
Diving Into Public Sentiments: SVM-Based Analysis of Water Service Opinions

Aldi Ridwansyah Hasugian^{1)*}, Ilka Zufria²⁾

^{1,2)} Computer Science Study Program, State Islamic University of North Sumatra, Indonesia

¹⁾aldiridwansyah85@gmail.com, ²⁾ilkazufria@uinsu.ac.id

ABSTRACT

Customer satisfaction is a crucial measure of a company's success, particularly for service-oriented businesses such as regional water utilities. In the case of PDAM Tirta Nauli Sibolga, key service aspects including pricing, water quality, water flow, and responsiveness to customer complaints significantly influence customer satisfaction. Despite ongoing efforts to provide quality services, challenges in areas such as water quality, quantity, and customer communication persist, leading to lower satisfaction levels among consumers. This study seeks to evaluate the implementation of regional regulations related to company management and clean water services, with a focus on improving service quality. Additionally, it applies the Support Vector Machine (SVM) algorithm to perform sentiment analysis of public opinions regarding the company's services. The analysis yielded an impressive accuracy of 98%, with precision, recall, and F1-score values consistently above 90%, indicating a high level of effectiveness in sentiment classification. By leveraging the results of this sentiment analysis, PDAM Tirta Nauli can gain valuable insights into the issues facing the community and implement targeted improvements. Ultimately, this research aims to provide a comprehensive evaluation of customer satisfaction and offer actionable recommendations to enhance service quality, ensuring better customer experiences and supporting the overall development of the region.

Keywords: Sentiment Analysis; Public Opinion; PDAM Tirta Nauli Sibolga; Support Vector Machine (SVM); Customer Satisfaction

INTRODUCTION

Perusahaan Daerah Air Minum (PDAM) is a regional company that provides clean water and is supervised by local authorities. PDAM is entrusted with the responsibility of managing and developing the clean water supply system, serving the community, and offering affordable prices (Amrulloh, 2022).

The success of a company is closely tied to customer satisfaction with the services it provides, and this principle applies to PDAM Tirta Nauli Sibolga. As a regional company responsible for providing clean water in Sibolga, PDAM Tirta Nauli Sibolga is expected to serve customers by maintaining affordable pricing, ensuring water quality, delivering consistent water flow, and responding promptly to customer complaints (Hasibuan et al., 2023). However, the services provided by PDAM Tirta Nauli Sibolga are currently falling short of customer expectations.

The quality of service provided by PDAM Tirta Nauli Sibolga does not fully meet customer expectations, leading to dissatisfaction among its users. As a result, there is a need for the company to address these issues and improve its overall service quality. The Constitution of the Republic of Indonesia 1945, Article 33, Clause 3, mandates that water resources, as a fundamental element of life, should be managed for the welfare of citizens, which further stresses the importance of providing clean water services to the community (Sapta, 2022).

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

In Sibolga, the local PDAM is known as PDAM Tirta Nauli, a service-oriented company located at Jl S.M. Raja No. 471, Sibolga City. As one of the local government-owned companies (BUMD) in Sibolga, PDAM Tirta Nauli Sibolga was established to contribute to local income. Therefore, PDAM Tirta Nauli Sibolga must improve productivity and service to support the region's development (Hasibuan et al., 2023).

Several issues have been identified in PDAM Tirta Nauli Sibolga's service delivery, including unsatisfactory water quality, uneven water distribution that causes shortages in certain areas, and poor communication with customers. These shortcomings are characterized by delayed responses to complaints and untimely service processes, which negatively impact the customer experience (Anastasya & Yuamita, 2022). Such issues emphasize the need for significant improvements in the quality of services provided to the public.

To address these challenges, data mining is proposed as an analytical tool to evaluate and improve service performance. Data mining involves analyzing data from various perspectives to extract valuable information and establish relationships between variables within large datasets (Noviyanto, 2020). This technique is particularly relevant in analyzing customer feedback, which often consists of unstructured textual data.

One effective method for sentiment analysis is the Support Vector Machines (SVM) algorithm, introduced by Boser, Guyon, and Vapnik in 1992. SVM is a supervised learning method capable of classifying customer opinions into positive and negative categories. It uses a hyperplane to separate the sentiment classes, making it highly suitable for analyzing textual feedback (Aruan et al., 2022).

To facilitate the collection and processing of textual feedback, tools like Google Forms can be utilized. Google Forms is widely used for surveys and feedback collection because it simplifies the process of gathering data digitally (Nurhaliza et al., 2020). By applying SVM to analyze the data collected through Google Forms, PDAM Tirta Nauli Sibolga can gain insights into public sentiment and identify areas for improvement. These insights can ultimately help the company enhance service quality and meet customer expectations more effectively.

LITERATURE REVIEW

Sentiment Analysis

Sentiment analysis is the process of extracting information about an entity and automatically identifying its subjectivity. The goal is to determine whether the text generated by users conveys a positive, negative, or neutral opinion. Sentiment classification can be performed at three levels of extraction: aspect or feature level, sentence level, and document level (Ikhsan, 2023). Sentiment analysis is a computational study to analyze people's feelings and opinions about an entity. In recent decades, sentiment analysis has become a widely researched topic. From the previous explanation, it can be concluded that sentiment analysis is a method for obtaining information regarding an individual's thoughts on an event. Sentiment analysis can be used to determine the general sentiment about the price, water quality, water service, water flow, and the responsiveness of PDAM Tirta Nauli Sibolga (Setiawan & Zufria, 2023).

Machine Learning

Machine learning can be defined as computer applications and mathematical algorithms adopted through learning from data to generate predictions for the future. The field of machine learning is concerned with the question of how to build computer programs that improve automatically based on experience. Recent research reveals that machine learning is divided into three categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning (Hermawan et al., 2023).

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

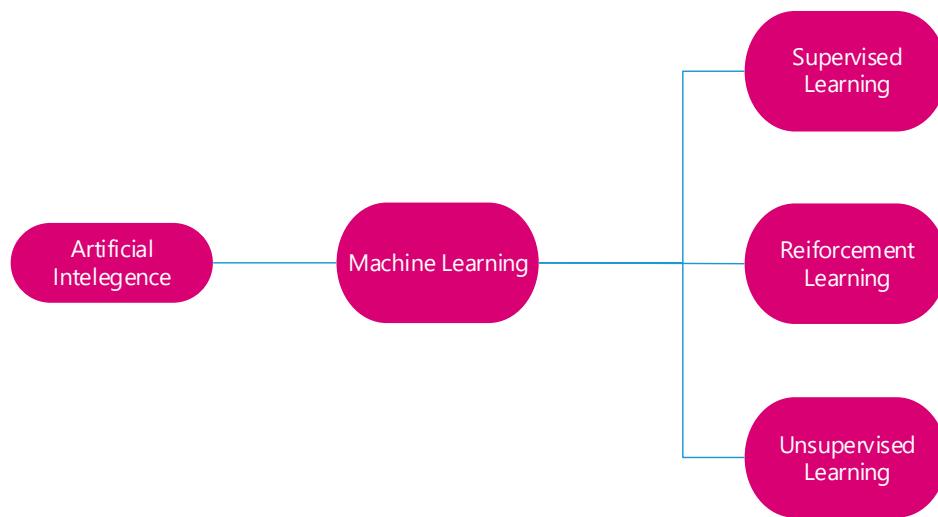


Figure 1. Artificial Intelligence and Machine Learning Scheme

Data Mining

Data mining is the process of discovering meaningful correlations, patterns, and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies, as well as statistical and mathematical techniques. Data mining is the process of discovering patterns and interesting knowledge from large datasets. The term data mining has several interpretations, such as knowledge discovery or pattern recognition. The term knowledge discovery is appropriate because the main goal of data mining is to uncover knowledge that is hidden within large datasets. Meanwhile, the term pattern recognition is appropriate because it focuses on finding patterns hidden in the data (Nabila et al., 2021).

The goal of data mining and Knowledge Discovery in Databases (KDD) is to use database data and process it to generate new, valuable information. Additionally, many other expressions, such as knowledge extraction, data archaeology, data scraping, pattern/data analysis, and knowledge discovery in databases, have connotations similar to data mining. A common misunderstanding is that data mining is identical to other terms widely used. The process of creating information from organized and unstructured sources is known as knowledge discovery in databases. Some view data mining as just one critical phase in the knowledge discovery process, while others prefer the knowledge generated to be in a machine-readable format that can be interpreted by machines and reflect knowledge discovery from data, or KDD (Lestari et al., 2023).

Text Mining

To analyze large amounts of data, you can use text mining techniques. Text mining is the process of discovering unknown knowledge through the automatic extraction of information from a large amount of unstructured text. Text mining is capable of extracting data in large volumes, which is then processed to obtain useful information from text collections. Data and information from the internet become the primary target of text mining research by companies for various purposes (Hermawan et al., 2023).

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

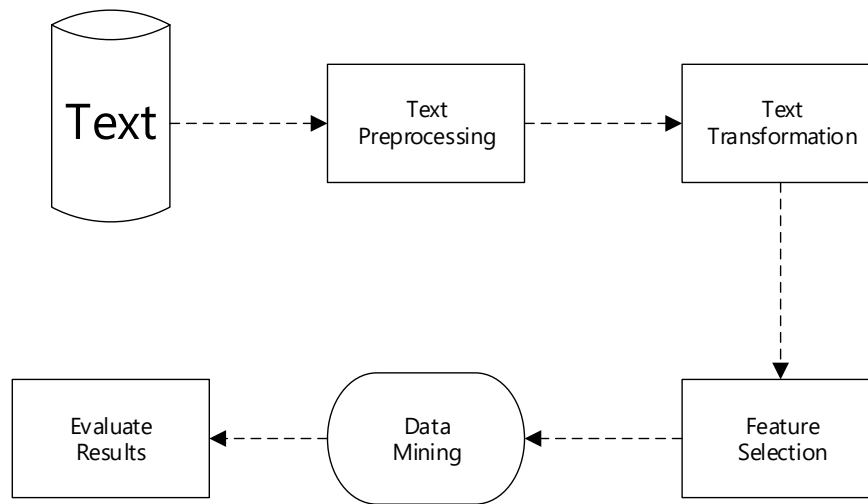


Figure 2. Text Mining Process

Support Vector Machine

In this study, the method used for classification is Support Vector Machine (SVM). SVM is a model derived from statistical learning theory that provides better results compared to other methods. In linear SVM, each training data is known as (x_i, y_i) , where $i = 1, 2, \dots, N$ and $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}$. T represents the attributes of training data I , and $y_i \{-1, +1\}$ is the class label. The question then is how Support Vector Machine can generate accuracy from the classification results of this test (Darwis et al., 2020). SVM attempts to find a hyperplane by maximizing the distance between classes, as illustrated in Figure 3 (a). In this way, SVM guarantees a high generalization ability for future data, as shown in Figure 3 (b) (Aruan et al., 2022).

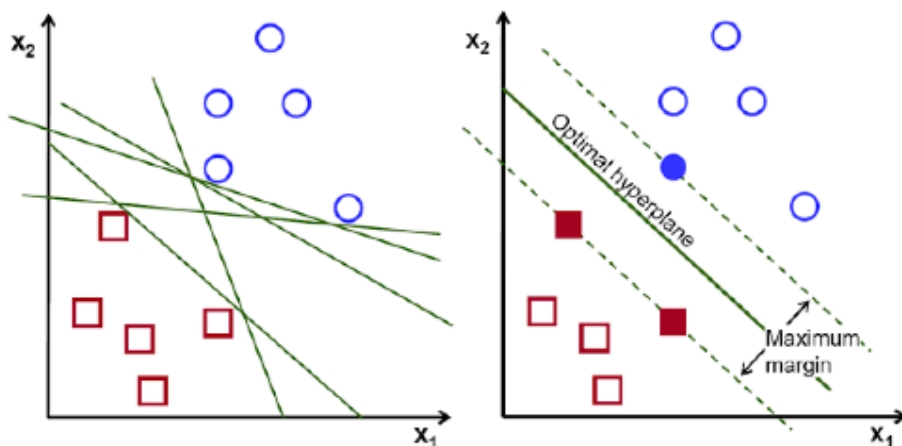


Figure 3. Best Hyperplane and Maximum Margin

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

METHOD

In this chapter, the author will explain how to conduct the research with details of the analytical tools and materials, as well as the systematic sequence of research procedures. This will serve as a guideline for solving the problems to be encountered. Below is the structure of the research framework.

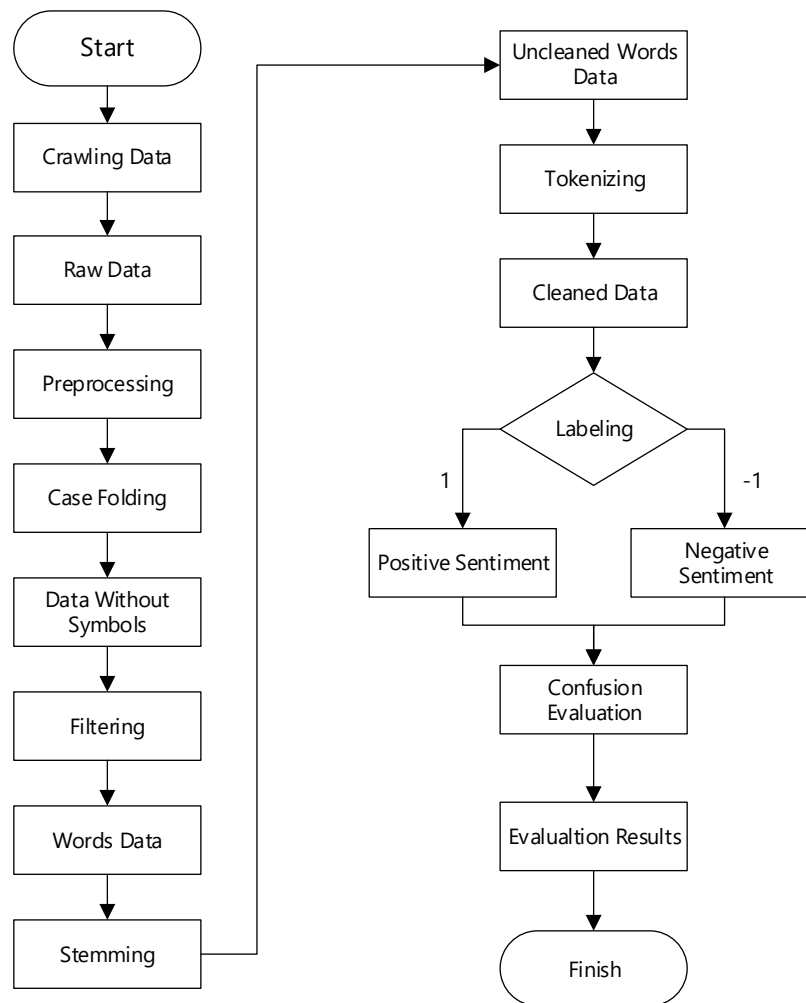


Figure 4. Research Framework

In this study, data analysis involves the preparation of survey data as the initial process for data processing. This process is used to describe the customer service survey of PDAM Tirta Nauli Sibolga. By displaying the data, it will be easier to analyze and organize it. Next, class labeling will be performed with the rule that if the data is less than 0, the class will be negative, and if it is greater than 0, the class will be positive. The stages of sentiment analysis are carried out using the SVM method, starting with input data consisting of training and testing data, which are then processed through the preprocessing stage until the classification process. The labeled data will be used for classification. This process is done to build a model that will be used to classify new data.

In the image above, the data, consisting of words that have undergone case folding and filtering, will be tokenized, labeled, and their weights calculated. Once the weights are determined, the classification process

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

begins. If the term has a value <0 , it is categorized as a negative comment, and if the value is >0 , it is classified as a positive comment.

This study begins by collecting data from a Google Form regarding the public's opinion, which will be classified. The words are then weighted, and sentiment classification is performed using the SVM method. The resulting dataset is classified using a confusion matrix to separate it into two categories: negative and positive classes.

The testing phase of this research aims to evaluate the accuracy of sentiment determination regarding the services of PDAM Tirta Nauli Sibolga using the confusion matrix through the SVM library. The process begins with training and testing models that have been previously classified, generating a matrix as a representative of actual and predicted classes. The results of the training model are tested with new data that was not previously used for training.

RESULT AND DISCUSSION

From the questionnaire results regarding the performance evaluation of PDAM Tirta Nauli, the data can be processed by mapping customer satisfaction as follows: satisfaction with service prices, represented by the variable `kepuasan_harga`; satisfaction with water service, represented by the variable `kepuasan_pelayanan_air`; satisfaction with water flow, measured through the variable `kepuasan_aliran_air`; satisfaction with the responsiveness of staff in handling complaints, represented by the variable `kepuasan_daya_tanggap`; and satisfaction with the quality of water provided, represented by the variable `kepuasan_kualitas_air`. Each of these variables can be rated using a Likert scale to obtain a clear picture of community satisfaction with PDAM Tirta Nauli, which will later be analyzed for performance evaluation and service improvement.

The total data consists of 402 entries. Before classification, each data point will go through a text preprocessing stage, which includes case folding, filtering, stemming, and tokenization. Once the text preprocessing process is completed, the data will be ready for classification.

Text Preprocessing

The data preprocessing stage is a process to make the dataset easier to handle in the subsequent steps. This stage includes steps to clean and prepare the data, such as removing stopwords, normalizing the text, and cleaning the data using regular expressions. These procedures ensure that the data is standardized, free from irrelevant elements, and ready for further analysis or classification tasks.

Table 1. Case Folding Stages

Before	After
Karena, harga yg di berikan masih terjangkau Karena menurut saya sesuai dengan yang dipakai	karena, harga yg di berikan masih terjangkau karena menurut saya sesuai dengan yang dipakai

Table 2. Filtering Stages

Before	After
karena, harga yg di berikan masih terjangkau karena menurut saya sesuai dengan yang dipakai	karena harga yg beri masih terjangkau karena menurut saya sesuai dengan yang dipakai

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Table 3.Stemming Stages

Before	After
karena harga yg beri masih terjangkau karena menurut saya sesuai dengan yang dipakai	karena harga yg beri masih terjangkau karena menurut saya sesuai dengan yang dipakai

Table 4.Tokenizing Stages

Before	After
karena harga yg beri masih terjangkau karena menurut saya sesuai dengan yang dipakai	karena, harga, yg, beri, masih, terjangkau karena, menurut, saya, sesuai, dengan, yang, dipakai

Data Labeling

After the data preprocessing stage, the next step is data labeling. In the data labeling stage, each opinion is evaluated to determine whether it carries a positive or negative sentiment. This process serves as a reference for assigning weights to each word, which will later be used to classify other opinions. Below is an example of sample data that has been labeled.

Table 5.Data Labelling Sample

tokenized_data	label
karena, harga, yg, beri, masih, terjangkau	POSITIVE
karena, pdam, beri, harga, yg, standar	POSITIVE
karena, masih, terjangkau	POSITIVE
sebab, harga, pdam, sibolga, masih, terjangkau, seluruh, lapisan, masyarakatkau, oleh, lapisan, masyarakat, khususnya, masyarakat, menengah, ke, bawah	POSITIVE
sangat, membantu	POSITIVE
karna, sesuai, pemakaian	POSITIVE
harga, terjangkau, kalangan, masyarakat, menengah, kebawah	POSITIVE
karna, harga, masih, terjangkau, dan, bisa, buat, kebutuhan, tsb	POSITIVE
sangat, puas	POSITIVE
karena, menurut, saya, sesuai, dengan, yang, dipakai	POSITIVE
karna, kadang, harga, tidak, sesuai, pakai	NEGATIVE
karna, belum, ada, program, air, gratis	NEGATIVE
mahal, tidak, sesuai, dgn, yg, saya, pakai	NEGATIVE
suka, lonjak, tidak, tentu	NEGATIVE
karna, kadang, harga, tidak, sesuai, pakai	NEGATIVE

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

TF-IDF Vectorizer

The TF-IDF Vectorizer stage is the process of transforming raw text into a numerical representation based on Term Frequency-Inverse Document Frequency (TF-IDF). This technique is used to extract features from text for various applications, such as sentiment analysis, topic modeling, or text classification.

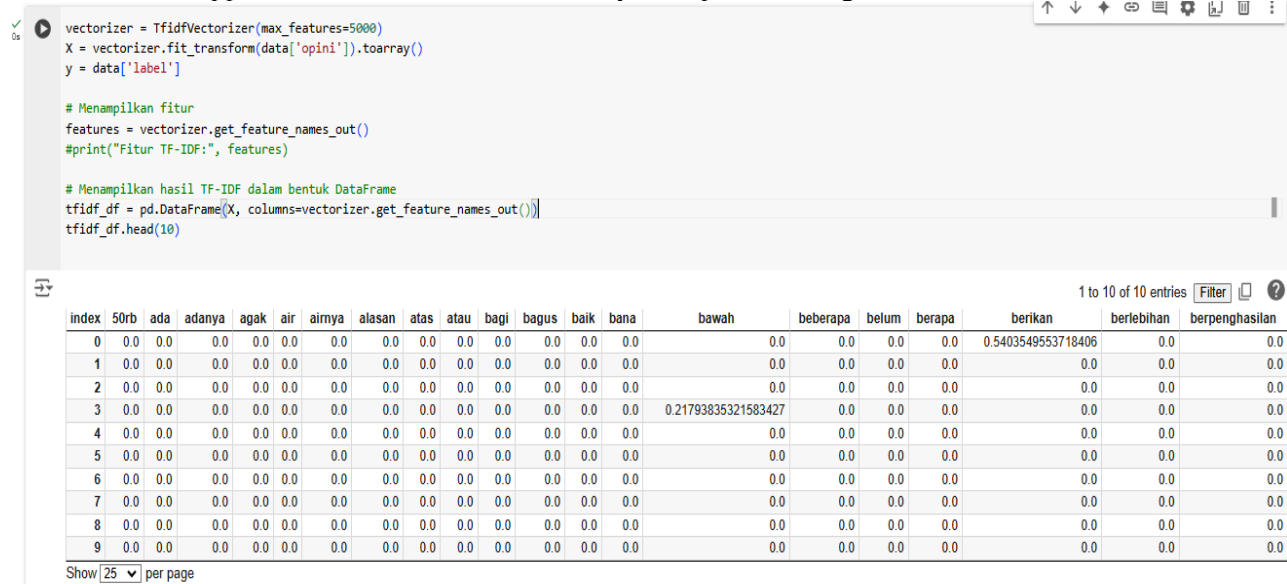


Figure 5. Google Colab TF-IDF Vectorizer Result

Split Data

In this stage, the data will be split into two parts: training data and testing data. The purpose of this step is to separate a portion of the data as training data, which will be used to train the model, while the remaining data will be used as testing data to evaluate the model's performance. By splitting the data, we ensure that the model not only performs well on the data it has already seen (training data) but also generalizes effectively on new or unseen data (testing data). The ratio used for the split is 70:30, with 70% of the data used for training and 30% used for testing.

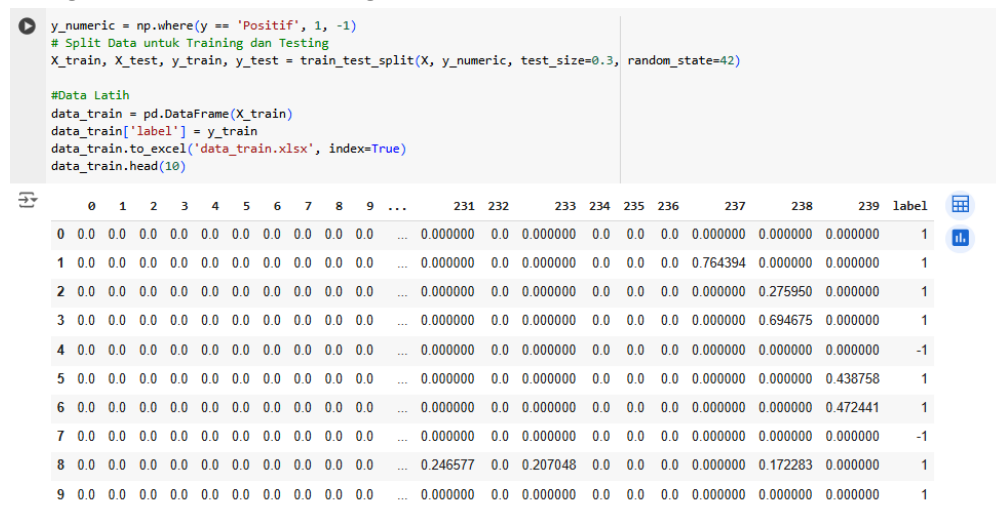


Figure 6. Google Colab Train Data Result

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).


```
#Data Uji
data_test = pd.DataFrame(X_test)
data_test['label'] = y_test
data_test.to_excel('data_test.xlsx', index=True)
data_test.head(10)
```

	0	1	2	3	4	5	6	7	8	9	...	231	232	233	234	235	236	237	238	239	label
0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
1	0.000000	0.0	0.0	0.0	0.334533	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.327342	0.0	1
2	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
3	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
4	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
5	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
6	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
7	0.000000	0.0	0.0	0.0	0.406034	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1
8	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.694675	0.0	1
9	0.394801	0.0	0.0	0.0	0.219829	0.0	0.0	0.360621	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	1

10 rows × 241 columns

Figure 7. Google Colab Test Data Result

Support Vector Machine

The calculation stage using Support Vector Machine (SVM) with a linear kernel is the process of building a classification or regression model that separates data using a hyperplane. The linear kernel is used when data can be separated linearly in the feature space. The SVM model is trained first before making predictions, which results in the formation of weight coefficients. The output of the SVM calculation is the Support Vector.

```
# Training model dengan data training
model = SVC(kernel='linear', C=1) # Anda bisa menyesuaikan nilai C untuk soft margin
model.fit(X_train, y_train)

# Ambil koefisien (w) dan intercept (b)
w = model.coef_[0]
b = model.intercept_[0]

final_w = pd.DataFrame({
    'Kata': tfidf_df.columns,
    'Koefisien Bobot (w)': w
})

# Menampilkan hasil
final_w = final_w.fillna(0) # Mengganti NaN dengan 0 jika tidak ada data untuk label tersebut
final_w = final_w.sort_values(by='Kata', ascending=True) # Menyortir berdasarkan Bobot Positif

# Menampilkan DataFrame
final_w.head(10)
```

	Kata	Koefisien Bobot (w)
0	50rb	0.164485
1	ada	0.134179
2	adanya	-0.731267
3	agak	-0.876231
4	air	-0.710943
5	airnya	0.150604
6	alasan	0.000000

Figure 8. Google Colab Training Data and Weight Coefficient results

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Training data in Google Colab involves a process where models, such as SVM, are trained using training data to find patterns that can be used for making predictions. During training, the model learns the relationship between input features and target outputs. The results of this training include weight coefficients, which indicate how much influence each feature has on the decisions made by the model. These weight coefficients are important for understanding how the model makes predictions, as larger weight values indicate more influential features in the model.

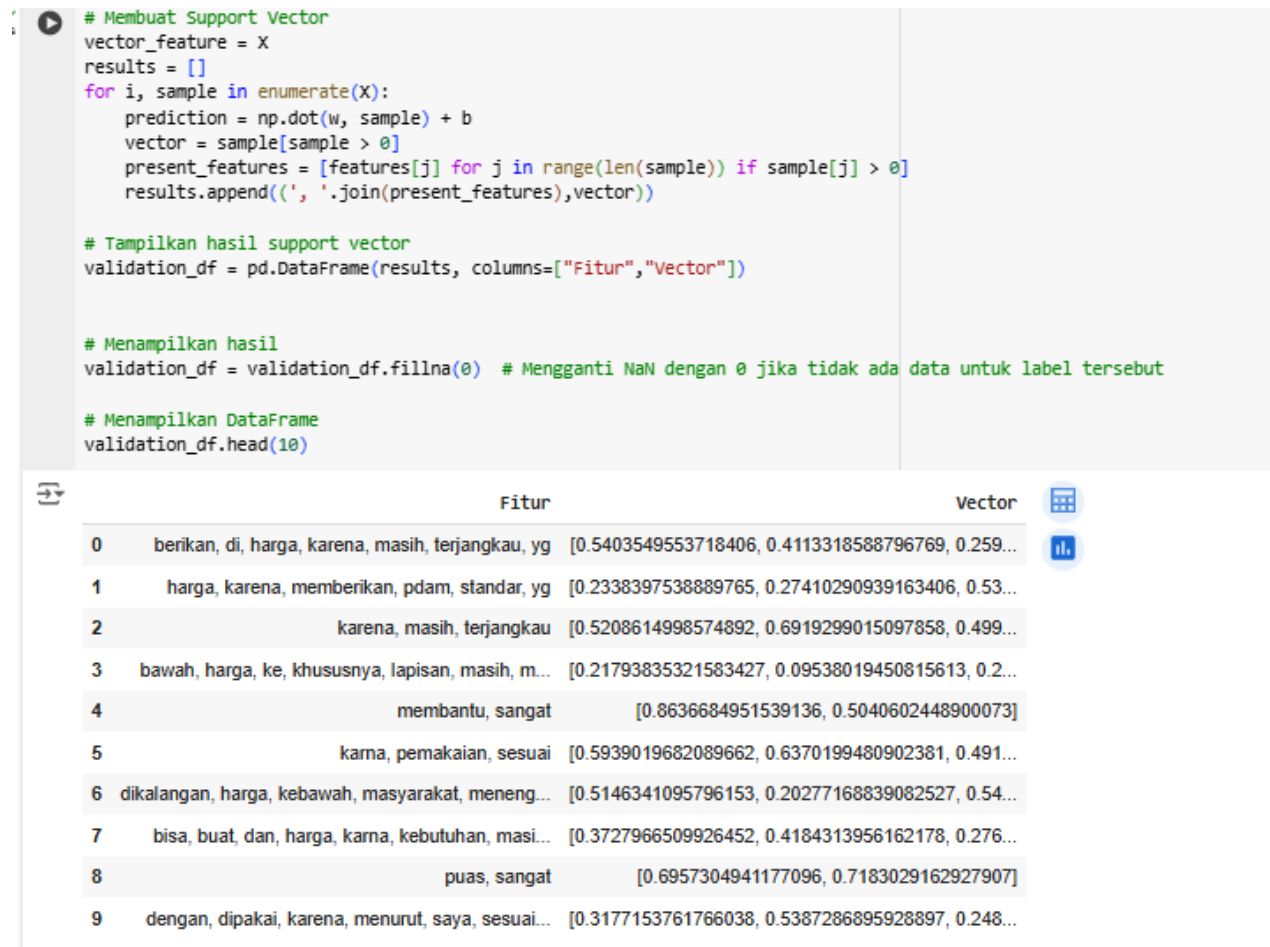


Figure 9. Google Colab Support Vectors results

Prediction using test data is the process in which a model trained on training data is used to predict outcomes on previously unseen data, known as the test data. The test data is used to assess the model's ability to generate accurate predictions beyond the data used during training. The predicted results are then compared with the actual labels to evaluate the model's performance.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).



Figure 10. Google Colab Prediction results

Confusion Matrix

Next, after obtaining the matrix from the support vector, the next step is to perform predictions to evaluate the accuracy of the model by calculating the values of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) according to the predicted values and their classes, where numerically 1 represents Positive and -1 represents Negative (Hokijuliandy et al., 2023).

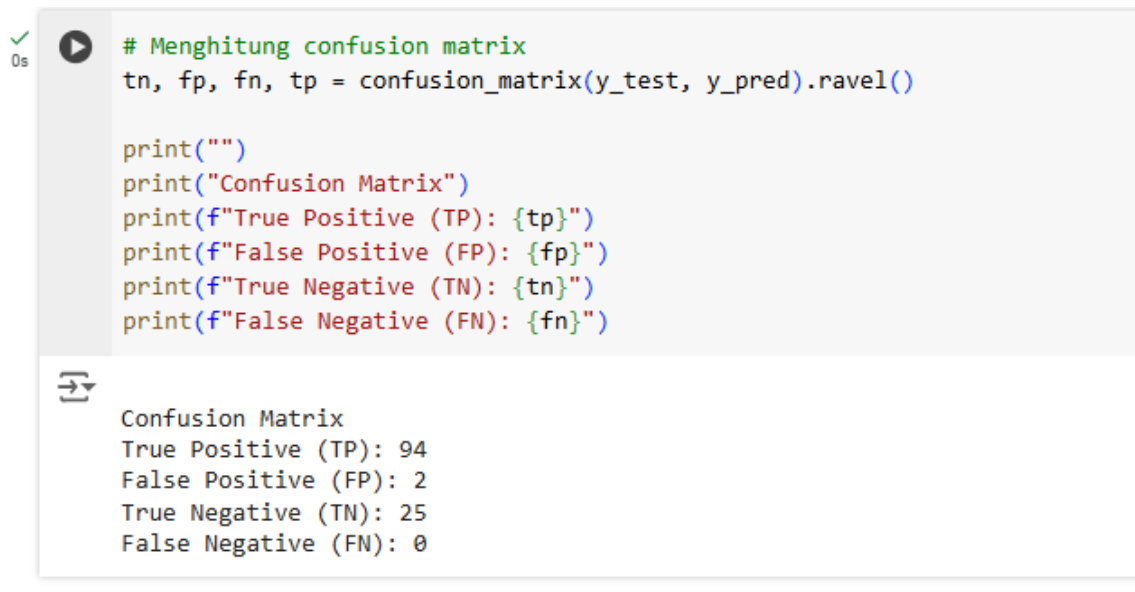


Figure 11. Google Colab Confusion Matrix result

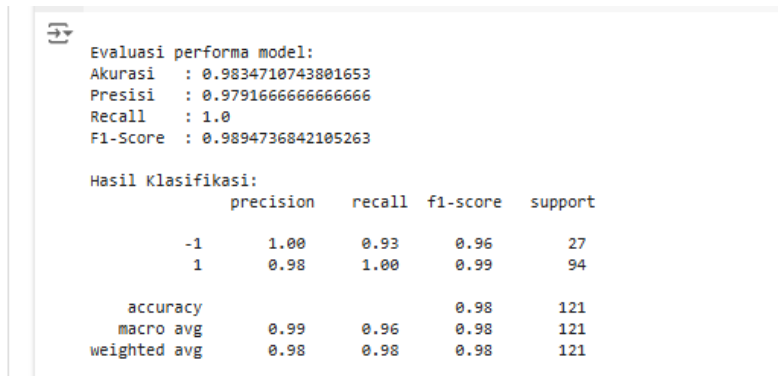
A Confusion Matrix in Google Colab is a tool used to evaluate the performance of a classification model. It displays the number of correct and incorrect predictions in a table format, with rows representing actual labels and columns representing predicted labels. The Confusion Matrix helps in calculating important

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

metrics such as accuracy, precision, recall, and F1-score, providing a deeper insight into how the model makes errors in classification (Singgalen, 2023).



```
Evaluasi performa model:
Akurasi : 0.9834710743801653
Presisi : 0.9791666666666666
Recall : 1.0
F1-Score : 0.9894736842105263

Hasil Klasifikasi:
      precision    recall  f1-score   support

-1      1.00      0.93      0.96         27
 1      0.98      1.00      0.99         94

accuracy          0.98         121
macro avg          0.99      0.96      0.98         121
weighted avg          0.98      0.98      0.98         121
```

Figure 12. Google Colab Performance Model result

Wordcloud

Further evaluation is conducted by assessing the classification accuracy and other aspects to evaluate the performance of the generated Wordcloud. The following evaluation results are generated: The positive Wordcloud highlights words that frequently appear for positive class words, while the negative Wordcloud highlights words that frequently appear for negative class words. This visualization helps in understanding the key terms associated with each sentiment class (Irvandi et al., 2023).



Figure 13. Google Colab Wordcloud result

CONCLUSION

After conducting research on PDAM Tirta Nauli and applying the data obtained in Google Colab, it was found that the use of wordcloud as a reference for sentiment classification showed very good performance. This is proven by the evaluation results, which yielded an accuracy of 98%, with precision, recall, and f1-score values consistently above 90% for each sentiment variable, particularly regarding service quality and pricing sentiment. By utilizing the generated wordcloud, it is hoped that PDAM Tirta Nauli will be able to identify the issues faced by the community and improve the services provided based on the problems identified by the public.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

ACKNOWLEDGMENT

I would like to express my sincere gratitude to all parties involved in this research, especially the management of PDAM Tirta Nauli Sibolga, and to my senior university, Satria Harum Nugraha, who has helped me by teaching how to apply the SVM method using Google Colab. My heartfelt thanks also go to my beloved mother, Yahnima Larossa, who played an essential role in writing this journal proposal and always provided support and motivation, which were invaluable to me in completing this journal. I would also like to express my deep gratitude to my father, Rahmad Hasugian, who has consistently supported me in every way possible so that I could finish this journal. Lastly, I want to thank my fellow comrades, seniors, and juniors at HMI who cannot be mentioned one by one, for all the support and encouragement that helped me finish this journal proposal.

REFERENCES

- Amrulloh, R. (2022). SURVEI KEPUASAN MASYARAKAT TERHADAP KUALITAS PELAYANAN PDAM INTAN BANJAR. *Al-KALAM : JURNAL KOMUNIKASI, BISNIS DAN MANAJEMEN*, 9(2), 128. <https://doi.org/10.31602/al-kalam.v9i2.6526>
- Anastasya, A., & Yuamita, F. (2022). Pengendalian Kualitas Pada Produksi Air Minum Dalam Kemasan Botol 330 ml Menggunakan Metode Failure Mode Effect Analysis (FMEA) di PDAM Tirta Sembada. *Jurnal Teknologi Dan Manajemen Industri Terapan*, 1(1), 15–21. <https://doi.org/10.55826/tmit.v1i1.4>
- Aruan, J. D. C., Rahayudi, B., & Ridok, A. (2022). Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machinedan Term Frequency – Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(5), 2072–2078.
- Hasibuan, N. ‘Ainun, Sihombing, R., & Tanjung, A. (2023). PENGARUH LINGKUNGAN KERJA TERHADAP PRODUKTIVITAS KERJA PEGAWAI PADA PERUSAHAAN DAERAH AIR MINUM (PDAM) TIRTA NAULI SIBOLGA. *Jesya*, 6(2), 1372–1381. <https://doi.org/10.36778/jesya.v6i2.1325>
- Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Hokijuliandy, E., Napitupulu, H., & Firdaniza, F. (2023). Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square. *SisInfo : Jurnal Sistem Informasi Dan Informatika*, 5(2), 40–49. <https://doi.org/10.37278/sisinfo.v5i2.670>
- Ikhsan, M. (2023). Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Long Short-Term Memory. *The Indonesian Journal of Computer Science Research*, 2(1), 31–41. <https://doi.org/10.59095/ijcsr.v2i1.29>
- Irvandi, Irawan, B., & Nurdian, O. (2023). NAIVE BAYES DAN WORDCLOUD UNTUK ANALISIS SENTIMEN WISATA HALAL PULAU LOMBOK. *INFOTECH Journal*, 9(1), 236–242. <https://doi.org/10.31949/infotech.v9i1.5322>

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

- Lestari, R. D., Hasibuan, M. S., & Wahyuni, S. (2023). Penerapan Data Mining Menggunakan Metode Teknik Klasifikasi Untuk Melihat Potensi Kepatuhan Wajib Pajak Kendaraan. *Journal of Computer Science and Informatics Engineering (CoSIE)*, 3(1), 1–14. <https://doi.org/10.55537/cosie.v3i1.710>
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108. <https://doi.org/doi.org/10.33365/jtsi.v2i2.868>
- Noviyanto. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma - Jurnal Komputer Dan Informatika*, 22(2), 183–188.
- Nurhaliza, A., Sudjani, D. H., & Maryani, N. (2020). Google Formulir Sebagai Alternatif Media Evaluasi Pembelajaran Bahasa Arab Pada Masa Pandemi Covid-19. *Tatsqify: Jurnal Pendidikan Bahasa Arab*, 3(1), 57–68. <https://doi.org/10.30997/tjpba.v3i1.4722>
- Sapta, C. A. (2022). *Tanggung Jawab Perusahaan Daerah Air Minum (PDAM) Terhadap Kerugian Konsumen dan Melonjaknya rekening Air Akibat Kebocoran Instalasi Dalam (Studi Pada PDAM Tirtanadi Provinsi Sumatera Utara)*. Universitas Medan Area.
- Setiawan, H., & Zufria, I. (2023). Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(3), 1003. <https://doi.org/10.30865/mib.v7i3.6144>
- Singgalen, Y. A. (2023). Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi Tripadvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM. *Journal of Computer System and Informatics (JoSYC)*, 4(2), 367–379. <https://doi.org/10.47065/josyc.v4i2.3092>