International Journal of Social Science, Educational, Economics, Agriculture Research and Technology (IJSET) | ISSN (e) : 2827-766X
Volume 5 No. 1 (2025)

SENTIMENT ANALYSIS OF TIX ID APPLICATION REVIEWS USING NAIVE BAYES AND SUPPORT VECTOR MACHINE

Salwa Nur. JB[1]*, Lola Astri Nadita[2], Sri Hidayati[3], Fachrurazy[4], Muhammad Amin[5]
Master of Information Technology Study Program, Panca Budi University, Medan
Email: salwanur2015@gmail.com[1]*, lolanadita@gmail.com[2],
srihidayati08guru.smp.belajar.id[3],oxxy.keren@gmail.com[4], mhdamin@dosen.pancabudi.ac.id5

Received	: December 26, 2025	Published	: December 26, 2025
Revised	: December 26, 2025	DOI	: https://doi.org/10.54443/ijerlas.v4i6.2061
Accepted	: December 26, 2025	Publish Link	: https://www.ijset.org/index.php/ijset/index

Abstract
TIX ID is an online cinema ticket purchasing application that plays on a smartphone platform. Where users can buy cinema tickets anywhere and anytime without having to wait in line. The concept of purchasing cinema tickets is integrated with a third party, namely DANA as a digital money concept that is integrated with several large applications such as Tokopedia and Shopee. This study aims to Conduct text preprocessing (NLP) on TIX ID application user review data so that the data is ready to be used in the sentiment analysis process, Extract text features using the Term Frequency–Inverse Document Frequency (TF-IDF) method to represent reviews in numeric form, Apply the Naive Bayes and Support Vector Machine (SVM) algorithms in classifying user review sentiments into positive and negative categories, Evaluate the performance of the Naive Bayes and Support Vector Machine (SVM) models using accuracy, precision, recall, and F1-score metrics, Provides an overview of user sentiment towards the TIX ID application as a consideration for developers in improving the quality and service of the application.

Keywords:: TIX ID Application, Natural Language Processing, Support Vector Machine

INTRODUCTION

The TIX ID app is a digital platform widely used by the public for online movie ticket booking. As the number of users increases, the app receives numerous reviews from users through app distribution platforms like the Google Play Store. These reviews reflect users' experiences, satisfaction, and complaints regarding the service provided. However, the sheer volume of user reviews makes manual analysis inefficient and prone to subjectivity. The TIX ID app is an online cinema ticket booking service released on March 21, 2018, by PT. Prosperous Archipelago Eagle. To use TIX ID, users can easily download the app from the Google Play Store for Android users or the App Store for iOS users. Once downloaded, the next step is to create a TIX ID account. TIX ID has become one of the fastest-growing entertainment apps this year, offering a wide selection of cinemas across Indonesia. Released on March 21, 2018, TIX ID has been downloaded by over five million users on both the App Store and Google Play Store. TIX ID aims to attract users to use the app when purchasing cinema tickets through attractive promotions, including discounts. This is one reason why TIX ID now has a strong competitive edge over its competitors. However, the sheer volume of user reviews makes manual analysis inefficient and prone to subjectivity. Natural Language Processing (NLP)-based sentiment analysis offers a relevant solution for understanding user opinions by extracting information from review text.

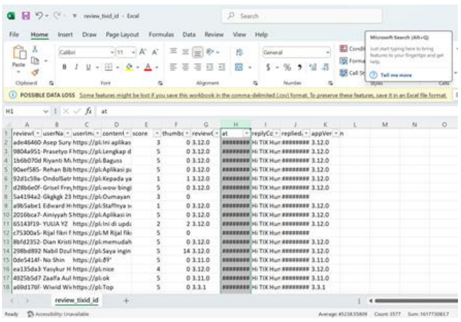
Naive Bayes and Support Vector Machine (SVM) are two classification algorithms frequently used in sentiment analysis, due to their simplicity, computational efficiency, and good performance in text data processing. Combined with text preprocessing techniques and TF-IDF feature extraction, the Naive Bayes algorithm is able to effectively classify reviews into positive and negative sentiments. Although sentiment analysis using TF-IDF combined with Naive Bayes and Support Vector Machine has been widely applied in previous studies, this research provides a context-specific empirical contribution by focusing on user reviews of the TIX ID application, one of Indonesia's most widely used digital cinema ticketing platforms. This study emphasizes real-world Indonesian-language user reviews, which present linguistic characteristics, informal expressions, and sentiment patterns that differ from datasets commonly used in prior research. Furthermore, this research highlights a comparative performance analysis between Naive Bayes and SVM under conditions of class imbalance, offering practical insights into algorithm selection for sentiment analysis in mobile application evaluation contexts. The findings are expected to serve as a reference for application developers and researchers in understanding user perception and improving service quality based on data-driven sentiment evaluation.

This study does not merely apply commonly used sentiment analysis methods, but also emphasizes the specific context of user reviews of the TIX ID application as one of Indonesia's widely used digital cinema ticketing platforms. The analyzed reviews employ informal Indonesian language that reflects real-world user expressions. Furthermore, this research compares the performance of Naive Bayes and Support Vector Machine algorithms under imbalanced data conditions, providing empirical insights into the strengths and limitations of each algorithm in classifying application review sentiments. The results are expected to serve as an evaluation reference for application developers and as a foundation for future research in digital application sentiment analysis.

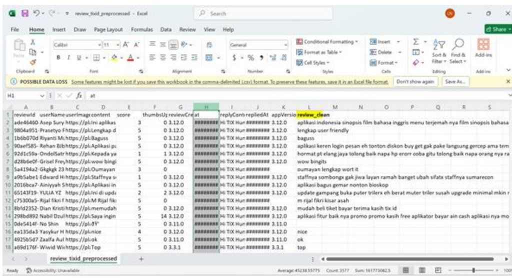
METHOD

In compiling this research, research methods are needed in the data collection process, firstly, the primary data source in this research is obtained from the results of data collection activities on the use of the TIX ID data application that uses NLP, secondly, secondary data in this research is data management from report data regarding the analysis of the number of uses of the TIX ID application in daily activities, in addition to supporting supporting data obtained from library media about the theories used in this research. The secondary data includes data mining, classification, naive bayes, support vector machines and other literature reviews. Literature studies consist of literature regarding the research being conducted. The literature in question includes several scientific journals and books related to this research case. Scientific journals and books are also used as references. The process of testing the application system that has been created will be tested to determine whether the TIX ID data application has run and is in accordance with the design carried out.

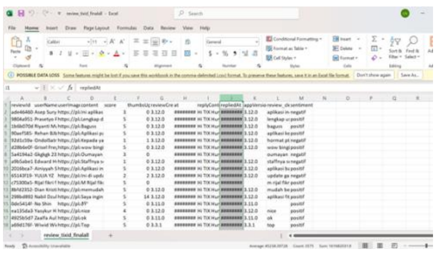
In the process of classifying review data using text mining with Naive Bayes Classifier and Support Vector Machine (SVM). Classification is divided into 2 criteria: Positive and Negative. Before classification, several stages of text processing are carried out using Natural Language Processing (NLP) methods such as case folding, stemming, tokenization and stopword removal. Tix ID data is taken from a scientific study with 3576 data points (2024-2023) from Kaggle: > (https://www.kaggle.com/datasets/ahmadseloabadi/tix-id-app-reviews-from-google-play-store) with the title "TIX ID app reviews from Google Play Store". Previous research:https://journal.eng.unila.ac.id/index.php/jitet/article/view/6250/2466



Raw Dataset



Dataset After Data Preprocessing



Final Dataset

Data Processing

The preprocessing stage is carried out using the Python programming language which includes case folding, cleaning, tokenizing, stopword removal, and stemming to produce clean text data before TF-IDF feature extraction and classification using the Naive Bayes algorithm. If the preprocessing is successful, the file will be automatically saved directly in the "Python Program" folder with the title "review_tixid_preprocessed.csv"

```
Contoh hasil preprocessing:

      Content                                          Review_clean
1758  ini kenapa ya mau di hubung kan ke dana tidak    hubung dana muncul tulis deteksi aman..
2899  Mantabss                                         mantabss
1977  Good, pengalaman pertama tiket online            good alam tiket online
2664  berharap cara pembayaran lebih banyak pilihannya harap bayar pilih
2502  Cepat dan mudah                                  cepat mudah
```

The following are the stages of the Data Preprocessing process:

1. Folding Case (Lower casing)
   The process of converting all letters to lowercase. In this process, the characters 'A'-'Z' in the data are converted to the characters 'a'-'z'.
   Example: “Complete and user friendly” >> “complete user friendly” (line 3)
2. Removing non-alphabetic characters
   Characters such as numbers, symbols, or emojis will be removed from the text.
   Example: "it's easy to watch without queuing" >> "it's easy to watch without queuing" (line 119)
3. Tokenization
   The tokenization process is performed using whitespace tokenization to improve system stability and avoid dependence on external resources.
4. Stopword
   Stop words are common words that often appear in large numbers and are considered meaningless. Examples of stop words in Indonesian are "yang," "and," "in," "of," etc. The purpose of using stop words is to remove low-information words from a text, allowing us to focus on important words instead.
5. Stemming
   Stemming is the process of mapping and parsing a word's form into its root form. To stem Indonesian, we can use the Sastrawi Python library we prepared earlier. The Sastrawi library applies the Nazief and Adriani algorithm to stem Indonesian.
   Example: Attachments: Attachments
   good: good

Automatic sentiment labeling (positive/negative score)

Sentiment labeling is a crucial step in sentiment analysis, providing the ground truth needed by supervised learning algorithms for learning and evaluating model performance. The sentiment labeling process is performed automatically based on user ratings (Score column). Ratings of 4–5 are categorized as positive sentiment, while ratings of 1–3 are categorized as negative sentiment. After the sentiment labeling process is successful, the file will be automatically saved as review_tixid_labeled.csv.

Output Sentiment Labeling:

```
Distribusi sentimen:
sentiment
positif    2891
negatif     685
```

TF-IDF Feature Extraction - Text to Numeric Transformation (Feature Extraction)

Naive Bayes cannot read text, so the data is converted to numbers using TF-IDF Feature Extraction. TF-IDF is used as a numerical representation of text for model training and is not stored as a separate dataset because it is temporary and is used directly in the classification process.

Python Implementation:

```python
X = data['review_clean']
y = data['sentiment']

tfidf = TfidfVectorizer(
    max_features=5000,
    ngram_range=(1, 2)
)

X_tfidf = tfidf.fit_transform(X)

print("\nJumlah dokumen:", X_tfidf.shape[0])
print("Jumlah fitur TF-IDF:", X_tfidf.shape[1])
```

```
Dimensi TF-IDF:
Jumlah dokumen : 3576
Jumlah fitur   : 5000
```

Data Classification Using Naïve Bayes Algorithm and Support Vector Machine (SVM)

After the data model is trained, the sentiment classification process is performed using the Naive Bayes algorithm due to its effectiveness in handling high-dimensional text data and imbalanced word distribution. The process classifies it into two criteria: Positive and Negative.

Evaluation of the metric model includes:

Metrik	Fungsi
Accuracy	Kebenaran keseluruhan
Precision	Ketepatan prediksi positif
Recall	Kemampuan menangkap sentimen positif
F1-Score	Keseimbangan precision & recall
Confusion Matrix	Detail kesalahan model

Training Data:
The dataset is divided into:

a. Training data (80%) >> The model can learn word patterns from sentiment
   80% data → training data
b. Test Data (20%) >> Generalizing model capabilities
   20% data → test data
   Confusion matrix only uses 20% of the data

RESULTS AND DISCUSSION

Results

Sentiment analysis

Sentiment analysis is the process of analyzing digital text to determine whether the emotional tone of a message is positive, negative, or neutral. Companies today have large volumes of text data, such as emails, customer support chat transcripts, social media comments, and reviews. Sentiment analysis tools can scan this text to automatically determine the author's attitude toward a topic. Companies use insights from sentiment analysis to improve customer service and enhance brand reputation. Sentiment analysis, also known as opinion mining, is a critical business intelligence tool that helps companies improve their products and services. The following describes the benefits of sentiment analysis:

1. Analyzing on a large scale
   Businesses continuously mine information from vast amounts of unstructured data, such as emails, chatbot transcripts, surveys, customer relationship management records, and product feedback. Cloud-based sentiment analysis tools enable businesses to scale the process of uncovering customer emotions in textual data at an affordable cost.
2. Providing objective insight
   Businesses can avoid the personal bias associated with human reviewers by using artificial intelligence (AI)-based sentiment analysis tools. As a result, companies obtain consistent and objective results when analyzing customer opinions. For example, consider the following sentence:

"I was impressed with the speed of the processor, but disappointed that it heated up quickly."
Marketers might overlook disappointing parts of a review and bias the processor's performance positively. However, accurate sentiment analysis tools sort and classify text to objectively capture emotions.

3. Building better products and services
Sentiment analysis systems help companies improve their products and services based on genuine and specific customer feedback. AI technology identifies real-world objects or situations (called entities) that customers associate with negative sentiment. In the example above, product engineers focused on improving the processor's heat management capabilities because the text analysis software associated "upset" (negative) with the processor (entity) and "heating" (entity).
4. Working time results
Businesses must respond quickly to potential crises or market trends in today's rapidly changing landscape. Marketers rely on sentiment analysis software to learn how customers feel about their brands, products, and services in real time and take immediate action based on their findings. They can configure the software to send alerts when negative sentiment is detected for specific keywords.

Sentiment analysis is an application of natural language processing (NLP) technology that trains computer software to understand text in a way that closely resembles humans. Here's how sentiment analysis works:
In the processing stage, sentiment analysis identifies keywords to highlight the core message of the text.

1. Tokenization breaks a sentence into elements and tokens.
2. Lemmatization converts words to their root form. For example, the root form of am is be.
3. Stop word removal filters out words that don't add meaning to a sentence. For example, with, for, at, and of are stop words.

NLP technology further analyzes the extracted keywords and assigns them a sentiment score. A sentiment score is a measurement scale that indicates the emotional element in a sentiment analysis system. It provides a perception related to the emotions expressed in the text for analytical purposes. For example, researchers use 10 to represent satisfaction and 0 to represent disappointment when analyzing customer reviews.

Naive Bayes Algorithm

The Naive Bayes algorithm, also known as Naive Bayes Classifiers, is a collection of differentiating algorithms based on Bayes' theorem. An algorithm is a systematic procedure or formula designed to help computers solve problems. Naive Bayes is an algorithm used by computers to perform specific tasks. The Naive Bayes algorithm's operation is governed by Bayes' theorem, which operates according to the principle of conditional probability. Conditional probability represents the probability or chance of an event, given that the related event has already occurred. One simple example of the concept of conditional probability is the toss of two coins. Each coin has two sides: heads and tails. Therefore, when two coins are tossed simultaneously, there is a possibility that they will show the same or different sides.
It is known that the sample space is as follows:

a. Coin 1= number; Coin 2= picture
b. Coin 1= number; Coin 2=number
c. Coin 1= picture; Coin 2= number
d. Coin 1= picture; Coin 2= picture

If we calculate the probability, then the probability (P) of both coins is:

a. The probability of both coins showing two heads = ¼
b. The probability of one of the coins showing a number = ¾
c. The probability of the second coin showing a number, and the first coin showing a head = ½
d. The probability of the second coin showing a head, and the first coin showing a tail = ½

Then, to find out what the probability is that the first coin shows a number (event A), and the second coin shows a picture (event B), you can use the following conditional probability formula:

$$P(A/B) = \frac{\boldsymbol{P}(B/A) * P(A)}{P(B)}$$

P(A/B)= Conditional probability of A given by B
P(B/A) =The conditional probability of B given by A
P(A) =Probability of event A
P(B) =Probability of event B

P(A/B) =[P (The first coin is a number and the second coin is a picture)*P(The second coin is a picture)] / P (The first coin is a number)
P(A/B)= [(1/2)*(1/2)]/(1/2)
P(A/B) =1/2 0.5
Based on the calculation, the conditional probability is ½ or if expressed as a decimal it becomes 0.5.
The benefits of the naive Bayes algorithm are:

1. Identifying faces and facial features, such as eyes, nose, mouth and eyebrows.
2. Predicting the weather.
3. M help doctors diagnose a patient's risk of disease, such as cancer or heart disease.
4. Used on Google News to group news, for example news about politics, entertainment, gaming or education.
5. In email, the naive Bayes algorithm is used to classify whether an incoming message is spam or not.
6. Using FMRI data, the Navie Bayes algorithm can predict various human cognitive conditions and detect brain injuries.

Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification and regression. SVM's working principle is based on Structural Risk Minimization (SRM), which is designed to process data into a hyperplane that classifies the input space into two classes. SVM theory begins with a linear clustering of cases that can be separated by a hyperplane and divided by class. The SVM concept begins with a two-class classification problem, requiring a positive and negative training set. SVM will attempt to find the best possible hyperplane (separator) to separate the two classes and maximize the margin between the two classes. The Support Vector Machine (SVM) method is used for automatic classification. Support vector machines can also solve classification and regression problems using both linear and nonlinear methods.

The support vector machine algorithm is used to find the best hyperplane in an N-dimensional space that clearly classifies data points. A hyperplane is a function used to separate one class from another. This function is used for classification within a higher-dimensional class space.

In 2D, the function used to classify between classes is called line similarity. In 3D, the function used to classify between classes is called plane similarity.

There are several advantages of using this algorithm:

1. Support vector machinevery helpful if you don't have much idea about the data.
2. Can be used for data that is not regularly distributed and whose distribution is unknown.
3. This algorithm has a very useful technique for solving any complex problem, called kernel.
4. It does not suffer from overfitting and works well when there is a clear indication of separation.
5. Capable of handling high-dimensional data.
6. The support vector machine algorithm has a better level of accuracy and makes predictions faster.
7. Can be applied to semi-supervised learning models.
8. It has better computational complexity compared to other algorithms.
9. Can be used even if the training data sample has bias.
10. Has the ability to normalize data.

Discussion

Interpretation of Naive Bayes results

1. The data distribution yielded 2,891 reviews with positive sentiment, while 685 reviews were negative. The predominance of positive sentiment indicates that users generally have a positive perception and experience of the analyzed service.
2. The classification model used produced an accuracy of 86.59%, Precision of 86.21%, Recall of 99.31% and F1-Score of 92.30%.

```
📊 HASIL EVALUASI NAIVE BAYES (OVERALL)
Accuracy  : 86.59%
Precision : 86.21%
Recall    : 99.31%
F1-Score  : 92.30%

📌 METRIK PER KELAS
Label     Precision    Recall      F1-Score
Positif   86.21%       99.31%      92.30%
Negatif   91.84%       32.85%      48.39%
```

```
📋 CLASSIFICATION REPORT
              precision    recall  f1-score   support

     negatif       0.92      0.33      0.48       137
     positif       0.86      0.99      0.92       579

    accuracy                           0.87       716
   macro avg       0.89      0.66      0.70       716
weighted avg       0.87      0.87      0.84       716
```

📌 CONFUSION MATRIX

	Pred Neg	Pred Pos
Actual Neg	45	92
Actual Pos	4	575

3. Based on the results of the confusion matrix on the test data, from approximately 579 reviews with positive sentiment, the model successfully classified 575 reviews correctly (True Positive), while 4 reviews were incorrectly classified as negative sentiment (False Negative).
4. Meanwhile, from approximately 137 reviews with negative sentiment in the test data, the model was able to classify 92 reviews correctly (True Negative), but there were still 45 reviews that were incorrectly classified as positive sentiment (False Positive).

These results indicate that the model has excellent capabilities in recognizing positive sentiment, but still has limitations in consistently identifying negative sentiment.

Model Performance on Positive Sentiment Labels (Dominance)

Classification reportThe results show that the model is highly effective in recognizing positive sentiment, with a Precision value of 86.21%, a Recall of 99.31%, and an F1-score of 92.30%. This indicates that the model correctly identified almost all positive reviews.

1. The positive sentiment label has a large Recall value reaching 99.31%,This means the model has a very strong ability to recognize language patterns, keywords, and sentence structures that represent positive sentiment. This is due to the predominance of positive sentiment data in the dataset. With a much larger amount of positive data than negative data, the model has more examples for learning the characteristics of positive sentiment during the training process. As a result, the model is more sensitive to the emergence of features associated with positive sentiment, thus minimizing the risk of false negatives (positive data predicted as negative).
2. The Precision value for the positive label is 86.21%, most of the data predicted as positive sentiment is indeed positive, although there is still a small number of negative reviews that are incorrectly classified as positive (false positive).
3. The F1-Score value for the positive label is 92.30%, meaning that the model's performance in the positive class is not only superior in capturing positive data, but also quite accurate in minimizing classification errors.

Model Performance on Negative Sentiment Labels

However, in the negative sentiment category, despite the high Precision value of 91.84%, the Recall value was still relatively low at 32.85% and the F1-Score was 48.39%. This indicates that some negative reviews are still not optimally detected and tend to be classified as positive sentiment.

1. The positive sentiment label has a low Recall value of 32.85%,This means that only about a third of all negative reviews were correctly detected by the model. This means that a large portion of negative data was misclassified as positive sentiment (false negatives). The low recall value for negative labels is primarily due to the imbalance in data volume, with negative data being significantly less than positive. As a result, the model is limited in learning the variety of language patterns that represent negative sentiment.
2. The precision value for positive labels is quite high at 91.84%, indicating that when the model predicts a review as negative, the prediction is almost always correct. False positive errors (positive reviews predicted as negative) are very minimal.
3. The F1-score for positive labels is 48.39%, reflecting a significant imbalance between precision and recall. The low F1-score is due to low recall, despite relatively high precision. This confirms that the model's performance on the negative class is unstable and requires improvement to detect more negative reviews without sacrificing prediction accuracy.

Interpretation of SVM results:

1. The data distribution yielded 2,891 reviews with positive sentiment, while 685 reviews were negative. The predominance of positive sentiment indicates that users generally have a positive perception and experience of the analyzed service.
2. The classification model used produced an accuracy of 89.80%, precision of 85.62%, recall of 79.49% and F1-Score of 82.03%.

```
HASIL EVALUASI MODEL SVM
Accuracy  : 89.80%
Precision : 85.62%
Recall    : 79.49%
F1-Score  : 82.03%

METRIK PER KELAS
Label     Precision   Recall     F1-Score
Positif   91.61%      96.20%     93.85%
Negatif   79.63%      62.77%     70.20%
```

```
CLASSIFICATION REPORT SVM:
              precision    recall  f1-score   support

     negatif       0.80      0.63      0.70       137
     positif       0.92      0.96      0.94       579

    accuracy                           0.90       716
   macro avg       0.86      0.79      0.82       716
weighted avg       0.89      0.90      0.89       716
```

```
CONFUSION MATRIX SVM
             Pred Neg    Pred Pos
Actual Neg   86          51
Actual Pos   22          557
```

3. Based on the results of the confusion matrix on the test data, from approximately 579 reviews with positive sentiment, the model successfully classified 557 reviews correctly (True Positive), while 22 reviews were incorrectly classified as negative sentiment (False Negative).
4. Meanwhile, from around 137 reviews with negative sentiment in the test data, the model was able to classify 51 reviews correctly (True Negative), but there were still 45 reviews that were incorrectly classified as positive sentiment (False Positive).

These results indicate that the SVM model has a fairly good ability to recognize positive sentiment, but still struggles to consistently distinguish negative sentiment. This may be due to an imbalanced class distribution in the dataset, where positive reviews significantly outnumber negative ones. This condition causes the model to tend to predict the majority class more often, thus increasing classification errors in the minority class.

Model Performance on Positive Sentiment Labels (Dominance)

The results show that the model is highly effective in recognizing positive sentiment, with a Precision value of 91.61%, a Recall of 96.20%, and an F1-score of 93.85%. This indicates that the model correctly identified almost all positive reviews.

1. The positive sentiment label has a large Recall value reaching 96.20%, indicatingA small proportion of positive reviews are misclassified as negative (false negative). This generally occurs because the word and phrase patterns in positive reviews tend to be more consistent, explicit, and easily recognized by bag-of-words-based models like TF-IDF combined with the Naive Bayes algorithm. Furthermore, the data distribution, which is dominated by positive sentiment, helps the model learn the characteristics of that sentiment more optimally.
2. The precision value for positive labels is 91.61%, indicating that the majority of positive sentiment predictions generated by the model are indeed positive reviews. In other words, the number of negative reviews incorrectly classified as positive (false positives) is relatively small. This indicates that the model is not only aggressive in capturing positive sentiment but also quite selective in ensuring that the positive predictions it provides have a high level of accuracy.
3. Meanwhile, the F1-score of 93.85% reflects an excellent balance between Precision and Recall. A high F1-score indicates that the model not only excels at finding positive reviews (high Recall) but also maintains its classification accuracy (high Precision). Thus, the model can be said to be stable and consistent in classifying positive sentiments, without producing too many classification errors.

Model Performance on Negative Sentiment Labels

However, in the negative sentiment class, despite the high Precision value of 79.63%, the Recall value was still relatively low at 62.77% and the F1-Score was 70.20%. This indicates that the model tends to be more cautious in predicting negative sentiment, resulting in only reviews with very strong negative characteristics being classified as negative.

1. The positive sentiment label has a low Recall value of 62.77%,This means that there are still a number of negative reviews that the model fails to detect and instead classifies as positive sentiment (false negatives). This phenomenon can be caused by several factors, including the use of ambiguous language, the presence of neutral or mixed-tone words in negative reviews, and the limitations of TF-IDF-based feature representation, which is not yet fully capable of capturing the semantic context and irony in review text.
2. The precision value for positive labels is high at 79.63%, indicating that the majority of reviews predicted as negative are indeed negative. In other words, the number of misclassifications, in which positive reviews are mistakenly predicted as negative (false positives), is relatively small. This indicates that the model has a high level of confidence in its negative sentiment predictions.
3. The F1-Score for positive labels is 70.20%, reflecting an imbalance between Precision and Recall in the negative sentiment class. This indicates that while the negative sentiment predictions are quite accurate, the model's ability to capture all negative reviews still needs improvement. Furthermore, the imbalanced data distribution—where positive reviews far outnumber negative reviews—also influences the model's tendency to focus more on studying positive sentiment patterns.

Comparison Result

Tabel Ringkasan Hasil Evaluasi Model Naive Bayes dan SVM

Algoritma	Akurasi (%)	Kelas Sentimen	Precision (%)	Recall (%)	F1-Score (%)
Naive Bayes	**86,59**	Positif	86,21	**99,31**	**92,30**
		Negatif	89,32	39,27	54,55
SVM	**lebih stabil**	Positif	**91,61**	96,20	**93,85**
	89.80	Negatif	**79,63**	**62,77**	**70,20**

1. Both algorithms were able to produce relatively high levels of accuracy, indicating that the text-based machine learning approach is effective for sentiment analysis of app reviews.
2. The Naive Bayes algorithm performed very well in classifying positive sentiment. This is demonstrated by the high Recall value, indicating that the model was successfully recognized with most positive reviews. The high Recall value in the positive class indicates that Naive Bayes tends to maximize positive sentiment detection, resulting in a relatively small number of errors in the form of positive reviews being misclassified as negative. Furthermore, the high Precision and F1-Score values indicate that the model's positive sentiment predictions are quite consistent and accurate.
3. However, in the negative sentiment class, Naive Bayes' performance tended to decline. While the Precision value remained relatively good, the lower Recall value indicated that some negative reviews were missed and instead classified as positive. This indicates that the model tends to be biased toward the majority class, which in this study is positive sentiment. The imbalanced data distribution also impacts Naive Bayes' ability to recognize distinctive patterns in negative reviews.
4. Meanwhile, the Support Vector Machine (SVM) algorithm demonstrated more balanced performance in separating positive and negative sentiment classes. SVM was able to construct a more optimal decision boundary in the high-dimensional feature space extracted from TF-IDF. This was reflected in the relatively stable Precision and F1-Score values across both classes, as well as its better ability to reduce cross-classification errors between positive and negative sentiment.

5. In the positive sentiment class, SVM maintained a high Recall value, although slightly lower than Naive Bayes. However, SVM showed improved performance in the negative sentiment class, particularly in Recall, indicating that it correctly recognized more negative reviews compared to the Naive Bayes model. Thus, SVM tends to be more robust in handling language variations and complex review contexts.

Overall, the comparison results show that Naive Bayes excels in classification speed and effectiveness for the majority class, while SVM provides more stable and balanced performance across sentiment classes. Therefore, choosing the best algorithm depends heavily on the research objective. If the primary focus is maximizing positive sentiment detection with minimal computation, Naive Bayes may be the right choice. Conversely, if the research demands a balance of performance across classes and more consistent accuracy, SVM is the more recommended algorithm.

Limitations of the Study

This study has several limitations that should be considered. First, the distribution of review data is imbalanced, with positive sentiment dominating negative sentiment, which may cause classification models to be biased toward the majority class. Second, sentiment labeling was performed automatically based on user ratings, which may not always accurately reflect the emotional content of the review text, particularly in cases of ambiguous or mixed sentiments. In addition, this study only applied binary sentiment classification, namely positive and negative, without considering neutral sentiment, which limits the overall representation of user opinions.

Future Work

Future research may extend this study by applying data balancing techniques, such as oversampling or undersampling, to reduce the impact of class imbalance. In addition, semantic-based feature extraction methods, such as word embeddings or deep learning-based language models, may be explored to enhance the model's ability to capture contextual meaning. Further studies may also expand sentiment classification by incorporating neutral or mixed sentiment categories to provide a more comprehensive understanding of user perceptions.

CLOSING

Conclusion

Based on the results of research and testing of sentiment analysis models on TIX ID user reviews, it can be concluded that the sentiment classification process using the TF-IDF method as a feature extraction technique, combined with Natural Language Processing (NLP)-based data pre-processing stages such as lower casing and stemming using the Sastrawi library, as well as the use of the Naive Bayes algorithm and Support Vector Machine (SVM) as a classification method, is able to classify review sentiments into two categories, namely positive sentiment and negative sentiment, with a total dataset of 3,576 review data. The Naïve Bayes classification model shows good performance with an accuracy of 86.59%. The precision value of 86.21% indicates that most of the positive sentiment predictions generated by the model are correct.

Meanwhile, the very high recall value of 99.31% indicates the model's excellent ability to detect almost all positive reviews. The F1-score of 92.30% confirms that the model has an optimal balance between precision and recall. However, in the negative sentiment class, Naive Bayes' performance tends to decline. Meanwhile, the Support Vector Machine (SVM) algorithm shows more stable and balanced performance between classes. The SVM model produces an accuracy of 89.80%, with a precision of 85.62%, a recall of 79.49%, and an F1-score of 82.03%. These values indicate that SVM is able to provide a more proportional performance distribution between positive and negative sentiment, making it more reliable in handling classification problems on datasets with class imbalance characteristics. The distribution of sentiment data shows that the number of reviews with positive sentiment is more dominant than negative sentiment, namely 2,891 positive data and 685 negative data.

Naive Bayes is superior in detecting positive sentiment and is suitable for use on data with an imbalanced class distribution., while SVM is more recommended for sentiment classification that requires a balance of performance between classes and robustness to review language variations.

The experimental results indicate that the Naive Bayes classifier achieved very high recall for the positive sentiment class, reaching 99.31%. This suggests that the model is highly sensitive to linguistic patterns associated with positive user feedback, which is likely influenced by the dominance of positive reviews within the dataset. As a result, Naive Bayes demonstrates strong performance in detecting favorable user perceptions but exhibits limitations in consistently identifying negative sentiment.

Conversely, the Support Vector Machine model produced a more balanced performance across sentiment classes. Although its recall for positive sentiment was slightly lower than that of Naive Bayes, SVM showed a notable improvement in detecting negative sentiment. This indicates that SVM is better suited to handling complex decision boundaries in high-dimensional feature spaces generated by TF-IDF, particularly in datasets with imbalanced class distributions.

These findings suggest that algorithm selection should be aligned with research objectives. If the primary goal is maximizing detection of positive sentiment, Naive Bayes offers an efficient solution. However, for applications requiring a more proportional recognition of both positive and negative user feedback, SVM provides a more robust alternative.

REFERENCES

Amin, M. (2020). *Employee Training on Employee Data Application System of PT. Indonesian Port I (Persero)*. Mantic Journal, 4(3), 1843-1848.

Amin, M. (2020). *Smart Water Faucet Control System Using Arduino Microcontroller and Ultrasonic Sensor. InfoTekJar*: National Journal of Informatics and Network Technology, 4(2), 245-249.

Amin, M., & Novelan, MS (2020). *Robot obstacle avoidance control system as a social distancing prototype using ultrasonic sensors and Arduino.* InfoTekJar: National Journal of Informatics and Network Technology, 5(1), 148-153.

Amin, M., & Rizal, C. (2025*). Development of Multimodal Generative AI Models for Adaptive Education Personalization in the Era of Quantum Machine Learning. Install: Computer Journal*, 17(09), 576-581.

Amin, M., Badawi, A., Hariyanto, E., & Sarif, MI (2025). *Socialization of Roles and Career Opportunities in Computer Science for High School Students. Journal of Community Service Results (JURIBMAS)*, 4(1), 207-213.

Amin, M., Irawati, N., Sinaga, HDE, Retnosari, D., Maulani, J., & Raja, HDL (2021). *Decision support system analysis for selecting a baby cream product with Preference Selection Index (PSI) Baby Sensitive Skin Under 3 Years.* Journal of Physics: Conference Series, 1933(1), 012035.

Amin, M., Raja, HDL, Nur, MNA, Prasetio, A., Sulaiman, OK, Karim, A., et al. (2022). *Wireless Network Technology.* Kita Menulis Foundation.

Amin, M., Tulus, T., & Ramli, M. (2016*). Robot Balancing Control Modeling Using Fuzzy Logic with Kalman Filter*. Technovation, 3(1), 39-44.

Batesda, B, dkk. (2014). *Analisa Sentimen Data Ulasan Pada Google Play Dengan Menggunakan Algoritma Navie Bayes dan Support Vector Machine.* Jurnal Sains Dan Teknologi ISTP, 22(01), 08-15. https://doi.org/10.59637/jsti.v22i01.423

Cynthia, E.P., Harahap, R., Amin, M., et al. (2025). *Strengthening the role of Gen-Z as digital content creators for promoting MSME products on social media.* Journal of Community Service Results (JURIBMAS), 4(2), 391-398.

Ginting, AN, & Amin, M. (2018*). Home Security Using PIR Sensor and Arduino GSM Module. Journal of Technology*, 5(1), 46-53.

KD Foote. "*A Brief History of Natural Language Processing (NLP)*. 2019

Liu., B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publisher

Novelan, MS, & Amin, M. (2020). *Monitoring system for temperature and humidity measurements with DHT11 sensor using nodeMCU.* International Journal of Innovative Science and Research Technology, 5(10).

PT. Nusantara Elang Sejahtera. (2013). TIX ID Google.com. Diakses pada tanggal 7 Januari 2026 dari https://play.google.com/store/apps/details?id=id/tix.android&hl=in&gl=US

Purb, Onno W. (2019). *Text Mining Analisis Medsos, Kekuatan Brand & Intelejen Di Internet.* Yogyakarta: Andi

Rozi, dkk. (2012), *Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi.* Jurnal EFECCIS. Universitas Brawijaya

Sukamto, R.A., dan Shalahuddin, M. *Rekayasa Perangkat Lunak.* Informatika. 2018

Suyanto. (2019). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data.* Bandung: Informatika

Wulandari, Retno Tri. (2017). *Data Mining Teori Dan Aplikasi Rapidminer.* Yogyakarta: Gava Media
Y Yunus, Muhammad. (2020). *"TF-IDF (Term Frequency-Inverse Document Frequency) : Representasi Vector Data Text."* Diakses pada tanggal 07 Januari 2026, dari https://medium.com/@yunusmuhammad007/tf-id-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda