



Sentiment Analysis of Electricity Company Service Quality Using Naïve Bayes

Yuli Astuti¹, Yova Ruldeviyani², Faris Salbari³, Aldiansah Prayogi⁴

^{1,2,3,4}Magister of Technology Information, Faculty of Computer Science, Universitas Indonesia

¹yuli.astuti11@ui.ac.id, ²yova@cs.ui.ac.id, ³faris.salbari11@ui.ac.id, ⁴aldiansah.prayogi11@ui.ac.id

Abstract

In facing the era of technological disruption, a large company providing electricity in Indonesia, namely PT PLN is transforming to digitize all business processes and improve the quality of customer service. PLN Mobile application was developed in December 2020, and 18 million users have downloaded it. PLN Mobile application provides various electrical services for users. There are a lot of online opinions today. Organizations need to know the public perception of their product or service, sales projections, and customer happiness. Our research will identify public opinion (positive and negative) about PLN Mobile Application using sentiment analysis by taking review data from Google Play Store. Sentiment analysis is classified using Naïve Bayes and analyzed based on the dimensions of the quality of electricity services: empathy, responsiveness, and reliability. The results of this study indicate that Naïve Bayes is quite well used for binomial labels (positive and negative) with an accuracy of 73%. Still, for service quality dimensions, the accuracy is 45%. Indonesian language datasets are quite difficult to process due to non-standard language, foreign words, mixed language variations, and abbreviations. Determination of ground truth or manual labeling requires consistency and skilled personnel to determine the context of the text data to obtain a model with optimal performance. This study informs the classification of each dimension of the quality of electricity services in Indonesia based on positive and negative sentiment data for PLN Mobile Application users. Reliability received the most negative sentiments. This can be used for PT PLN to improve the quality-of-service reliability to customers.

Keywords: sentiment analysis; naïve bayes; electricity; service quality

1. Introduction

Organizations in the public sector encounter numerous difficulties in a highly competitive economy. The public sector should likewise emphasize service quality and include it in framing strategic decisions. The aforementioned factors necessitate that public sector enterprises in developing nations upgrade their operations and perform more effectively [1].

In facing the era of technological disruption, a large company providing electricity in Indonesia, namely PT PLN is transforming to digitize all business processes and improve the quality of customer service. One of them is by developing applications that are closer to customers, providing the best service and experience as well as convenience for customers, such as the convenience of new electricity installations, additional power, easy token purchases, electricity payments, and submitting complaints on the one hand. The application is named PLN Mobile which was inaugurated in December 2020.

Based on data from Google Play Store and App Store, PLN Mobile Application has been downloaded by 18 million users (android and iOS) with a rating of 4.5. PLN Mobile Application features currently include electricity services (New Installs, Power Changes, Simulations), Payments (Tokens and Electricity Bills), Meter Reading and SwaCam (Monthly Billing Simulations), Complaints (Interference and Complaints related to electricity) with outage management, Information Provision Subsidies, SPKLU (General Electric Vehicle Charging Station), and internet installation.

For businesses to learn about general perceptions of their goods and services, sales projections, and actual customer happiness, this internet opinion is crucial data [2]. Companies can find possibilities to raise the caliber of their goods or services based on this information.

Sentiment analysis is a crucial field that lets us understand users' general attitudes about certain topics. Organizations can determine consumer happiness

using this information. Several authors have provided definitions of sentiment analysis. However, Liu's definition is the one that is most frequently adopted in the scientific community [3], which describes sentiment analysis as a branch of research that examines people's attitudes, evaluations, and judgments of various objects, including goods, services, organizations, people, situations, events, subjects, and qualities. Many methods for sentiment analysis have been put out in recent years. Semantic orientation and machine learning are the two fundamental techniques on which most of these strategies are built. Although both strategies produced good results, several studies in the literature have indicated that machine learning produces superior outcomes. Deep learning, a revolutionary methodology that greatly outperforms conventional approaches, has, nonetheless, gained the attention of researchers recently [4], [5].

Most users provide comments related to PT PLN services through comments on the Google Play Store. Previous research has analyzed PLN Mobile Application to assess acceptance and usefulness [6], to analyze user opinions using Twitter data [7], and the grouping user comment types [8]. Our research will identify public opinion (positive and negative) about the PLN Mobile Application using sentiment analysis by taking review data from Google Play Store. Then, we analyze public opinions that can be used to improve the quality of products or services at PT PLN. User comments through the Google Play Store became management KPIs for service improvement, business process improvement, and exploring new services that can improve the customer experience.

The formulation of sentiment classification is a classification of two types of issues, namely positive and negative [3]. The online reviews of the PLN Mobile Application were used as the training and testing data. The positive and negative categories are identified using ratings because online reviews typically include a rating score that is awarded by reviewers, such as 1–5 stars. For instance, reviews with four or five stars are seen as positive, but reviews with one to two stars are regarded as negative. Text classification issues include sentiment analysis. The word sentiment or opinion in the sentiment classification denotes a positive or negative opinion, such as great, extraordinary, dreadful, bad, terrible, etc. The algorithm used in this research is Naïve Bayes.

Research related to improving service quality using sentiment analysis has been conducted by [2]. The study used Twitter data to determine the level of customer satisfaction using the SVM and Naïve Bayes algorithms. The study by [8] grouped the types of application complaints using Naïve Bayes. Research

by [9] used Naïve Bayes to determine the Quality of Service on Wifi networks. Our research chose Naïve Bayes to extract the opinion of PLN Mobile Application users to improve the quality of products or services at PT PLN.

The components of the quality of electricity services have been identified through the prior study. Five components or dimensions were found from 21 variables using exploratory factor analysis, which was used to determine service quality dimensions. In the context of Sri Lanka, the five qualities—tangibility, empathy, responsiveness, reliability, and assurance—are categorized as dimensions of the quality of electrical services [1]. Research by [10] mentions three determinants of service quality in Gambian power companies: assurance, responsiveness, and empathy. Research by [11] uses five service quality dimensions, reliability, responsiveness, assurance, empathy, and tangible, to determine customer satisfaction in Nigerian power companies.

Our research takes three factors that will be examined according to the service context at PLN Mobile Application to analyze sentiments like empathy, responsiveness, and reliability. From this research, PT PLN develops and implements a strong strategy and business with professional service delivery focused on consumerism.

2. Research Methods

This study uses a research methodology, as shown in Figure 1.

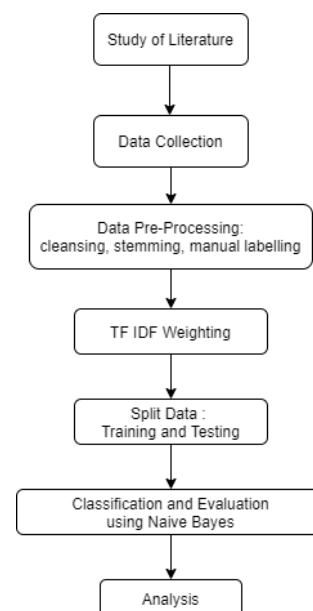


Figure 1. Research Methodology

2.1 Study of Literature

Based on the research of [1] in the Sri Lankan electricity company, research of [10] in the Gambian

power company using exploratory factor analysis (EFA), and research of [11] in the Nigerian power company, three dimensions of electricity service quality are taken as follows.

Empathy is the attention given to customers during or after service delivery [10]. Customers are satisfied when the business completely understands their expectations and the customer service offered can address their issues. Empathy is a process that businesses can use to improve communication with their clients to ensure and offer appropriate solutions to unmet consumer wants from the services supplied as a strategy to promote customer happiness.

The company's responsiveness describes employees' spirit in providing services as quickly as possible whenever customers need them [10]. Employees have high enthusiasm for providing services to customers in a timely and hassle-free manner. Employees respond to customer complaints when needed and provide solutions to any customer problems as quickly as possible so that customers are satisfied with the services provided.

Reliability is to provide consistent service as expected by customers [11]. Reliability results in customer retention when a service provider can supply services accurately and continuously by client expectations. Overall, service quality is significantly impacted by reliability. The ability of the business to overcome any potential obstacles to expected service and maintain a high error-free rate serves as the foundation for reliability in service quality for customer satisfaction.

Service quality is described from various perspectives by several researchers. According to [11], service quality is the connection between client perceptions and expectations of certain service offerings. According to [12], consumer attitudes, perceptions, expectations, and satisfaction with services are all influenced by service quality. According to [13], [14], the customer's overall assessment of service excellence is service quality. The user's overall perception of the company and its services' relative dominance is also considered about service quality.

A study conducted by [15] in an Indian electric utility firm specifies seven items, including power quality, supply, mode of utility bill payment, complaint processing, new connectivity, security, inconsistent voltage, and time management for bill payments. The quality of services in the electrical sector needs to be improved, and suitable policy changes must be made. They place more emphasis on the crucial significance that offering excellent customer service has in raising the caliber of their services.

2.2 Data Collection

The internet is one of the sources of data in the era of Big Data development that can be acquired cheaply and used for a variety of reasons. Data collection from this research is based on one of the Google Play Store reviews of the PLN Mobile Application in September 2021 via the internet. The entity obtained from the review is in the form of the username used by the customer for the review, the date of the review, the star rating given, and comments from the reviewer. One of the tools used for data retrieval is Webharvy [16], [17] which is then used to retrieve PLN Mobile application review data on the Google Play Store.

When retrieving data using Webharvy, JavaScript is required to display all full reviews from users who wrote longer than the limit to display text. The Inspect button element informs the button class code that can be used to generate the JavaScript exemplified as a code.

```
[a] var links =  
document.getElementsByClassName('LkJZd  
ScJHi OzU4dc');  
  
[b] for(var i=0; i < links.length; i++)  
{  
  
    links[i].click();  
  
}
```

Rating data is needed to make it easier to categorize user comments into positive and negative. Then to be able to retrieve rating data, it is necessary to capture HTML and then apply an Aria Regular Expression (Regex) label with the sentence before the star number, as shown in Figure 2.

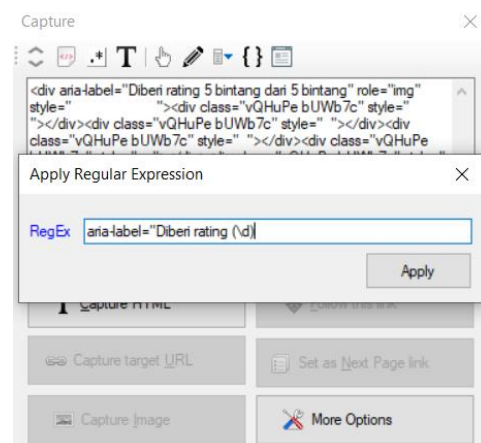


Figure 2. Display to Retrieve Rating Data

2.3 Data Pre-Processing

We did unnecessary manual deletion to remove irrelevant words and make the words uniform, and we got 1500 data sets. Then to break the sentence into tokens or per word, it is necessary to use the

Tokenization method [18]. Data cleaning is required to remove unnecessary punctuation and delete Stop Words after the Tokenization process. Stemming is used for pre-processing data to obtain information by tracing the words that are affixed back to the root. In Indonesia, the Stemming method has been researched and proven with high accuracy of results [19] to get word compaction into basic words to make it easier to conduct the analysis, and it is necessary to use the Stemming method. After getting the basic words from the Stemming results, we try to do manual labeling in a positive or negative direction which will later try to be weighted and classified. Labeling is divided into two classes "Positive" and "Negative". The class "Positive" denotes a positive user review, whereas the word "Negative" denotes a negative user review. Search engines typically evaluate or rate the degree of connection between documents and user queries using different TF-IDF weighting methods [18] so that after the text data has been successfully transformed into uniform and basic words, they start to weigh as much importance as words using the TF-IDF method. The Naïve Bayes technique will be used to do sentiment analysis on user comments or opinions to categorize the judgments they make about this series.

2.4 Data Training and Testing

Training data and testing data are split into two sets with a ratio of 80%:20%. The training data is used to generate a Naïve Bayes classification model. Then the model will be applied to the testing data. The splitting of the dataset into training and testing uses stratified sampling to ensure that each subset has a proportional and balanced distribution of data.

2.5 TF-IDF Weighting

Utilizing the TF-IDF approach, the review text data was weighted. A shorthand for the terms "term frequency" and "inverse document frequency," respectively, is TF-IDF. The first topic to be covered deals with the term's frequency; TF is used to calculate how frequently a term appears in a text [18]. If we have a document named "Doc1" with 1500 words and exactly 10 instances of the word "Good," we can use this as an example. Since a document might be somewhere between extremely short and very long overall, any term can appear more frequently in longer texts than in shorter ones. To solve this issue, the frequency of any phrase is determined by dividing its occurrence in the document by the total number of terms therein. Therefore, formula 1 may be used to determine the frequency of the word "Good" in the document "Doc1".

$$TF = 10/1500 = 0,006 \quad (1)$$

Second, we will talk about document frequency inversion. When the document term frequency is determined, the algorithm evaluates every keyword

similarly, regardless of whether it is a stop word like "from," which means it may be inaccurate [18]. Varying keywords have different weights. The IDF exists for circumstances like this, where the word "from" has no use or has very little value while appearing 2000 times in the document. Those that frequently appear in documents are given less weight by document frequency inversion, whereas words that do not are given greater weight. For instance, if we have 10 documents and 5 of them contain the word "application," we may use Formula 2 to determine the document frequency inversion.

$$IDF = \log_e (10/5) = 0,3010 \quad (2)$$

It is clear from TF and IDF that the more frequently words occur in a document, the higher the term frequency, and the less frequent words occur in a document, the higher the degree of importance (IDF) for the keywords searched for in certain documents.

TF-IDF will generate numeric data in the form of vectors that represent each text data point into numbers. In this research, examples of text data vectorization results can be seen in Table 1. Each word has its weight (TF-IDF calculation) for each document or in this case, is a review.

Table 1. Sample of TF-IDF Text Vectorization Results of PLN Mobile Application

Review	aplikasi	gajelas	mudah
gajelas banget padahal baru download daftar padahal baru guna aplikasi	0.04	0.38	0.0
gajelas aplikasi sangat mudah guna semua kalangan layanan sangat cepat	0.05	0.0	0.15
aplikasi sangat bantu pelanggan mudah moga lebih baik depan	0.06	0.0	0.15

2.6 Classification Using Naïve Bayes

Information extraction about product reviews ran into several issues. The issue might be viewed as an uncertainty issue. The Naïve Bayes classifier is one of the machine learning techniques that can be utilized to solve these issues. Naïve Bayes is an approach to uncertainty reasoning that makes use of an inference-based probabilistic model. The Naïve assumption—that characteristics are expected to be conditionally independent given a class—is enforced by Naïve Bayes. However, it has been demonstrated that Naïve Bayes performs well for many classification issues. It is also well known that Naive Bayes performs exceptionally well with straightforward calculations [20].

It is easy to assume that sentiment words and phrases might be employed for sentiment classification in an unsupervised manner since sentiment words are frequently the dominant element in the categorization of sentiment. The unsupervised learning strategy is

like Turney's (2002) research, which categorizes based on several fixed syntactic patterns used to express thoughts. Part-of-speech (POS) tags serve as the foundation for the syntactic pattern [21].

A dictionary of sentiment words and phrases with related orientations is used in the lexicon-based method, another unsupervised approach, to combine intensification and negation to calculate a sentiment score for each document. This approach was first applied to the sentiment classification of sentences and aspects [22].

Aspects and sentiments are two factors that are classified. We may state that both factors have an impact on how words are used in sentences by characterizing the system in terms of a generative model. As a result, each variable's value affects the word probability distribution.

Figure 3 depicts the procedure for creating the classification model in this study.

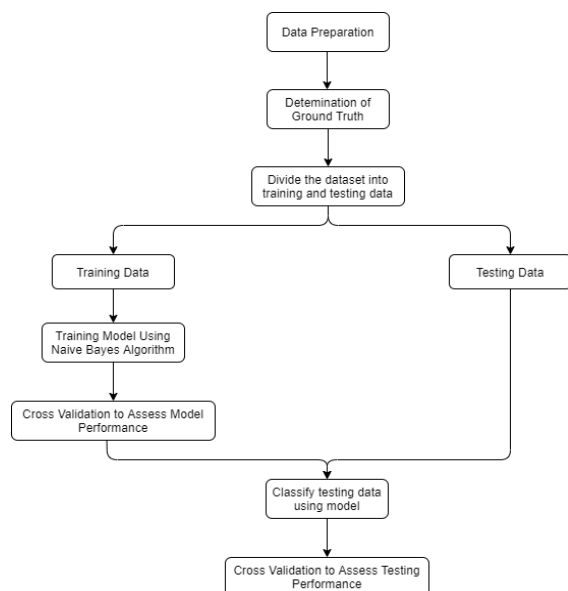


Figure 3. Step of Making a Naïve Bayes Classification Model

Implementation of the Naïve Bayes algorithm for data classification in this research is using Rapidminer. After the data is divided into two sets and weighted in the form of a numerical vector, the Naïve Bayes classification algorithm will be implemented in the training data and produce a classification model.

After that, the performance of the model will be validated using the cross-validation method. Cross-validation applies a classification model to the training data itself to validate whether the model can predict the suitable class for each data point.

The evaluation of the models was done using a variety of indices, including the confusion matrix, accuracy, precision, and recall. The accuracy of actual and predicted class is represented by the confusion matrix

by comparing the predicted class and actual class. The accuracy shows the percentage of actual events in all examined data. Precision shows the comparison between true event and total true event predicted by classifier. Recall shows the comparison between true predictive event and total number positive event. The formulas for this evaluation of the models are provided in Formula 3, 4 and 5 [23].

$$Accuracy = \frac{true\ positive + true\ negative}{total\ number\ of\ predictions} \quad (3)$$

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (4)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (5)$$

After the classification model gets optimal performance results, the model can be applied to the testing data to get the class prediction results made by the model on the labels in the testing data. Then the classification results are validated using cross-validation to see their performance.

2.7 Analysis

After the classification, an analysis of the service quality of 1500 PLN Mobile Application users' opinion data was carried out. Each opinion is grouped into the dimensions of empathy, responsiveness, and reliability. In addition to being grouped into individual dimensions, user comments are also grouped into positive and negative. Further negative comments will be made to PT PLN for improvements to its services.

3. Results and Discussions

The experiment was conducted to find out whether the Naïve Bayes classification method is appropriate for the user review dataset against the PLN Mobile Application on the Google Play Store. In this case study, the experiment was carried out with a multi-label classification approach. Each review data is given two independent labels, namely sentiment (negative or positive) and service quality dimensions (empathy, responsiveness, or reliability).

Table 2 informs the amount of review data taken from the Google Play Store and has been grouped into 3 dimensions of electrical service quality.

Table 2. Classification of PT PLN Service Quality

Dimensions	Number of Opinions	Percentage (Number of Opinion/ Total Opinions)
Empathy	367	24%
Responsive	410	27%
Reliability	723	48%
Total	1.500	100%

The performance results of training and testing data can be seen in Table 3. The results showed relatively good testing performance on binomial labels (sentiment labels) but relatively poor on multinomial

labels (category or dimension labels). Based on training and testing data for sentiment classification, it was found that the Naïve Bayes algorithm has an accuracy of 73% for sentiment labels, and 45% for service quality dimensions labels

Table 3. Naïve Bayes Algorithm Accuracy

Label	Accuracy
Sentiment (Positive and Negative)	73,08%
Service Quality Dimensions	44,79%

The experiment's results also showed that the model did not get overfitting if the training data's accuracy was much better than the accuracy of the testing data. In this case study, overfitting was avoided with balanced data, a model that was no more complex than the number of datasets, and cross-validation of the model against the training data.

Tables 4 and 5 show the confusion matrix. The findings show that the resulting model is less able to recognize data with empathy labels. This is most likely due to unbalanced data and inconsistent ground truth. Datasets with Indonesian have their challenges, non-standard words, the use of foreign words, mixed language variations, and the frequent use of abbreviations for a word or phrase. The libraries in programming languages available to pre-process Indonesian datasets are also very limited. This requires more effort at the pre-processing stage of text data, such as creating rule-based cleaning by hardcoding.

Table 4. Confusion Matrix on Sentiment Labels

Sentiment	True Negative	True Positive	Precision
Predictive Negative	112	39	74.17%
Predictive Positive	44	104	70.27%
Recall	71.79%	72.73%	

Table 5. Confusion Matrix on Quality Dimensions

Dimensions	True Reliability	True Responsive	True Empathy	Precision
Predicted Reliability	73	27	27	57.48%
Predicted Responsive	39	38	13	42.22%
Predicted Empathy	53	16	14	16.87%
Recall	44.24%	46.91%	25.93%	

Table 6 informs the classification of each dimension of service quality based on PLN Mobile Application user sentiment data. Reliability gets the most negative sentiment at 28%. This shows that reliability is the most complained about by customers compared to other dimensions.

Examples of customer comments on each dimension with positive sentiments can be seen in Table 7.

Table 6. Classification of Sentiments on Each Dimension of Quality

Dimensions	Number of Positive Sentiment	Number of Negative Sentiment	Percentage of Positive Sentiment	Percentage of Negative Sentiment
Empathy	186	181	12,4%	12,1%
Responsive	264	146	17,6%	9,7%
Reliability	307	416	20,5%	27,7%
Total	757	743	50,5%	49,5%

Table 7. Customer Comments with Positive Sentiments

Dimensions	Customer Comments
Empathy	<i>Sangat bagus dan cepat pelayanannya hitungan menit langsung direspon operator trima kasih ya</i> (Very good and fast service in minutes immediately responded by the operator thanks).
Responsive	<i>Aplikasi sangat membantu setelah pengaduan beberapa jam kemudian langsung eksekusi pengerjaan perbaikan gangguan listrik.</i> (The application is very helpful after a complaint a few hours later directly the execution of repairing electrical disturbances).
Reliability	<i>Membantu proses pembayaran listrik apalagi di masa pandemi.</i> (Helping the electricity payment process, especially during a pandemic). <i>Mempermudah dalam laporan gangguan pelanggan.</i> (Make it easier to report customer outages).

Table 8 shows examples of negative sentiment related to each dimension of service quality.

Table 8. Customer Comments with Negative Sentiments

Dimensions	Customer Comments
Empathy	<i>Mohon untuk pasang baru jangan dipersulit.</i> (Please do not make it difficult for new installation services)
Responsive	<i>Masalah belum ditangani bahkan belum ada petugas yang datang tapi di aplikasi tertulis sudah sampai tahap selesai.</i> (The matter has not been dealt with even though no officers have arrived, but in the written application, it has reached the stage of completion).
Reliability	<i>Mati listrik melulu.</i> (Frequent power outages). <i>Semenjak kemaren mati lampu. Sekarang listrik di rumah kami mati nyala mati lagi nyala lagi.</i> (Since yesterday, the lights went out. Now the electricity in our house is off and on again).

Figure 4 shows the words that often arise from the sentiment analysis of PLN Mobile Application users, namely 'aplikasi, sangat, listrik, layan, bantu, cepat, mobile, baik, mudah, tugas, lapor, baru, bayar, kasih, langgan, respon, token, gangguan'.

Things that companies can improve in response to these customer comments are simplifying the process of new installation services, synchronizing the process of repairing power failures in the field with the status of work in the application, and repairing the power grid that often experiences outages.

- doi: 10.47119/ijrp100591820201389.
- [18] J. H. Jaman, R. Abdulrohman, A. Suharso, N. Sulistiowati, and I. P. Dewi, "Sentiment analysis on utilizing online transportation of indonesian customers using tweets in the normal era and the pandemic covid-19 era with support vector machine," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 5, pp. 389–394, 2020, doi: 10.25046/AJ050549.
- [19] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J. Big Data*, vol. 8, no. 1, pp. 1–16, 2021, doi: 10.1186/s40537-021-00413-1.
- [20] M. S. Mubarak, A. Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes," *AIP Conf. Proc.*, vol. 1867, no. August, 2017, doi: 10.1063/1.4994463.
- [21] Peter D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pages 417–424, Morristown, NJ, USA, 2001. Ass," *Society*, vol. 12, no. 3, pp. 417–424, 2002, [Online]. Available: <http://www.google.com>.
- [22] M. Taboada, J. Brooke, and K. Voll, "Lexicon-Based Methods for Sentiment Analysis," no. September 2010, 2011.
- [23] A. I. Pandesenda, R. R. Yana, E. A. Sukma, A. Yahya, P. Widharto, and A. N. Hidayanto, "Sentiment Analysis of Service Quality of Online Healthcare Platform Using Fast Large-Margin," *Proc. - 2nd Int. Conf. Informatics, Multimedia, Cyber, Inf. Syst. ICIMCIS 2020*, pp. 121–125, 2020, doi: 10.1109/ICIMCIS51567.2020.9354295.
- [24] M. S. Alrajak, I. Ernawati, and I. Nurlaili, "Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K- Nearest Neighbor (K-NN)," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 110–122, 2020.
- [25] S. A. S. Herjunaa, G. A. Fatonib, and A. Yakub, "Penerapan Algoritma Naïve Bayes Classifier Menggunakan R-Programming Untuk Pengelompokan Jenis Keluhan Aplikasi Pln Mobile," vol. 2, no. 1, pp. 19–30, 2022.