Journal of Advanced Health Informatics Research (JAHIR)
Vol. 1, No. 2, August 2023, pp. 83-94
ISSN: 2985-6124, DOI: 10.59247/jahir.v1i2.44

Understanding User Sentiment: Analysis of SATUSEHAT Application Reviews on Google Play Store

Rian Ardianto[1], Hamzah M. Marhoon [2]
[1]*Universitas Harapan Bangsa, Purwokerto, Indonesia*
[2] *Al-Nahrain University, Jadriya, Baghdad, Iraq*

ARTICLE INFO

Article history:

Received July 28, 2021
Revised August 28, 2021
Published September 28, 2021

Keywords:

SatuSehat;
Text Mining;
Sentiment Analysis;
Google Play Store;
Naïve Bayes;

ABSTRACT

This study looks into user opinions on the SATUSEHAT Mobile application, formerly known as PEDULILINDUNGI until being renamed by the Ministry of Health on March 1, 2023. Textual information relating to various issues is analyzed and manipulated using Data Mining methods, with an emphasis on the SATUSEHAT application. The goal is to do sentiment analysis on user opinions derived from Play Store reviews, with a focus on positive and negative comments. The Naive Bayes Classifier (NBC) is used in the study to categorize user reviews from the Google Play Store's SATUSEHAT application. In compared to earlier research, the Naive Bayes technique outperforms them, obtaining an accuracy of 92%. This outperforms previous methods; for example, prior study using Naive Bayes achieved 80% accuracy. Furthermore, when paired with other test data characteristics, Naive Bayes reached 87% accuracy, while integration with N-Gram yielded 89%. This study reveals the Naive Bayes algorithm's higher performance when dealing with unbalanced data proportions, showing the possibility for additional investigation. By adding bigger and more diverse datasets, accuracy might be improved. Furthermore, improving preprocessing stages may help to correct misspelled words and align them with standard forms. The results demonstrate the effectiveness of the Naive Bayes method combined with the TF-IDF Vectorizer for sentiment analysis in the SATUSEHAT application. Because of its great accuracy and flexibility to skewed data proportions, this algorithm is a promising choice for further research. Larger datasets and better preprocessing are two possible areas for development.

Corresponding Author:
Rian Ardianto, Universitas Harapan Bangsa, Purwokerto, Indonesia
Email: rianardianto@uhb.ac.id

1. INTRODUCTION

The Covid-19 pandemic has caused significant changes in Indonesia [1], and the rapid development of software and computer applications during this pandemic era has led the government to develop a mobile health application called SATUSEHAT Mobile, formerly known as PEDULILINDUNGI, officially renamed by the Ministry of Health on March 1, 2023 [2]. This program monitors the spread of the Covid-19 virus, provides vaccination information, and identifies the zones of different areas. It also allows citizens to exchange their location data for travel purposes, allowing Covid-19 patients to follow their contact history [3]. Users of this program receive notifications when they are in a Covid-19 red zone, which is defined as an area where positive Covid-19 instances or patients under surveillance have been recorded [4].

Sentiment analysis is used in this study to examine the opinions of SATUSEHAT application users based on user reviews acquired from the Play Store. Sentiment analysis is a kind of opinion mining, and this study focuses on both positive and negative opinions. Data Mining methods are used in the analysis to analyze and modify textual data connected to certain objects or entities, such as services, goods, public personalities, events, or selected subjects [5]. The researchers picked the SATUSEHAT application as the target of investigation in this example.

Mardiana investigated five Data Mining methods in her 2019 study: Naive Bayes, Neural Network, Decision Tree, Support Vector Machine, and k-Nearest Neighbor. The results revealed that the Naive Bayes method fared the best in textual data processing [6]. As a result, the Naive Bayes method is used for comparison in this study, with an extra comparison for data weighting utilizing the Term Frequency-Inverse Document Frequency Vectorizer inside the algorithm. In her 2019 study. Ni Putu used Naive Bayes to assess Facebook sentiment and ran seven accuracy tests with varying sample sizes to discover the optimum data for sentiment analysis in that dataset [7]. Fitriyani, in her 2021 research, used Support Vector Machine to evaluate the sentiment of the Sambara application and ran five tests with varied sample sizes [8]. The researcher used the same method in this study to identify the appropriate sample size for sentiment analysis using a dataset collected from user evaluations of SATUSEHAT on Google Play Store.

The appearance of the SATUSEHAT application has attracted public reactions and criticism. Each person has their own views and beliefs, some in favor and some opposed to this program. Given the current challenges, a solution, such as sentiment analysis of public comments, is required to obtain information on the feelings associated with the Indonesian government's SATUSEHAT application. This research will assist the Indonesian government in addressing the issues encountered by application users. There are several approaches or procedures that may be used in the process of retrieving information. Document clustering is accomplished by labeling document categories. Document classification requires a significant amount of time. As a result, a mechanism for quick and precise document classification or grouping is required. Reverend Thomas Bayes pioneered the Nave Bayes classification. The Nave Bayes technique was first used between 1702 and 1761. According to Lewis, Hand, and Yu, Naive Bayes (also known as basic Bayes) is a relatively basic and very successful strategy for categorization learning [9][10]. Meanwhile, Kononenko and Langley define Nave Bayes as a probabilistic label of class data or as a labeled class attribute [11][12]. Nave Bayes provides various advantages, according to Hamzah, including a simple method, quicker calculation, and high accuracy [13]. The Nave Bayes Method [14] is one of the most successful and efficient classification methods, according to Zaidi. Furthermore, this approach comes under machine learning methods with probabilistic answers. It is also noted that the Nave Bayes approach can execute text categorization in huge datasets at fast speed [15]. For this study, the researcher will rely on user evaluations and comments from the Google Play Store's SATUSEHAT application. The Nave Bayes Classifier (NBC) method will be employed in this study.

2. METHODS

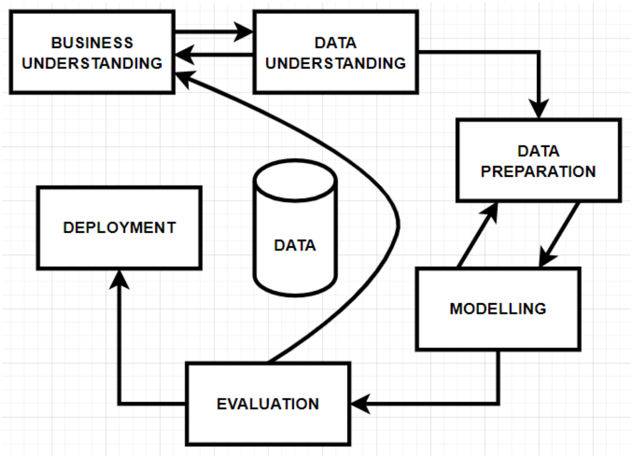

Fig. 1. Model CRISP-DM

This study adopts the CRISP-DM approach (Cross-Industry Standard for Data Mining), as shown in Figure 1. Business Understanding, Understanding, Data Preparation, Data Modeling, Evaluation, and Deployment are the CRISP-DM phases [3].

2.1. Business Understanding

The emphasis in the business understanding stage is on grasping the study object. Understanding the research object is accomplished in this study by gathering information from the Google Play Store using a scraper tool on Google Collab, especially targeting the SATUSEHAT application. This step is motivated by the fact that the evaluations offered often consist of textual information in digital media, which is grouped based on the conversation content in each remark category. Online media is more than just a place to read story headlines; it can also be used to monitor developing concerns and even measure public health electability. Sentiment analysis is used to find a categorization approach that can help determine positive and negative comments on news stories. A understanding of the best classification algorithms is also formed during this step to help further data processing.

2.2. Data Understanding

The procedure of getting raw data based on the relevant properties is carried out at the data understanding step. The data is gathered from the Google Play Store website at the following URL: https://play.google.com/store/apps/details?id=com.telkom.tracencare&hl=en-ID, under the program SATUSEHAT. The information is derived from reviews with scores ranging from 1 to 5. The Google Play Store yielded a total of 10,000 primary data reviews. Following data collecting, a data purification procedure is carried out, yielding a total of 9,303 reviews. This study makes use of review data in Indonesian.

2.3. Data Preparation

The data preparation stage is a step that involves the data preparation process, with the goal of obtaining clean and ready-to-use data for research purposes. The text pre-processing step will be carried out at the early stage of text mining, with the researcher utilizing the Google Collab tools. Several text pre-processing techniques will be performed on the review dataset at this point, including filtering, labeling, cleaning, case folding, stop word removal, tokenizing, and stemming. The topic of these stages will be expanded on in the next section.

2.4. Modeling

This is the step in which data mining strategies are chosen by deciding the algorithm to be employed. This study employs modeling tools in accordance with the established approach, and the tool of choice is Google Collab. As its model, the study used a classification technique, especially the Nave Bayes algorithm. After evaluating each model, the best accuracy value for each method is obtained by classifying positive and negative review articles.

2.5. Evaluation

The purpose of the assessment stage is to determine the utility of the model developed in the preceding modeling step. This study applies an evaluation stage with classification on review data of the SATUSEHAT application on the Google Play Store, aided by TF-IDF weighting, with a data set of 9,303 reviews of the SATUSEHAT application. During the testing phase, the data is separated into 80% for training and 20% for testing.

2.6. Deployment

The deployment step is used to develop an implementation model within a tool that may be written in a variety of computer languages. The outcomes of the testing and assessment procedures were used as reference data in the development of this implementation model. The deployment employed in this study is the implementation stage of the used method, with accuracy numbers that may be used for future development in the usage of the SATUSEHAT application. Furthermore, it aids appropriate government agencies in assessing the operation of capabilities inside the application, such as tracking to prevent the spread of Covid-19.

2.7. Weighting Word

Weighting The practice of giving values to each characteristic based on their relevance and effect on categorization results is known as wording. Based on the minimal weight obtained for each feature, these values can be used as a foundation for feature selection. The TF-IDF (Term Frequency-Inverse Document Frequency) approach is used for weighting. The TF-IDF algorithm is one of the algorithms used in text mining feature weighting approaches [16]. The following equation is used to compute the weight of each word in order to calculate TF-IDF (1):

$$W_{i,j} = \frac{n_{i,j}}{\sum_{j=1}^{p} n_{i,j}} \log_2 \frac{D}{d_j} \tag{1}$$

Explanation:

$W_{i,j}$ = TF-IDF weighting for term “j” in document “i”

$n_{i,j}$ = Number of occurrences of term “j” in document “i”

p = Total number of terms formed

$\sum_{j=1}^{p} n_{i,j}$ = Total occurrences of all terms in document “i”

D = Total number of documents

d_j = Number of documents containing term “j”

2.8. Data processing

The Naive Bayes algorithm is a classification method based on statistical data that is used to categorize the likelihood of membership in a class. When used to vast amounts of data, Naive Bayes displays good accuracy and speed [4]. The Naive Bayes formula in its most generic form is as follows (2):

$$P(X) = \frac{P(H)P(H)}{P(X)} \tag{2}$$

Explanation:

X = Data with an unknown class

H = Hypothesis that data X belongs to a specific class

$P(H|X)$ = Probability of hypothesis H given condition X

$P(H)$ = Probability of hypothesis H

$P(X|H)$ = Probability of data X given hypothesis H

The Confusion Matrix is a matrix with true positive and false positive prediction values. The levels of recall, precision, and accuracy will be examined during accuracy testing [5]. Precision measures the system's capacity to locate the most relevant ranks and is defined as the percentage of documents returned that are actually relevant to the query. Recall measures the system's capacity to locate all relevant items in the document collection and is defined as the percentage of relevant documents to the query that are returned. The accuracy ratio is the ratio of properly recognized cases to the total number of cases, whereas the error rate is the percentage of mistakenly identified cases out of the total number of instances [6]. The four components of the confusion matrix are shown in Table 1, which are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) [17].

Table 1. Example Confusion Matrix

Confusion Matrix	**Prediction**	
	Positive	**Negative**
Actual Positive	TP	FP
Actual Negative	FN	TN

The parameters TP, FP, TN, and FN are utilized to determine the values of accuracy, precision, and recall [18]. These parameters are utilized to assess the outcomes of the tested model in equations (3), (4), and (5). The following equation is used to calculate the accuracy value:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} x100\% \tag{4}$$

$$Recall = \frac{TP}{TP + FN} x100\% \tag{5}$$

3. RESULTS AND DISCUSSION

As noted in the data comprehension section above, this study uses data taken from evaluations of the SATUSEHAT application in the Google Play Store. The total number of reviews collected is 10,000. Following that, data purification is conducted in the initial step, and each remark is categorized as good or negative. This initial step yields a total of 9,303 labeled reviews, which will be the dataset used in this investigation. As this research falls within the domain of text mining, preliminary steps must be taken in processing the data to obtain a model suitable for this research case, which is sentiment analysis of the SATUSEHAT application using the Nave Bayes classification algorithm with Google Collab as the tool.

3.1. Pre-processing

The discussion at this point focuses on the first step of preparing the dataset for classification using the Nave Bayes method. For the text dataset of reviews, multiple pre-processing techniques are used in this study. The stages are as follows:

3.1.1. Filtering Data

The discussion at this point focuses on the first step of preparing the dataset for classification using the Nave Bayes method. For the text dataset of reviews, multiple pre-processing techniques are used in this study. The stages are as follows:

Table 2. Before Filtering

reviewId	userName	userImage	content	score	thumbsUp Count	version	at
51572765-74…	Asep Ayi	https://play-lh.g...	Lebih tingkatkan ke akuratan nya	5	160	5.7.1	2023-01-1…
fface8c8-4c…	AW [ARI WAHYU]	https://play-lh.g...	Sangat buruk, gabisa login padahal data sudah benar	1	22	5.7.1	2023-07-2…
fef5b4d0-69…	Triandi Riyadi	https://play-lh.g...	Alhamdulillah bisa diakses dan bisa di download sertifikatnya	5	1055	5.7.1	2023-05-1…

Table 3. After Filtering

content	score
Lebih tingkatkan ke akuratan nya	5
Sangat buruk, gabisa login padahal data sudah benar	1
Alhamdulillah bisa diakses dan bisa di download sertifikatnya	5

3.1.2. Labelling Data

The data labeling method is required to establish whether a review in a document belongs to the positive or negative labeled class. Positive attitudes such as user pleasure, enjoyment, and contentment with the application are included in the positive class. The negative labeled class, on the other hand, includes complaints, objections, protests, critiques, and negative emotions such as wrath, frustration, and disappointment [8]. This study uses crowdsourcing for data labeling, enlisting the help of numerous labelers to optimize labeling outcomes [19] based on the emotion of SATUSEHAT application assessments. At this point, the Crowdsourced Labeling approach makes use of the Google Spreadsheet application. The choice of Google Spreadsheet is based on various factors, including the fact that all labelers are already familiar with it. Because the instrument requires internet access, real-time data tagging is possible. The usage of Google Spreadsheet provides efficient and concurrent labeling, which speeds up the process. As a consequence, the received data is expected to correspond to the planned results. For labeling, each labeler is given 200 data contents. Labelers can then apply positive, neutral, or negative labels to the data. Labelers are not required to apply labels to all content in each review. If a labeler does not assign a label to a piece of material, the labeling result is termed neutral, with a value of 0 reflecting the acquired label. This is based on the weighting of the Weighted Majority Voting technique used after the labeling results were obtained. This implementation involves 50 labelers, including 20 healthcare professionals and 30 members of the general public with a high school diploma or equivalent. Following that, weights are calculated based on the Crowdsourced Labeling findings. Weights are assigned to each label in the Weighted Majority Voting implementation, with the agree label receiving a weight of +1, the neutral label receiving a weight of 0, and the disagree label receiving a weight of -1. The final labeling results are determined based on the calculations performed using the Weighted Majority Voting technique, as shown in Figure 2 and Table 4 shows an example of labeling results.

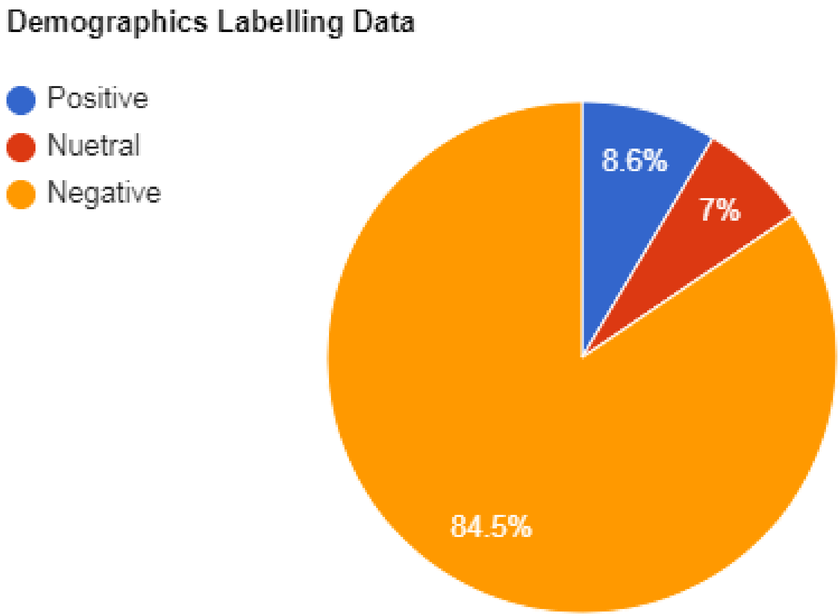

Fig. 2. Labelling Data

Table 4. After Labeling

content	score	sentiment
Lebih tingkatkan ke akuratan nya	5	Positive
Sangat buruk, gabisa login padahal data sudah benar	1	Negative
Alhamdulillah bisa diakses dan bisa di download sertifikatnya	5	Positive

3.1.3. Case Folding

Case folding is the process of cleaning and tidying data by converting all letters to lowercase [20] and deleting punctuation and non-alphabetic characters (in this case, only characters from 'a' to 'z', as shown in Table 5.

Table 5. Process Case Folding

content	
Before	**After**
Lebih tingkatkan ke akuratan nya	lebih tingkatkan ke akuratan nya
Sangat buruk, gabisa login padahal data sudah benar	sangat buruk, gabisa login padahal data sudah benar
Alhamdulillah bisa diakses dan bisa di download sertifikatnya	alhamdulillah bisa diakses dan bisa di download sertifikatnya

3.1.4. Stopword

Table 6. Process Stopword

content	
Before	**After**
lebih tingkatkan ke akuratan nya	tingkatkan akuratan nya
sangat buruk, gabisa login padahal data sudah benar	buruk gabisa login data
alhamdulillah bisa diakses dan bisa di download sertifikatnya	alhamdulillah diakses download sertifikatnya

To do sentiment analysis, unnecessary or meaningless words are removed during the filtering or stopword removal phase. A list of stopwords will be supplied, and if any of them are discovered in the text being processed, they will be eliminated [21]. Table 6 shows an example of the result following the stopword procedure.

3.1.5. Tokenizing

Tokenization is a means of breaking down a string input into words, which may be regarded as a method of breaking sentences into words by splitting them in order to evaluate and discover a set of syntactic word structures [20], as demonstrated in Table 7 sample findings.

3.1.6. Stemming

As shown in Table 8, the stemming step converts words with affixes into their base form, corrects words with spelling problems, and deals with specific short words in the existing data template format [21].

Table 7. Process Tokenize

content	
Before	**After**
tingkatkan akuratan nya	['tingkatkan', 'akuratan', 'nya']
buruk gabisa login data	['buruk', 'gabisa', 'login', 'data']
alhamdulillah diakses download sertifikatnya	['alhamdulillah', 'diakses', 'download', 'sertifikatnya']

Table 8. Process Stemming

content	
Before	**After**
['tingkatkan', 'akuratan', 'nya']	tingkat akurat nya
['buruk', 'gabisa', 'login', 'data']	buruk gabisa login data
['alhamdulillah', 'diakses', 'download', 'sertifikatnya']	alhamdulillah akses download sertifikat

3.2. Processing

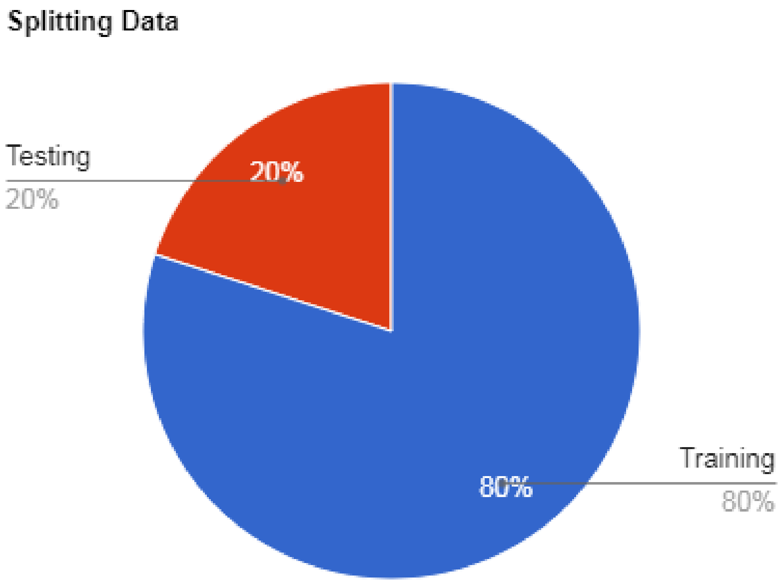

Fig. 3. Process Splitting

The data for the reviews came from the SATUSEHAT application reviews on the Google Play Store platform. There are 10,000 items in the comment data, with 858 favorable remarks (ratings 4 and 5), 697 neutral comments (rating 3), and 8,445 bad comments (ratings 1 and 2). However, neutral data was eliminated during classification model processing since neutral remarks are unable to offer a definitive judgement (considered uncertain), resulting in less accurate and useful data owing to confusing context. The comment data was reduced to 9,303 entries after the data purification procedure. Following that, the data was separated into 80% training data (7,442 entries) and 20% testing data (1,861 entries). The data is divided into two columns: one for comments and the other for the labels applied to the comments. Each remark is labeled with one of two sorts of labels: “Positive” or “Negative” Figure 3 shows the graphic for splitting training and testing data.

3.3. Post-processing

The research findings were assessed using a confusion matrix, as shown in Figure 3, and are further explained in Table 9. In all, the algorithm accurately categorized 43 good comments and 1662 negative comments. There were 1705 valid classifications out of 1861 comments.

Table 9. Confusion Matrix

Confusion Matrix	Prediction	
	Positive	**Negative**
Actual Positive	1662	35
Actual Negative	121	43

```
confusion_matrix:
 [[1662   35]
 [ 121   43]]
```

Fig. 4. Result Confusion Matrix

Data from the confusion matrix may be used to determine the accuracy, precision, and recall of the prediction results. The test findings shown in Table 10 and depicted in Figure 5 demonstrate an accuracy of 92%, an average precision of 93%, and an average recall of 98%.

Table 10. Value Methods

	Precision	Recall	F1-Score	Support
Negative	0.93	0.98	0.96	1697
Positive	0.55	0.26	0.36	164
Accuracy			0.92	1861
Macro AVG	0.74	0.62	0.66	1861
Weighted AVG	0.90	0.92	0.90	1861

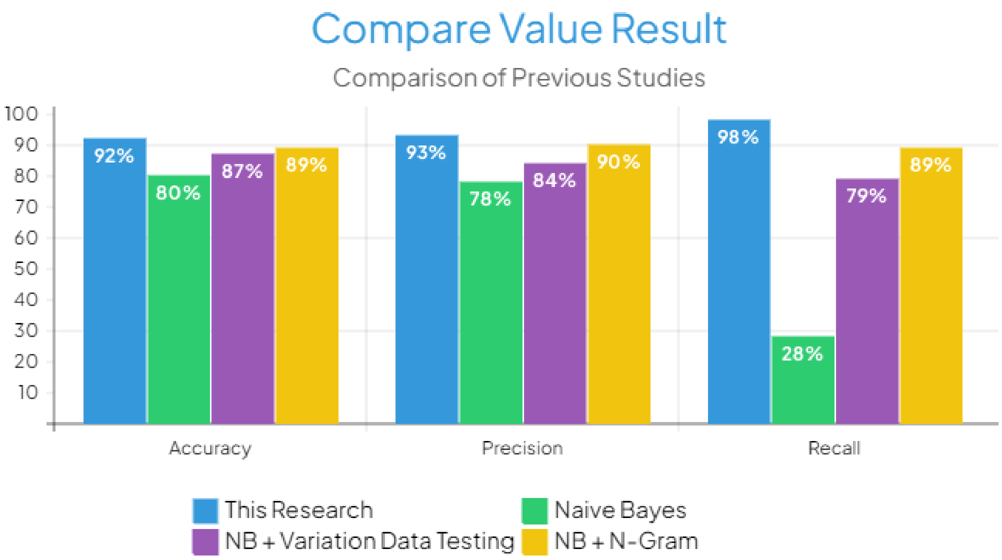

Fig. 5. Graphic Result

When the confusion matrix from a previous study using the same dataset and referencing the testing of accuracy, precision, recall, and F1-score is considered, the performance yielded is not significantly different from previous research utilizing the Naive Bayes method [6], Naive Bayes with different test data features [7], and Naive Bayes with N-Gram [8]. The performance of each strategy produced different results. In terms of accuracy, Naive Bayes proved to be more competitive than previous studies. It beat its competition in accuracy, scoring 92%, whereas the prior research, also employing Naive Bayes, scored 80%. Furthermore, utilizing Naive Bayes with other test data features yielded an accuracy score of 87%, while combining Naive Bayes with N-Gram yielded an accuracy score of 89%. As a result, our study outperformed others in the field of machine learning approaches. Figure 5 depicts the prior study's performance outcomes and accomplishments.

4. CONCLUSION

According to the findings of the research conducted on the implementation of the sentiment analysis classification algorithm for the SATUSEHAT application on the Google Play Store, the Nave Bayes algorithm with TF-IDF Vectorizer is the best algorithm for providing positive and negative labels with 92% accuracy, even when dealing with imbalanced data with unequal 1:1 proportion for each label. This implies that, when compared to other algorithms, the Naive Bayes method with TF-IDF Vectorizer achieves a high degree of precision, which might serve as a possible area of additional study in this research. More importantly, the algorithm's performance may be improved by including larger datasets with variances in comments. Furthermore, enhancements can be performed at the preprocessing step to repair misspelled words and transform them to their standard forms.

REFERENCES

[1] P. Purwono, E. Setyawati, K. Nisa, and A. Wulandari, "Strategi Gamifikasi Sebagai Peningkatan Motivasi Kuliah Pemrograman Website Pada Masa Pandemi Covid19," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 6, no. 3, p. 129, 2021, doi: 10.31328/jointecs.v6i3.2459.

[2] Kementrian Kesehatan, "PeduliLindungi Resmi Berubah Menjadi SATUSEHAT," *https://promkes.kemkes.go.id/pedulilindungi-resmi-berubah-menjadi-satusehat*, Mar. 02, 2023.

[3] M. S. Brown, "Data Mining For Dummies," *Embracing the data-mining process*, pp. 73–88, 2014.

[4] H. Annur and H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *ILKOM Jurnal Ilmiah*, vol. 10, no. 2, pp. 160–165, Aug. 2018, doi: 10.33096/ilkom.v10i2.303.160-165.

[5] F. -, M. Zarlis, and E. B. Nababan, "Analisis Perbandingan Akurasi dalam Identifikasi Autism dengan SVM dan Naive Bayes," *Jurnal SIFO Mikroskil*, vol. 17, no. 2, pp. 137–144, Oct. 2016, doi: 10.55601/JSM.V17I2.384.

[6] G. K. Locarso, "ANALISIS SENTIMEN REVIEW APLIKASI PEDULILINDUNGI PADA GOOGLE PLAY STORE MENGGUNAKAN NBC," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 2, 2022.

[7] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697–711, Sep. 2021, doi: 10.30645/J-SAKTI.V5I2.369.

[8] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, pp. 157–163, Jan. 2021, doi: 10.30865/MIB.V5I1.2580.

[9] D. D. Lewis, "Naive(Bayes)at forty: The independence assumption in information retrieval," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1398, pp. 4–15, 1998, doi: 10.1007/BFB0026666/COVER.

[10] D. J. Hand and K. Yu, "Idiot's Bayes: Not So Stupid after All?," *International Statistical Review / Revue Internationale de Statistique*, vol. 69, no. 3, p. 385, Dec. 2001, doi: 10.2307/1403452.

[11] I. Kononenko, "Semi-naive bayesian classifier," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 482 LNAI, pp. 206–219, 1991, doi: 10.1007/BFB0017015/COVER.

[12] P. Langley and S. Sage, "Induction of Selective Bayesian Classifiers," *Uncertainty Proceedings 1994*, pp. 399–406, Feb. 2013, doi: 10.1016/b978-1-55860-332-5.50055-9.

[13] A. Hamzah, "Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis | PROSIDING SNAST," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, 2012. https://journal.akprind.ac.id/index.php/snast/article/view/1744 (accessed Aug. 24, 2023).

[14] N. A. Zaidi, J. Jes´, J. Cerquides, M. J. Carman, and G. I. Webb, "Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting," *Journal of Machine Learning Research*, vol. 14, no. 60, pp. 1947–1988, 2013.

[15] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 3, pp. 279–285, Dec. 2019, doi: 10.26418/JP.V5I3.34368.

[16] A. Gaydhani, V. Doma, S. Kendre, and L. Bhagwat, "Detecting Hate Speech and Offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach," Sep. 2018.

[17] R. Umar, I. Riadi, and P. Purwono, "Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial dengan Metode Support Vector Machines," *Cybernetics*, vol. 4, no. 01, p. 32, 2020, doi: 10.29406/cbn.v4i01.2042.

[18] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697–711, Sep. 2021, doi: 10.30645/J-SAKTI.V5I2.369.

[19] I. Ihsan, D. Nurjanah, and H. Nurrahmi, "Sentiment Analysis Rkuhp Pada Twitter Menggunakan Metode Support Vector Machine," *eProceedings of Engineering*, vol. 8, no. 2, Apr. 2021.

[20] N. Haqqizar and T. N. Larasyanti, "Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes," *Prosiding TAU SNAR-TEK Seminar Nasional Rekayasa dan Teknologi*, vol. 1, no. 1, pp. 30–33, 2019.

[21] S. KHAIRUNNISA, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," 2021.