

SENTIMENT ANALYSIS OF NANOVEST INVESTMENT APPLICATION USING NAIVE BAYES ALGORITHM

Lelianto Eko Pradana¹, Yova Ruldeviyani²

¹Faculty of Computer Science, Universitas Indonesia

²Faculty of Computer Science, Universitas Indonesia

email: lelianto.eko11@ui.ac.id¹, yova@cs.ui.ac.id²

Abstract

Various applications provide simple ways for individuals interested in investing in crypto assets or stocks - both domestic and international - to do so. One of the companies in this industry, Nanovest, has launched the Nanovest investment application. Since its release in 2022, numerous positive and negative responses have been on Google Play, the App Store, and Twitter. However, Nanovest faces two main problems regarding the use of its application. First, they often receive complaints submitted to the operational team, indicating dissatisfaction or problems faced by users. Second, Nanovest has never conducted formal research regarding user experience in using their application. This indicates a lack of understanding of the perspectives, needs and challenges faced by users. This study tries to find out how the public responds to the Nanovest application through a sentiment analysis. This study used tweet and review data from January 1, 2022, to February 17, 2023. The data underwent sentiment analysis, employing the Naïve Bayes algorithm, and were classified into positive and negative sentiments. The findings revealed that 96.07% of the sentiments expressed towards Nanovest were positive, while 22.11% were negative, with these percentages calculated based on the total number of sentiments detected in the data. To evaluate the model's performance, a 10-fold cross-validation approach was utilized alongside the Naïve Bayes algorithm, resulting in an impressive accuracy rate of 94.8391%. This positive sentiment suggests that users are highly favorable towards the crypto assets and global stock investment services offered by the Nanovest application. Nevertheless, 3.93% of users still expressed dissatisfaction with the app due to some flaws that existed when Nanovest was initially launched. Based on the results that have been obtained and analyzed for the development team, it is recommended to make three improvements, namely reducing application size to minimize memory usage, increasing overall application performance, and increasing access speed across all features to allow application users to access more efficiently. It is recommended for the product team and stakeholders to consider developing the Candlestick chart feature into the application. This also increases the competitiveness of the Nanovest application against other applications.

Keywords : Nanovest, Sentiment Analysis, Twitter, Google Play, Apps Store, Naïve Bayes

Received: 29-05-2023 | **Revised:** 22-06-2023 | **Accepted:** 24-07-2023

DOI: <https://doi.org/10.23887/janapati.v12i2.62302>

INTRODUCTION

Today, many applications offer easy investment methods for users who wish to invest in crypto assets or stocks, both domestic and global stocks. According to Yeong, crypto assets are digital currency that operates independently from banks or financial institutions [1]. Leveraging blockchain technology, these assets offer advanced security features and serve as the bedrock for establishing a secure public key infrastructure. As digital assets, market participants predominantly rely on internet platforms, including social networks and specialized forums, to gather market information, as highlighted by Aslanidis et al. [2]. Therefore,

crypto assets must be unique, secure, and well-encrypted digital assets to facilitate this information-gathering process. Trade and investment in global stocks have helped to improve the relationship between global equities, leading to a more diversified set of benefits for investors who trade developed country stocks in particular [3]. With technology advancing every day, investing in stocks is becoming more accessible. Investing means committing funds for a certain period to receive future compensation, considering factors such as the expected inflation rate and the uncertainty of future payments [4]. People have greater access to global equity holdings, and as technology continues to develop, the level of

public ownership of shares, including global shares, is likely to increase.

The pandemic has led to a significant increase in investment in stocks and crypto assets in Indonesia. Nanovest, an Indonesian-based company established in 2021, aims to make investing accessible and appealing to everyone, especially in its primary market of Indonesia. To revolutionize how young people invest and pursue financial freedom, Nanovest has developed the Nanovest platform, which focuses on investing in crypto assets and global stocks [5]. Nanovest is currently the only provider of an application that allows users to invest for free in crypto assets and global stocks, with over 2,000 US shares already available on the platform. These shares can be traded on various stock exchanges, such as NASDAQ, NYSE, and BATS, and purchased in small fractions. The Nanovest application began its services at the end of 2021 and had a grand launch in September 2022. Nanovest has created an ecosystem and community for its application users. However, Nanovest faces two main problems regarding the use of its application. First, they often receive complaints submitted to the operational team, indicating dissatisfaction or problems faced by users. These complaints can be related to platform performance, transaction difficulties, technical problems, or unsatisfactory customer service. The high number of complaints indicates the need to identify root causes and take appropriate remedial action. Second, Nanovest has never conducted formal research regarding user experience in using their application. This indicates a lack of understanding of the perspectives, needs and challenges faced by users.

Sentiment analysis has been conducted on the Nanovest app using Twitter, Google Play, and the App Store data to gain insights into customer responses. By analyzing sentiments expressed in tweets, reviews on Google Play, and the App Store about the Nanovest app, investment firms can assess customer sentiments and identify potential issues or areas for improvement. Suppose there are many negative reviews regarding the app's specific feature or function. In that case, Nanovest can utilize this feedback to address the problems and enhance the app accordingly. Similarly, if there are numerous positive tweets regarding a particular aspect of the app, Nanovest can use this feedback to improve further and enhance that aspect to increase customer engagement. Additionally, sentiment analysis gives Nanovest a better understanding of customer needs, preferences, and insights that can inform

product development and future marketing strategies. By monitoring sentiment around their app on Twitter, Nanovest can gain valuable insights into customer experiences and take necessary steps to improve customer satisfaction and retention.

Table 1. Several Cases were Reported by the Operations Team to the Nanovest Developer Team.

No	Cases
1	The user transaction history is missing under the "ALL" category, but it is available for other categories such as Crypto, Stock, Top Up, and so on.
2	The user with user ID 6139590561759232 placed a double top-up order of Rp5,000 with order ID E0000383127.
3	The admin user is still pending, and the email for password reset is not being delivered.
4	The transactions failed on the third-party side but were successful on our end.
5	The holdings of Google stocks experienced a sudden increase following a stock split.
6	The asset liquidation for user ID 6097888719405056 has been in progress since July 1st, 2022, and is still ongoing.

A previous study analyzed three online investment applications: HSB Investment, Seedlings, and Stockbit [6]. The study provided recommendations for new online investors and suggested ways for application owners to improve the quality of their services. The data was collected through an online questionnaire and processed using the Random Forest method. This study, however, centers its focus on the recently released Nanovest application, which has garnered over one million downloads from both the Play Store and App Store. The study has been conducted by reviewing the use of Nanovest on Twitter, a social media platform that enables users to interact and socialize with each other through text classification. Social media platforms like Twitter have a significant influence on our daily lives, and they provide valuable and timely information. In this study, comments from Twitter users of the Nanovest application have been analyzed, not just from the Google Play and App Store reviews.

This study analyzes all tweets containing the "nanovest" keyword, Google Play reviews, and App Store reviews, using the Naïve Bayes method to classify the sentiment. The Naive Bayesian classification algorithm is widely employed in the analysis of big data and diverse fields owing to its simple yet efficient algorithmic structure [7]. In addition to the Naïve Bayes method, a text mining process has been conducted to extract and process information, patterns, and knowledge that were previously unknown [8]. Text mining encompasses activities such as searching for information, gathering data, conducting statistical analysis, applying machine learning techniques, and utilizing computing capabilities [9]. Sentiment analysis was carried out, which involves extracting and analyzing data, views, and emotions in various contexts such as situations, events, products, or services. This study used the text-mining method with the Naive Bayes Classifier [10]. Specifically, this study analyzed sentence details when applying sentiment analysis [11]. The Naive Bayes Classifier, a widely utilized and straightforward approach for text classification that relies on Bayes' theorem and independent features, was employed in this research endeavor [12]. To enhance the Naive Bayes method, the text mining was adapted with improved distribution [13]. This method has been applied to various well-known studies and documents for text classification and sentiment analysis, demonstrating its reliability in predicting whether texts are negative or positive. Furthermore, the algorithm's performance was assessed using multiple evaluations [14].

METHOD

Data Collection and Processing

All collected data is processed through a series of steps. Initially, the problems and solutions are identified, followed by selecting a suitable method for sentiment analysis. The Naïve Bayes Algorithm, which combines probability and statistical methods, is one of the techniques used for classification. The Naive Bayes Algorithm operates under the assumption that the attributes are independent of each other [15]. This approach is utilized to predict the probability of a specific word belonging to a particular class [16]. The previous data processing results train the input set using the Naïve Bayes Classifier. The trained model is subsequently employed to classify the sentiments as positive or negative during the testing stage. The Naive Bayes Classifier formula is as follows.

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{H}{X}\right)P(H)}{P(X)} \quad (1)$$

Within the Bayesian statistics, the posterior probability $P(H|X)$, denoting the likelihood of hypothesis H being true based on the observed value of X , holds significance [16]. Here, $P(H)$ refers to the probability of H without considering the value of X , while $P(X)$ represents the probability of X having a specific value. The utilization of this classification approach offers several advantages, including the assumption of independence that facilitates efficient computation and the incorporation of a probabilistic hypothesis [17].

Based on Figure 1, the process starts with crawling data from Google Play, App Store, and Twitter. Reviews from the Google Play are taken using the google-play-scraper python library. By utilizing one of the

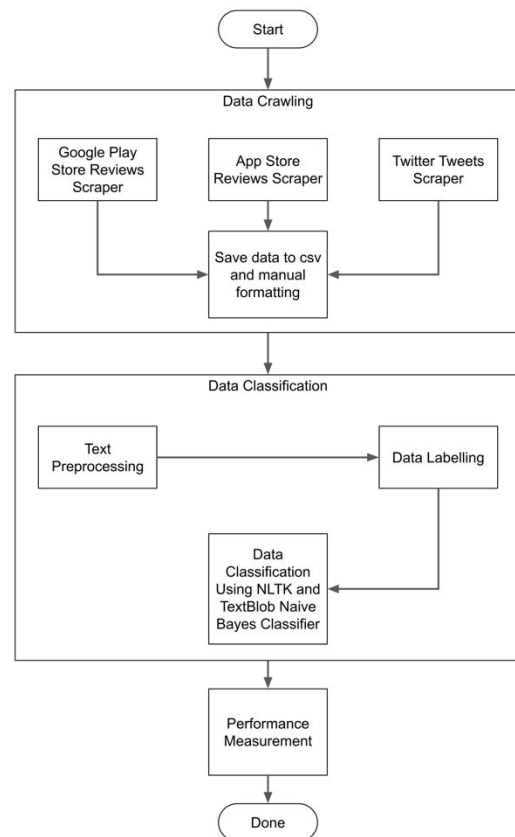


Figure 1. Data Processing Flow

functions in the library, namely `Sort` and `reviews_all`, review data can be retrieved involving the `'com.nanovest.prod'` parameter, `sleep_milliseconds=0` which is equal to 0 by default, `lang='id'` to set Indonesian language because by default it is the same as `'en'` or english language, set `country='id'`, which by default has the value `'us'` which means USA,

and is sorted according to the newest, sort=Sort.NEWEST, because it is sorted from the relevant or using Sort by default.MOST_RELEVANT. Here is the pseudocode of the explanation above.

```
import numpy as np
import pandas as pd

# Set the target app ID
target_app_id = 'com.nanovest.prod'

# Set the desired parameters for reviews collection
parameters = {
    'sort': Sort.NEWEST # defaults to Sort.MOST_RELEVANT,
    'app_id': target_app_id,
    'lang': 'id', # defaults to 'en'
    'sleep_milliseconds': 0, # defaults to 0
    'country': 'id', # defaults to 'us'
}

# Collect reviews for the Nanovest app
reviews = reviews_all(parameters)

# Convert the Nanovest reviews into a pandas DataFrame
df_nanovest = pd.DataFrame(np.array(reviews), columns=['review'])

# Expand the 'review' column into separate columns
expanded_reviews = pd.DataFrame(df_nanovest.pop('review').tolist())
df_nanovest = df_nanovest.join(expanded_reviews)

# Print the head of the DataFrame
print(df_nanovest.head())

# Save the DataFrame as a CSV file
df_nanovest.to_csv('nanovest.csv', encoding='utf-8')
```

Reviews from the App Store are taken using the app_store_scraper python library. It is known that by utilizing one of the functions in the app_store_scraper, namely, the AppStore function, review data can be retrieved by involving the country='id' parameter to set the focus on reviews written from Indonesia, app_name='nanovest-saham-aset-digital', app_id='1580892310', both app_name and app_id can be accessed directly from the App Store. Here is the pseudocode of the explanation above.

```
import numpy as np
```

```
import pandas as pd

# Set the target app details
target_country = 'id'
target_app_name = 'nanovest-saham-aset-digital'
target_app_id = '1580892310'

# Create an instance of the AppStore class
nanovest = AppStore(country=target_country, app_name=target_app_name, app_id=target_app_id)

# Fetch reviews for the target app
nanovest_reviews = nanovest.review(how_many=10000)

# Convert the Nanovest reviews into a pandas DataFrame
df_nanovest_reviews = pd.DataFrame(nanovest_reviews)

# Print the head of the DataFrame
print(df_nanovest_reviews.head())
```

Tweets from Twitter can be retrieved using the Python snsrape library by using one of its functions, snsrape. TwitterSearchScrapper involving the parameter 'nanovest since:2022-01-01 until:2023-02-18', then tweet data from Twitter from January 1, 2022, to February 18, 2023, was successfully obtained. All data obtained is then stored in a CSV format file which is then formatted as needed for data classification. Here is the pseudocode of the explanation above.

```
import snsrape.modules.twitter as sntwitter
import csv

# Creating a list to store tweet data
data = []

# Define the header for the CSV file
header = ['Date', 'ID', 'Content', 'Username']

# Leveraging the TwitterSearchScrapper to extract data and add tweets to the list
for i, tweet in enumerate(sntwitter.TwitterSearchScrapper('nanovest since:2022-01-01 until:2023-02-18').get_items()):
    if i > 100000:
        break
    data.append([tweet.date, tweet.id, tweet.content, tweet.user.username])

# Open a 'data_crawling.csv' CSV file for writing
```

```
with open('./data_crawling.csv', 'w',
encoding='UTF8', newline='') as f:
    writer = csv.writer(f)
```

```
# Write the header of the CSV file
writer.writerow(header)
```

```
# Write the data rows to the CSV file
writer.writerows(data)
```

The data collected for this study have undergone the text preprocessing stage before being classified. During this stage, the text from the obtained data underwent a cleansing process, which involved the removal of certain characters such as punctuation, special characters, numbers, and whitespace. Case folding was applied to convert all letters in the text document to either lowercase or uppercase. Tokenization was performed to divide the text document into smaller parts known as tokens. Normalization replaced words that were like a common word. Filtering removed words that were not important or irrelevant in the text document. Lastly, stemming removed affixes from words, leaving only the base forms. Figure 1 provides a detailed overview of the data crawling and classification process, indicating the completion of this research.

After going through the above process, the data is then classified using the NLTK (Natural Language Tool Kit) python library and the TextBlob python library, which involves the Analyzer parameter in the form of the NaiveBayesAnalyzer() from textblob.sentiment to generate a classification using the Naive Bayes Classifier for each sentence entered.

NLTK, a prevalent Python tool, is extensively employed for processing human language data in research studies. This comprehensive framework provides intuitive interfaces to over 50 corpora and lexical resources, including the renowned WordNet. It encompasses a multitude of libraries dedicated to text processing, facilitating essential tasks such as classification, tokenization, stemming, tagging, parsing, and semantic reasoning. Notably, NLTK also offers convenient wrappers for advanced NLP libraries, widely utilized across various industries, demonstrating its significance and relevance in research endeavors.

The subsequent stage encompasses evaluating the program's performance. The processed data serves as the basis for conducting k-fold cross-validation, which serves to analyze and validate the program's performance level. This methodology involves partitioning the data into K equally sized folds (K1, K2, ...Kn) and repeating the training and

evaluation process n times. In each iteration, one fold is assigned as the test data, while the remaining folds serve as the training data [15]. This approach ensures that each data point is utilized for prediction at some point [18]. The model's accuracy is assessed by using the test data in each fold, and this process is reiterated until the model is finalized. The model's accuracy was calculated by dividing the total accuracy by the number of K folds.

Data Crawling Process

After generating all review data from the Google Play and App Store, and all tweets related to the keyword 'nanovest', then, the following pseudocode utilizes k-fold cross-validation to measure the accuracy of previously classified data.

```
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
from sklearn.model_selection import
StratifiedKFold
import pandas
```

```
# Function to import the dataset
```

```
def import_data():
    # Import the total dataset from
    'improved_results.csv'
    data =
    pandas.read_csv('improved_results.csv')
```

```
# Get a list of column names
headers = list(data.columns.values)
```

```
# Separate the dataset into independent and
dependent variables. headers[3:] to retrieve
values in the 'Positive' and 'Negative' columns,
while headers[2] to retrieve values in the
"Classification" column.
```

```
x = data[headers[3:]]
y = data[headers[2]].values.ravel()
```

```
return x, y
```

```
# Main program
```

```
if __name__ == '__main__':
    # Import the data
    x, y = import_data()
```

```
# Set the number of folds to 10 (Data will be
retrieved for 2-Fold to 10-Fold)
```

```
skf = StratifiedKFold(n_splits=10)
```

```
# Empty lists were utilized to store the
predicted values and actual values
```

```
predicted_y = []
expected_y = []
```

```

# Partition the data
for train_index, test_index in skf.split(x, y):
    # Split the data into training and testing
    sets
    x_train, x_test = x.loc[train_index],
x.loc[test_index]
    y_train, y_test = y[train_index],
y[test_index]

# Create and fit the Gaussian Naive Bayes
classifier
classifier = GaussianNB()
classifier.fit(x_train, y_train)

# Predict the target variable for the test set
predicted_y.extend(classifier.predict(x_test))
# Store the expected target variable for this
specific fold
expected_y.extend(y_test)

# Calculate and print the accuracy
accuracy =
metrics.accuracy_score(expected_y,
predicted_y)
print("Accuracy: " + accuracy.__str__())

```

To carry out this research, in Figure 2, the k-fold cross-validation technique was implemented using a 10-fold approach. This technique involves running the model ten times, with each iteration using a randomly selected training set that comprises 90% of the total training data. A validation set comprising the remaining 10% of the training set is set aside to assess the model's performance. The primary advantage of employing k-fold cross-validation is its ability to yield a more precise evaluation of the model by assessing it on multiple subsets of the data. By repeating the fitting process with different training and validation sets, any bias that may have been introduced by a particular subset of data is reduced. This process ensures that the model is not overfitting to any specific subset of the data and instead can generalize well to new, unseen data.

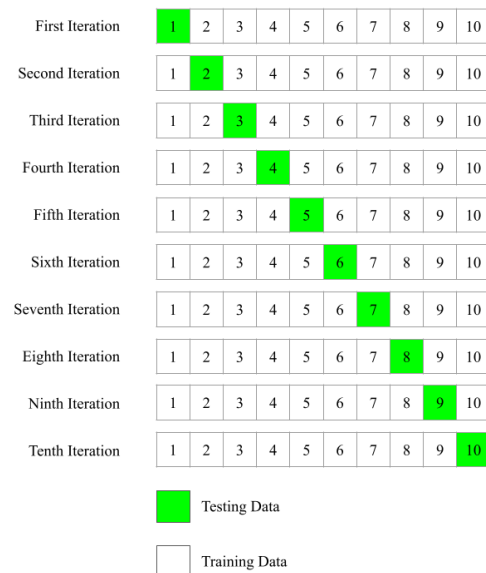


Figure 2. An overview of iterations in k-fold cross-validation

To evaluate the performance of the model or algorithm being used, the dataset was divided into ten approximately equal-sized subsets, also known as folds. This approach is commonly used in cross-validation techniques to ensure the evaluation is as accurate and unbiased as possible.

During the cross-validation process, nine subsets are used for training the model or algorithm, while the remaining subset is reserved for testing its performance. This process is repeated for each of the ten subsets, with a different subset reserved for testing each time. Using ten folds in this process ensures that the model or algorithm is evaluated on a variety of different subsets of the data, helping to minimize the impact of any subset on the results. This approach helps to provide a more accurate and robust assessment of the model's or algorithm's performance and is commonly used in machine learning research.

By leveraging this approach, researchers can gain valuable insights into the performance of models or algorithms and make informed decisions about their suitability for different applications or data sets. This information can improve the accuracy and reliability of the model or algorithm, which ultimately leads to better results. The results provided will be analyzed and discussed to provide suggestions for improving the Nanovest application by the developer team as well as additional strategies or services that need to be considered by stakeholders.

RESULT AND DISCUSSION

To gather the data needed for the study, a crawling process was conducted, resulting in a total of 8855 data points. This data was collected from various sources, including Twitter, Google Play, and the App Store. The Twitter data was collected over a specified time period, starting from January 1, 2022, up to February 17, 2023. After the data was collected, it was organized and processed to facilitate analysis at the subsequent study stage. The relevant data columns were extracted and formatted to allow for further processing.

The data obtained from Twitter, Google Play, and the App Store was stored in a CSV file, then analyzed using the Naive Bayes Classifier, as illustrated in Figure 7. This analysis method is commonly used in sentiment analysis to classify text data into positive or negative sentiments based on the probability of the data having either sentiment.

By utilizing the Naive Bayes Classifier, the researchers gained valuable insights into the overall sentiment of the data, allowing them to identify common themes and patterns in the data. This information can be useful for understanding the attitudes and opinions of users towards a particular product or service, which can be used to inform decision-making and improve the user experience.

The sentiment analysis results are presented in Table 2, which shows the classification of the text into positive or negative sentiments. The classification was performed using Naive Bayes, a machine learning algorithm that calculates the probabilities of positive and negative sentiment in the data. "Positive" column



Figure 3. Word Cloud of the Positive Sentiment Column

represents the probability of the data having positive sentiment, while "Negative" column represents the probability of having negative sentiment. The sentiment of the data was determined based on the presence of certain words in the data, and if the "Positive" column's score was higher than "Negative" column's score, the data was classified as having positive sentiment. Conversely, if the "Negative" column's score was higher than "Positive" column's score, the data was classified as having negative sentiment.

Table 2. Data Crawling and Classification Results

No	Reviews	Edited Reviews	Classification	Positive	Negative
1	Aplikasi nanovest ini enak dipakainya aman buat tanam saham kt jd nggak prlu takut"buat investasi, simpel gak ribet	aplikasi nanovest ini enak dipakainya aman tanam saham kita jadi tidak perlu takut buat investasi simpel tidak ribet	pos	0.5	0.5
2	Investasi di sini aman banget karena aplikasinya terenkrpsi dengan baik dan udah terdaftar di BAPPEBTI	investasi aman banget aplikasinya terenkrpsi sudah terdaftar bappebti	pos	0.5	0.5
3	Aplikasi sampah ini uang yang saya top up senilai 500k hilang begitu saja sudah diemail berkali kali tapi tidak ada tanggapan sama sekali dari pihak nanovest !!!!!	aplikasi sampah uang top up senilai 500k hilang diemail berkali kali tanggapan nanovest	neg	0.2704654896	0.7295345104
4	Aplikasi sampah ngadain event misi natal, pas mau selesaikan misinya yang serok crypto 50k order selalu di gagalkan bahkan sampai puluhan kali, kalau bisa kirim gambar ku ss kan order gagalnya, nanovest sampah	aplikasi sampah ada event misi natal pas selesaikan misinya serok kripto 50k order gagalkan puluhan kali kirim gambar ku ss order gagalnya nanovest sampah	neg	0.3476484782	0.6523515218
5	aplikasi terbaik menurutku buat belajar crypto sama invest, tampilan aplikasinya ga ngebosenin juga	aplikasi terbaik menurutku belajar kripto invest tampilan aplikasinya tidak ngebosenin	pos	0.7083333333	0.2916666667
6	Gak bisa jual aset kriptonya apk ga jelas Sebagai pemula yang mau nyoba investasi,	tidak jual aset kriptonya aplikasi tidak	neg	0.4705623314	0.5294376686
7	aku nggak ragu sama keamanan yang disediakan Nanovest. Apalagi selama ini kendalaku ya takut banget soal keamanannya. So far so good lah Nanovest ini!	pemula nyoba investasi tidak ragu keamanan disediakan nanovest kendalaku ya takut banget keamanannya so far so good nanovest ini	pos	0.516393555	0.483606445

Figure 4. Word Cloud of the Negati

To provide further insight into sentiment analysis, a word cloud has been created based on the data that has been obtained and processed. The word cloud displays the most frequently occurring words in the data and visually represents the overall sentiment. This analysis can help to identify common themes and patterns in the data, which can be useful for understanding the underlying sentiment and making informed decisions based on the insights gained from the analysis.

In Figure 3, there are several words that appear very often, including “aset kripto”, “event nanolympic”, “aplikasi nanovest”, “aplikasi bagus”, and “investasi saham”. This is enough to represent that as an investment application for crypto assets and stocks, the Nanovest application has succeeded in making it easier for users to access, buy and sell and even respond to one of the trading activities at Nanovest, namely Nanolympic, which has also been classified as successful in encouraging user interest to be more comfortable in transacting on the Nanovest application. In addition to the above, some users suggest that the Nanovest application provides candlestick charts to assist the trading process on stocks and crypto assets.

In Figure 4, there are several words that need to be highlighted to improve the Nanovest application in the future, such as the words "update", "withdraw", "kyc", and "data". Based on some of the review data obtained, the frequency of application updates which is quite frequent, is felt uncomfortable by users, especially for users who have limited internet quota and smartphone capacity. There is also a problem with the system for withdrawing funds from the user's balance. Users often fail to fail when they want to withdraw funds while always succeeding when topping up their balance. This has an impact on user trust in the Nanovest application. However, judging from the data obtained, this problem occurred when the Nanovest application had just been launched. Meanwhile, problems related to the KYC (Know Your Customer) system are related to the difficulty of the data verification process that needs to be carried out, and users admit that they often experience rejection, so some of them ask that their data be deleted because they don't end up using the Nanovest application. However, like the problem with withdrawing funds, this problem also only occurs when the Nanovest application has just been launched. The problem that still needs attention is related to the size of the



Table 2. Positive and Negative Sentiment Classification

Classification	Total
Positive	8507
Negative	348
Grand Total	8855

Based on the previously outlined data processing, the researchers obtained results in Table 2 above. The data was analyzed using the Naive Bayes classifier to classify it into positive and negative sentiments. According to the results, most of the data points, 8507, were classified as having a positive sentiment. This indicates that most of the feedback about Nanovest services and applications was positive or enthusiastic. On the other hand, only 348 data points were classified as having a negative sentiment, indicating a relatively small percentage of negative feedback or disinterest towards Nanovest services and applications.

The researchers gained valuable insights into users' attitudes and opinions toward Nanovest services and applications by categorizing the data into positive and negative sentiments. This information can inform decision-making and identify areas where improvements can be made to enhance the user experience and increase customer satisfaction. The results suggest that most users positively perceive Nanovest's services and applications, which is a positive sign for the company's future growth and success.

After analyzing all the data obtained from different sources, it was found that 96.07% of the data had a positive sentiment, while 3.93% had a negative sentiment. This percentage is obtained by comparing the total positive and negative sentiments with the overall data.

application, which is considered quite large, and some parts of the application, which are quite slow to access.

Based on the negative sentiments of Nanovest application users, it can be concluded that Nanovest needs to optimize application performance and improve user experience. For the development team, it is recommended to make three improvements, namely reducing the size of the application to minimize memory usage, increasing overall application performance, and increasing access speed across all features to allow application users to access more efficiently. Based on the findings of this study, it is recommended that the product team and stakeholders consider developing the Candlestick chart feature into the application. This feature provides users with powerful tools to analyze market trends and make well-informed trading decisions. This also increases the competitiveness of the Nanovest application against other applications.

The components of the generated data have been discussed and explored, and text analysis processing steps have been applied. The database in the form of a CSV file is used to store large amounts of review and tweet data.

All parts of the text preprocessing have been done. All data that has been processed has also been classified using the help of the Natural Language Toolkit (NLTK) and the TextBlob python library.

Table 3. The Results of Accuracy Testing for Each Fold

Fold	Accuracy (%)
2	95.1892%
3	94.9407%
4	94.8617%
5	94.8617%
6	94.8278%
7	94.8504%
8	94.8504%
9	94.8504%

10	94.8391%
----	----------

The accuracy of the Naive Bayes Classifier was measured using a Python plugin called sklearn, which has a StratifiedKFold extension function on model_selection. The accuracy measurement results ranging from 2-Fold to 10-Fold have been measured and the results are presented in Table 3. The results showed that the classification system was 94.8391% accurate for 10-fold. The high accuracy rate suggests that the Naive Bayes Classifier effectively categorizes the data into positive and negative sentiments. This can be attributed to the thorough data processing and analysis techniques, including the k-fold cross-validation method.

CONCLUSION

Based on the research that has been done, the positive sentiments of the Nanovest application outweigh the negative sentiments. This shows that many users of the Nanovest application feel comfortable using the application. This high positive sentiment, 96.07%, shows how pro-expression is towards crypto asset and global stock investment services in the Nanovest application. Existing application services need to be improved and repaired because there are 3.93% of users who are still dissatisfied. From the data obtained, most people are not satisfied when Nanovest has just launched and has a defect in the application. The model performance is made to have an accuracy of 94.8391%.

Based on the results that have been obtained and analyzed for the development team, it is recommended to make three improvements, namely reducing application size to minimize memory usage, increasing overall application performance, and increasing access speed across all features to allow application users to access more efficiently. It is recommended for the product team and stakeholders to consider developing the Candlestick chart feature in the application. This also increases the competitiveness of the Nanovest application against other applications.

For future research, researchers can compare applications that serve the sale and purchase of crypto assets and stocks to be able to assess the comparison of customer satisfaction levels for each of these applications.

REFERENCES

[1] Y.-C. Yeong, K. S. Kalid, K. S. Savita, M. N. Ahmad, and M. Zaffar, "Sustainable cryptocurrency adoption assessment among

it enthusiasts and cryptocurrency social communities," *Sustainable Energy Technologies and Assessments*, vol. 52, p. 102085, 2022. doi:10.1016/j.seta.2022.102085

[2] N. Aslanidis, A. F. Bariviera, and Ó. G. López, "The link between cryptocurrencies and google trends attention," *Finance Research Letters*, vol. 47, p. 102654, 2022. doi:10.1016/j.frl.2021.102654

[3] O. S. Yaya, L. A. Gil-Alana, O. B. Adekoya, and X. V. Vo, "How fearful are commodities and US stocks in response to global fear? persistence and cointegration analyses," *Resources Policy*, vol. 74, p. 102273, 2021. doi:10.1016/j.resourpol.2021.102273

[4] D. Tambunan, "Investasi Saham di Masa Pandemi covid-19," *Widya Cipta: Jurnal Sekretari dan Manajemen*, vol. 4, no. 2, pp. 117–123, 2020. doi:10.31294/widyacipta.v4i2.8564

[5] "Investasi Saham Amerika & Aset Kripto legal Dan Aman," *Nanovest*, <https://www.nanovest.io/> (accessed May 29, 2023).

[6] W. A. Wahyuni, S. Saepudin, and F. Sembiring, "Sentiment Analysis of Online Investment Applications on Google Play Store using Random Forest Algorithm Method," *Jurnal Mantik*, vol. 5, Feb. 2022.

[7] R. Hua, songhua hu, H. Chen, and X. Zhao, "Improved naive Bayes classification algorithm for Traffic Risk Management, 2021. doi:10.21203/rs.3.rs-355037/v1

[8] R. Talib, M. Kashif, S. Ayesha, and F. Fatima, "Text mining: Techniques, applications and issues," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 11, 2016. doi:10.14569/ijacsa.2016.071153

[9] J. E. Chire Saire and A. Pineda-Briseno, "Analysis of covid-19 impact in Mexico City using text mining and Twitter," 2020 International Conference of Digital Transformation and Innovation Technology (Incodtrin), 2020. doi:10.1109/incodtrin51881.2020.00018

[10] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and Trends," *Knowledge-Based Systems*, vol. 226, p. 107134, 2021. doi:10.1016/j.knosys.2021.107134

[11] S.-U. Hassan et al., "Predicting literature's early impact with sentiment analysis in Twitter," *Knowledge-Based Systems*, vol. 192, p. 105383, 2020. doi:10.1016/j.knosys.2019.105383

- [12] H. Parveen and S. Pandey, "Sentiment analysis on Twitter data-set using naive Bayes algorithm," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016. doi:10.1109/icatcct.2016.7912034
- [13] M. Abbas, K. Ali, S. Memon, and A. Ahmed, "Multinomial Naive Bayes Classification Model for Sentiment Analysis," International Journal of Computer Science and Network Security, vol. 19, Mar. 2019. doi:10.13140/RG.2.2.30021.40169
- [14] B. Seref and E. Bostanci, "Sentiment analysis using naive Bayes and complement naive Bayes classifier algorithms on Hadoop framework," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018. doi:10.1109/ismsit.2018.8567243
- [15] R. Riyanto, "Application of the vector machine support method in twitter social media sentiment analysis regarding the COVID-19 vaccine issue in Indonesia," Journal of Applied Data Sciences, vol. 2, no. 3, pp. 102–108, 2021. doi:10.47738/jads.v2i3.40
- [16] M. Vadivukarassi, N. Puviarasan, and P. Aruna, "Sentimental Analysis of Tweets Using Naive Bayes Algorithm," World Applied Sciences Journal, 2017. doi:10.5829/idosi.wasj.2017.54.59
- [17] H. Krishnan, M. S. Elayidom, and T. Santhanakrishnan, "Emotion Detection of Tweets using Naïve Bayes Classifier," International Journal of Engineering Technology Science and Research, vol. 4, no. 11, Nov. 2017.
- [18] M. R. Nasution and M. Hayaty, "Perbandingan Akurasi Dan Waktu proses algoritma K-nn Dan SVM Dalam Analisis Sentimen twitter," Jurnal Informatika, vol. 6, no. 2, pp. 226–235, 2019. doi:10.31311/ji.v6i2.5129
- [19] S. A. Saputra, B. Rahmatullah, and P. Budiyo, "Sentiment Analysis User Ajaib Application Using Naïve Bayes Algorithm," Journal of Information System, Informatics and Computing, pp. 497–505, Dec. 2022.