SENTIMENT ANALYSIS OF SERVICE PROVIDER ON TWITTER TWEET USING NAIVE BAYES CLASSIFIER WITH PHP

Juri Pebrianto
Teknik Informatika, Universitas Pamulang
Jl. Surya Kencana No.1, Pamulang, Kota Tangerang Selatan, Banten
juripebrianto@gmail.com

Abstract

The purpose of this research is to undertake a complete sentiment analysis of Twitter users' opinions and attitudes about internet services provided by Indosat, Ooredoo, and XL Axiata in Indonesia. This study applies the Naive Bayes Classifier algorithm to efficiently categorize attitudes into positive, negative, and neutral groups, using the Twitter API for data collecting and the PHP programming language for data processing. The results of sentiment analysis reveal striking trends: 56% of Indosat Ooredoo service-related tweets contain unfavorable attitudes, whereas 50% of XL Axiata service-related tweets have comparable negative sentiments. Significantly, the sentiment analysis system constructed using PHP has a remarkable accuracy rate of 78% when compared to manually classified findings. This study adds vital information by throwing light on customer sentiments towards cellular internet service providers. This study allows each provider to fine-tune and optimize their internet services based on a data-driven understanding of consumer sentiment by illuminating user opinions. In conclusion, this study fills a significant information gap by analyzing user attitude towards prominent cellular internet service providers. Its novel methodology, which incorporates PHP and the Naive Bayes Classifier algorithm, not only provides an effective way of sentiment analysis but also gives providers with practical information for improving service quality.

***Keywords**: Naïve Bayes, Php, Sentiment Analysis, Twitter*

INTRODUCTION

The increase in the volume of user-generated content on Twitter has resulted in tweet sentiment analysis becoming an essential tool for the extraction of information about Twitter users' emotional state (H. T. Phan et al., 2020).

This shift in communication patterns is visible on platforms such as Twitter, where people chronicle their daily lives, express their thoughts on numerous topics, and participate in discussions about popular themes. Unlimited freedom of expression, coupled with the versatility of various online platforms, has accelerated the transition from conventional blogs and mailing lists to compact microblogging media. As a result, people are increasingly sharing insights about their product and service experiences, as well as their thoughts on a wide range of topics such as politics, religion, personality, products, and more.

Twitter has become a major social media platform and has attracted considerable interest among researchers in sentiment analysis (Wang et al., 2022). There are more than 300 million active Twitter users, making it one of the most popular microblogging services (Reyna et al., 2022). Companies perform sentiment analysis to examine feedback on products, government and other agencies use it for public-health monitoring and predicting political trends, and so on (M. Bibi et al., 2020). Prior to the emergence of social networks, manual mechanisms were usually employed for this purpose. Companies used to manually analyze the popularity of their products by surveying customers. However, with the advent of social networks, e.g., twitter, manual analysis of data has become a challenging problem.

According to the researcher, numerous concerns and comments from internet users about various services are expressed through social media channels. Twitter, in particular, has emerged as a significant medium for customer service interactions, alongside traditional communication channels such as phone and email. These distinct qualities give a rich data set for this study, allowing businesses to examine the benefits and drawbacks of their goods and develop strategic strategies to retain public participation. This is referred to as sentiment analysis, and is at the core of this research.

P-ISSN: 2656-1719
E-ISSN: 2656-2774

Currently, sentiment analysis is the study of the processing of views, sentiments, and emotions communicated through textual to intentionally convey affective states and thus are suitable indications of sentiment and emotion in texts (Chen et al., 2021) and expressions pertaining to entities and characteristics are mostly done by hand. As a result, boosting the sentiment value of user opinions automatically remains a hard challenge. This study focuses on the normal linguistic style utilized by Twitter users in communicating their point of view. To address this issue, text mining must be implemented, which includes preprocessing, feature extraction, and the use of the Naive Bayes Classifier as the primary classification tool to classify polarity opinions as positive or negative.

Sentiment analysis, often known as opinion mining, is a burgeoning field that deals with the processing of individual sentiments and emotions as expressed through the textual medium and involves entity detection, sentiment analysis (F. K. Sufi et al., 2021). This approach determines if attitudes are positive, negative, or neutral by distinguishing the polarity of text fragments in phrases or texts (Catelli et al., 2022). The technology of text Mining can discover, retrieve and extract information from a text corpus, which is usually too complicated for manual work . To be more specific, text mining combines technologies such as natural language processing, artificial intelligence, information retrieval, and data mining to help understand complex written analytical processing systems (P. Wang et al., 2020).. Text mining in natural language processing is a research topic based on big data analysis, which includes many related topics: sentiment analysis (Do et al., 2019), (Yadav et al., 2020), (Drus et al., 2019), text classification (Minaee et al., 2021), (H. T. Phan et al., 2020); and text clustering (Ghonge et al., 2019).

In this study methodology, the Naive Bayes Classifier, a leading classification algorithm, emerged as a prime contender. However, before it can be used in a more specific setting, the Naive Bayes Classifier must be rigorously tested. This is predicated on the realization that Twitter conversations, as reflected in tweets, exhibit different characteristics than more traditional types of writing, such as newspapers or scientific studies.

As a result, the purpose of this research is to discover the significant impact of Twitter dialogue on service and product reputation perceptions. The major purpose is to discern the range of public response elicited by Twitter chats, which encompasses both positive and negative. This affirmation is extremely important since brand reputation has a significant influence on consumer perceptions, influencing purchasing decisions across a wide range of items. Notably, existing research has demonstrated the genuine impact of Twitter discourse on brand reputation. The precise directional impact of this influence, however, is dependent on elements such as rumors spread through word-of-mouth, ratings and positive comments sourced from sites beyond the Twitter realm, or even offline locations.

To capture this dynamic connection, researchers exploit vast amounts of data using advanced programming techniques and Twitter's Application Programming Interface (API). In this investigation, a detailed evaluation of the PHP programming language: Hypertext Preprocessor is conducted, utilizing its built-in string manipulation capabilities, which are vital for research purposes. As a result, this study ventures into unexplored terrain by investigating the practicality of PHP for tasks such as text mining, sentiment analysis, and the use of the Naive Bayes Classifier approach.

While Python is generally linked with sentiment analysis and text mining because of its large ecosystem of tools, frameworks and pre-processes using the natural language toolkit provided by the Python (P. Gupta et al., 2021), using PHP in this context gives a unique and new perspective. The choice to investigate PHP as a tool for data mining, analysis, categorization, and sentiment analysis was motivated by various factors, including its untapped potential: Python has unquestionably established itself as a data science and analytics powerhouse, with to packages as diverse as NLTK, Scikit-learn, and TensorFlow. However, the potential of PHP, which is frequently hailed for its web development abilities, has gone relatively untapped in the world of text-based analysis. This study delves into PHP's latent capabilities and sheds light on its feasibility for tasks beyond its conventional application and language string manipulation capabilities, honed over years of web development, proving to be extremely advantageous in processing and dissecting textual data, aspects that are an important part of sentiment analysis.

P-ISSN: 2656-1719
E-ISSN: 2656-2774

Building on previous research, the majority of which followed well-trodden paths, this pilot study charts uncharted territory by utilizing the power of PHP—a language known for its web development capabilities—to uncover the complexities of sentiment analysis in Indonesian-language tweets relating to a specific product. The dataset is on a revolutionary journey after being polished and rigorously tuned utilizing a number of powerful text mining techniques functioning within the PHP ecosystem. This complex procedure not only corrects intrinsic flaws but also increases the data set's robustness, making it ideal for further research.

Importantly, this study takes advantage of PHP's natural characteristics to begin a deep sentiment categorization, classifying them into a fundamental triad of positive, negative, and neutral. The sophisticated algorithm of the Naive Bayes Classifier takes center stage in this complicated choreography, expertly orchestrating the classification process. This symbiotic relationship between PHP and the Naive Bayes Classifier not only simplifies classification efforts but also enables careful adjustment of sentiment dynamics. These intricate connections demonstrate PHP's previously untapped potential in the field of sentiment analysis, paving the path for future academics to discover new insights and approaches in this dynamic subject.

RESEARCH METHODOLOGY

2.1. Data Set

The Twitter Application Programming Interface (API) facilitates data collection on Twitter. This enables PHP to retrieve tweets relevant to the keyword "internet services offered by Indosat, Ooredoo, and XL Axiata" in real time and save them to a database. To establish a representative sample, collections were carried out methodically within a predetermined time span.

2.2. Using PHP to Preprocess Data

After receiving the data, a crucial step is to use PHP-based preprocessing techniques to improve data quality. PHP's intrinsic string handling capabilities are used for tasks such as case folding (the process of equalizing cases in a document), cleansing (which is a word cleaning stage that has no effect on sentiment classification results at all), converting emoticons (this greatly influences the sentiment of a document, because emoticons can describe the feelings of the text), converting negation (the process of converting negation words), tokenization or tokenization (the process of converting tokenization words), and tokenization or tokenization (the process of token , Normalization or normalization (is a technique for returning words in a dataset that are abbreviated in writing or words that are not in the large Indonesian dictionary), stopwords (may add data dimensions to the classification process), and stemming (is a procedure for changing the words) are all techniques for returning words in a dataset that are abbreviated in writing or words that are not in the large Indonesian dictionary. This procedure simplifies the textual material so that it can be analyzed further.

2.3. Sentiment analysis with the Naive Bayes Classifier

The Naive Bayes Classifier is well-known for its success in text-based sentiment classification, utilizing a probabilistic model to assess the polarity of a tweet's sentiments. In this phase, the PHP programming employs the Naive Bayes Classifier method by reading the database that has passed the preprocessing stage and directly reading the labeled data. The first label is supplied manually, and the reading function uses PHP to demonstrate the accuracy of employing the Naive Bayes algorithm in the PHP programme, which allocates each tweet to one of three categories: positive, negative, or neutral.

P-ISSN: 2656-1719
E-ISSN: 2656-2774

2.4. Justification Method

Our chosen technique is supported by the specific strengths of PHP and the Naive Bayes Classifier. Because of its string processing features, PHP, being a multipurpose programming language, is well suited for preprocessing tasks. This makes it easier to convert raw Twitter data into an organized and refined data set suitable for sentiment analysis.

The Naive Bayes classifier was chosen because it is compatible with text-based analysis and can manage the intricacies of sentiment categorization in social media information. By exploiting PHP's pre-processing capabilities and including the Naive Bayes Classifier, our approach handles the complicated challenges of sentiment analysis in the context of microblogging platforms in a synergistic manner. Designing or proposing and implementing the model will be carried out in this study, the following is a diagram depicting the model proposed in this study.

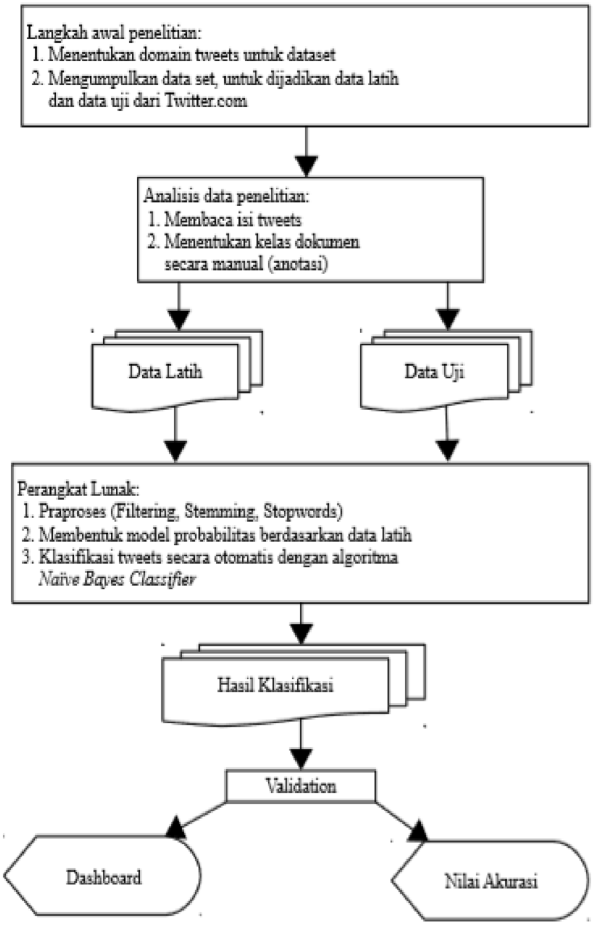

Image 1. Flow of Research Method

The explanation of the above is as follows:

1. Determine the Domain, the Tweet Domain used in this research is XL, Indosat. Tweets containing these keywords will be automatically generated.
2. Data collection in this study was sourced from the Twitter.com microblog. Data retrieval uses the Twitter API which is integrated into the PHP programming language.
3. Annotate or define a class on a collection of datasets. This annotated dataset is used as training data. To get data with the golden standard, the annotations are done manually.

P-ISSN: 2656-1719
E-ISSN: 2656-2774

4. The preprocessing processes that will be carried out at this stage are Case folding, cleaning, converting emoticons, converting negations, tokenizing, normalization, stop word removal and Steaming.
5. Formation of a probability model, at this stage all the training data Tweets that have been collected will be converted into a collection of terms that have a probability value of the specified class.
6. Classification, this stage will group tweets into positive, negative and neutral classes. The algorithm used is the Naive Bayes classification. After classification, the quality of the resulting model will be tested by calculating the accuracy.
7. The output data that has been classified will have information value and then be displayed on the dashboard so that users can see sentiment in the specific Domain.

2.5. Data Set

The text dataset comes from publicly uploaded tweets taken using the Twitter Application Programming Interface. The crawling process on Twitter is carried out by utilizing the provided Application Interface (API) facility. By first registering the program on Twitter developers. The limit given by Twitter in accessing its API is 100 Tweets, the author crawls the Twitter data and enters the data into a database with a different id_tweet then the data stored in the database will be used as a dataset for analysis. The keywords that will be searched on Twitter are "Indosat internet and XL internet", these keywords are used as categories in this analysis. To access the Twitter API, the Twitter API link used to integrate with the PHP programming language is https://api.twitter.com/1.1/search/tweets.json. The return data from the query is in json format which will then be directly processed in PHP programming language scripts.

The results of the analysis and testing that have been carried out using 342 training and test data from public opinion taken from Twitter using the Twitter API regarding the topic of Indosat internet services and XL internet services using the Naive Bayes Classifier method, can provide good accuracy percentage values, because the Naive method Bayes Classifier in carrying out document classification tests can have a high level of classification accuracy although it is very simple and efficient, namely only by utilizing the use of the occurrence and frequency of features in each opinion. In addition, the results of the classification accuracy using the Naive Bayes Classifier method often depend on the amount of training data used in the system. The use of training data in testing is 24 positive, 153 negative and 165 neutrals.

Looking at the data from sentiment analysis and accuracy in the results sub-chapter, the authors conclude that social media Twitter is widely used as a medium of communication between service providers and consumers to express complaints or appreciation for the services or products they use, especially cellular providers for Internet Service Providers in Indonesia. Namely Indosat Ooredoo and XL Axiata.

Most Twitter users upload negative complaints about the cellular provider's services which they feel are lacking in service, an example of some of the negative complaint data can be seen in the table below:

Table 1. Example of a negative sentiment Tweet

ID Tweet	ID User	Text
1165870108838785024	Hadyafn	@myXLCare siang mau tanya, no xl mama saya gak bisa di paketin internet sudah seminggu ini. Sudah dicoba matikan da… https://t.co/vOFmqfFYFI

P-ISSN: 2656-1719
E-ISSN: 2656-2774

1165544048028860417	deLoucious	Hi @myXLCare kenapa XL tidak bisa koneksi internet, jika sambungan data hidup selalu keluar motif "sign in to network" ??!
1163840945495740416	Agugege1	@IndosatCare hy saya dari pengguna 085787016977 saya benar sangat kecewa dengan layanan Indosat, terkadang saya d… https://t.co/tSUNiJdF2a

Apart from Tweets with negative sentiments, most Tweets only have neutral sentiments. Judging from the sentences of the Tweets, most of the Tweets contain promos. An example of some of the tweet data with a neutral sentiment can be seen in the table below:

Table 2. Example of a neutral sentiment Tweet

ID Tweet	ID User	Text
1161614861433884675	MNCVisioncare	@linborneng Baik, sesuai pembicaraan dengan Ibu Evelin melalui telepon layanan TV berlangganan + Internet XL kuota… https://t.co/PeK8QDEIZW
1163375159647956992	esther100419	RT @kompascom: Penawaran ini berupa gratis layanan XL Prio Pass selama di Singapura, menariknya program ini sudah bisa dimanfaatkan sejak 1…
1164291456531456001	bandarpulsa95	Kuota Paketan Internet INDOSAT terbaru. Mulai dari 250MB. Harga termurah! Cek harga dan order https://t.co/6B5ejjTvwA

The positive sentiment that the writer found was the least in number, users rarely appreciated or uploaded their positive sentiments in using the internet service provided by the cellular provider, an example of some of the positive sentiment Tweet data can be seen in the table below:

Table 3. Example of a positive sentiment Tweet

ID Tweet	ID User	Text
1165921663764041729	klikmagdotcom	XL Axiata Perluas Jaringan 4G di Sumbawa dan Berikan Donasi Internet Cepat https://t.co/HyOgzTA3g8 https://t.co/hnuLgHYnXq
1161935794350616577	DedySuwadha	Pakai Kartu XL PRIO Pass Gratis Layanan Internet Selama di Malaysia dan Singapura – https://t.co/d9l94JeOEy https://t.co/7zmsz4zE7L
1165591726557282305	dokyeomchu	Maketin kuota internet pake indosat enak bener cuy, murahhh ❤ love bgt deh 1164338754657501185

P-ISSN: 2656-1719
E-ISSN: 2656-2774

2.6. Analysis

The dataset in this study uses the JSON format collected from Twitter by the Crawling method from social media Twitter which is Integration with the PHP programming language. The data taken is only tweets in Indonesian, namely Tweets with the keywords Internet and Internet XL. The data is taken randomly either from Tweets that are public on Twitter. The dataset used is 342 Tweets, 165 for Indosat internet keywords and 177 for XL internet keywords.

Sentiment label weighting is first done manually and then after being given a sentiment label, the data becomes training data for an automated sentiment analysis system. Sentiment label weighting on Tweets contained in the system is given with reference to positive and negative weighting words derived from sentiment label data which is done manually by the author.

The preprocessing process is important for the next stage, namely reducing enhanced attributes that have less influence on the classification process. The data entered at this stage is still dirty Raw Data, so the result of this process is a quality document which is expected to facilitate the classification process.

The preprocessing process consists of several stages, namely, cleaning data, filling in missing values, smoothing noisy data, recognizing or eliminating outliers, and solving inconsistencies, data transformation, normalization and aggregation, data reduction, obtaining representations that are reduced in volume but produce the same or similar analytical results, data discretization, part of data reduction but with special importance, especially numerical data.

RESULT

The data being tested is data resulting from manual labeling by the author, testing with the same data was carried out due to the limited amount of Tweet data that can be obtained and also in addition to that, in order to know directly the accuracy of the system created in labeling sentences on Tweets automatically. Automation. The greatest value of accuracy when testing is 78% with the following calculation:

3.1. Sentiment Analysis of Internet Services

Sentiment analysis of Twitter data provides an in-depth view of user sentiment towards internet services supplied by Indosat, Ooredoo, and XL Axiata. The findings highlight significant patterns in sentiment dispersion, offering information on the overall sentiment landscape.

3.2. Provider Comparative Analysis

Analyzing the distribution of sentiment among the three providers reveals that negative sentiment dominates discussions regarding Indosat Ooredoo services, with a score of 56%. Similarly, talks about XL Axiata services revealed a substantial negative sentiment score of 50%. These data not only give a picture of the current user mood, but also highlight the importance of strategic adjustments by these providers.

3.3. Insights and Implications

When compared to previous research, our findings provide a fresh perspective by combining PHP-based preprocessing with the Nave Bayes Classifier for sentiment analysis in the context of mobile internet services. This method enables a more in-depth assessment of user sentiment, particularly in the context of Twitter microblogging. The use of PHP in the preprocessing phase adds a novel layer to sentiment analysis, contributing to a larger discussion concerning the usage of unusual programming languages in sentiment research.

P-ISSN: 2656-1719
E-ISSN: 2656-2774

3.4. Naive Bayes Classifier

The results of this study will discuss the results of the classification using the Naive Bayes Classifier algorithm in analyzing sentiment on Twitter tweets taken using the Twitter API integrated with the PHP programming language as follows:

Table 4. Confusion Matrix keyword analysis "internet Indosat"

ACTUAL	POSITIVE	NEGATIVE	NEUTRAL
POSITIVE	4	1	5
NEGATIVE	0	79	12
NEUTRAL	2	4	58
TOTAL	6	84	75
	165		

From the data table above, it can be calculated and determined the sentiment value of Twitter Tweets with the keyword "internet Indosat" are:

$$\frac{Match\ Value}{True\ Positive + True\ Negative + True\ Neutral}$$

$$Positive = \frac{4}{4 + 79 + 58} = \frac{4}{141} = 0{,}03 = 3\ \%$$

$$Negative = \frac{79}{4 + 79 + 58} = \frac{79}{141} = 0{,}56 = 56\ \%$$

$$Neutral = \frac{58}{4 + 79 + 58} = \frac{58}{141} = 0{,}41 = 41\ \%$$

Based on the calculation above, the biggest sentiment rating is negative for Tweets with the keyword "internet Indosat".

Table 5. Confusion Matrix keyword analysis "internet XL"

ACTUAL	POSITIVE	NEGATIVE	NEUTRAL
POSITIVE	15	2	1
NEGATIVE	2	63	42
NEUTRAL	1	4	47
TOTAL	18	69	90
	177		

P-ISSN: 2656-1719
E-ISSN: 2656-2774

From the data table above, it can be calculated and determined the sentiment value of Twitter Tweets with the keyword "internet XL" are:

$$\frac{Match\,Value}{True\ Positive + True\ Negative + True\ Neutral}$$

$$Positive = \frac{15}{15 + 63 + 47} = \frac{15}{125} = 0{,}12 = 12\ \%$$

$$Negative = \frac{63}{15 + 63 + 47} = \frac{63}{125} = 0{,}50 = 50\ \%$$

$$\text{Neutral} = \frac{47}{15 + 63 + 47} = \frac{47}{125} = 0{,}38 = 38\ \%$$

Based on the calculation above, the biggest sentiment rating is negative for Tweets with the keyword "internet XL".

3.5. Validation and Accuracy

After obtaining these data, the author wants to get the accuracy value from sentiment analysis for the keyword "internet Indosat and Internet XL" with the formula and accuracy calculations used are as follows:

$$\frac{(TP\ +\ TN\ +\ TNT)}{(TP\ +\ FP\ +\ TN\ +\ FN\ +\ TNT\ +\ FNT)}$$

Explanation:
TP is True Positive, which means positive predictions are positive.
TN is True Negative, which means negative predictions are negative.
TNT is True Neutral, which means the neutral prediction is neutral.
FP is False Positive, which means a positive prediction whose value is wrong.
FN is False Negative, which means a negative prediction whose value is wrong.
FNT is False Neutral, which means a neutral prediction whose value is wrong

Based on the formula above, it can be calculated the value of the accuracy for the Twitter Tweet sentiment analysis system with the keyword "internet Indosat" which is made using the PHP programming language as follows:

$$\frac{(4\ +\ 79\ +\ 58)}{(4\ +\ (0 +\ 2)\ +\ 79\ +\ (1\ +\ 4) +\ 58\ +\ (5\ +\ 12))}$$

Simplified to:

$$\frac{141}{6\ +\ 84\ +\ 75} = \frac{141}{165} =\ 0{,}85$$

Meanwhile, the same calculation is performed for the keyword "internet XL" with the following calculations and results:

$$\frac{(15 +\ 63\ +\ 47)}{(15 +\ (2\ +\ 1)\ +\ 63\ +\ (2 +\ 4)\ +\ 47\ +\ (1\ +\ 42))}$$

P-ISSN: 2656-1719
E-ISSN: 2656-2774

Simplified to:

$$\frac{125}{18+69+90} = \frac{125}{177} = 0{,}71$$

$$Avg = \frac{141+125}{165+177} = \frac{266}{342} = 0{,}78 = 78\ \%$$

When compared to manual classification results, the sentiment analysis system gets a noteworthy accuracy rate of 78%. This validation highlights the resilience of our process and increases the trustworthiness of sentiment analysis results.

DISCUSSION

The provided text discusses the results of sentiment analysis on Twitter tweets using the Naive Bayes Classifier algorithm. The author manually labeled a limited amount of tweet data and tested it with the same data to determine the accuracy of the system they created for automatically labeling sentences on tweets. The highest accuracy achieved during testing was 78%.

The discussion begins by presenting a confusion matrix for keyword analysis of "internet Indosat." The matrix shows the actual sentiment (positive, negative, or neutral) and the corresponding predicted sentiment. The sentiment values are then calculated based on the matrix, and it is determined that the most prevalent sentiment for tweets with the keyword "internet Indosat" is negative.

Next, a similar confusion matrix is provided for keyword analysis of "internet XL." The sentiment values are again calculated based on the matrix, and it is found that the dominant sentiment for tweets with the keyword "internet XL" is also negative.

The text then introduces the concept of accuracy and explains the formula for calculating it. The accuracy is calculated separately for the sentiment analysis system with the keywords "internet Indosat" and "internet XL." For "internet Indosat," the accuracy is found to be 0.85, and for "internet XL," the accuracy is 0.71.

Finally, the average accuracy for both keywords is calculated, resulting in an overall accuracy of 0.78 or 78%.

In summary, the discussion focuses on the results of sentiment analysis using the Naive Bayes Classifier algorithm for tweets related to "internet Indosat" and "internet XL." It provides the sentiment values, accuracy calculations, and determines the dominant sentiment for each keyword.

CONCLUSION

Sentiments were broadcasted by Twitter users by uploading tweets with internet keywords Indosat where the services provided by Indosat Ooredoo, namely positive sentiments got a score of 3%, negative sentiments got a score of 56% and neutral sentiment got a score of 41% with this score Sentiment for Indosat Ooredoo is negative while sentiment for tweets with the internet keyword XL where the service is provided by XL Axiata gets a positive sentiment score of 12%, negative sentiment scores 50% and neutral sentiment scores 38% with this calculation, Sentiment towards the Internett service provided by XL Axiata is negative.

PHP programming language can be used to carry out the text mining process.

The accuracy value of the sentiment analysis system built with the PHP programming language is obtained by comparing the results obtained by manual classification by the author and obtaining the accuracy value above, the value for accuracy is 78%.

REFERENCE

Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian. *Electronics*, 11(3), 374. https://doi.org/10.3390/electronics11030374

P-ISSN: 2656-1719
E-ISSN: 2656-2774

Chen, Z., Cao, Y., Yao, H., Lu, X., Peng, X., Mei, H., & Liu, X. (2021). Emoji-Powered Sentiment and Emotion Detection from Software Developers' Communication Data. ACM Trans. *Softw. Eng. Methodol*., 30(2).

Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications*, 118, 272-299.

Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707-714.

F. K. Sufi and M. Alsulami, Automated Multidimensional Analysis of Global Events With Entity Detection, Sentiment Analysis and Anomaly Detection, *IEEE Access*, vol. 9, pp. 152449-152460, 2021, doi: 10.1109/ACCESS.2021.3127571.

Ghonge, A. H., & Kamble, V. (2019). A Review on Improving the Clustering Performance in Text Mining. *International journal of scientific research in science, engineering and technology*, 6, 380-385.

H. T. Phan, V. C. Tran, N. T. Nguyen and D. Hwang, Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model, *IEEE Access*, vol. 8, pp. 14630-14641, 2020, doi: 10.1109/ACCESS.2019.2963702.

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, 52(1), 273-292.

M. Bibi, W. Aziz, M. Almaraashi, I. H. Khan, M. S. A. Nadeem and N. Habib, A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis, *IEEE Access*, vol. 8, pp. 68580-68592, 2020, doi: 10.1109/ACCESS.2020.2983859.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, 54(3), 1-40.

P. Gupta, S. Kumar, R. R. Suman and V. Kumar, "Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter," *in IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 992-1002, Aug. 2021, doi: 10.1109/TCSS.2020.3042446.

P. Wang et al., Classification of Proactive Personality: Text Mining Based on Weibo Text and Short-Answer Questions Text, *IEEE Access*, vol. 8, pp. 97370-97382, 2020, doi: 10.1109/ACCESS.2020.2995905.

Reyna, N. S., Pruett, C., Morrison, M., Fowler, J., Pandey, S., & Hensley, L. (2022). Twitter: more than tweets for undergraduate student researchers. *Journal of Microbiology & Biology Education*, 23(1), e00326-21.

Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment Analysis of Twitter Data. *Applied Sciences*, 12(22), 11775. https://doi.org/10.3390/app122211775

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335-4385.