

# Analysis of eFootball Game User Sentiment Using the Support Vector Machine (SVM) Method

Caca Agustian<sup>1</sup>, April Lia Hananto<sup>2</sup>, Fitria Nurapriani<sup>3</sup>, Baenil Huda<sup>4</sup>

1,2,3,4 Information Systems Study Program, Universitas Buana Perjuangan Karawang  
Copyright © 2019 Puseurjaya. All Rights Reserved. All rights reserved 413611.  
[si21.cacaagustian@mhs.ubpkarawang.ac.id](mailto:si21.cacaagustian@mhs.ubpkarawang.ac.id)<sup>1\*</sup>, [aprilia@ubpkarawang.ac.id](mailto:aprilia@ubpkarawang.ac.id)<sup>2</sup>

## Abstract

This study examines the analysis of user review sentiment for eFootball games on the Google Play Store using the Support Vector Machine (SVM) method. A total of 900 reviews written in Indonesian were taken and collected, and divided based on user ratings. The research process includes data exploration, text cleanup (preprocessing), sentiment labeling based on rankings, modeling using SVM, and model evaluation with confusion matrix and accuracy metrics. The results of the analysis showed that the majority of reviews conveyed positive sentiment (48.7%) followed by negative sentiment (44.9%) and neutral sentiment (6.4%). The SVM-based model built in this study achieved an accuracy of 76%, with adequate precision, memory, and F1 scores, especially in the positive and negative sentiment categories. These findings suggest that SVM is effective in classifying digital game review sentiment, but performance in the neutral category requires significant improvement. The study contributes to the use of machine learning to analyze user perceptions of eFootball games and provides recommendations for developers to improve product quality through automated sentiment analysis.

**Keywords:** Sentiment analysis, eFootball, Google Play Store, Support Vector Machine, user reviews

## 1. Introduction

The digital gaming industry is one of the fastest growing especially in this contemporary era. Due to the ever-increasing number of users, gaming has turned into one of the main sources of entertainment in various societies. Sports games, especially soccer, are some of the most popular categories. One of the most popular soccer games around the world is eFootball, which was developed by Konami. The game offers a realistic experience of playing soccer through a digital arena with features such as multiplayer mode, regular player updates, and excellent graphics.[1]

This study chose SVM (Support Vector Machine) analysis for sentiment analysis. Support Vector Machine, or better known as SVM, is one of the algorithms in machine learning based on classification, including the classification of text data. This method is very effective and reliable for classifying positive, negative, and neutral sentiments as it tries to find the best hyperplane that can separate the data into specific categories.[2]

Although the SVM method has been widely used for sentiment analysis in various domains, there is still a lack of implementation in the field of gaming, especially in analyzing eFootball game reviews. Therefore, this study aims to explore the use of SVM in analyzing user sentiment of eFootball games, determining the factors that influence user sentiment, and assessing the performance of the model in how accurately it classifies the data.[3]

## 2. Research Methodology

The steps in this description follow a flowchart that describes each stage performed during the study. Figure 1 below is:

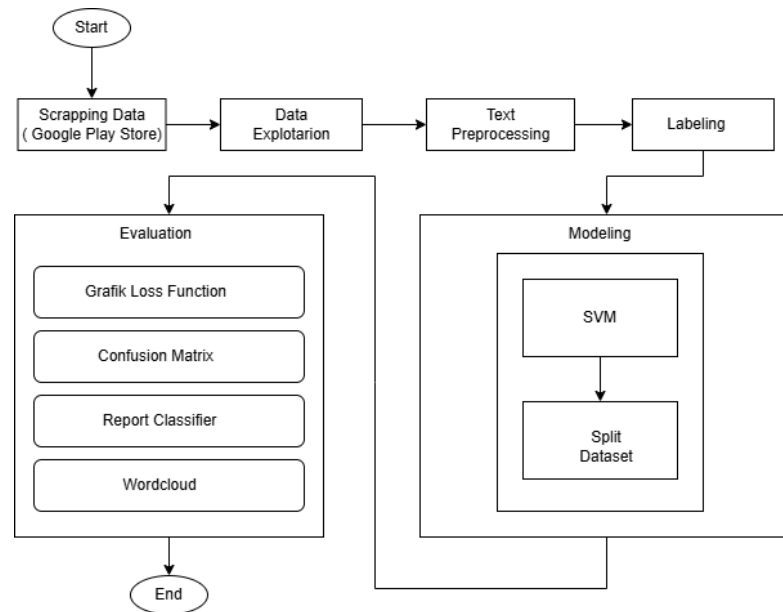


Figure 1: Research Stages

### 2.1. Data Collection

In this study, the data to be taken comes from the Google Play Store platform on the eFootball game in Indonesian. The dataset is obtained using a google-play-scraper package that provides an API in the Python programming language, and the data will be stored in CSV format. The number of datasets used in this study is 1,000 records. This number is taken based on the distribution of ratings or scores. So each rank or score (1-5) is taken per 200 records. So the amount of data generated is 1,000 records consisting of ranks 1-5 on the Google Play Store platform.

### 2.2. Data Exploration

The next step is to understand each column/feature in the dataset, there are several columns/features in the dataset including Date, Username, Rating, and Comments. The Date column will be useful for finding values that are considered duplicate in reviews, the Username column for sentiment models, the Rating column for sentiment labels (positive, neutral, and negative), and the Comment column for data validation. In this case, a review will be considered relevant if it is supported by a lot of likes/support.

### 2.3. Pre-Processing of Text

The initial stage of text preprocessing aims to clear the dataset of noise to make it more structured. This process uses the NLTK module and the Literary library in Python. The steps taken include:

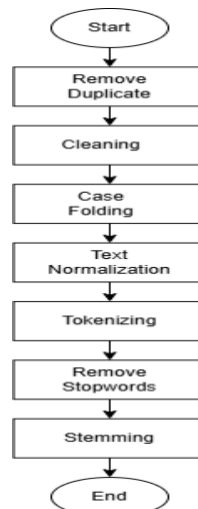


Figure 2: Text Preprocessing Stage

- Remove Duplicates*: Delete duplicate review data based on reviewId to avoid duplicate reviews from the same user.
- Cleanup*: Remove invalid characters such as punctuation, numbers, symbols, and emojis to make data processing easier.
- Uppercase Folding*: Converts all text to lowercase

- d. *Text normalization*: Corrects words, typos, abbreviations, and incorrect words out of context.
- e. *Tokenizing*: Separating words in sentences
- f. *Remove stop words*: Eliminate words that appear frequently but have no significant meaning, such as conjunctions or pronouns.
- g. *Stemming*: Change the word to its basic form by removing the affix.

## 2.4. Pelabelan

In this study, the marking process was carried out based on the score given in the review. The number (1-2) is considered negative, the number (3) is neutral, and the number (4-5) is positive.

## 2.5. Modeling

The modeling stage includes the process of entering the data that has been prepared from the collection stage into the model. At the modeling stage, an SVM model is used that applies the entire dataset.

## 2.6. Evaluation

The purpose of the model evaluation process is to assess the extent to which the model is able to classify various categories. One way to evaluate in this study is to use a confusion matrix. It is a tool used to measure model performance in machine learning classification tasks by generating outputs consisting of two or more classes. Next, the authors examine the graph of the loss function to determine if the model is too fit. At this stage of evaluation, the results of the sentiment analysis prediction for each sentence in the dataset will be reviewed. The highest accuracy achieved during the training of the previous model is used as the accuracy for the model.

# 3. Results and Discussion

## 3.1. Scraping Data

**Data Source** The data obtained in this study refers to data taken from user reviews on the Google Play Store about eFootball games. This research is an experimental study, so data was collected using data crawling techniques with the help of the Google Play Scraper library. Where, the library allows researchers to automatically retrieve review data from the platform. The range of data collection in this study is from January to May 2025. The number of datasets is determined based on ratings 1-5, a maximum of 200, if there are only 5 ratings, the researcher takes  $200 \text{ data} \times 5 = 1000$ . However, this study was created based on only 900 reviews based on a lack of data available within the exact time frame from the date of data collection. The data in this study was obtained with two main limitations, namely: the order of the LATEST reviews and the order of relevance of the MOST RELEVANT. Furthermore, this data will be cleaned to improve its quality before being used for more in-depth sentiment analysis.



	Tanggal	Nama Pengguna	Rating	Komentar
0	05-05-2025 09:42	Wahyudin Galuh	1	game konyol masih banyak bug sinyal anj
1	05-05-2025 09:42	Pengguna Google	3	tolong pas main online jaringan bagus , dan TP...
2	05-05-2025 09:26	Adnan Januzaj	1	game babi nge lag melulu tiap main divisi cuihhh
3	05-05-2025 09:24	ilham purnama sakti	4	game bagus, tapi masih kurang karena ga suppor...
4	05-05-2025 09:23	Vhian Saputra	5	Di pergerakan nya kurang maksimal, virtualnya ...
...	...	...	...	...
995	02-05-2025 12:11	Evos Kenzo	5	game nyaa sudaa baguss tapi biar lebih bagus d...
996	02-05-2025 12:08	Mas Yusuf	1	Kenapa malah makin berat nih game
997	02-05-2025 12:06	Soni Panggabean	3	kalau bisa saran..buat master liga nya..biar b...
998	02-05-2025 12:05	Fiki Andrian	5	mantap
999	02-05-2025 12:04	Beno Benoo	4	ku kasih bintang 4 karna tiba-tiba bug karna y...

1000 rows x 4 columns

**Figure 3:** Results on Data Scraping

## 3.2. Data Exploration

The scraping result has many attributes created in the data frame. Here, the features I use are Date, Username, Rating, Comments. The Date column is used to search for Reviews that are considered to have duplicate values, the Username column for sentiment models, the Rating column is used for sentiment labels are positive, neutral, and negative, the Comment column is used for data validation. We can enter the review as a relevant review if it gets a lot of likes/support.

## 3.3. Pre-Processing of Text

Text Preprocessing is the process of removing all irrelevant or unnecessary data and the necessary data is properly prepared to create machine learning models on text data. It helps in removing duplicates, cleaning, converting text to lowercase, normalizing text, tokenizing, removing stopwords and giving a limean. Here are some of the steps that have been implemented:

a. *Remove Duplicates*

In this stage, the dataset has been cleaned of duplicate reviews based on the "date" column. The elimination of identical data is crucial to avoid bias in the model.

```
[4] # 1. Load dataset
df = pd.read_csv('/content/komentar_efootball_rapih.csv')

# 2. Konversi kolom Tanggal ke datetime (format: dd-mm-yyyy HH:MM)
df['Tanggal'] = pd.to_datetime(df['Tanggal'], format='%d-%m-%Y %H:%M')

# 3. Urutkan data berdasarkan Tanggal (ascending = False untuk mengambil entri terbaru)
df = df.sort_values('Tanggal', ascending=False)

# 4. Hapus duplikat berdasarkan kolom Tanggal (pertahankan entri terakhir)
df_clean = df.drop_duplicates(subset=['Tanggal'], keep='first')

# 5. Tampilkan hasil
print(f"Jumlah data sebelum remove duplicate: {len(df)}")
print(f"Jumlah data setelah remove duplicate: {len(df_clean)}")
print(f"\nContoh 5 data terbaru setelah cleaning:")
display(df_clean.head())
```

Jumlah data sebelum remove duplicate: 1000  
Jumlah data setelah remove duplicate: 867

Figure 4: Remove Duplicated Process Results

Figure 3 shows the amount of data after deleting entries that have duplicate values based on the date column. From the initial total of 1000 data, after the deduplication process, 867 data remained. This shows that there are 133 duplicated review data.

b. *Cleaning*

This step includes the process of removing special characters such as punctuation, numbers, emojis, and other symbols that are not the letters of the alphabet. The goal is to ensure that the text consists only of words that have meaning.

Table 1: Cleaning Stages in Reviews

Not	Before Cleaning	After Cleaning
1	Please play online	Please play online
	It's a good one, and TPI is still lagging.	It's a good idea and it's still lag
2	Indonesian commentator	Indonesian Commentator
	Dong...	Dong
3	Hopefully efootball will get better	Hopefully eFootball will get better

c. *Folding Case*

The first step in processing is to convert the entire text of the review to lowercase.

Table 2: Case Folding Stage

Not	Before CF	After CF
1	In his movements there is less maximum, virtual...	in his movements less maximum, virtual...
2	Please don't play online good network, and TP...	Please don't play online good network, and tp...
3	There are so many silly bugs...	There are so many silly bugs...

d. *Normalization of Text*

Normalization of texts, known as suspects, has the purpose of normalization, namely non-standard words to be standard, providing a substitute for misguided or non-standard words. Often the replacement is in the form of slang, abbreviations, or informal words with a more general form

Table 3: Text Normalization Stage

Not	Before Normalizing Text	Setelah Text Normalization
1	In his movements are less than optimal,	Less Than Enough Virtual Movement

	virtually...	good
2	It's a good game, but it's still lacking because not support...	It's a good game with no gamepad support. Instal apk j...
3	It's a good game, but it's still lacking support...	It's a good game with no gamepad because not Instal apk j...

e. *Tokenization*

At the tokenizing stage, the text will be broken down into parts.

**Table 4:** Tokenizing Yield Stage

Not	Before Tokenizing	After Tokenizing
1	Please play online a good network, and TP Still Alive	(Please, please, please, network, good, tp, lag, trus]
2	It's a good game, but it's still lacking because of the lack of support. gamepad and having to install the 3rd apk so it's complicated	(Game, Good, Less, GA, Support, gamepad, install, apk, so, Sigh]
3	If you don't want to be able to use a mobile app like the FCC... Such an emotion sticker. Humans	(Scott, mobile, sticker, emot, abandoned, tau, lwan, ny, indeed, humans]

f. *Remove the stoppage*

This study eliminated stopwords by utilizing an Indonesian special text processing library called Sastrawi.

**Table 5:** Stages of Remove Stopwords

Not	Before Removing Breakdown	After Remove Stopwords
1	In the movement it is less than optimal, the virtual	less than optimal movement less
2	Why after the update is worse, the response is less	After the update it gets worse Lack of response
3	Indonesian commentator dong	Indonesian commentator

g. *Mood*

The researchers applied the Literary literature in this study to perform stemming operations for Indonesian language processing.

**Table 6:** Voting Stage Results

Not	Before the Vote	After Stemming
1	In his movements are less than optimal, Virtual	At the very least, it is virtual. Less good
2	Why is it that after the update, it gets worse? Less response	How do you make updates worse? Less response

3	Indonesian commentator dong	Commentator in Indonesian
---	-----------------------------	---------------------------

### 3.4. Labeling

The sentiment labeling process involves three categories, namely negative, neutral, and positive. Labeling depends on the rating or score given to the review. Labels are given based on a review score between 1 and 2 for negative sentiment and 3 for neutral reactions as well as a score between 4 and 5 for positive categories.

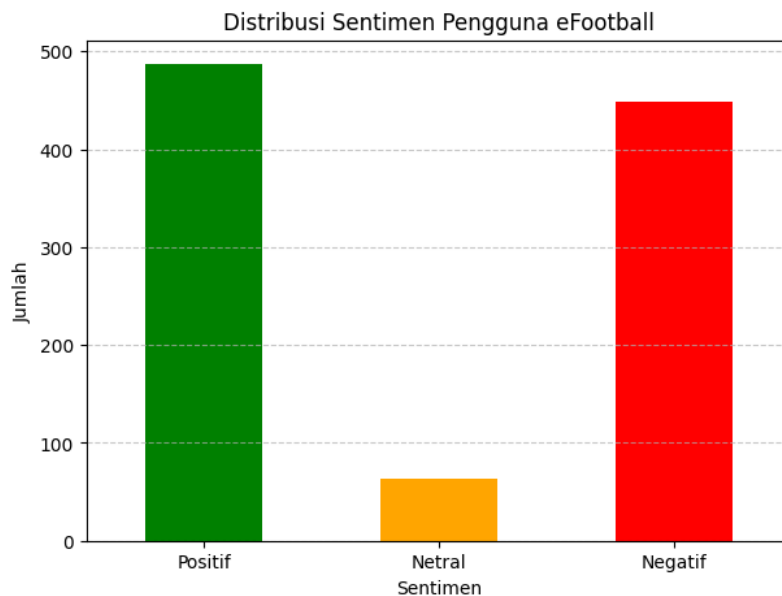


Figure 5: Number of eFootball User Label Distributions

Figure 5 shows the rating distribution for eFootball reviews. A total of 487 reviews were labeled positive (48.7%), indicating that the majority of users were satisfied with the game. There were 449 negative reviews (44.9%), reflecting a significant proportion of dissatisfied users. Meanwhile, 64 reviews were given a neutral rating (6.4%), the smallest number of the three categories, indicating users with opinions that were neither too positive nor negative.

### 3.5. Modeling

The modeling process will be carried out using the Support Vector Machine scenario.

### 3.6. Support Vector Engine

This dataset has 1000 reviews, of which there are 487 positive reviews, 449 negative reviews, and 64 neutral reviews. The results of both scenarios will be used to evaluate the model's performance on accuracy, precision, recall, f1-score, and other evaluation metrics and to analyze the impact of the relevance of the review to the sentiment analysis conducted.

### 3.7. Separate Datasets

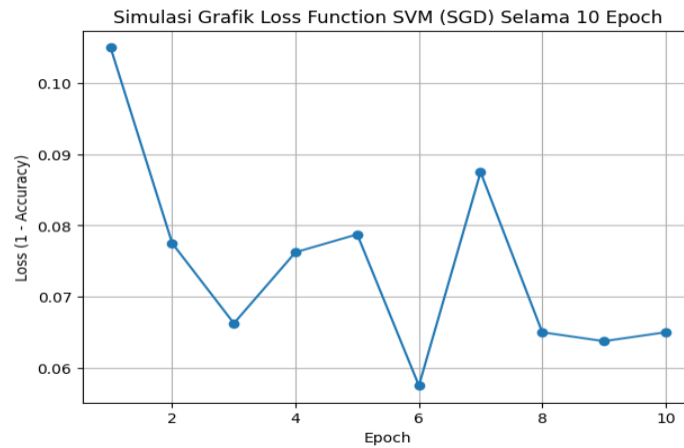
The data sharing process in this study is divided into three parts, namely data training, data validation, and data test. With a proportion of 80:10:10 where 80 percent is for data train, 10 percent for data validation, and 10 percent for test data.

### 3.8. Evaluation

In the evaluation stage, the analysis of the SVM model with several metrics such as: accuracy-loss function graph, confusion matrix, and wordcloud visualization.

#### 1. Loss Chart Function

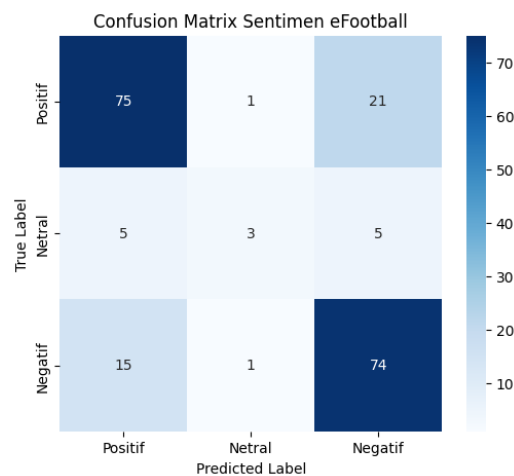
To assess the consistency of the model's performance, we will examine the extent of the graph of the *Loss Function* function on the validation data. As for the number of epochs, we will set them to 10 epochs, applying an early stop to prevent the model from overfitting.



**Figure 6:** Loss Function Chart Results

## 2. Confusion Matrix

The *confusion matrix method* is used to evaluate the model's performance after it is applied. The result of the confusion matrix in the *Support Vector Machine* scenario is as follows.



**Figure 7:** Confusion Matrix Results on Support Vector Machine

In Figure 7, we can see that out of the 90 positive reviews, the model correctly classified 75 True Positive (TP) out of 90, while 1 review was misclassified as neutral (False Neutral) and 21 were classified as negative (False Negative). In addition, out of 13 neutral reviews, the model correctly classifies 3 as True Neutral. However, he misclassified 5 as positive (False Positive) and 5 as negative (False Negative). Likewise, out of 97 negative reviews, the model correctly classified 74 as True Negatives (TN); however, 15 were misclassified as positive (FP) and 1 as neutral (FN).

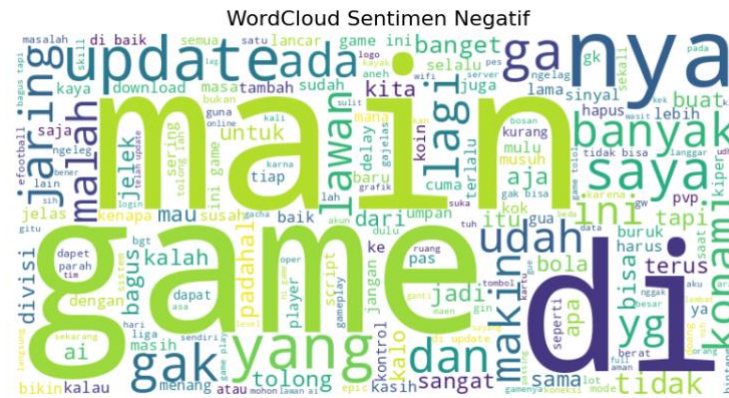
Classification Report:				
	precision	recall	f1-score	support
Positif	0.74	0.82	0.78	90
Netral	0.60	0.23	0.33	13
Negatif	0.79	0.77	0.78	97
accuracy			0.76	200
macro avg	0.71	0.61	0.63	200
weighted avg	0.75	0.76	0.75	200

**Figure 8:** Classification Report Results

Referring to Figure 8, the results of the classification report are also presented which show that the model managed to achieve an accuracy of 76% overall, or in other words, the performance of the model can be said to be satisfactory. Furthermore, the







As a result of the analysis from the visualization in eFootball, we can pull five keywords that represent positive sentiment, namely play, good, game, please, and steady. The next five keywords represent neutral sentiment, namely, play, update, league, can, good. And the last five keywords on negative sentiment are, network, opponent, game, bad, and Konami.

## 4. Conclusion

From the above process, it can be concluded that this research successfully applied the Support Vector Machine method to conduct sentiment analysis. The research process includes data collection, exploration, text preprocessing, sentiment labeling, modeling, and model evaluation. Based on the results of the analysis, most of the reviews were obtained from users with a positive sentiment of 48.7%, followed by negative sentiment of 44.9% and neutral 6.4%. The SVM model that was built was able to achieve an accuracy of 76% with good precision, recall, and f1-score values, especially in the positive and negative sentiment categories. This proves that SVM is effectively used for sentiment classification in digital game review data, although performance in the neutral category needs to be improved.

## 5. Suggestion

1. Improved data quality  
To improve the performance of the model, the amount of data in the neutral category can be multiplied so that the distribution of the data is more even, so that the model will more easily recognize the pattern of the neutral sentiment pattern
2. Feature enrichment  
In addition to the text feature, other features such as review time, the number of likes on comments, or the analysis of certain aspects of the sentiment level of the game's features, performance, graphics can provide more specific details and insights into user sentiment.
3. Implementation on a larger scale.  
This analysis can be extended by taking data from different platforms, such as App Store, forums, social media to provide a more comprehensive sentiment forecast.
4. Application of analysis results  
The results of sentiment analysis can be used by eFootball game developers, to review features that have not been maximized, and increase user satisfaction based on feedback that is automatically classified as analysis results.

## Reference

- [1] Al Ayyubi, Bahaestic. "Millennial Digital Fiqh: Gacha System in the Online Game Efootball Champion Squads." *At-Ta'awun: Journal of Mu'amalah and Islamic Law* 2.2 (2023): 230-242.
- [2] Ahmad, Kiki. "Analysis of Akulaku and Kredivo Online Loan Sentiment Using the Support Vector Machine (SVM) Method." *Journal of Mandalika Literature* 4.4 (2023): 323-332.
- [3] Atmajaya, Dedy, Annisa Febrianti, and Herdianti Darwis. "SVM and Naive Bayes Method for ChatGPT Sentiment Analysis on Twitter." *The Indonesian Journal of Computer Science* 12.4 (2023).
- [4] Alkindi, Aditia Fikri, and Nurliana Nasution. "Analysis of User Review Sentiment on Roblox Games Using Support Vector Machine and Naive Bayes Methods." *J-Com (Journal of Computer)* 4.2 (2024): 164-177.
- [5] Arrafiq, Muhammad Sunni. "User Sentiment Analysis of Seabank Indonesia's Application Services on Instagram Using the Support Vector Machine Method." *Journal of Information System Research (JOSH)* 5.4 (2024).
- [6] Darwis, Dedi, Eka Shintya Pratiwi, and A. Ferico Octaviansyah Pasaribu. "Application of the SVM Algorithm for Sentiment Analysis on Twitter Data of the Corruption Eradication Commission of the Republic of Indonesia." *Journal of Edutic Science: Education and Informatics* 7.1 (2020): 1-11
- [7] Fide, Sola, Suparti Suparti, and Sudarno Sudarno. "Analysis of tiktok application review sentiment on google play using the support vector machine (svm) and association method." *Gaussian Journal* 10.3 (2021): 346-358
- [8] Fitriyana, Vina, et al. "Sentiment Analysis of Jamsostek Mobile Application Reviews Using the Support Vector Machine Method." *Journal of Informatics* 14.01 (2023): 40-49.
- [9] Gusrananda, Muhammad Raufan. *Fanatical and Consumptive Behavior of Online Efootball Game Players*. Diss. Islamic University of Indonesia, 2023.
- [10] Haris, Muhammad, Aries Suharso, and E. Haodudin Nurkifli. "SENTIMENT ANALYSIS ON EFOOTBALL GAMES ON GOOGLE PLAY STORE

- USING THE INDOBERT ALGORITHM." JATI (Journal of Informatics Engineering Students) 8.6 (2024): 12108-12121.
- [11] Hananto, April Lia, et al. "Digital twins and 3D digital twins: concepts, applications, and challenges in industry 4.0 for digital twins." *Computer* 13.4 (2024): 100.
  - [12] Hananto, April Lia, et al. "Elman and Cascade Neural Network with Conjugate Gradient Patterns Restarted to Predict Diesel Engine Performance and Butanol-Triggered Emissions as a Sustainable Biofuel." *Engineering Results* 19 (2023): 101334.
  - [13] Hananto, April Lia, et al. "Analysis of drug data mining using the K-Means algorithm." *Journal of Physics: Conference Series*. Vol. 1908. No. 1. IOP Publishing, 2021.
  - [14] Haris, Muhammad, Aries Suharso, and E. Haodudin Nurkifli. "SENTIMENT ANALYSIS ON EFOOTBALL GAMES ON GOOGLE PLAY STORE USING THE INDOBERT ALGORITHM." JATI (Journal of Informatics Engineering Students) 8.6 (2024): 12108-12121.
  - [15] Kusumah, Andika Juda. *Analysis of Fiqh Al-Bai' on the buying and selling of players on the eFootball Game application with a Random System*. Diss. UIN Sunan Gunung Djati Bandung, 2024.
  - [16] Lidinillah, Edo Ridho, Tatang Rohana, and Ayu Ratna Juwita. "Analysis of twitter sentiment towards steam using logistic regression algorithm and support vector machine." *TECHNOSCIENCE: Journal of Science, Technology and Informatics* 10.2 (2023): 154-164.
  - [17] Mubaroroh, Hindun Habibatul, Hasbi Yasin, and Agus Rusgiyono. "Sentiment Analysis of Ruangguru Application Review Data on Google Play Site Using Naïve Bayes Classifier Algorithm with Normalization of the Word Levenshtein Distance." *Gaussian Journal* 11.2 (2022): 248-257.
  - [18] Pamungkas, Adji Surya, and Nuri Cahyono. "Analysis of ChatGPT Review Sentiment on Play Store using Support Vector Machine and K-Nearest Neighbor." *Edumatic: Journal of Informatics Education* 8.1 (2024): 1-10.
  - [19] Pangestu, Adhi, Yoseph Tajul Arifin, and Rizky Ade Safitri. "Analysis of Public Review Sentiment of Online Game Users on the Steam Platform Using Naïve Bayes Algorithm." JATI (Journal of Informatics Engineering Students) 7.6 (2023): 3106-3113.
  - [20] Rahman, Rohim Nur, Abdul Rahim, and Wawan Joko Pranoto. "Sentiment Analysis of eFootball Game Reviews 2024 On Playstore using Naïve Bayes Algorithm." *SCIENTIFIC JOURNAL OF INFORMATICS* 13.01 (2025): 38-44.
  - [21] Saputra, Hendi. "Sentiment analysis on booster vaccines using the support vector machine multiclass algorithm on Twitter." *Journal of Smart Technology* 3.10 (2023).
  - [22] Syah, Herwin, and Arita Witanti. "Analysis of Public Sentiment Towards Covid-19 Vaccination on Twitter Social Media Using Support Vector Machine (SVM) Algorithm." *Journal of Information Systems and Informatics (Simika)* 5.1 (2022): 59-67.
  - [23] Tukino, Tukino, and Fifi Fifi. "Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Layanan Ojek Online." *Jurnal Desain Dan Analisis Teknologi* 3.2 (2024): 104-113.
  - [24] Ulfah, Amanny, and Ina Najiah. "Implementation of Web Scraping on the Sinta Journal Site Using the Selenium Webdriver Python Framework." *JIKA (Journal of Informatics)* 7.1 (2023): 29-36.
  - [25] Utami, Mar Atul Aji Tyas, Pika Silvianti, and Muhammad Masjkur. "Support Vector Machine Algorithm for Aspect-Based Sentiment Analysis of Mobile Legends: Bang-Bang Online Game Review." *Xplore: Journal of Statistics* 12.1 (2023): 63-77.