Journal of Applied Informatics and Computing (JAIC)
Vol.8, No.2, December 2024, pp. 516~522
e-ISSN: 2548-6861

Sentiment Analysis of Online Vehicle Tax Renewal Application Users Using Support Vector Machine Algorithm

Muhamad Ilham Fauzy [1]*, Ferian Fauzi Abdulloh [2]*
* Informatika, Universitas Amikom Yogyakarta
ilhamfauzy123@students.amikom.ac.id [1], ferian@amikom.ac.id [2]

Article Info

Article history:

Received 2024-10-07
Revised 2024-11-12
Accepted 2024-11-14

Keyword:

Sentiment Analysis,
Support Vector Machine,
Online Tax Renewal,
Data Review,
Machine Learning.

ABSTRACT

This study examines user sentiment towards online vehicle tax renewal applications by utilizing the Support Vector Machine (SVM) algorithm. The data was collected from user reviews on the Google Play Store for three major applications: New Sakpole, Sapawarga, and Timsalut. The reviews were preprocessed through steps including normalization, case folding, tokenization, and stopword removal. The SVM algorithm was then applied to classify the reviews into positive or negative sentiments. A comparative analysis was performed with K-Nearest Neighbors (KNN) and Naïve Bayes, with SVM demonstrating the best performance, achieving an accuracy of 76.5%. In addition to accuracy, metrics such as precision, recall, and F1-score were also evaluated to provide a more comprehensive assessment of the models. The results indicate that while these applications help facilitate vehicle tax payments, there remains significant user dissatisfaction, particularly related to technical issues and usability concerns. This study offers valuable insights for application developers, highlighting areas for improvement in functionality and user experience to better meet public expectations.

This is an open access article under the CC–BY-SA license.

I. INTRODUCTION

To increase regional revenue derived from Local Own-Source Revenue (PAD), particularly from local taxes used to finance household expenditures and development projects [1], several efforts are required. These include improving services, enhancing tax collection performance, expanding the types of taxes (tax base broadening), and granting local governments the authority and flexibility to explore each region's potential [2]. One of the most promising sources of local revenue is the Motor Vehicle Tax, which holds significant potential [3]. Therefore, local governments or regional revenue agencies play a vital role in monitoring the increasing number of motor vehicles, which continues to rise in various regions.

With the constant growth in the number of motor vehicles, efficient and systematic management becomes increasingly crucial. From administrative aspects like tax processing and vehicle registration renewals to safety and vehicle monitoring, the demand for an integrated platform has become more urgent. According to data from the Central Statistics Agency, in 2022, the number of motor vehicles in Indonesia reached 148 million units, with an annual growth trend [4]. This situation presents significant challenges for both the government and society regarding effective management and services.

Although the government has launched various applications aimed at facilitating tax payments [5], many issues persist, as reflected by the abundance of negative feedback in user reviews of these tax renewal apps. This indicates that the application developers have not yet fully understood user needs and preferences.

Given these challenges, which may explain the lack of progress in app development, this study seeks to analyze user reviews of tax renewal applications [6]. The researcher collected sentiment data from three key apps New Sakpole, the tax renewal app for Central Java [7]; Sapawarga, used for tax renewal in West Java [8]; and Timsalut, the tax renewal app for North Sulawesi [9]. The sampled reviews were categorized into two sentiments: positive and negative [10].

In previous studies, the Support Vector Machine (SVM) algorithm proved to be the most effective method for sentiment analysis [11]. This research compares the performance of three algorithms SVM, KNN, and Naïve Bayes (NB) with respective accuracy rates of 76.5%, 59.1%, and 72.3% [12]. By utilizing the SVM algorithm, which has demonstrated high accuracy in sentiment analysis, the researcher aims to provide insights for developers to improve and enhance the applications to meet user expectations.

II. Method

The research methodology serves as a guide, outlining the plans and steps to be taken in the research to achieve the specified objectives [13]. This study employs the Support Vector Machine (SVM) method for sentiment analysis. The research flowchart is illustrated in Figure 1.

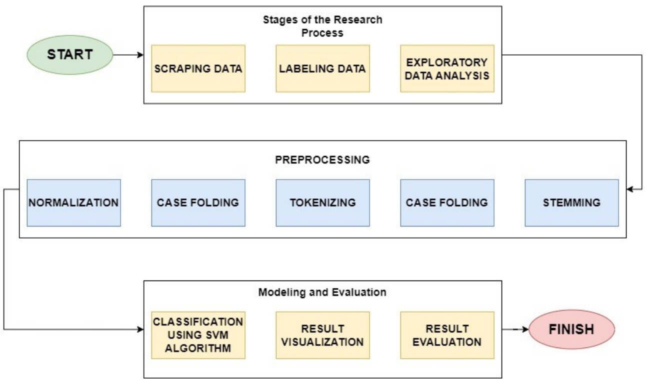

Figure 1. Research Methodology

A. Scraping Data

This study utilizes user reviews data, which were collected through data scraping techniques on the Google Play Store [14]. The researcher employed the *google-play-scraper* library to conduct this step. A total of 3,088 randomly selected data points were gathered in stages.

B. Labeling Data

The data obtained from scraping were labeled as either positive or negative. The researcher collaborated with several labelers to ensure optimal results [13]. Scores were not used as a reference, as some reviews were considered irrelevant to the given score. Reviews labeled as negative typically include complaints, protests, or expressions of dissatisfaction with the application.

C. Exploratory Data Analysis (EDA)

The purpose of Exploratory Data Analysis (EDA) is to gain initial insights into the dataset before proceeding with more detailed research processes [15]. EDA allows for the discovery of initial details related to the data, such as data volume, completeness, the number of positive-labeled data, and the number of negative-labeled data.

D. Data Preprocessing

The data preprocessing stage involves crucial steps that need to be completed before building a text mining model [16]. The researchers conducted a brief analysis of the dataset's characteristics and implemented several preprocessing steps, including normalization, case folding, tokenization, and stemming.

1) Normalization: This step corrects non-standard words in user reviews to their standard form [10]. In addition to addressing abbreviations like "gk" (for "tidak"), the normalization process also handles common spelling errors and removes any unnecessary special characters, such as punctuation marks. These corrections ensure uniformity and reduce noise in the data. Examples of normalized words and corrections are shown in Table 1.

TABLE I
EXAMPLE OF NORMALIZE WORDS

No	Kata tidak baku/singkatan	Kata baku
1	gk, g, gak, ngak, nggak, gx, ga, tdk	tidak
2	tp, tapi, tpi, tetapi	namun
3	sdh, udh, udah	sudah
4	eror	error
5	dpt, dpat, dapet	belum
6	apk	aplikasi
7	jgn, jngn, jngan	jangan
8	krn, krna	karena
9	jos, mantap, mntp	sangat baik
10	utk,untk	untuk

2) Case Folding: This process converts all text in user reviews to lowercase [17], to ensure uniformity when processed by the model. the text was also cleaned for common typographical errors. Misspelled words were corrected using a predefined dictionary of standard words and common typos, ensuring that the text data was as clean as possible for sentiment analysis. The case folding process is depicted in Figure 2.

Figure 2. Case Folding Process

3) Tokenization: This step breaks down the text into individual words [18]. The goal is to represent the text in a way that the model can easily understand, allowing the algorithm to recognize patterns within the text. Figure 3 illustrates the tokenization process.

Figure 3. Tokenization Process

4) Stopword Removal: In software development, removing stopwords has been proven to enhance tool performance more effectively compared to using a standard stopword list, as demonstrated in previous research [19]. Stopwords are words that frequently appear but do not carry significant information, such as "and," "or," "is," and others.

The process of stopword removal aims to eliminate these words from the text, allowing the model to focus on more meaningful and relevant terms. By reducing this noise, the model can analyze the text more efficiently. This process is illustrated in Figure 4.

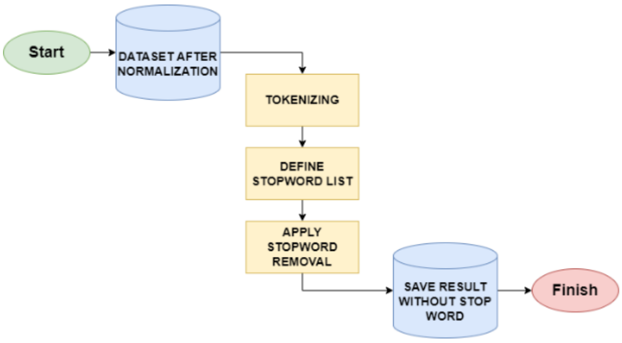

Figure 4. Stopword Process

5) *Stemming*: This process reduces inflected or derived words to their root forms [20], helping to minimize the number of unique words the model needs to process, thereby improving performance. Figure 5 shows the stemming process.

Figure 5. Stemming Process

E. *Classification Using SVM Algorithm*

Support Vector Machine (SVM) is a classic model for binary classification that works by finding the optimal hyperplane to separate available data samples [21]. In this study, the SVM algorithm is used to classify user reviews of the tax renewal application.

The choice of SVM in this study is due to its advantages in handling high-dimensional text data, which is typical in sentiment analysis. SVM is effective in maximizing the margin between different classes, making it robust for text classification tasks where the data might not be linearly separable. The use of kernel functions allows SVM to project data into higher dimensions, enabling better classification of complex data patterns. Given these strengths, SVM is well-suited for this study's goal of classifying user sentiment based on textual reviews.

The dataset is split into two parts: 80% for training data and 20% for testing data [22]. The training data are then transformed into a numerical representation using Count Vectorizer, which learns the vocabulary from the training data and generates features based on word frequency [23]. The same process is applied to the testing data, using the vocabulary learned from the training set. Hyperparameter optimization of the SVM algorithm is performed using *GridSearchCV* to test various parameter combinations, such as the value of C and the type of kernel, aiming to find the combination that yields the highest accuracy [24]. This process is shown in Figure 6.

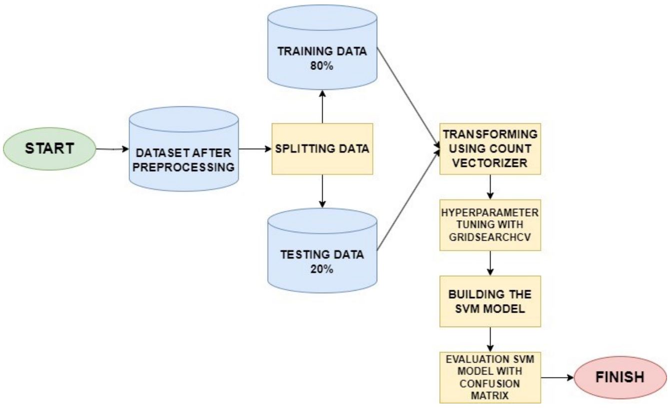

Figure 6. Classification Process Using SVM

F. *Visualization of Result*

For visualizing the results, the researcher used a word cloud, a visual tool that displays frequently occurring words [25]. The use of word clouds provides a clear picture of word frequency in user reviews.

G. *Evaluation of Result*

This study employs a Confusion Matrix to calculate accuracy by comparing the correct and incorrect predictions from the classification method against the actual target or predicted data [19]. The formulas for calculating accuracy, precision, and recall are as follows:

$$Accuracy = \frac{TP+TN}{Total} \tag{1}$$

$$precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The Confusion Matrix is used to calculate accuracy, precision, and recall, as outlined in the following formulas [26][27].

Explanation :

- TP (*True Positive*) = The number of positive class data correctly classified as positive.
- TN (*True Negative*) = The number of negative class data correctly classified as negative.
- FP (*False Positive*) = The number of negative class data incorrectly classified as positive.
- FN (*False Negative*) = The number of positive class data incorrectly classified as negative.

III. RESULT AND DISCUSSION

In this section, the researcher will present the data from the study's findings to demonstrate the success of this research.

A. Scraping Data

This study will utilize user review data obtained through data scraping techniques from the Google Play Store. The researcher will employ the *google-play-scraper* library to carry out this process. The data will be collected in a random and incremental manner, resulting in a total of 3,088 entries.

TABEL II
RESULT OF SCRAPING DATA

username	score	content
Rega A	2	Aplikasinya cukup mempermudah yang cuma mau bayar pajak kendaraan atas nama diri sendiri, tinggal masukan NIK, kemudian bayar online. Tapi ternyata tetap harus MANUAL DATANG ke SAMSAT untuk pengesahan. Jadi fungsi aplikasi ONLINE ini kesannya tidak ada gunanya? Semoga kedepannya ada perbaikan. Terima kasih.
Deden Andri Kurniawan	1	tolong diperbaiki ya,masa sinyal wifi bagus tapi di aplikasi nya *tidak ada koneksi internet* jadikan ribet mau pencet info tentang PPDB Jawa Barat,malah ke menu utama aja terus
Irvan Rudianto	2	Saya sudah melakukan pembayaran pajak ken-daraan, karena motor saya blm balik nama, saya menggunakan fitur cek kendaraan lain., pembayaran Menggunakan fitur qris di aplikasi dan Tertulis berhasil. Nominal saldo pun terpotong, ada email qris dan bapenda yg menerangkan bahwa pembayaran berhasil. Tetapi di histori pem-bayaran aplikasi , ada tulisan KODE BAYAR GAGAL. agak aneh sih, padahal pembayaran sudah berhasil dan saldo pun berkurang. Mungkin bisa di tingkatkan lagi aplikasi nya.

B. Labeling Data

The scraped data will be labeled as either positive or negative. The researcher collaborates with multiple annotators to ensure optimal labeling accuracy. Scores are not used as a reference point, as some reviews are considered irrelevant to the given scores. Reviews labeled as negative include complaints, protests, and negative sentiments such as disappointment and dissatisfaction with the application.

TABEL III
RESULT OF LABELING DATA

username	content	sentiment
Rega A	Aplikasinya cukup mempermudah yang cuma mau bayar pajak kendaraan atas nama diri sendiri, tinggal masukan NIK, kemudian bayar online. Tapi ternyata tetap harus MANUAL DATANG ke SAMSAT untuk pengesahan. Jadi fungsi aplikasi ONLINE ini kesannya tidak ada gunanya? Semoga kedepannya ada perbaikan. Terima kasih.	negatif
Deden Andri Kurniawan	tolong diperbaiki ya,masa sinyal wifi bagus tapi di aplikasi nya *tidak ada koneksi internet* jadikan ribet mau pencet info tentang PPDB Jawa Barat,malah ke menu utama aja terus	negatif
Irvan Rudianto	Saya sudah melakukan pembayaran pajak ken-daraan, karena motor saya blm balik nama, saya menggunakan fitur cek kendaraan lain., pembayaran Menggunakan fitur qris di aplikasi dan Tertulis berhasil. Nominal saldo pun terpotong, ada email qris dan bapenda yg menerangkan bahwa pembayaran berhasil. Tetapi di histori pem-bayaran aplikasi , ada tulisan KODE BAYAR GAGAL. agak aneh sih, padahal pembayaran sudah berhasil dan saldo pun berkurang. Mungkin bisa di tingkatkan lagi aplikasi nya.	negatif

C. Exploratory Data Analysis

The purpose of Exploratory Data Analysis (EDA) is to gain initial insights into the dataset before proceeding with more in-depth research processes. EDA helps uncover various preliminary details related to the data being examined, such as data volume, completeness, the number of positive labels, and the number of negative labels. During this phase, the researcher conducted an initial assessment of the dataset.

The dataset used in this study consists of 1,683 negative and 1,405 positive reviews, showing a relatively balanced distribution of sentiments. To ensure the classifier performs well for both classes, additional evaluation metrics such as precision, recall, and F1-score were used to provide a more comprehensive understanding of the model's performance beyond overall accuracy.

D. Preprocessing Data

The data preprocessing stage involves essential steps that must be completed before building a text mining model. Researchers conduct a brief analysis of the data characteristics by performing several preprocessing steps, including normalization, case folding, tokenization, and stemming.

TABEL IV
INITIAL DATA

Stage	Preprocessing Result
Initial Data	Aplikasinya cukup mempermudah yang cuma mau bayar pajak kendaraan atas nama diri sendiri, tinggal masukan NIK, kemudian bayar online\. Tapi ternyata tetap harus MANUAL DATANG ke SAMSAT untuk pengesahan\. Jadi fungsi aplikasi ONLINE ini kesannya tidak ada gunanya? Semoga kedepannya ada perbaikan\. Terima kasih

1) Normalization: This step involves converting non-standard words into their standard forms within user reviews. During the labeling process, the researcher observed that some words were either inaccurately spelled or abbreviated, prompting the use of normalization to standardize those words.

TABEL V
NORMALIZATION RESULT

Stage	Normalization Result
Normalization	Aplikasinya cukup mempermudah yang cuma mau bayar pajak kendaraan atas nama diri sendiri, tinggal masukan NIK, kemudian bayar online. Tapi ternyata tetap harus MANUAL DATANG ke SAMSAT untuk pengesahan. Jadi fungsi aplikasi ONLINE ini kesannya tidak ada gunanya? Semoga kedepannya ada perbaikan. Terima kasih

2) Case Folding: This process involves converting all text in the user reviews to lowercase. The purpose of this step is to ensure that the text processed by the model is in a uniform format, facilitating consistent analysis.

TABEL VI
CASE FOLDING RESULT

Stage	Case Folding Result
Case Folding	aplikasinya cukup mempermudah yang cuma mau bayar pajak kendaraan atas nama diri sendiri tinggal masukan nik kemudian bayar online tapi ternyata tetap harus manual datang ke samsat untuk pengesahan jadi fungsi aplikasi online ini kesannya tidak ada gunanya semoga kedepannya ada perbaikan terima kasih

3) Tokenization: This step involves breaking down the text into individual words or tokens. This process aims to represent the text in a way that is easier for the model to understand, allowing the algorithm to better recognize patterns within the text.

TABEL VII
TOKENIZING RESULT

Stage	Preprocessing Result
Tokenizing	[‘aplikasinya’, ’cukup’, ’mempermudah’, ’yang’, ‘Cuma’, ’mau’, ’bayar’, ’pajak’, ’kendaraan’, ‘atas’, ’nama’, ’diri’, ’sendiri’, ’tinggal’, ‘masukan’, ’nik’, ’kemudian’, ’bayar’, ‘online’, ’tapi’, ’ternyata’, ’tetap’, ‘harus’, ’manual’, ‘datang’, ‘ke samsat’, ‘untuk’, ‘pengesahan’, ‘jadi’, ‘fungsi’, ‘aplikasi’, ‘online’, ‘ini’, ‘kesannya’, ‘tidak’, ‘ada’, ‘gunanya’, ‘semoga’, ‘kedepannya’, ‘ada’, ‘perbaikan’, ‘terima’, ‘kasih’]

4) Stopword Removal: In this step, common words that do not contribute to the meaning have been filtered out to improve the model's focus and accuracy.

TABEL VIII
STOPWORD RESULT

Stage	Stopword Result
Stopword Removal	[‘aplikasinya’, ‘cukup’, ‘mempermudah’, ‘yang’, ‘Cuma’, ‘mau’, ‘bayar’, ‘pajak’, ‘kendaraan’, ‘atas’, ‘nama’, ‘diri’, ‘sendiri’, ‘tinggal’, ‘masukan’, ‘nik’, ‘kemudian’, ‘bayar’, ‘online’, ‘tapi’, ‘ternyata’, ‘tetap’, ‘harus’, ‘manual’, ‘datang’, ‘ke’, ‘samsat’, ‘untuk’, ‘pengesahan’, ‘jadi’, ‘fungsi’, ‘aplikas’i, ‘online’, ‘ini’, ‘kesannya’, ‘tidak’, ‘ada’, ‘gunanya’, ‘semoga’, ‘kedepannya’ ‘ada’, ‘perbaikan’, ‘terima’, ‘kasih’]

5) Stemming: This step involves reducing words with affixes to their base or root forms. The goal is to minimize the number of unique words the model needs to handle, which can enhance the model's performance.

TABEL IX
STEMMING RESULT

Stage	Preprocessing Result
Stemming	[‘aplikasinya’, ‘cukup’, ‘mempermudah’, ‘yang’, ‘Cuma’, ‘mau’, ‘bayar’, ‘pajak’, ‘kendaraan’, ‘atas’, ‘nama’, ‘diri’, ‘sendiri’, ‘tinggal’, ‘masukan’, ‘nik’, ‘kemudian’, ‘bayar’, ‘online’, ‘tapi’, ‘ternyata’, ‘tetap’, ‘harus’, ‘manual’, ‘datang’, ‘ke’, ‘samsat’, ‘untuk’, ‘pengesahan’, ‘jadi’, ‘fungsi’, ‘aplikas’i, ‘online’, ‘ini’, ‘kesannya’, ‘tidak’, ‘ada’, ‘gunanya’, ‘semoga’, ‘kedepannya’ ‘ada’, ‘perbaikan’, ‘terima’, ‘kasih’]

E. Classification Using SVM Algorithm

Support Vector Machine (SVM) is a classical model for binary classification that operates by finding the optimal hyperplane to separate the given data samples. In this study, the SVM algorithm is used to classify user reviews of tax renewal applications. The process begins by splitting the dataset into two parts: 80% for training data and 20% for testing data. The training data is then transformed into a numerical representation using Count Vectorizer, which learns the vocabulary from the training data and generates features based on word frequency. This process is also applied to the test data using the vocabulary learned from the training set. Subsequently, hyperparameter optimization is performed on the SVM algorithm using GridSearchCV to test various combinations of parameters, such as the value of C and the kernel type. This optimization aims to find the parameter combination that produces the best accuracy.

After data preprocessing, the dataset is divided into training and testing sets, with 80% allocated to training (2,161 samples) and 20% to testing (927 samples). Count Vectorizer is then used to learn the vocabulary from the training data and to create features based on the frequency of word occurrences. This same vocabulary is applied to the test data.

The next step involves hyperparameter optimization using GridSearchCV to identify the best parameter combination. The values for the parameter C tested include 0.01, 0.05, 0.25, 0.5, 0.75, 1, and 10, while the kernel types tested are linear, RBF, and polynomial. The results of the hyperparameter optimization process using GridSearchCV are illustrated in Figure 7.

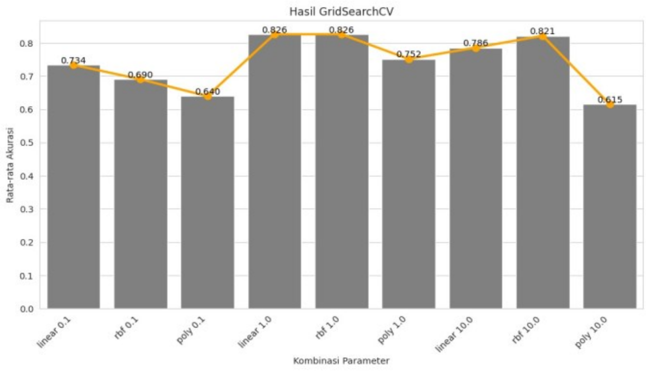

Figure 7. GridsearchCV Illustrated

SVM achieved an accuracy of 76.5% on the test data, demonstrating its effectiveness in classifying user reviews for the online tax renewal applications. In addition to accuracy, precision, recall, and F1-score were also calculated to provide a more comprehensive evaluation of the model's performance. The precision for the positive class was 0.84, while the precision for the negative class was 0.83 . The recall and F1-scores for both classes are also provided in Table X.

TABEL X
RECALL AND F1-SCORE

Class	Recall	F1-Score
Negative	0.86	0.85
Positive	0.81	0.82

F. Visualization of Result

For result visualization, the researcher utilizes a word cloud, a visual tool that highlights frequently occurring words. The use of the word cloud is intended to provide a clear representation of word frequency in user reviews. The word cloud is displayed in Figure 8.

Figure 8. Wordcloud Result

IV. CONCLUSION

The findings of this study emphasize the capability of the Support Vector Machine (SVM) algorithm in accurately analyzing user sentiments towards online vehicle tax renewal applications. Despite the development of these digital tools aimed at simplifying tax payment processes, the results indicate a significant amount of user dissatisfaction, as evidenced by the negative reviews. Among the three algorithms compared SVM, KNN, and Naïve Bayes the SVM algorithm emerged as the most effective, achieving an accuracy rate of 76.5%. This highlights the potential of SVM as a robust method for sentiment analysis in applications where understanding user feedback is critical.

Moreover, it is important to note that precision, recall, and F1-score metrics were employed alongside accuracy to evaluate the classifier's performance, providing a more comprehensive view of its capabilities. While SVM performed well overall, this study also highlights several areas of improvement for the applications being analyzed.

In light of the increasing number of motor vehicles and the growing demand for efficient tax management, it is critical for developers to address the key challenges identified in this sentiment analysis. Specifically, improvements in application stability, simplification of the payment process, and enhancements to the user interface are necessary to meet users' expectations. Common issues, such as technical problems and complicated user experiences, should be prioritized.

By leveraging the insights derived from this analysis, developers can design more user-centric and responsive applications that align more closely with the needs and preferences of the public. Ultimately, addressing these issues could lead to improved user satisfaction, thereby contributing to the overarching goal of enhancing local tax collection and increasing regional revenue.

REFERENCES

[1] M. S. Nasir, "Analisis Sumber-Sumber Pendapatan Asli Daerah Setelah Satu Dekadeotonomi Daerah," *J. Din. Ekon. Pembang.*, vol. 2, no. 1, p. 30, 2019, doi: 10.14710/jdep.2.1.30-45.

[2] D. Mulyono and M. Riyanto, "Optimalisasi Aplikasi Administrasi Kendaraan Pajak Online (Sakpole) Dalam Upaya Meningkatkan Pelayanan Pajak Kendaraan," *Cita Huk. Indones.*, pp. 174–187, 2023.

[3] N. Fajriyanti, R. Kusumadewi, and F. P. Wahyu, "Efektivitas Sistem Pemungutan Pajak Kendaraan Bermotor Melalui Online (E-Samsat) Pada Pusat Pengelolaan Pendapatan Daerah Wilayah Kabupaten Bandung I Rancaekek," *J. Dialekt. J. Ilmu Sos.*, vol. 20, no. 2, pp. 95–109, 2022, doi: 10.54783/dialektika.v20i2.74.

[4] A. W. Finaka, "Indonesia Produsen Kendaraan Bermotor TERBESAR."

[5] A. Mulyawan and D. Novia, "Aplikasi Pembayaran Pajak Kendaraan Bermotor Online Berbasis Web (Studi Kasus di Samsat Soreang Kab. Bandung)," *J. Comput. Bisnis*, vol. 10, no. 1, pp. 30–39, 2016.

[6] M. I. Ramdhan, M. I. Nurdiansyah, and N. N. Amalia, "Analisis Sentimen Komentar Pengguna Mobile JKN di Google Play Store," vol. 39, pp. 92–100, 2024.

[7] P. L. B. P. Jateng, "New Sakpole," Google Play Store.

[8] P. P. J. Barat, "Sapawarga," Google Play Store.

[9] B. P. Sulut, “Timsalut,” Google Play Store.

[10] T. Safitri, Y. Umaidah, and I. Maulana, “Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine,” *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 28–35, 2023, doi: 10.30871/jaic.v7i1.5039.

[11] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, “Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS,” *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 207–212, 2023, doi: 10.30591/jpit.v8i3.5064.

[12] A. Salma and W. Silfianti, “Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors,” *Int. Res. J. Adv. Eng. Sci.*, vol. 6, no. 4, pp. 158–162, 2021.

[13] H. P. Doloksaribu and Yusran Timur Samuel, “Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi,” *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 16, no. 1, pp. 1–11, 2022, doi: 10.47111/jti.v16i1.3747.

[14] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC,” *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.

[15] R. Merdiansah, S. Siska, and A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.

[16] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, “Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine,” *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.

[17] G. R. Ditami, E. F. Ripanti, and H. Sujaini, “Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 3, p. 508, 2022, doi: 10.26418/jp.v8i3.56478.

[18] “Sentiment Analysis of pegipegi.com Review on Google Play Store with Naïve Bayes,” *Progr. Stud. Sist. Inf. Fak. Tek. dan Ilmu Komput.*, vol. Vol 13, No, 2024.

[19] N. Cahyono and Dewi Setiyawati, “Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia,” *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 262–272, 2023, doi: 10.33022/ijcs.v12i1.3154.

[20] D. Safryda Putri and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine,” *J. Ilm. Inform.*, vol. 11, no. 01, pp. 32–40, 2023, doi: 10.33884/jif.v11i01.6611.

[21] B. Gaye, D. Zhang, and A. Wulamu, “Improvement of Support Vector Machine Algorithm in Big Data Background,” *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/5594899.

[22] R. Tineges, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM),” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.

[23] T. Turki and S. S. Roy, “Novel Hate Speech Detection Using Word Cloud Visualization and Ensemble Learning Coupled with Count Vectorizer,” *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136611.

[24] M. A. Muslim *et al.*, “Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy,” *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020, doi: 10.52465/joscex.v1i1.3.

[25] B. F. S. Supriyanto and S. Rosalin, “Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency,” *J. Minfo Polgan*, vol. 12, no. 1, pp. 25–32, 2023, doi: 10.33395/jmp.v12i1.12312.

[26] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.

[27] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, “Multi-label Classifier Performance Evaluation with Confusion Matrix,” pp. 01–14, 2020, doi: 10.5121/csit.2020.100801.