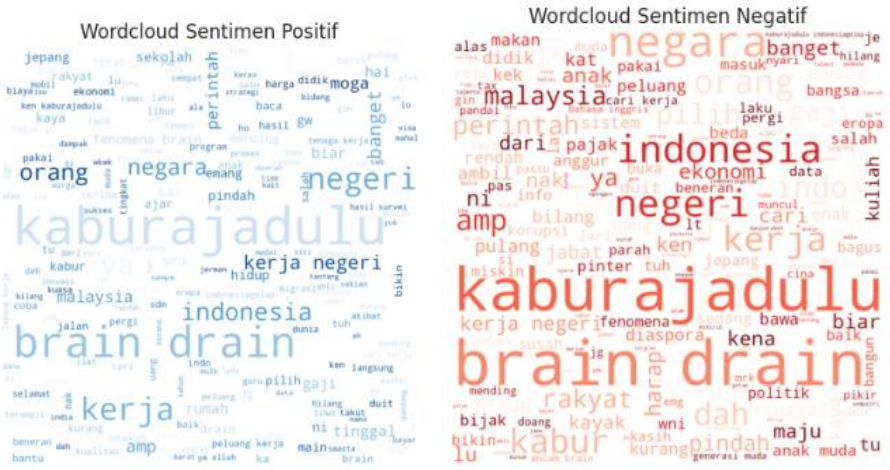

Figure 3. (a) Positive Wordcloud, (b) Negative Wordcloud

As seen in Figure 3, words such as “karier,” “pengalaman,” and “luar negeri” appear more frequently in positive sentiment tweets, reflecting public optimism toward career development and global opportunities. Meanwhile, the negative wordcloud highlights words like “kehilangan,” “talenta,” and “negara,” which indicate concerns about the outflow of skilled professionals and its potential impact on national progress. This visualization supports the quantitative results, showing that both positive and negative perspectives coexist in public discourse about the brain drain issue. After performing sentiment classification using the Naive Bayes algorithm on public opinion regarding the brain drain phenomenon on social media platform X, the next step was to evaluate the model's performance using a confusion matrix. This evaluation aimed to assess the model's ability to accurately differentiate tweets into positive and negative sentiments. The confusion matrix was used to compare the model's predictions with the actual labels on a test dataset of 233 tweets (20% of the total 1,161 tweets).

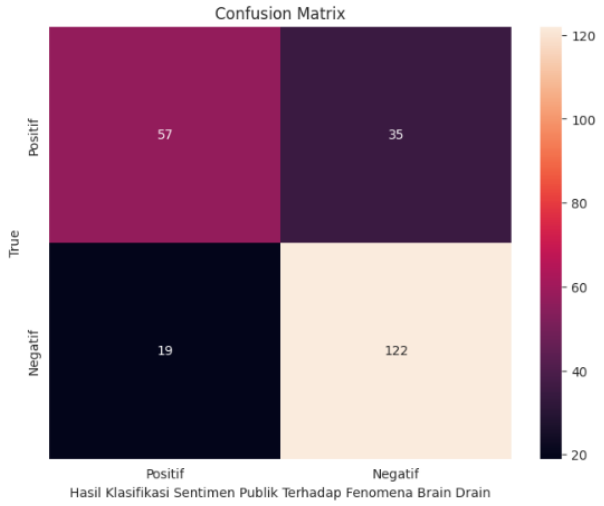

Figure 4. *Confusion Matrix*

As seen in Figure 4, the confusion matrix summarizes the number of correctly and incorrectly classified tweets for each sentiment category. The diagonal cells represent the correctly predicted values, while the off-diagonal cells show misclassifications. From this matrix, evaluation metrics such as accuracy, precision, recall, and F1-score are calculated to quantitatively assess the performance of the Naïve Bayes model. The results of these metrics are presented and discussed in the following figure. The comment accuracy for the dataset is shown in the image below where it is 0.768, precision is 0.75, recall is 0.619 and F1-score is 0.678.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 5,Oktober 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i5.9131
Hal 743-755
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

```
Hasil Akurasi: 0.7682403433476395
              precision    recall  f1-score   support

     Negatif       0.75      0.62      0.68        92
     Positif       0.78      0.87      0.82       141

    accuracy                           0.77       233
   macro avg       0.76      0.74      0.75       233
weighted avg       0.77      0.77      0.76       233
```

Figure 5. Confusion Matrix Results

Based on the evaluation results, the Naïve Bayes model used in this study had an accuracy of 76.8%, meaning the model was able to correctly classify sentiment on most of the test data. A precision value of 75% indicates that the prediction of positive sentiment is quite accurate, while a recall of 61.9% indicates that the model is still less than optimal in capturing all actual positive data. The F1-score value of 67.8% indicates a moderate balance between precision and recall. Thus, the Naïve Bayes model proved quite effective in analyzing public sentiment related to the brain drain phenomenon, although there is still room for performance improvement, particularly in the recall aspect to make the model more sensitive to positive data.

4. CONCLUSION

Based on the research results, sentiment analysis of public opinion regarding the brain drain phenomenon was successfully conducted using the Naïve Bayes Classifier algorithm. The data used came from 1,170 tweets collected from social media platform X (Twitter) containing relevant keywords related to the brain drain topic. Similar data was then reduced to 1,161 items. All data underwent preprocessing stages including cleaning and case folding, tokenizing, normalization according to the Indonesian Dictionary (KBBI), stopword removal, stemming, and detokenizing. The data was then automatically labeled using the InSetLexicon dictionary to identify word polarity, allowing for categorization into positive and negative sentiments. The labeling results revealed that 728 tweets were positive and 433 were negative. The classification process was performed using the Naïve Bayes algorithm, using an 80:20 split between training and test data. The evaluation results using a confusion matrix showed an accuracy of 76.8%, a precision of 75%, a recall of 61.9%, and an F1-score of 67.8%. These evaluation values illustrate that the model has a fairly good performance in classifying, although there are still limitations in the recall aspect in optimally capturing all positive data. In general, the results of this study indicate that public opinion regarding the brain drain phenomenon is mostly positive, which indicates public support or acceptance of overseas job opportunities as a means of self-development. However, negative opinions are also quite significant, indicating public concerns about the potential loss of domestic experts and the long-term impact on national development. Thus, this study successfully captured the dynamics of diverse public opinion on the brain drain issue through a Naïve Bayes-based sentiment analysis approach.

REFERENCES

[1] I. P. Sari, N. F. Al-Phasa, A. I. Muhammad, R. A. Feriansyah, M. Rifqi, and M. Z. Zulfika, "Ketika Talenta Pergi: Menelisik Dampak Brain Drain Terhadap Kemajuan Indonesia," *Arini: Jurnal Ilmiah dan Karya Inovasi Guru*, vol. 2, no. 1, pp. 99–111, Jun. 2025, doi: 10.71153/arini.v2i1.344.

[2] V. Erlisya, Aisyah Aulia, Ferik Clodya, Nursyidah, and Wahjoe Pangestoeti, "FENOMENA #KABURAJADULU: ANALISIS DAMPAK BRAIN DRAIN TERHADAP PEREKONOMIAN INDONESIA," *JURNAL ILMIAH EKONOMI DAN MANAJEMEN*, vol. 3, no. 6, pp. 171–177, Jun. 2025, doi: 10.61722/jiem.v3i6.5083.

[3] M. Sholeh, "Data Jumlah Pengangguran Terbuka di Indonesia Berdasarkan Pendidikan Terakhir Tahun 2024." [Online]. Available: https://data.goodstats.id/statistic/data-jumlah-pengangguran-terbuka-di-indonesia-berdasarkan-pendidikan-terakhir-tahun-2024-zYgTu

[4] K. D. Nathania, "Ramai Tagar Kabur Aja Dulu, Pakar UGM: Bentuk Sikap Kritis dan Sindiran Anak Muda atas Situasi di Tanah Air," Gusti Grehenson. Accessed: Feb. 20, 2025. [Online]. Available: https://ugm.ac.id/id/berita/ramai-tagar-kabur-aja-dulu-pakar-ugm-bentuk-sikap-kritis-dan-sindiran-anak-muda-atas-situasi-di-tanah-air/

[5] C. N. Harjadi, "Fenomena Brain Drain: Keputusan Migran Berpendidikan Tinggi Tinggalkan Negara Asal," GoodStats. Accessed: Apr. 23, 2024. [Online]. Available: https://goodstats.id/article/tren-brain-drain-keputusan-migran-berpendidikan-tinggi-meninggalkan-negara-asal-UJkF8

[6] L. Uliana, A. Desmi, L. A. Widari, Y. Afrillia, and Fadlisyah, "Analisis Sentimen Terhadap Islamophobia Di Twitter Menggunakan Algoritma Decision Tree C4.5," *Jurnal Teknologi Terapan & Sains 4.0*, vol. 2, no. 1, pp. 26–27, 2023.

[7] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 115–122, 2023, doi: DOI : https://doi.org/10.55338/saintek.v4i3.1368.

[8] F. Muslihatinningsih, Zainuri, and E. Santoso, "Brain Drain Indonesia Dan Dampaknya Bagi Indonesia," *JAE: JURNAL AKUNTANSI DAN EKONOMI*, vol. 7, no. 1, pp. 42–52, 2022, doi: DOI: 10.29407/jae.v7i1.17702.

[9] M. D. S. Mustika and N. N. Yuliarmi, "The Determinants of The Brain Drain Phenomenon in Educational Migration Activities," *JEKT JURNAL EKONOMI KUANTITATIF TERAPAN*, vol. 17, no. 1, pp. 76–89, 2024.

[10] M. R. Fajriansyah and M. M. Ir Siswanto, "Analisis Sentimen Pengguna Twitter Terhadap Partai Politik Pendukung Calon Gubernur Di Jakarta Menggunakan Algoritma C4.5 Decision Tree Learning," *SKANIKA*, vol. 1, no. 2, pp. 697–703, 2020.

[11] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, Feb. 2024, doi: 10.57152/malcom.v4i2.1206.

[12] M. Muharrom, "Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas," *Bulletin of Information Technology (BIT)*, vol. 4, no. 4, pp. 430–438, 2023, doi: 10.47065/bit.v3i1.

[13] M. Audina Rambe, I. Komputer, S. dan Teknologi, and U. Islam Negeri Sumatera Utara, "Analisis Sentimen Pengguna Aplikasi Dana Menggunakan Metode Naïve Bayes," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY*, vol. 4, no. 2, 2024.

[14] Yulindawati, S. Lailiyah, and A. Yusnita, "REKOMENDASI PEMILIHAN JUDUL TUGAS AKHIR MENGGUNAKAN METODE NAÏVE BAYES," *Journal of Information System Management (JOISM) e-ISSN*, vol. 5, no. 2, pp. 2715–3088, 2024, doi: 10.24076/joism.2024v5i2.1383.

[15] M. Asfi and N. Fitrianingsih, "Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi," *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, pp. 45–50, 2020, doi: 10.30743/infotekjar.v5i1.2536.

[16] F. Azmi, M. K. Gibran, I. Fawwaz, R. Anugrahwaty, and A. Saleh, "Intelligent Actuator Control in Smart Agriculture through Machine Learning and Sensor Data Integration," *ZERO J. Sains, Mat. dan Terap.*, vol. 9, no. 1, pp. 150–161, 2025, doi: http://dx.doi.org/10.30829/zero.v9i1.24421.

[17] J. Kristovani Siagian, "ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP RENCANA KENAIKAN PPN MENJADI 12% DI MEDIA SOSIAL X DENGAN METODE NAÏVE BAYES," *5th Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 3, no. 2, 2024.

[18] M. R. B. Keliat and M. Ikhsan, "Komparasi Algoritma Support Vector Machine dan Naïve Bayes pada Klasifikasi Jenis Buah Kurma berdasarkan Citra Hue Saturation Value," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 2540–9719, 2025, doi: 10.18495/generic.v14i1.122.

[19] M. K. Gibran, M. I. Rifki, A. H. Hasugian, A. T. A. A. Siahaan, A. Sahputra, and R. Ong, "Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes," *Instal J. Komput.*, vol. 16, no. 03, pp. 210–220, 2024, doi: https://doi.org/10.54209/jurnalinstall.v16i03.240.

[20] Darmastyo Bagas Prabowo, "PREDIKSI HASIL PERTANDINGAN SEPAKBOLA ENGLISH PREMIER LEAGUE DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS DAN NAÏVE BAYES CLASSIFIER," Universitas Islam Indonesia Yogyakarta, Yogyakarta, 2020.