Sentiment Analysis of Society Towards the Childfree Phenomenon (Life Without Children) on Twitter Using Naïve Bayes Algorithm

Siti Nurhaliza[1]\*, Dimas Febriawan[2], Firman Noor Hasan[3]

Department of Informatics, Faculty of Industrial Technology and Informatics[1], [2], [3]

University of Muhammadiyah Prof. Dr. HAMKA

Jakarta, Indonesia

lizzaaxo46@gmail.com@gmail.com[1] , dimas.febriawan@uhamka.ac.id[2] , firman.noorhasan@uhamka.ac.id[3]

***Abstract*—** **Childfree has become a topic that is discussed quite often in almost all regions of Indonesia, especially in urban areas. Not only negative stigma, the choice to live a life without children in Indonesia also carries positive stigma. Views on childfree in Indonesia are highly diverse, considering the many differences in social environments and personal experiences of each individual. In this research, the Naïve Bayes algorithm is used as a sentiment classifier in the form of textual data collected through Twitter using the Rapid Miner. The aim of this research is to analyze and present data regarding public sentiment towards the childfree phenomenon in Indonesia. The results of this research reveal the presence of 319 positive sentiments and 181 negative sentiments, and the accuracy value of the Naïve Bayes algorithm in conducting sentiment analysis on the childfree phenomenon reached 95.02% in sentiment analysis towards the childfree phenomenon. This is attributed to varying perspectives on personal well-being and understanding of lifestyle choices, where some consider that they can attain greater personal well-being by opting for a childfree life. They prioritize focusing on their careers, education, or other activities that they believe can bring them happiness and greater well-being in their lives.**

***Keywords—*** ***Childfree, Naïve Bayes, Sentiment Analysis, Twitter***

# Introduction

Along with the development of technology, particularly on social networks, Indonesia is currently preoccupied with a new phenomenon called “childfree” that emerges amidst community life. Childfree itself is a form of life choice without being blessed with a child or in other words, childfree is a voluntary lifestyle choice to not have children with enviromental considerations [1]. The childfree phenomenon, orginating from foreign cultures is slowly present and developing in Indonesia. However, given the diversity of people’s perspectives and the numerous differences in people’s social life environments, this phenomenon is still widely debated. Some consider this phenomenon is suitable to be implemented in Indonesia, and vice versa, considering the cultural context here in Indonesia. Additionally, there are some individuals within society that still struggle to filter the presence of this phenomenon properly. This is shown by the existence of some people who simply follow this trend without fully understanding it [2].

One of the social platforms where a lot of people share their comments or express their opinions is Twitter. Tweets written by the public are valuable because they provide feedback on a particular matter. Moreover, these tweets can also be used or serve as a basis for understanding the sentiments or perception of the public regarding of what is being discussed. [3]. This is due to Twitter’s expansive user base, exceeding 140 million active users, enabling the dissemination of over 400 million tweets or expressions each day. In addition, Twitter users have the capacity to engage in discussions on trending topics, facilitating the sharing insights and gathering feedback from fellow users [4].

The tweet data will undergo processing using text mining methodology to facilitate comprehensive analysis. Text mining, as a discipline, is geared towards the identification and previous disclosure or revelation of heretofore undiscovered yet potentially valuable insights embedded within unstructured or semi-structured textual data [5]. Given the various tweets contained on Twitter, the analysis process naturally requires a considerable amount of time. Therefore, several methods can be employed with the aim of shortening the time of sentiment analysis process on the tweet data. One such method that can be uses to analyze the sentiment of a matter is the Vaive Bayes method or algorithm [6], [7]. The Naïve Bayes algorithm is based on the Bayes principle which explains that all events have contributions of equal importance or have independence concerning the selection of a particular class. This algorithm is method employed in the text mining process to visualize societal sentiment [8].

Sentiment analysis is a method of identifying a sentiment in the form of text data and how it can be categorized as either positive or negative sentiment. This analytical approach aims to discern and evaluate the emotional tone, attitudes, or opinions that expressed within the text, providing insights into the overall subjective context [9]. In other words, sentiment analysis also referred to as opinion mining. Opinion mining itself is a combination of text mining and natural language processing. The purpose of text mining is to augment textual data originating from specific files and identify the words that represent the content of those files. Consequently, this enables an interconnected analysis between the files [10]. The analysis process will be performed utilizing the Rapid Miner tool, where the Naïve Bayes algorithm will be applied and implemented. In the context of this research, Rapid Miner plays a dual role as both a data analyzer and data mining engine. It serves as a comprehensive tool for not only analyze the data but also serves as an engine for the data mining process within the research framework[11]. Within the Rapid Miner platform, several stages are undertaken to align with the objectives of this research. It starts by extracting and aggregating data in the form of opinions from Twitter. After completing the data crawling phase, several stages are implemented to cleans the data for optimal analysis. This involves important stages, such as tokenization, case folding, and stopword removal to improve the quality of dataset. Following data preprocessing, the next steps involve labeling and classification, an important component in the analytical pipeline [12][13]. Following these stages, the analysis is carried out using Naïve Bayes, which allows classification based on the assumption that each predicted attribute has an independent conditional relationship in each class. Consequently, this algorithm proves to be a highly effective classification method, yielding robust classification outcomes [14]. This method has proven to be effective in classifying and performing better than the other methods. Naïve Bayes is considered better because of its speed and simplicity in classifying text data. [15]. Another advantage provided by the Naïve Bayes algorithm is that there is no need for a large amount of training data, so the text classification process to be predicted can be done easily and quickly. To calculate the classification of this method is considered through probability calculations [16].

Similar research was conducted by Pristiyono entitled "Sentiment Analysis of COVID-19 Vaccine in Indonesia using Naïve Bayes Algorithm" [17] which discusses sentiment analysis of vaccines for COVID-19 in Indonesia by making the Naïve Bayes algorithm as a classifier. This research aims to assess the sentiment of the Indonesian through social network analysis of the COVID-19 vaccine as of January 2021. The results of this research sentiment measurement show that there are more than 56% negative tweets, more than 39% positive tweets, and 1% neutral tweets. Other similar research was also conducted by Charlyn Villavicencio entitled “Twitter Sentiment Analysis towards COVID-19 vaccines in the Philippines Using Naïve Bayes” [18] which discusses the analysis of the Philippines’s sentiment utilizing the Naïve Bayes algorithm regarding the COVID-19 vaccine in Philippines with positive, negative, and neutral polarity. The results of this research present that the Naïve Bayes algorithm produces an accuracy value for 81,77% of analyzing the sentiment of the Philippines towards the COVID-19 vaccine in Philippines. Another similar research was conducted by Lopamudra Dey entitled “Sentiment Analysis of Review Datasets using Naïve Bayes and K-NN Classifier” [19] which discusses evaluating the performance of sentiment classification in terms of accuracy, and also the precision value. Both algorithms used to analyze the topic of the research were Naïve Bayes and K-Nearest Neighbor. The experimental results in this research present that the results with the Naïve Bayes approach are better than the K-NN approach by producing an accuracy value of 80%. The differennce between this research and previous research is the measurement of sentiment, where previous research measured positive, negative, and neutral sentiment. While this research only measures two types of sentiment, namely positive and negative. Additionally, there are also differences in the use of classifier algorithms. In previous research, there was a comparison between the Naïve Bayes and K-Nearest Neighbor algorithms. While this research only uses the Naïve Bayes algorithm as a sentiment analyzer.

In this research, the data collected comes from a platform called Twitter, the social media which contains comments or opinions of the public regarding childfree. After that, the data that has been collected will be carried out to a data cleaning process, subsequently it will be analyzed using the Naïve Bayes algorithm. This research itself has a goal, which is to find out the positive and negative sentiments of the Indonesian regarding the childfree phenomenon using tweets that uploaded in Twitter. Afterward, the data that has been identified whether positive or negative will be analyzed by applying the Naïve Bayes algorithm [20]. This research focuses on public opinion about childfree that found on Twitter as many as 500 data. The process of data retrieval and sentiment analysis is carried out using Rapid Miner tool.

# Methodology

##  Research Stages

Figure 1 is an explanation about the stages of this research that begin with the data collection (crawling) phase, and proceed to the analysis stage of the data that has gone through the previous stages.



Figure 1. Research Stages

1. Crawling Data from Twitter

This stage marks the initial step, namely, the stage of collecting data in the form of opinions from Twitter through Rapid Miner using the Twitter Search operator. The data gathered for this research amounted to 2000 data entries, collected within the timeframe from May 3rd to May 10th, 2023.



Figure 2. Data Collecting Stage (Crawling Data from Twitter)

In Figure 2, there is a data collection process through Rapid Miner which is utilizing two main operators, namely, the Retrieve operator and the Search operator. Since this research aims to gather data from Twitter, the Twitter-connection Retrieve operator employs to connect with Twitter in order to collect the data that relevant to this research. On the other hand, the Search Twitter operator is an operator that utilized to search for tweets on Twitter based on specify keywords, timeframe, and others entered into the parameters of this operator. In this research, the authors collected data using the Search Twitter operator and based on the keyword “childfree” entered into the parameters as shown in Figure 3.



Figure 3. Keywords in Twitter Search Operator

1. Pre-processing

After the data is collected, a pre-processing stage is carried out which includes data cleansing, tokenizing, and filtering (stopword removal). This stage is carried out with the aim that the classification or analysis process can be processed easily. [21]. Below is the operator used for performing the data cleaning process. This operator is the Replace operator, which has a function to cleanse the data from several symbols, numbers, links, and other irrelevant elements. In the data cleaning process of this research, the Replace operator is employed to eliminate terms like “RT”, links, mentions, hashtag symbols, and other miscellaneous symbols.



Figure 4. Data Cleansing

The next stage involves tokenizing, case folding, and stopword filtering. The Tokenize operator shown in Figure 5 below is an operator that utilized as a tool to break or separate text, which may be in the form of paragraphs, sentences, or words that contain symbols into basic words. The Transform Case operator is an operator is an operator that employed to convert all letters in the text into lower case or upper case letters [22]. Meanwhile, the Stopword Filtering operator is an operator used to eliminate words contained in stopwords. The words contained in the stopword list are those deemed to lack meaningful or significant contributions to the analysis process. The three operators shown in Figure 5 are included in the operator, namely, Process Document from Data. Process Document from Data is an operator who is designed to prepare the textual data that will be further analyzed using the chosen algorithm.



Figure 5. Tokenizing, Case Folding, Stopword

The pre-processing stage involves cleansing, tokenizing, and filtering the data to prepare it for analysis. In addition, the removal of duplicate entries ensures that the dataset is free of redundant information, resulting in a more focused and efficient analysis. In this case, the dataset, which was originally 2000 data, has been refined to 500 data that ready for further processing.

TABLE 1. Pre-processing Result

|  |  |
| --- | --- |
| **Before** | **After** |
| RT @fatimah\_rusalka: @skripssweet Orang tua problematik melahirkan dan membesarkan anak yg problematik juga. Tolongnkalian childfree aja da… | orang tua problematik melahirkan dan membesarkan anak yg problematik juga tolong kalian childfree aja da |
| itu pilihan pastinya, dan kita ga bisa nilai sepihak kalau itu baik ataupun tidak, pasti di balik itu semua ada alasan tertentu. Jadi kalau emg pasangan milih childfree selama mereka bahagia itu baik tp kalau beda pendapat itu perlu dibicarakan kembali. https://t.co/4QY9vduvSO | itu pilihan pastinya dan kita ga bisa nilai sepihak kalau itu baik ataupun tidak pasti di balik itu semua ada alasan tertentu jadi kalau emg pasangan milih childfree selama mereka bahagia itu baik tp kalau beda pendapat itu perlu dibicarakan kembali |
| tidak masalah, kalau emang childfree karena masih kurangnya finansial itu lebih baik, daripada anak gak terpenuhi gizi dan lain sebagainya https://t.co/PaJyKv1rna | tidak masalah kalau emang childfree karena masih kurangnya finansial itu lebih baik daripada anak gak terpenuhi gizi dan lain sebagainya |
| @tanyakanrl Aku salah satu yg kelahiran 2000 yg milih childfree bahkan ga berharap sama pernikahan karena rasa trauma masa lalu ??? | aku salah satu yg kelahiran 2000 milih childfree bahkan ga berharap sama pernikahan karena rasa trauma masa lalu |
| @tanyakanrl ngga sih, pasangan yg memutuskan untuk childfree banyak tapi yg masih pgn punya anak jg ga kalah banyak | ngga sih pasangan yg memutuskan untuk childfree banyak tapi masih pgn punya anak jg ga kalah banyak |
| @risasasasamiya @tanyakanrl Di kota berlaku mungkin. Kalo di desa kemungkinan besar ga bakal childfree. Orang baru nikah beberapa bulan aja udah pada berdoa supaya isi padahal ekonomi masih belum stabil. Terus juga takut di nyinyirin kalo ga segera hamil. | di kota berlaku mungkin kalo di desa kemungkinan besar ga bakal childfree orang baru nikah beberapa bulan aja udah pada berdoa supaya isi padahal ekonomi masih belum stabil terus juga takut di nyinyirin kalo ga segera hamil |
| tetap pengen punya banyak anak di tengah gempuran gen z childfree ?? | tetap pengen punya banyak anak di tengah gempuran gen z childfree |
| RT @tsubakiee: @tanyakanrl Gak semua anak 2000an milih childfree nder. Kebanyakan org yg milih childfree tu biasanya dari kalangan yg punya lebih banyak akses ke pendidikan dan informasi dan emg cenderung punya lifestyle yg open-minded dan tbh jumlahnya ga sebanyak itu | anak milih childfree nder kebanyakan milih childfree kalangan akses pendidikan informasi cenderung lifestyle open minded |

1. Splitting and Labeling Data

After the pre-processing stage, splitting the data into 2 parts, namely, training data and testing data. Subsequently, the training data will go through the labeling stage independently, and later the data is used as a reference for the analysis of the testing data which will be conducted directly in Rapid Miner tool. Examples of labeled training data, categorized into positive and negative sentiments, is illustrated in Table 2.

TABLE 2. Training Data Labeling

|  |  |
| --- | --- |
| **Text** | **Label** |
| orang tua problematik melahirkan dan membesarkan anak yg problematik juga tolong kalian childfree aja da | Pos |
| itu pilihan pastinya dan kita ga bisa nilai sepihak kalau itu baik ataupun tidak pasti di balik itu semua ada alasan tertentu jadi kalau emg pasangan milih childfree selama mereka bahagia itu baik tp kalau beda pendapat itu perlu dibicarakan kembali | Pos |
| tidak masalah kalau emang childfree karena masih kurangnya finansial itu lebih baik daripada anak gak terpenuhi gizi dan lain sebagainya | Pos |
| aku salah satu yg kelahiran 2000 milih childfree bahkan ga berharap sama pernikahan karena rasa trauma masa lalu | Pos |
| ngga sih pasangan yg memutuskan untuk childfree banyak tapi masih pgn punya anak jg ga kalah banyak | Neg |
| di kota berlaku mungkin kalo di desa kemungkinan besar ga bakal childfree orang baru nikah beberapa bulan aja udah pada berdoa supaya isi padahal ekonomi masih belum stabil terus juga takut di nyinyirin kalo ga segera hamil | Neg |
| tetap pengen punya banyak anak di tengah gempuran gen z childfree | Neg |

The data presented in Table 3 below is one of the testing datasets that have been partitioned. Later on, the data will be classified by the Naïve Bayes algorithm to determine the sentiment prediction. The prediction results that will include calculated probability values and other relevant metrics.

TABLE 3. Sample of Testing Data

|  |  |
| --- | --- |
| **Text** | **Label** |
| anak milih childfree nder kebanyakan milih childfree kalangan akses pendidikan informasi cenderung lifestyle open minded | ? |

1. Testing Data Classification (Naïve Bayes)

After carefully labeling all the training data, the labeled dataset becomes an important reference point for performing sentiment prediction and testing on the specified test dataset. This process involves applying the Naïve Bayes algorithm within the Rapid Miner platform.



Figure 6. Data Classification (Naïve Bayes)

In Figure 6 there are Split Data, Naïve Bayes, Apply Model, and Performance operators. The Split Data operator is used to split the dataset into 2 parts, namely, training data and testing data. In this research, researchers used a 60:40 ratio in splitting the data, namely 60% for training data and 40% testing data. Subsequently, the Naïve Bayes classifier algorithm which is employed to classify or analyze sentiment. Then, the Apply Model operator is used as machine learning tool to predict the testing data. The final operator, the Performance operator, is an operator that assesses the performance of the machine learning model. To elaborate, these four operators are the operators that used in the process of data preparation, model development, model application on test data, and evaluating model performance in terms of analyzing data. Through the data classification process, the sentiment predictions for the testing data are obtained as shown in Table 4.

TABLE 4. Number of Sentiment Data

|  |  |  |
| --- | --- | --- |
| **Data Type** | **Positive** | **Negative** |
| **Training Data** | 221 | 78 |
| **Testing Data** | 98 | 103 |
| **Total** | 319 | 181 |

1. Evaluation of Test Results

After completing the testing stage and predicting the sentiment, the next step is to calculate the probabilities. There are two main probability calculations at play, namely, prior probability and posterior probability. It is very important to note that the calculation of the prior probability precedes the determination of the posterior probability. This sequential process is fundamental as it establishes the foundational probabilities required for subsequent analysis and ensures a comprehensive understanding of the sentiment prediction results.

1. Prior Probability

Calculation of prior probabilities is essential for each class present in the dataset that employed in this research. This approach is designed to reduce or mitigate bias in classification by taking into account for the distribution of classes, ranging from frequently observed to less frequent occurrences. The prior probability calculation generates both positive and negative class probabilities, derived from the identification process related to the childfree phenomenon. This ensures a balanced consideration of class occurrences and increases the robustness of the subsequent analysis [23].

* Positive Prior Probability

$\left(P\left(Positive\right)\right)=\frac{Number of Positive Data}{Total Amount of Data}$ (1)

* Negative Prior Probability

$\left(P\left(Negative\right)\right)=\frac{Number of Negative Data}{Total Amount of Data}$ (2)

1. Posterior Probability

After obtaining the prior probabilities, the subsequent stage involves the calculation of posterior probability. The purpose of calculating these posterior probabilities is to ascertain the class for new cases identification. This step becomes relevant when there are newly cases that identified during the data processing stage in this research. The calculation of posterior probabilities in this research is essential for extending the model's capability to accurately classify and handle new cases in the dataset. [24].

* Positive Posterior Probability

$Posterior Pos=Class Positive ×(P(Positive)$ (3)

* Negative Posterior Probability

$Posterior Neg=Class Negative ×(P(Negative)$ (4)

Beside of the probability values, the values of accuracy, precission, and recallwere also analyzed and calculated. These values only can be calculated by the first calculating the confusion matrix that listed in Table 5 and shown in Figure 7.

TABLE 5. Confusion Matrix

|  |  |
| --- | --- |
| **Prediction** | **True** |
| **Positive** | **Negative** |
| **Positive** | 89 (TP) | 1 (FN) |
| **Negative** | 9 (FP) | 102 (TN) |



Figure 7. Confusion Matrix Visualization

The confusion matrix shown in Table 5 and Figure 7 above contain a statement of the classification of the amount of data that is correctly predicted as positive sentiment or True Positive (TP), incorrectly predicted as positive sentiment or False Positive (FP), correctly predicted as negative sentiment or True Negative (TN), and incorrectly predicted as negative sentiment or False Negative (FN). Later, the confusion matrix will be used to calculate the accuracy value, recall or True Positive Rate (TPR), precision or Positive Predictive Value (PPV), True Negative Rate (TNR), and Negative Predictive Value (NPV).

The accuracy value represents the percentage of testing data whose class has been successfully indentified or classified correctly by the system based on their original class [25]:

$Accuracy=\frac{TP+TN}{TP+FN+FP+TN}$ (5)

The recall value, also known as True Positive Rate (TPR) which has another name, namely, the sensitivity value represents the percentage of the system’s success in classifying positive class as positive class. The formula for calculating the recall value (TPR) is given in equation (6) below [25]:

$Recall (True Positive Rate)=\frac{TP}{TP+FN}$ (6)

The precision value, also known as Positive Predictive Value (PPV), represents the percentage of data that predicted as a positive class by the classification algorithm which is the actual positive data of all those predicted as positive class. The formula for calculating the precision value (PPV) is given in equation (7) below [25]:

$Precision (Positive Predictive Value)=\frac{TP}{TP+FP}$ (7)

The True Negative Rate (TNR) value, which has another name, namely, the specificity value, represents the percentage of the system’s success in classifying negative class as negative class correctly. The True Negative Rate (TNR) value calculation formula is given in equation (8) below [26]:

$True Negative Rate=\frac{TN}{FP+TP}$ (8)

The NPV or Negative Predictive Value represents the percentage of data that predicted as a negative class by the classification algorithm that is correctly negative data out of all class that predicted as negative class. The formula for calculating the Negative Predictive Value (NPV) is given in equation (9) below [26]:

$Negative Predictive Value=\frac{TN}{FN+TN}$ (9)

# Result and Analysis

In this section, the results of the research and various tests are presented. The sentiment analysis of the data using Naïve Bayes algorithm is followed by the calculation of probabilities. The probability calculations that employed in this research include prior probability and posterior probability. After calculating the values for both prior and posterior probability values, furthermore the accuracy, recall (TPR), precision (PPV), TNR, and NPV values are also calculated by utilizing the confusion matrix as a reference in performing the five calculations after the probability calculations.

To calculate both probability values, namely, prior and posterior probabilities, it is necessary to identify the most frequently occurring words from the data collected. Below is the wordcloud visualization of the most commonly appearing words in the data collected for this research:



Figure 8. Wordcloud Visualization of the Most Commonly Appearing Words

From the wordcloud presented in Figure 8 above, three frequently occuring words will be selected as examples and references for calculating both prior and posterior probabilities. These selected words include "anak" which appears 125 times, "nikah" which appears 37 times, and "pasangan" which appears 30 times. The sentiment examples that chosen for the calculation of prior and posterior probabilities are shown in Table 6 below.

TABLE 6. Data Classification

|  | **Anak** | **Nikah** | **Pasangan** | **Label** |
| --- | --- | --- | --- | --- |
| **P2** | 1 | 0 | 0 | Pos |
| **P62** | 0 | 0 | 1 | Pos |
| **P69** | 1 | 0 | 0 | Pos |
| **P99** | 0 | 1 | 0 | Pos |
| **N36** | 1 | 0 | 1 | Neg |
| **N48** | 0 | 1 | 0 | Neg |
| **N67** | 1 | 0 | 0 | Neg |
| **U169** | 1 | 0 | 0 | ? |

Here is the calculation of the prior and posterior probabilities of some data that has been classified:

1. Prior Probability

$\left(P\left(Positive\right)\right) =\frac{Number of Positive Data}{Total Amount of Data}=\frac{4}{7}=0,57$

$\left(P\left(Negative\right)\right)=\frac{Number of Negative Data}{Total Amount of Data}=\frac{3}{7}=0,42$

**Class Positive (P(U169|Positive))**

= P(anak=2|Pos)$ × $P(nikah=1|Pos)$ × $P(pasangan=1|Pos)

= 0,5$ × $0,25$ × $0,25

= 0,03125

**Class Negative (P(U169|Negative))**

= P(anak=2|Neg)$ × $P(nikah=1|Neg)$ × $P(pasangan=1|Neg)

= 0,66$ × $0,33$ × $0,33

= 0,07187

1. Posterior Probability

**Posterior Positive** = (P(U169|Positive)$×$(P(Positive))

 = 0,03125$×$0,57

 = 0,0178125

**Posterior Negative** = (P(U169|Negative)$×$(P(Negative))

= 0,07187$×$0,42

= 0,0301854

After calculating these probability value, it can be classified that the 169th testing data falls into the negative sentiment category. This is because the resulting negative posterior probability value is greater than the positive posterior probability.

TABLE 7. Classified Testing Data

|  |  |
| --- | --- |
| **Text** | **Label** |
| anak milih childfree nder kebanyakan milih childfree kalangan akses pendidikan informasi cenderung lifestyle open minded | Negative |

From Table 4, the confusion matrix, calculations are made for the accuracy, recall (TPR), precision (PPV), TNR, and NPV values. Table 7 shows an example of classified testing data that classified as negative sentiment based on the calculation of the both probabilities value, namely, prior probability and posterior probability. The accuracy value calculation is carried out with the aim of assessing the extent to which the Naïve Bayes classification algorithm can correctly predict the entire testing dataset. Furthermore, the calculation of the recall value or True Positive Rate (TPR) is done to assess how far the Naïve Bayes classifier algorithm can detect all positive sentiments correctly. Subsequently, the calculation of the precision value or Positive Predictive Value (PPV) is calculated to measure how far the Naïve Bayes classifier algorithm can provide positive predictions accurately. Following that, the calculation of the True Negative Rate (TNR) value is used to assess how far the Naïve Bayes algorithm can accurately detect all negative sentiments. Lastly, the calculation of the Negative Predictive Value (NPV) value is calculated to evaluate how accurately the Naïve Bayes classification algorithm can provide negative predictions. Below is the calculation for accuracy, recall, precision, TNR, and NPV values:

* **Accuracy**

 $Accuracy=\frac{89+102}{89+9+1+102}$

 $Accuracy=0,9502 or 95,02\%$

* **Recall (True Positive Rate)**

 $Recall=\frac{89}{89+1}$

 $Recall=0,9889 or 98,89\%$

* **Precision (Positive Predictive Value)**

 $PPV=\frac{89}{89+9}$

 $PPV=0,9082 or 90,82\%$

* **TNR (True Negative Rate)**

 $TNR=\frac{102}{9+102}$

 $TNR=0,9189 or 91,89\%$

* **NPV (True Negative Rate)**

 $NPV=\frac{102}{102+1}$

 $NPV=0,9903 or 99,03\%$

From these calculations it can be explained that the Naïve Bayes algorithm operates effectively in analyzing the sentiment. This is evidenced by the calculation of the accuracy value which reaches 95.02%, this means that the algorithm or classification model has a high level of accuracy in predicting positive and negative sentiments correctly and accurately. Furthermore, the results of the calculation of the recall (TPR) and TNR values are 98.89% and 91.89%, respectively. This implies that the recall (TPR) and TNR values clarify that the classification algorithm or model is able to detect positive and negative sentiments correctly. Subsequent calculation is the calculation of the precision (PPV) and NPV values, yielding 90.82% and 99.03%, respectively. From these calculations, it can be seen that the classification algorithm or model indicates that each predicted positive and negative sentiment is correct or accurate.



Figure 9. Calculation Results on Rapid Miner

# Conclusion

From the results of testing and analysis that has been carried out in this research on sentiment analysis of Indonesian opinions about the childfree phenomenon whose data is drawn from Twitter in the amount of approximately 2000 data, which then the data goes through several stages of cleaning to make it easier to analyze, resulting in a dataset of 500 data, and finally the data is analyzed using the Naïve Bayes classifier algorithm. It can be concluded that the Naïve Bayes classification algorithm applied in this research is considered effective in analyzing or classifying the sentiment of public opinions about childfree whose data is collected from Twitter. The results of the accuracy value calculated from testing the evaluation of the Naïve Bayes classification algorithm or model of community sentiment about childfree reached 95.02%. This indicates that the Naïve Bayes classification algorithm can predict the sentiments or opinions of the society about the childfree phenomenon correctly and accurately. This is because the high accuracy value suggests the effectiveness of the algorithm. In addition, it can also be concluded that public opinion about the childfree phenomenon is more inclined towards those who respond negatively than positively. This can be proven by the positive sentiments that outnumber the negative ones with the total number of sentiments being 319 sentiments, and the negative sentiments totaling 181 sentiments. Thus, the conclusion regarding people's views in Indonesia regarding the childfree phenomenon dominantly leads to a positive response. This is inluenced by perspectives regarding the approach to personal well-being and understanding of lifestyle choices, where some of individuals consider that they can achieve greater personal well-being by choosing to be childfree. They prefer to focus on career, education, or other activities that provide them more happiness and prosperity to their lives.

# REFERENCES

[1] E. Nakkerud, “Ideological Dilemmas Actualised by the Idea of Living Environmentally Childfree,” *Hum. Arenas*, no. 0123456789, 2021, doi: 10.1007/s42087-021-00255-6.

[2] D. Siregar, F. Ladayya, N. Z. Albaqi, and B. M. Wardana, “Penerapan Metode Support Vector Machines ( SVM ) dan Metode Naïve Bayes Classifier ( NBC ) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter,” vol. 7, no. 1, pp. 93–104, 2023.

[3] P. W. Ratiasasadara, S. Sudarno, and T. Tarno, “Analisis Sentimen Penerapan Ppkm Pada Twitter Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi-Square,” *J. Gaussian*, vol. 11, no. 4, pp. 580–590, 2023, doi: 10.14710/j.gauss.11.4.580-590.

[4] G. A. A.Jabbar Alkubaisi, S. S. Kamaruddin, and H. Husni, “Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers,” *Comput. Inf. Sci.*, vol. 11, no. 1, p. 52, 2018, doi: 10.5539/cis.v11n1p52.

[5] F. N. Hasan, F. Sidik, and P. Afikah, “Sentiment Analysis of Community Response on Cooking Oil Price Increase Policy with Naïve Bayes Classifier Algorithm,” *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 71–76, 2022, [Online]. Available: https://inacl.id/journal/index.php/jlk/article/view/99.

[6] M. Fahmi, A. Puspita, and Y. Yuningsih, “Sentiment Analysis of Online Gojek Transportation Services on Twitter Using the Naïve Bayes Method,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 8, no. 2, pp. 84–90, 2023, doi: 10.33480/jitk.v8i2.4004.

[7] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, “Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance,” *J. Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: 10.14710/j.gauss.v11i2.35472.

[8] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.

[9] R. Novendri, A. S. Callista, D. N. Pratama, and C. E. Puspita, “Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes,” *Bull. Comput. Sci. Electr. Eng.*, vol. 1, no. 1, pp. 26–32, 2020, doi: 10.25008/bcsee.v1i1.5.

[10] R. Sari and R. Y. Hayuningtyas, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website,” *Indones. J. Softw. Eng.*, vol. 5, no. 2, pp. 51–60, 2019, doi: 10.31294/ijse.v5i2.6957.

[11] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, “Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.

[12] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier,” *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: 10.26418/jlk.v5i1.79.

[13] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.

[14] N. S. Wardani, A. Prahutama, and P. Kartikasari, “Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial,” *J. Gaussian*, vol. 9, no. 3, pp. 237–246, 2020, doi: 10.14710/j.gauss.v9i3.27963.

[15] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, “Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier,” *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012024.

[16] J. LING, I. P. E. N. KENCANA, and T. B. OKA, “Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square,” *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014, doi: 10.24843/mtk.2014.v03.i03.p070.

[17] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, “Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021, doi: 10.1088/1757-899x/1088/1/012045.

[18] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, “Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes,” *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.

[19] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, “Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier,” *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016, doi: 10.5815/ijieeb.2016.04.07.

[20] Z. Bokaee Nezhad and M. A. Deihimi, “Twitter sentiment analysis from Iran about COVID 19 vaccine,” *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 16, no. 1, p. 102367, 2022, doi: 10.1016/j.dsx.2021.102367.

[21] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asiimetrik J. Ilm. Rekayasa Inov.*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asiimetrik.v4i1.3577.

[22] I. R. Afandi, F. H. Noor, A. A. Rizki, N. Pratiwi, and Z. Halim, “Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes,” *Jlk*, vol. 5, no. 2, pp. 63–70, 2022, [Online]. Available: https://t.co/2HAdwg1drL.

[23] Duwi Cahya Putri Buani, “Optimasi Algoritma Naïve Bayes dengan Menggunakan Algoritma Genetika untuk Prediksi Kesuburan (Fertility),” *J. Evolusi*, vol. 4, no. 01, pp. 55–64, 2016, doi: 10.31294/evolusi.v4i1.3397.

[24] N. Nuraeni, “Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier: Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC,” *J. Tek. Komput.*, vol. 3, no. 1, pp. 9–15, 2017, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/jtk/article/view/1337.

[25] Y. I. Kurniawan and T. I. Barokah, “Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor,” *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: 10.33557/jurnalmatrik.v22i1.843.

[26] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.