Proceeding of the International Conference on Management, Entrepreneurship, and Business
Volume 1 Nomor 1 Tahun 2024
e-ISSN: XXXX-XXX; p-ISSN: XXXX-XXX, Hal 228-231
DOI: https://doi.org/xx.xxxx
Available online at: https://xxxx.xxx

Classification Methods in Sentiment Analysis of Customers Statisfaction as a Services Improvement Strategy

Wiwin Windihastuty1, Yani Prabowo2, M.N. Farid Thoha3
1,2Technology Information Faculty, Universitas Budi Luhur, Jakarta, Indonesia,
wiwin.windihastuty@budiluhur.ac.id
3Business Management Faculty, Universitas Budi Luhur, Jakarta, Indonesia, mn.faridthoha@budiluhur

Abstract. *Customer satisfaction is a crucial indicator in assessing the quality of a company's products, services and overall experience. This research aims to identify the level of customer satisfaction and optimize the available data for effective use in sentiment analysis. In this study, we analyzed 4,353 customer reviews collected over the past year, with 3,481 reviews used as training data and 871 reviews as testing data. The analysis process was conducted using the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach and leveraged the Logistic Regression algorithm to build a predictive model. Model evaluation using the confusion matrix yielded an accuracy of 94.60%, a precision of 94.26%, and a recall of 94.60%. The analysis was conducted using Jupyter Notebook and the Python programming language. The results indicate that sentiment analysis is effective in identifying and predicting customer satisfaction levels, which in turn can help a company's products improve its service strategies. The optimization of previously underutilized data now provides deeper insights into customer perceptions and expectations, enabling the company to make more targeted decisions and enhance overall customer satisfaction*.

Keywords *Sentiment Analysis, Confusion Matrix, Customer Satisfaction, Automotive*

INTRODUCTION

Data mining is the process of finding new useful correlations, patterns and trends by mining large amounts of data (A.E. Pratama., 2022). Data mining aims to improve traditional techniques so that they can handle very large amounts of data and high data dimensions (L. Gaol., 2019). This study uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, a framework that aims to convert business problems into tasks and carry out data mining projects independently without being related to the application area or technology used (S. Huber, 2019). The method (CRISP-DM) consists of several stages, namely: business understanding, data understanding, data preparation, modeling and evaluation (Singgalem, 2023).

This study uses the Logistic Regression algorithm, which is one of the algorithms for classification in sentiment analysis with output in the form of positive and negative classes (Amin, 2022). Sentiment analysis is a technique in natural language processing that aims to identify and classify opinions or emotions expressed in text (Wibowo, 2019).

Received: August 29, 2023; Accepted: November 22, 2023; Published: Desember 28, 2024
*Corresponding author, e-mail address

This study also uses the Term Frequency-Inverse Document Frequency (TF-IDF) method, which is the process of calculating or extracting words into a number in the form of a vector that is used to determine the weight of a word in a document that is useful for determining the classification matrix (K.T Putra, 2022).

The Term Frequency (TF) method is a way to give weight to the relationship of a word with a document (M. Lestandy, 2021) while Inverse Document Frequency is the number of documents containing a word based on all documents in the dataset [9]. Confusion Matrix is one of the metrics used to measure the performance of a classification model, especially in the context of pattern recognition and machine learning. Confusion matrix is useful for analyzing how well a classifier can recognize tuples from different classes (M.H, 2023). Jupyter Notebook is an open-source application for creating and sharing documents with code, visualizations, and text, while python is a programming language widely used in data science due to its simple syntax and strong library support.

This study differs from previous studies in several key aspects that enrich the understanding of sentiment analysis in the automotive service industry (Utami, 2029). This study implements Logistic Regression, this approach was chosen because of its efficiency in managing large datasets and is suitable for diverse and unstructured review data. The majority of previous studies also rely on data taken from general consumer review platforms, this study uses data from freshchat, a platform that provides direct communication between customers and service providers (Willy Novianto, 2023). This not only enriches the quality of the data with real-time information but also increases the relevance of the findings for practical applications in business. This study focuses on samples of customers who use services at a company that focuses more on the automotive market compared to previous studies that did not limit the types of services or products reviewed. Unlike previous studies that focused more on measuring sentiment in general, this study aims to directly link the results of sentiment analysis with indicators of business success, such as customer satisfaction and positive reviews that provide added value for the company management to implement more informed business strategies.

METHODS

This study aims to analyze customer satisfaction sentiment towards services using text-based sentiment analysis methods. The research process consists of several systematic and measurable steps to obtain in-depth information about user responses and views on the application as shown in Figure 1.

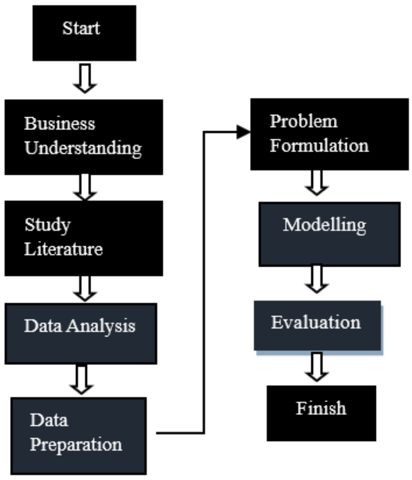

Figure 1. Research Step

Figure 1 shows a flow diagram of the research process starting from the start, business understanding, data understanding, data preparation, modeling, evaluation, problem identification stages, problem formulation, literature study, data collection and completion. This flow is the reference in the sentiment analysis process carried out in this study. This study provides the highest level of accuracy of the results of customer sentiment who have serviced their vehicles at the company to find out whether customers are satisfied with the service provided or not. This study adopted the CRISP-DM method. CRISP-DM helps simplify the sentiment analysis process from the beginning to the end with a special focus on applications in a business context. The stages of research using the CRISP-DM method that will be carried out in this study are explained as follows:

a. Business Understanding

The unit studied in this study is customer satisfaction of the company that the business condition is quite good, but there is a significant business problem, namely a lot of suboptimal data. This problem makes it difficult to conduct sentiment analysis effectively, thus hampering the company's efforts to identify and improve customer satisfaction.

b. Data Analysis

In this phase, we use a customer review dataset consisting of 4,353 raw data. This dataset includes conversation_id, sender, rating, and customer_review attributes that provide important information for sentiment analysis of customer satisfaction.

c. Data Preparation

Data preparation is done using jupyter notebook as the main tool, with python as the programming language used to process and analyze data. The output produced is a confusion matrix that displays the accuracy, precision, and recall matrices of the model built. The TF-IDF stages include calculating the frequency of word occurrences in documents and measuring how important the word is overall which is then used to weight the words in the text before modeling.

d. Modeling

At this stage, the results of the best model developed with the logistic regression algorithm for sentiment analysis are displayed. The main output is a probability graph that shows the visualization of positive and negative sentiment predictions for each text analyzed.

e. Evaluation

At the evaluation stage, metrics such as accuracy, precision, recall and F1-score are used, all of which are measured through a confusion matrix. To ensure the level of model trust, testing is carried out by dividing the data between training data and test data and cross-validation. The level of trust is determined based on the results of this evaluation with values that indicate how well the model can predict the results accurately and consistently.

RESULTS

In an effort to develop a deeper understanding of sentiment analysis of services and optimize data usage at the company, this study analyzes customer interaction data stored in the freshchat application. Freshchat is a communication platform used by the company to interact with customers, this makes the freshchat application a very relevant data source for conducting sentiment analysis research on the services that the company has provided to customers. Table 1 is an example of a raw data sample taken from the freshchat application that will be subjected to sentiment analysis, a total of 4,353 raw csv data, in the table there are several attributes such as conversation_id, sender, rating, and customer_review.

Table 1. Data Row Sample

conversation_id	*sender*	*rating*	*customer_review*
0	Customer	5	I bought oil at this workshop very cheap
1	Costumer	4	Satisfied with the service
2	Costumer	5	Yokohama tire brand is cheap!
3	Costumer	5	Friendly to me
4	Costumer	5	I like the service
5	Costumer	5	Brake pad replacement is very fast
6	Costumer	5	Friendly and I am satisfied with the service
7	Costumer	5	I am impressed with his expertise
8	Costumer	5	Gives the best recommendations
9	Costumer	3	Oil changes take quite a long time
10	Costumer	5	Very Expert in repairing cars
11	Costumer	5	Gives good advice for Service
12	Costumer	5	Fast in working
13	Costumer	5	The results are not disappointing
14	Costumer	5	Very impressed with the service

3.1 Data Preprocessing

a. Data Cleaning

At the data cleaning process stage, the commands provided by the python library are used. Data will be processed using the Jupyter Notebook application to be cleaned by deleting unnecessary and irrelevant data/attributes such as hashtags, mentions, symbols, emoticons, and whitespace and eliminating duplicate data.

Data before cleaning:
[Yokohama tire brand is cheap!], [Very good at repairing cars □],
[Fast at work □], [Very impressed with the service :D].

Data after cleaning:
[Yokohama tire brand is cheap], [Very good at repairing cars], [Fast at work], [Very impressed with the service].

b. Tokenization

Tokenization is the process of dividing text into smaller units called tokens. At this stage, tokenization is run using a python nltk library that has been installed in the jupyter notebook. TF-IDF is a method for evaluating how important a word in a document is to the collection of documents. To calculate tf-idf, this study uses the python library that has been provided, namely scikit-learn. Scikit-learn will help evaluate a word in the dataset whose output is in the form of a word list that describes the weight of each word in the document.

Data before tokenization:
[Yokohama tire brand is cheap], [Very Expert in repairing cars], [Fast at work], [Very impressed with his service].

Data after tokenization:
[Yokohama tire brand, cheap], [Very, Expert, in, repairing, cars], [Fast, in, working], [Very, impressed, with, his, service].

c. TF-IDF

Table 2 shows the weighting results that have been carried out using the tf-idf method, the results of the tf-idf process are obtained from the calculation process carried out using a python library called scikit-learn. In the process, scikit-learn will identify important words in a sentence. TF-IDF (Term Frequency-Inverse Document Frequency) is used to help identify important words in a document by giving higher weight to words that often appear in a data. Common words such as "and", "or", "adalah" will have low weight because they often appear in many documents.

Table 2. TF - IDF

Number	Will Be	Good	Can Be	Bad
1	0.000000	0.614189	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.707107
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.447214	0.000000
5	0.463693	0.000000	0.000000	0.000000

d. Stop Words Removal

Stopword removal is used to remove words that do not affect sentiment such as conjunctions, prepositions, and articles. This study uses an Indonesian stopword dictionary or corpus from a python library called nltk.corpus. In the process, nltk.corpus performs text analysis and natural language processing and provides easy access to various curated resources that can be used to build models.

Data before stop words:

[Satisfied, with, service], [Brand, tire, Yokohama, cheap], [I, like, with, service, nya]. Data after stop words:

[Satisfied, service], [Brand, tire, Yokohama, cheap], [like, service].

e. Text Normalization

Text normalization is the stage of converting text from capital letters to lowercase letters. This is done to make it easier to read and for equalization with other data in the dataset. This command is written in the python programming language using the lower library. Lower is one of the libraries for normalizing text to ensure that differences in capitalization do not affect further text analysis.

Data before stop words:

[Satisfied, with, service], [Brand, tire, Yokohama, cheap], [I, like, with, the service, nya]. Data after stop words:

[satisfied, service], [brand, tire, yokohama, cheap], [like, the service].

f. Words Count

In Figures 2, 3 and 4 are the results of the word cloud process. Word Cloud is a visualization of the results of preprocessing data that has been done word cloud has the aim of finding out words that often appear in the dataset. Words that appear in the word cloud are limited to only 500 words based on the frequency of the most words that appear from positive, negative, and neutral sentiments. To create a word cloud visualization, this study uses a library from python and a way to find words with frequent occurrences using the scikit-learn library, where the process has been carried out in the tf-idf process.

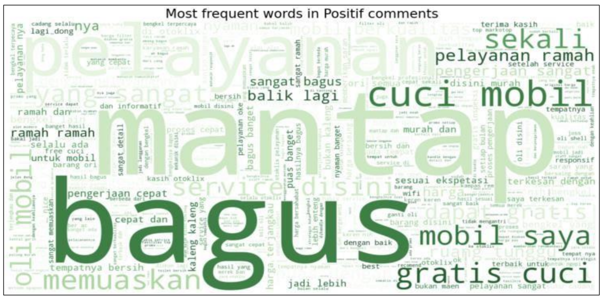

Figure 2. Positive Sentences That Often Appear

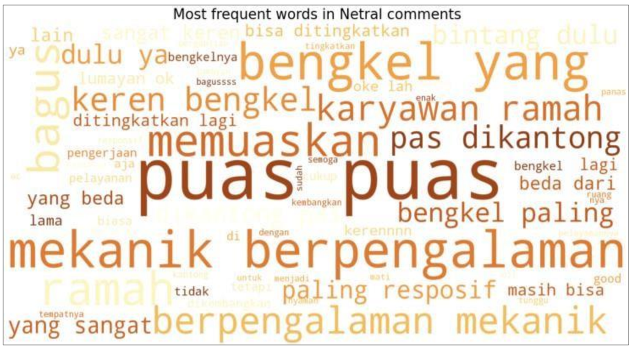

Figure 3. Neutral Sentences That Often Appear

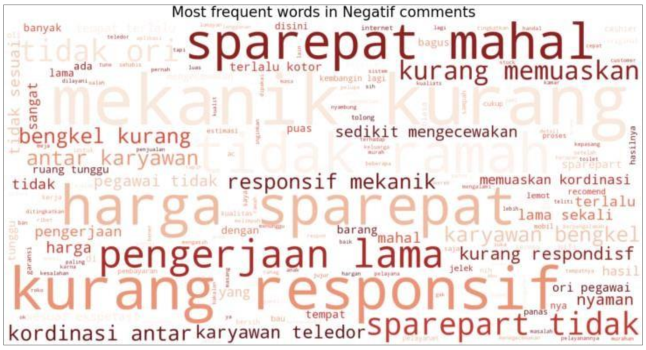

Figure 4. Negative Sentences That Often Appear

3.2. Testing

In the testing section, the model performance measurement mechanism is presented with the aim of evaluating the effectiveness of the model in analyzing sentiment from customer review data at the company. The testing method used is a confusion matrix, which is equipped with test parameters such as accuracy, precision, and recall. The test results are presented in tabular form to facilitate interpretation and at this stage the test results will be interpreted in detail to match the topic of the sentiment analysis being carried out. In the research conducted, this study used a supervised learning approach with a division of 80% training data and 20% test data with an input feature containing a text column that will be analyzed in this study, the text analyzed is 'customer review' and the output label that shows the sentiment column or text assessment in this study the output label is the 'rating' column.

Table 3 shows the results of the training data sample after the training data division of 80% which has 3,481 raw data after the dataset division.

Table 3. Sample Training Data

conversation_id	***sender***	***rating***	***customer_review***
1	Customer	4	great service
2	Customer	4	shell oil is cheap
3	Customer	5	really great
4	Customer	5	good results
5	Customer	5	always be successful

Table 4 shows the results of the test data sample that has been divided into 20% of the test data which has 871 raw data after the division.

Table 4. Sample Test Data

conversation_id	***sender***	***rating***	***customer_review***
1	Customer	5	abundant promo
2	Customer	4	original goods
3	Customer	5	comfortable workshop
4	Customer	5	friendly service
5	Customer	5	reliable technicians

In the data sharing process helps in assessing how well the model built can predict new data based on previously trained data. Training data is a collection of data used to train the model. The model is learned using this data so that it can make predictions or decisions with data that has never been seen before and for the process of dividing training data and test data there is no data that fails to process. In this study, the model was trained with a series of data and then tested to measure the level of accuracy.

Table 5. Accuracy Results of Training Data and Test Data

Iterasi	**Hasil**
1	0.8723
2	0.8376
3	0.8433
4	0.8548
5	0.8433
Akurasi Final	**0.8369**

Table 5 shows the accuracy results from five training iterations. Model testing is done to assess how well the model predicts labels from data that has never been seen before. Testing is done using k-fold cross-validation which divides training data and test data. For k-fold cross-validation, it is processed in a jupyter notebook using the python programming language. The accuracy value is obtained from the results of measuring model performance, this value is obtained by calculating the accuracy which is the ratio of the number of correct predictions to the total number of predictions. The data before testing displays data that has been preprocessed and has not been converted into a numeric format that can be processed as a model, while the data that has been tested on the data model has been converted into numeric numbers and has been trained with training data and tested on test data. Model accuracy is calculated based on how well the model predictions match the actual labels on the test data. The formula for calculating accuracy:

$$\text{Accuracy} = \frac{\text{Total True Prediction}}{\text{Total Prediction}} \tag{1}$$

Tabel 6. Summary Metrics

0	Metric	Value
1	Accuracy	0.946039
2	Precision	0.942568
3	Recall	0.946039

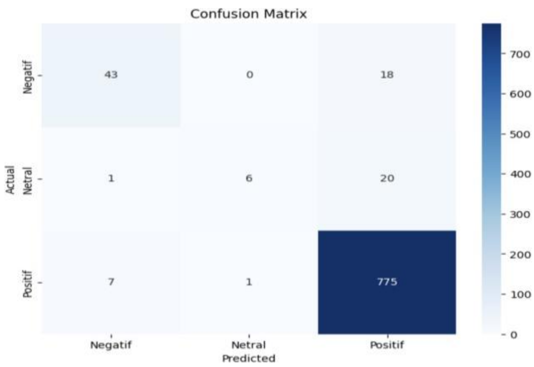

Figure 5. Sentiment Analysis Results in the Confusion Matrix Table

Table 6 and Figure 5 are summary metrics and sentiment analysis results in the confusion matrix table obtained from the test results. In the summary metrics, it can be ascertained that the level of customer satisfaction with company's services is positive, indicating that customers are satisfied with the services provided. With an accuracy level reaching 94.6%, a precision of 94.3%, and a high recall of 94.6%. This sentiment analysis model has proven effective in identifying customers who are satisfied with the services they receive, as seen in the sentiment analysis results in Figure 5, which shows 775 reviews that are truly expressed in positive sentences. To obtain the accuracy, precision and recall values in this study, the following formula was used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Note:
TP = *True Positives*
TN = *True Negatives*
FP = *False Positives*
FN = *False Negatives*

All of these formulas are done and processed in the jupyter notebook application using the python programming language and assisted by the sklearn.metrics library. The library provides various functions to evaluate model performance including sentiment analysis classification.

DISCUSSION

From the results of the analysis and design that have been carried out, several conclusions can be drawn regarding the research on the application of sentiment analysis to find out customer satisfaction at the company :

a. This study successfully answered questions about customer satisfaction with company's services through sentiment analysis. By using the classification method on a total of 4,353 customer reviews, where 3,481 reviews were used as training data and 871 reviews as test data. This study shows that sentiment analysis techniques can provide in-depth insights into customer perceptions and expectations. The results of the analysis using the logistic regression model produced an accuracy of 94.60%, a precision of 94.26% and a recall of 94.60%

which indicates that customer satisfaction can be predicted well through this analysis.

b. Data that was previously under-optimized can now be processed effectively through the classification method. With a total of 4,353 reviews and dividing the data into 80% for training and 20% for testing, sentiment analysis provides in-depth insights into customer experience. This allows the company to make more targeted strategic decisions based on a better understanding of customer feedback.The implementation of sentiment analysis techniques in this study has succeeded in providing significant results in assessing customer satisfaction. The tests conducted showed that with a deep understanding of customer sentiment with an accuracy of 94.60%, a precision of 94.26%, and a recall of 94.60% the company can evaluate whether the service provided is optimal or needs further improvement.

c. The implementation of sentiment analysis techniques in this study has succeeded in providing significant results in assessing customer satisfaction. The tests conducted showed that with a deep understanding of customer sentiment with an accuracy of 94.60%, a precision of 94.26%, and a recall of 94.60% the company can evaluate whether the service provided is optimal or needs further improvement.

CONCLUSION

From the results of the analysis and design that have been carried out, several conclusions can be drawn regarding research into the application of sentiment analysis to find out customer satisfaction.

a. This study successfully answered questions about customer satisfaction with the company's services through sentiment analysis. By using the classification method on a total of 4,353 customer reviews, where 3,481 reviews were used as training data and 871 reviews as test data. This study shows that sentiment analysis techniques can provide in-depth insights into customer perceptions and expectations. The results of the analysis using the logistic regression model produced an accuracy of 94.60%, a precision of 94.26% and a recall of 94.60%

which indicates that customer satisfaction can be predicted well through this analysis.

b. Previously under-optimized data can now be processed effectively through classification methods. With a total of 4,353 reviews and dividing the data into 80% for training and 20% for testing, sentiment analysis provides deep insights into customer experience. This allows company to make more targeted strategic decisions based on a better understanding of customer feedback.
c. The implementation of sentiment analysis techniques in this study has succeeded in providing significant results in assessing customer satisfaction. The tests conducted showed that with a deep understanding of customer sentiment with an accuracy of 94.60%, a precision of 94.26%, and a recall of 94.60% the company can evaluate whether the service provided is optimal or needs further improvement.

REFERENCES

A.E. Pratama., A. A. (2022). Analisis Sentimen Masyarakat Terhadap Tim Nasional Indonesia pada Piala AFF 2020 Menggunakan Algoritma K-Nearset Neighbors. *TICOM*.

Amin, S. A. (2022). Analisis Sentimen Evaluasi PEmbelajaran Tatap Muka 100 % Pada Pengguna Twitter MEnggunakan Metode Logistic Regression. *Jurnal Pendidikan Tambusai*, 13217-13227.

K.T Putra, M. H. (2022). Perbandingan Feature Extraction TF-IDF dan Bow Untuk Analisis Sentimen Berbasis SVM. *Jurnal Cahaya Mandalika*, 1449-1463.

L. Gaol., S. S. (2019). Implementasi Data Mining Dengan Metode Regresi Linear Berganda Untuk Memprediksi Data PErsediaan Buku Pada PT. Yudhistira Ghalia Indonesia Area Sumatera Utara. *KOMIK*.

M. Lestandy, A. A. (2021). Analisis Sentimen Tweet Vaksin COVID-10 Menggunakan Recurrent Neural Network and Naive Bayes. *Jurnal Rekayasa Sistem dan Teknology Informasi*, 802-808.

M.H, M. D. (2023). Analisa Sentimen Tweet COVID-19 Menggunakan MEtode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan Count Vectorizer. *Jurnal Ilmu Multidisiplin*, 37-43.

S. Huber, H. W. (2019). DMME : Data Mining Methodology for Engineering Application a Holistic Extension to the CRISP-dm mODEL. *pROCEDIA cirp, eLSEVIER*, 403 - 408.

Singgalem, Y. (2023). Analisis Sentimen dan Sistem Pendukung Kepuasan Menginap di Hoel Menggunakan Metode CRISP=DM dan SAW. *JOurnal Information System Research*, 1343-1353.

Utami, S. (2029). Analisis Sentimen PEngguna Twitter MEngenai 'Sedotam Plastik' Dengan Metode K-Nearest Neighbor (KNN) dan Neighbor-Weightes (NWKNN) . *Tugas Akhir*.

Wibowo, F. V. (2019). Analisis Sentimen PElanggan Toko Online JD-Id MEnggunakan Metode Naive Bayes Classifier Berbasis Kobversi Ikon Emosi. *Jurnal SIMETRIS*, 2019.

Willy Novianto, W. W. (2023). ANALISIS SENTIMEN TERHADAP KEPUASAN PELANGGAN MENGGUNAKAN METODE KLASIFIKASI PADA. *SENAFTI*, 281 - 288.