

## CLASSIFICATION OF MYPERTAMINA APP REVIEWS USING SUPPORT VECTOR MACHINE

Alwan Fadlurohman<sup>\*1</sup>, Novia Yunanita<sup>2</sup>, Febrian Hikmah Nur Rohim<sup>3</sup>, Amelia Kusuma Wardani<sup>4</sup>, Ariska Fitriyana Ningrum<sup>5</sup>

<sup>1,2,3,4,5</sup>Undergraduate of Data Science Study Program, Universitas Muhammadiyah Semarang  
Jl. Kedungmundu Raya No. 18, Kedungmundu, Subdistrict of Tembalang, Semarang, Central Java

*Corresponding Author's Email: alwan@unimus.ac.id*

---

**Abstract:** Indonesia is rich in natural resources, including oil and gas, and it manages these strategic assets through state-owned enterprises, one of which is PT Pertamina. Pertamina is responsible for domestic fuel production, distribution, and price stabilization. To improve efficiency and transparency, Pertamina developed the MyPertamina application that enables cashless fuel purchases, stock monitoring, and up-to-date price information. The application aims to streamline distribution and control fuel prices, thus helping to stabilize the cost of goods and services. MyPertamina also ensures subsidized fuel distribution is more effective and targeted by identifying and verifying subsidy recipients, reducing the potential for abuse. A sentimental analysis of subsidized fuel user reviews using this application is needed to understand the public's views. This research uses the Support Vector Machine (SVM) method to analyze the sentiment of MyPertamina app reviews. This research produced a stable model. Out of 200 reviews, 190 were negative, and nine were positive, with an SVM model accuracy of 97%. Wordcloud visualization shows the words that appear frequently in each sentiment. Positive reviews appreciated the photo verification feature, easy payment, and good service. Negative reviews included verification difficulty, app error, and feature failure.

**Keywords:** Classification, MyPertamina, Sentiment Analysis, Support Vector Machine.

---

### 1. INTRODUCTION

Fuel oil (BBM) is an essential necessity for society in the modern era. Fuel use in various sectors, such as households, industry, and transport, plays a vital role in determining fluctuations in the prices of goods and services. A spike often follows any increase in fuel prices in the price of necessities as fuel is a major component in the cost of production and distribution of goods [1]. Conversely, decreasing fuel prices can reduce inflationary pressures and lower the prices of goods the community needs. This is particularly relevant in Indonesia, which is rich in natural resources, including oil and gas mines, which are considered strategic assets. The state manages these resources through State-Owned Enterprises (SOEs). One of Indonesia's SOEs managing oil and gas mining is PT Pertamina. PT Pertamina is responsible for the production and distribution of fuel in the country and plays a vital role in fuel price stabilization. PT Pertamina has developed a digital application called MyPertamina to improve efficiency and transparency. This application allows consumers to access various services, such as cashless fuel purchases, fuel stock monitoring, and the latest price information. With this application, it is expected that fuel distribution will be more efficient and fuel prices more controlled, which can help stabilize the prices of goods and services in the market [2].

PT Pertamina created the MyPertamina application to ensure that the distribution of subsidized fuel, such as diesel and pertalite, is more effective and targeted. This application requires subsidy recipient vehicles to register their personal and vehicle data. With this system, Pertamina can monitor and ensure that fuel subsidies are received by those entitled [2], [3]. PT Pertamina has also introduced various innovations in the payment system to provide convenience to consumers in transactions. MyPertamina allows users to make digital payments with loyalty point bonuses, which can be used to get attractive offers. The application also provides a monthly fuel purchase recording system so users can easily track their fuel consumption [3]. In addition to these features, MyPertamina also provides a service to find the location of the nearest Pertamina petrol station, making it easier for users to find a place to refuel.

The features in MyPertamina serve as a complete solution for fuel users, providing convenience, speed, and security when carrying out vehicle fuel purchase transactions. This application allows PT Pertamina to collect data valuable for planning and managing future fuel distribution and improving customer service with more structured feedback. In addition to being a medium for transactions, the MyPertamina application is also a means for consumers to get a reward and vouchers that can be used for transaction benefits at various Pertamina petrol stations. MyPertamina application can be downloaded on Android and IOS operating systems through Google PylaStore and App Store [4].

Through this application, the government can identify and verify subsidy recipients more accurately and reduce the potential for subsidy abuse. To understand the public's opinion about this policy, sentiment analysis of subsidized fuel user reviews using the MyPertamina application is required. This research was conducted due to the urgent need to ensure that the fuel subsidy policy is effective and well-targeted [5]. User reviews of the MyPertamina app show that many users give unfavorable feedback, indicating problems in policy implementation and app usage.

To analyze this problem more accurately and effectively, this research uses the Support Vector Machine (SVM) method, which has the advantage of producing smoother and more stable models [6]. SVM generates smoother, more stable models for handling large and varied review data. The main advantage of SVM is its ability to cope with the problems of narrow margins and irregular data, which are often found in user reviews [7], [8]. By minimizing a smoother objective function, SVM can reduce overfitting and produce more accurate and consistent predictions.

Research conducted by Wahyudi and Kusumawardana [6], which analyzed sentiment on Grab application review data using the SVM algorithm, showed an accuracy rate of 85.54%. The word 'ovo' was identified as a word that often appears in positive reviews, while the word 'driver' often appears in negative reviews. Another study involving sentiment analysis using the SVM method was conducted by Ilmawan and Mude [9], comparing the performance of the SVM and Naïve Bayes methods. The results showed that the SVM method had the highest accuracy rate, 81.46%, compared to the Naïve Bayes method, which reached 75.41%. In addition, sentiment analysis of Zoom Cloud Meetings application review data using the SVM method resulted in an accuracy rate of 81.22%, while the Naïve Bayes method only reached 74.37% [10].

Based on the above results, this research will focus more on analyzing the sentiment of MyPertamina application user reviews using the SVM method. The results are expected to provide valuable insights for the government and PT Pertamina in improving and optimizing services so that the distribution of fuel subsidies can run more effectively and application users feel more satisfied.

## **2. METHODOLOGY**

### **2.1. Data Description**

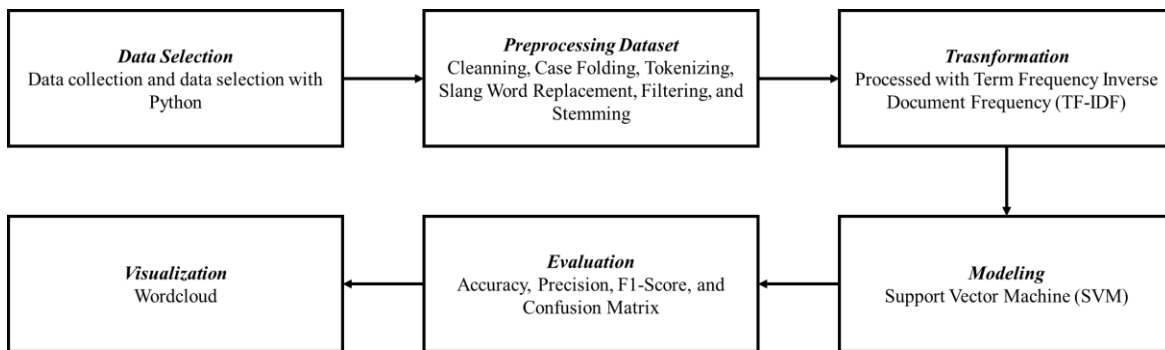
The data used in this study were taken from Indonesian reviews of the MyPertamina application. The selected reviews are then separated into two categories: positive reviews and negative reviews. Positive reviews are responses or those with a rating of 4-5 on the MyPertamina application, while negative reviews are responses with a rating of 1-3 on the MyPertamina application.

### **2.2. Research Methods**

In this research, data collection techniques are carried out through web scraping of the MyPertamina application on the Google Play Store to collect information about downloading the application. This process aims to identify the usage pattern of a person's MyPertamina application. After the data is collected through web scraping from the Google Play Store, the next step is to preprocess it to ensure its quality and usefulness for further analysis. The preprocessing process starts with data cleaning, which removes duplicates, fills in missing values, and fixes inconsistencies in the data [11]. After that, additional relevant features are added to enrich the information. Then, the processes of tokenization and lemmatization are applied to the text data to break down sentences into individual words and convert the words to their base form, which is useful for further text analysis [12]. Stopwords or common words that do not provide significant meaning, such as "and," "which," and "in," are

also removed to focus on essential words [13]. Finally, exploratory data analysis (EDA) is performed to understand the distribution of data and identify patterns, anomalies, and relationships between features [14]. By performing this comprehensive preprocessing, the resulting data is ready for further analysis and the construction of more accurate predictive models, ensuring that all categories of generated data are treated fairly and informatively.

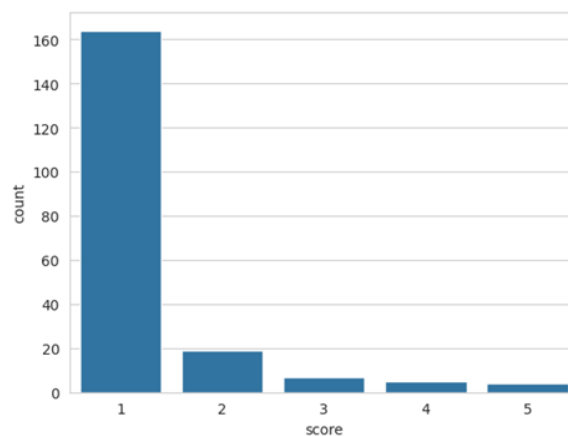
Once the preprocessing stage is complete, the next step is to build a predictive model to analyze the processed data. The modeling process begins by using the SVM algorithm. The next step is to train the model using the training data and adjust the hyperparameters to optimize model performance. Model evaluation is performed using various metrics such as accuracy, precision, recall, and F1-score to evaluate model performance and ensure that the model can predict accurately [15]. Accuracy measures how often the model gives correct predictions overall, while precision measures how often the model's optimistic predictions are accurate. Recall measures how the model can usually identify positive classes overall. F1-score is the harmonic mean of precision and recall, which gives an overall picture of the model's performance. The Confusion Matrix provides a visual representation of the model's performance in correctly predicting each class [16]. By conducting a comprehensive evaluation using these metrics, we can better understand how well our model predicts MyPertamina user data from the Google Play Store. Careful evaluation will help ensure that the developed model can provide accurate and valuable insights for further app development. The research flow described above can be found in Figure 1.



**Figure 1. Stages of Research**

### 3. RESULTS AND DISCUSSION

The data used are reviews from users of the MyPertamina Application. In the cloud stage, descriptive analysis is carried out to show satisfaction reviews from users of the MyPertamina application.



**Figure 2. User Rating on MyPertamina App**

Figure 2 shows that the use of the MyPertamina application is not fully maximized; it could be noticed that some users give a rating of 1 or feel dissatisfied with the application provided.

The training data that is tested is taken through the scraping process shown in Figure 3. Comments include several columns such as username, score 1-5, date, and user comment content on a MyPertamina application using the Python programming language, and the data obtained is 200 data. Review data obtained will be labeled Positive and negative labels, scores 1-3 will be labeled negative, and scores 4-5 will be labeled positive.

	userName	score	at	content
121	Rizal Fadhil	1	2024-03-22 08:15:33	Aplikasi apa ini... Daftar subsidi tepat 5x ga...
122	agny ikh	3	2024-03-22 06:13:10	SUGGESTION: Pada halaman ketika mengisi biodat...
123	Djumiddin A6	2	2024-03-22 05:30:10	Lama betul verifikasi kendaraan nya, kasian ki...
124	Totok Riyanto	1	2024-03-21 03:33:38	App nya diperbaiki disaat mau masuk barcode h...
125	Adhy Bee	1	2024-03-20 08:54:25	Susah payah antri tiba-tiba barcode tidak bisa di...
126	Mutiara Putri Allolangi	1	2024-03-19 11:21:03	Tolong kalau buat aplikasi jangan menyusahkan ...

Figure 3. Scrapping MyPertamina App Rating Data

After the data is obtained, data preprocessing will be carried out in various stages, such as eliminating columns that are not needed, case folding, tokenizing, normalization, and stemming, to get the results we want. The results of preprocessing MyPertamina application comment review data can be clearly understood by visualizing it in positive and negative opinions. This visualization using a word cloud will display the words that appear the most or words that often occur in a sentiment review. The more the word appears or is used when giving a review, the larger the size of the word displayed in the word cloud visualization. The results of the positive sentiment word cloud are shown in Figure 4.



Figure 4. Positive Sentiment Wordcloud Results

Based on the visualization of Figure 4, it is noticed that the positive sentiment reviews have some prominent words such as can, appropriate, photo, me, payment, service, listening, and others that show positive opinions towards MyPertamina app reviews. Therefore, it could be hypothesized that this application makes it easy for users to do various things ("can"), appreciate the verification or documentation feature through photos, easy payment, and good officer service to users. The visualization results for negative sentiment can be seen in Figure 5.



Figure 5. Negative Sentiment Wordcloud Results

The results of negative sentiment reviews produce several prominent words such as can't, no, error, update, please, long transaction, always, only, and several other words that show negative opinions on MyPertamina application reviews. Therefore, it can be hypothesized that this application has many complaints from users, such as difficulties in the verification process, frequent errors, and "can't," indicating a failure of the function or feature of the application that users want to use.

The accuracy of the SVM is tested in the model evaluation process. Table 1 shows the results of the SVM model's precision, recall, and F1-Score measurements.

**Tabel 1. Classification Report**

	Precision	Recall	F1-Score
Negative	0.97	1.00	0.98
Positive	0.00	0.00	0.00
Accuracy			0.97

The precision value indicates the number of correctly classified 0 (negative) value category data against the total positive classified data. Recall shows the percentage of positive category data the system correctly classified. The F1-Score value combines the recall and precision values. Table 1 shows that the accuracy value in sentiment analysis of MyPertamina application comment reviews using the SVM model is 97%. Positive sentiments that can be used as a hypothesis reference are that this application allows users to do various things ("can"), appreciate the verification or documentation feature through photos, easy payment, and good worker service to users. As for the negative sentiment, it states that this application has many complaints from users, such as difficulties in the verification process, frequent errors, and "can't" indicates a failure of the function or application feature that users want to use.

#### 4. CONCLUSIONS

The sentiment analysis of user comment review data on the MyPertamina application using the SVM algorithm is carried out at stages ranging from Data Understanding, Data Preprocessing, Word visualization, Data Preparation, and Modeling to Testing. The accuracy of the model produced using the SVM method gets a value of 97%. Positive sentiments that can be used as a hypothesis reference are that this application allows users to do various things ("can"), appreciates the verification or documentation feature through photos, easy payment, and good worker service to users. As for the negative sentiment, it states that this application has many complaints from users, such as difficulties in the verification process, frequent errors, and "can't" indicates a failure of the function or application feature that users want to use.

#### REFERENCES

- [1] R. Islamia, I. R. Al Faizy, A. Aqilla, R. F. Ahmad, A. zahra P. Arum, and G. Pratama, "Dampak Kenaikan Harga Bahan Bakar Minyak (BBM) Terhadap Sembilan Bahan Pokok (Sembako) di Toko Sani Kabupaten Cirebon," *Jurnal Ekonomi Manajemen*, vol. 17, no. 2, pp. 1–7, 2022.
- [2] C. G. indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 7, no. 3, pp. 1131–1139, 2023.
- [3] F. M. Oktaviana, D. Wijayanto, and T. Wahyudi, "Pengaruh Social Marketing Campaign Terhadap Keputusan Konsumen Bertransaksi Menggunakan App MyPertamina Di Pontianak," *Jurnal Teknik Industri*, vol. 5, no. 1, pp. 23–29, 2021.
- [4] R. M. Ibrahim and N. N. K. Moeliono, "Pengaruh Manfaat, Kepercayaan, Efikasi Diri, Kemudahan Penggunaan, Keamanan Terhadap Persepsi Konsumen Pada My Pertamina (Studi Pada Penggunaan My Pertamina Kota Bandung)," *Jurnal Ilmiah Mahasiswa Ekonomi Manajemen*, vol. 5, no. 2, pp. 396–413, 2020.

- [5] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 42–48, 2023.
- [6] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *Jurnal Informatika*, vol. 8, no. 2, pp. 200–207, 2021.
- [7] I. Santoso, W. Gata, and A. B. Paryanti, “Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum,” *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 3, no. 3, pp. 364–370, 2019.
- [8] N. Arifin, U. Enri, and N. Sulistiyowati, “Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, pp. 129–136, 2021.
- [9] L. B. Ilmawan and M. A. Mude, “Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store,” *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 154–161, 2020.
- [10] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and Samudi, “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine,” *CESS (Journal of Computer Engineering System and Science)*, vol. 5, no. 2, pp. 293–298, 2020.
- [11] V. Kotu and B. Deshpande, *Data science: concepts and practice*. Cambridge: Morgan Kaufmann, 2018.
- [12] J. Li, A. Sun, J. Han, and C. Li, “A Survey on Deep Learning for Named Entity Recognition: Extended Abstract,” in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 2023, pp. 3817–3818. doi: 10.1109/ICDE55515.2023.00335.
- [13] H. Jelodar et al., “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimed Tools Appl*, vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.
- [14] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel, and K. C. Lichtendahl, *Data mining for business analytics: concepts, techniques, and applications in R*. New Jersey: John Wiley & Sons, 2017.
- [15] S. Raschka and V. Mirjalili, *Python machine learning: Machine learning and deep learning with Python, sci-kit-learn, and TensorFlow 2*. Birmingham: Packt Publishing Ltd, 2019.
- [16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*, 3rd ed. Morgan Kaufmann, 2011.