Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

Comparison of Algorithms for Sentiment Analysis of Operator Satisfaction Level for Increasing Neo Feeder Applications in PDDikti Higher Education LLDIKTI Region VI Semarang Central Java

M. Ulil Albab[1)*], Ema Utami[2)], Dhani Ariatmanto[3)]
[1,2,3)]Universitas AMIKOM Yogyakarta, Indonesia
[1)]m.ulil.albab05@gmail.com, [2)] ema.u@amikom.ac.id [3)]dhani@amikom.ac.id

Submitted : Aug 19, 2023 | **Accepted** : Aug 20, 2023 | **Published** : Oct 1, 2023

Abstract: Sentiment analysis on the satisfaction level of PDDikti operators is very important to find out how PDDikti operators feel after the version of the academic reporting application for higher education was upgraded, namely Neo Feeeder. The increase in the version of this application causes some of the features in it to not function properly. So some academic reporting activities from tertiary institutions experience problems. As a result of this condition, the most felt impact is students, where students experience delays in graduation. Then it is necessary to evaluate through sentiment analysis from PDDikti operators to find out the response from operators and be able to provide positive suggestions to developers from the PDDikti reporting application. This study applies several classification methods for sentiment analysis at once, including the Random Forest algorithm, the Support Vector Machine algorithm, the Multinomial Naïve Bayes algorithm, the Decision Tree algorithm, and the K-Nearest Neighbor algorithm. Of the 5 methods applied, the results of their performance accuracy will be compared. The performance of the highest classification algorithm is the K-Nearest Neighbor (K-NN) algorithm which produces an accuracy value when testing data, which is up to 90% using the oversampling technique in unbalanced classes. While the lowest classification accuracy performance value is in the Multinomial Naïve Bayes (MNB) algorithm with a value of 76%. It is proven that oversampling can help the performance of the classification algorithm to be more optimal. Thus, it should be noted that the balance of data classes is an important factor when applying the classification method.

Keywords: Decision Tree; K-Nearest Neighbor; Multinomial Naïve Bayes; Neo; Oversampling; Random Forest; Sentiment Analysis; Support Vector Machine;

INTRODUCTION

PDDikti operators are the spearhead of universities in reporting academic and non-academic data to the Central Higher Education Database (PDDikti). In reporting it, a PDDikti operator is equipped with academic knowledge about college data and college data reporting techniques. This is very important for operators to understand because the data reported is important data for each tertiary institution in the LLDIKTI VI working area which will be entered into PDDikti. The reported data becomes the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

background of each tertiary institution and provides evidence that the tertiary institution has academic data in such a way. As for reporting procedures, PDDikti operators are provided with an official application from the Ministry of Education and Culture that functions as a client. This application is embedded in each tertiary institution which runs as a client for reporting containers which will later be synchronized and entered into the PDDikti data center. The name of the PDDikti reporting client application is the PDDikti Feeder.

On March 1 2022, PDDikti has released a new application which is an improvement from the previous application. The application is called Neo Feeder. The advantages of the Neo Feeder application compared to the previous version are a fresher appearance, technology upgrades from monolithic to microservices, multi-platform Operating Systems, updates directly from within the application and integration with the Merdeka Campus. However, in using the Neo Feeder application from the initial release time, PDDikti operators encountered various problems and application bugs that had been experienced. The obstacle that is often experienced is not being able to synchronize data from the Neo Feeder application to the central PDDikti. This resulted in delays in reporting university data to the central PDDikti. Starting from here, there are many issues that say that this version upgrade is problematic and not ready and should be tested properly before being published and applied to all higher education service providers. Various comments and responses from operators were conveyed regarding this incident. Therefore, to find out in detail the responses, comments and opinions of PDDikti operators, a sentiment analysis is needed. This study is intended to determine the response to their level of satisfaction with the increase and update of the PDDikti Neo Feeder application. One method that can be implemented for this case is to analyze the sentiments of the operators to find out the comments and responses of the operators.

With the responses or opinions that appear, it will give a positive value to the neofeeder application developer to fix problems that occur such as errors and bugs in the application. So that in the future the neofeeder application will be even better and operators as users will feel good satisfaction. One of the places or social media where operators discuss and provide input suggestions is WhatsApp. Through the WhatsApp group of PDDikti Operators at LLDikti Region IV Central Java, all responses, comments in the form of chats sent to the group were collected. The research that will be carried out is the processing of comment and opinion data by PDDikti operators in the form of text. Therefore, sentiment analysis research is felt to be able to meet the needs of this research. Because sentiment analysis is research about processing and analyzing text-based opinion data that contains polarity so that it is able to produce positive, negative and neutral information (Laurensz & Eko Sediyono, 2021). From the data which was originally in an unstructured condition, a conclusion can be drawn using sentiment analysis (Giovani et al., 2020). Machine Learning (*supervised learning*) is a method that can be used to apply sentiment analysis with the ability to classify text (Sulastomo et al., 2022). One of the uses of the supervised learning method for text classification is to use the SVM (Support Vector Machine) algorithm (Oktafani & Prasetyaningrum, 2022). One of the advantages of the SVM algorithm compared to other classification algorithms is in determining the distance using a support vector so that the computation process will be faster (Desiani et al., 2022). Another classification algorithm is the Multinomial Naïve Bayes Classifier which is used to collect comments and classify positive or negative meanings for analysis with the aim of identifying attributes between status and certain behavioral and feeling relationships (Kewsuwun & Kajornkasirat, 2022). The K-Nearest Neighbor algorithm can also be used for classification by following the principle of similarity by calculating the distance (*Euclidean distance*) between points (Shah et al., 2022). The Random Forest algorithm is a classification algorithm that produces more than 1 decision tree (Morama et al., 2022). Then the Decision Tree decision algorithm can change facts to form decision trees that can represent rules (Prasetyo & Pahlevi, 2019). Of the 5 algorithm models, sentiment analysis research will be carried out on the satisfaction level of PDDikti operators on the increase in the Neo Feeder application to get the best model.

There have been many previous studies on sentiment analysis, including research conducted by (Wijaya, 2022) on Customer Satisfaction Sentiment Analysis for Food Delivery Service Providers by comparing 3 classification algorithm models. The results obtained by the three models can be said to be quite good with accuracy and F values above 78%. Using TF-IDF to extract features can be said to have succeeded in improving the evaluation results of the three models used. However, at the data

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

preprocessing stage carried out by (Wijaya, 2022) it still uses manual methods and takes time. Then (Mukarramah et al., 2021) conducted research using a polynomial kernel function with an accuracy value of 0.512, a precision of 0.437, a recall of 0.45, and an f-measurement of 0.512. However, the test results obtained are still relatively low for accuracy. Therefore, based on previous research, it is necessary to conduct research using several models to get better accuracy results.

LITERATURE REVIEW

Much research has been done on sentiment analysis. Such as research conducted by (Nurmawiya & Harvian, 2021) with an accuracy of reaching 97.87%. Then (Mardiana et al., 2022) conducted a study by comparing models Multinomial Naive Bayes, Complement Naive Bayes, XGBoost, Random Forest, Balanced Random Forest, Multi-Level Ensemble where on positive sentiment the best classification model is Balanced Random Forest, negative sentiment is a model Balanced Random Forest and neutral are models Balanced Random Forest. (Oktafani & Prasetyaningrum, 2022) succeeded in carrying out the classification by comparing the effect of using the RBF and Linear kernels and the number of positive labeled data was 4036 data and 1917 negative labeled data which resulted in the RBF kernel getting better accuracy than the linear kernel. Then research conducted by (Sulastomo et al., 2022) sentiment analysis using the SVM method resulted in a classification accuracy of 96%. Then (Turjaman & Budi, 2022) conducted research on Marketing Mix Aspect-Based Sentiment Analysis for Digital Wallet Application Reviews (Case Study: Linkaja Application on Twitter) using a 4-aspect scenario with the results The price aspect has the best accuracy value with a score of 80.0%, followed by product aspect 77.5%, promotion aspect 76.2%, place aspect 71.4%. Then (Hendriyanto et al., 2022) conducted research using the RBF (*Radial Basis Function*) kernel which produced 92.31% accuracy, 96.3% precision, 89.66% recall, and 92.86% f1-score.

METHOD

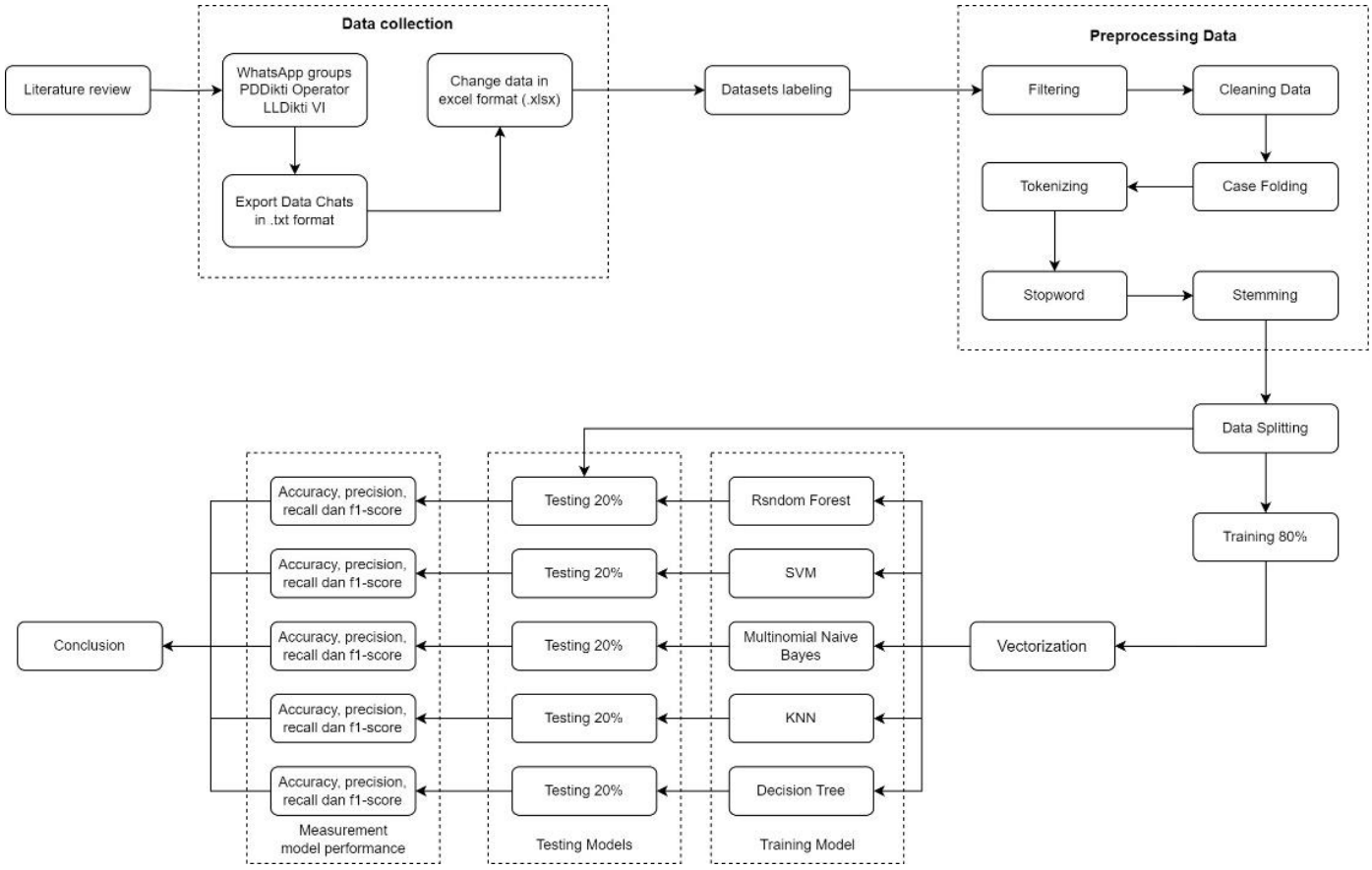

Fig 1. Research Flow

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

The flow in making a research step is by starting to examine the problem to be studied and followed by determining the method to overcome the problem. The data to be processed is chat data in the PDDIKTI Operator WhatsApp group Region VI Semarang, Central Java. The data will be grouped into 3 categories or classes, namely Positive, Negative and Neutral. All data that is considered to meet the requirements is then labeled by an expert or psychologist to get results. The initial steps of the research conducted by the author can be seen in Figure 1.

Dataset

The data collected and retrieved is from the social media group WhatsApp PDDikti Operator Region VI Semarang, Central Java. The data taken is in the period from 2 February 2022 to 30 September 2022 with a total of 319 data with the keyword "*neo*". Data is retrieved with the command export chats which can then be downloaded into .txt format. Then the data is entered into Microsoft Excel and saved in (xlsx) format. Then the data will be labeled with three classes of labels, namely positive, negative, and neutral. After labeling, the next step is data in Excel format will be retrieved on the *Google colab* platform using the *Python* programming language. This data will be ready for text preprocessing so that the data is clean and ready to be processed using the proposed method.

Data Labeling

Data that has been successfully collected is given a sentiment label as a condition for carrying out the classification stage. Labeling is carried out by human resources who are experts in the field of psychology, in this case he is named Su'ad Jauharoh, S.Psi., S.Pd.I., M.A. He is a lecturer who teaches at 3 (*three*) higher education institutions in Central Java. Labeling is done by human labor with the assumption that only living things can feel emotions or sentiments, while machines do not (Asro'i & Februariyanti, 2022). Categories or classes in data labeling consist of positive, negative, and neutral classes. After labeling, there were 166 data for the Neutral class, 71 data for the negative class and 82 data for the positive class. Illustratively the class distribution data after the labeling process can be seen in Figure 2.

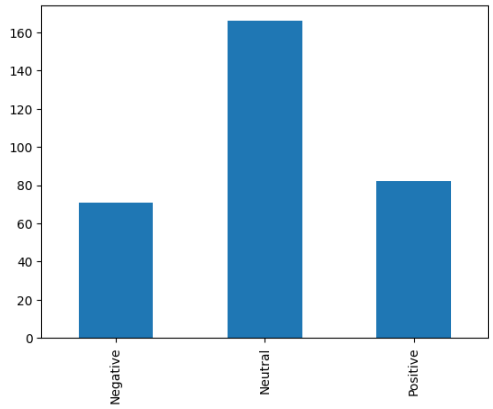

Fig 2. Visualization of Data Labeling Results

Preprocessing

It is an important stage in processing data in text form. The hope is that by doing a preprocessing of this text, the data to be used in this research is really clean and ready to use. As for the stages themselves, Filtering Data is the stage for *filtering* data to filter the data used for research needs. Then continue the process *Data Cleaning* is a step to remove unused words such as hashtags, usernames and reduce noise when analyzing sentiment. Stage *Case Folding* is a way to make all data into lowercase form (lowercase) with the intention of converting all sentences into all lowercase forms in a standard state ready to be processed. Then continue the stage *Tokenization*, is the stage of separating sentences into words meaning that all data in the form of sentences is broken down into separate forms per word. Stage *Stopword*, is the stage in removing words that do not have a descriptive meaning, meaning that affixes that do not mean anything will be deleted because they will be useless. When processed, they will not mean

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

anything. Last is the stage *Stemming*, which is the process of mapping and decomposing various forms of words into their basic forms so that words that already have affixes return to their initial or original forms to be ready for word processing.

Vectorization

Vectorization is the conversion of textual data into numeric data, where the value is the number of occurrences of a word in a document. This emphasis is necessary because computers only understand and process numerical information (Sastypratiwi et al., 2022). This study uses TF-IDF weighting for vectorization. TF-IDF is a way to emphasize the relationship of a word (term) with a document. This method combines two concepts of weight, namely the occurrence of words in certain documents and the occurrence of documents containing inverted words. The number of occurrences of a word in a particular document indicates the importance of the word in that document. The frequency of documents containing a word indicates how often the word appears. Therefore, the weight of the word-document relationship is high when the frequency of the word in the document is high and the frequency of all documents containing that word in the document is low. The TF-IDF formula is as follows:

$$tf = 0{,}5 + 0{,}5\ x \frac{tf}{max\ (tf)} \quad (1)$$

$$idf_t = log\ \left(\frac{D}{df_t}\right) \quad (2)$$

$$W_{d,t} = tf_{d,t}\ x\ IDF_{d,t} \quad (3)$$

Keterangan :

tf	= many words to search in a document
$max(tf)$	= the highest number of occurrences of a term in the same document
Nilai *D*	= document totals
df_t	= number of documents containing the term t
IDF	= inversed Document Frequency (log2 (D/df))
d	= d document
t	= the t-word of the keyword
W	= the weight of the d-th document to the t-word

Random Forest

The Random Forest algorithm is a way of grouping data that gets more than one result tree result (*decision tree*) uses a collection of randomly selected sample and training data attributes to improve performance (Morama et al., 2022). The process in deciding tree that is, determining the amount tree to be built, followed by calculations *gain impurity*, then do the calculations information gain at every possible moment question splitting. Next is the process of changing the value information gain highest be root. The last step repeats from the calculation *gain impurity* until k-tree.

Support Vector Machine

The Support Vector Machine (SVM) Classification Algorithm is an intelligent machine learning system that uses a hypothetical place as a linear function in a high-dimensional feature space trained by a learning algorithm based on optimization theory applying learning curves developed from statistical learning theory (Sulastomo et al., 2022). Support Vector Machine (SVM) is one of the classification methods that uses machine learning (*supervised learning*) which predicts the model class or models based on the results of the training process. Evaluation is carried out by looking for hyperplanes or decision boundaries that separate classes from the others, which in this case plays a role in differentiating positive mood sentences (*marked with +1*) from negative emotions (*marked with -1*) (Husada & Paramita, 2021). SVM looks for hyperplane values using support vectors and limit values. In this study,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

the input data has a vector representation obtained from the weighing process. With in-depth training the SVM classifier then creates emerging values or patterns that are used in the SVM testing process to characterize data sentiment. To obtain the optimal hyperplane line to separate data from the two classes, the hyperplane limit calculation is used and the maximum point is sought (Khotimah et al., 2022; Ramon et al., 2022). How to get a hyperplane with a support vector machine using the equation:

$$(w.x_i) + b = 0 \quad (4)$$

In the data, which belongs to class -1 can be formulated as the equation:

$$(w.x_i + b \leq 1, y_i = -1 \quad (5)$$

While the data on , which belongs to class +1 can be formulated as the equation:

$$(w.x_i + b \leq 1, y_i = 1 \quad (6)$$

Multinomial Naïve Bayes

Multinomial Naïve Bayes Classifier is one of the variations of the Naïve Bayes classifier algorithm. This algorithm uses the multinomial distribution of the conditional probability function. Even though it uses a polynomial distribution, this algorithm can be used in cases of text mining by converting text data into a nominal form that can be calculated using integer values (Verawati & Audit, 2022). On multinomial naïve bayes classifier, the document class is not only calculated from the words that appear but the number of occurrences of the word itself.

K-Nearest Neighbor

K-Nearest Neighbor is a similarity-based artificial learning algorithm that works based on the closest value. K-nearest neighbor is an algorithm that takes training data and gives a K value to find the closest K based on the calculated distance (Sholeha et al., 2022). So the KNN algorithm is very dependent on distance calculations. In this case the calculated distance between nearest neighbors uses Euclidean distance. In another sense, KNN is a supervised method that requires training data to group data using the shortest distance. The working principle of KNN is to find the shortest distance between the data to be executed in the training data. In KNN all data that is owned must have a label, so when there is new data it is compared to existing data and the data that is most similar is taken and looks at the label of the data (Isnain et al., 2021; Simanjuntak et al., 2017)..

Decision Tree

Decision Tree or Decision tree is a popular and effective method of prediction and classification. Tree decisions can change facts to form decision trees that can represent rules. The regulations that apply are easy to understand. The procedure performed in the decision tree method transforms the data into a decision tree, changing tree into roles, and simplify roles (Prasetyo & Pahlevi, 2019). In the decision tree method, inner node, node root, and the final node is part of tree. For a while, a variable or property is root and inner node, and class name is leaf node. The classifier keeps track root and an inner node with query data to the end node. The class identifier in this request is based on an existing identifier on the internal node

Confusion Matrix

Measurement of the performance of a grouping algorithm is something that needs to be taken into account. Group system performance describes how well the system classifies data (Siringoringo, 2018). The Confusion matrix is one way that can be used to calculate the performance of a grouping method. Basically, confusion matrix contains information that compares the results of the classification performed by the system with the results of the classification that should be. An example of a confusion matrix for binary classification is shown in Table 1.

Table 1. Grouping Measurements

Class	Positive Classification	Negative Classification
Positive	True Positive	False Negative
Negative	False Positive	True Negative

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

Based on the table above, it can be explained that True Negative is negative data that is detected correctly, while False Positive is positive data but is detected incorrectly. Meanwhile, True Positive is positive data that is correctly detected. False Negative is negative data that is detected incorrectly. In other words, the performance value is the comparison between data that is correctly classified, namely true negative and true positive. The performance value can be found by the equation

$$Akurasi = TP + TNTotal\ X\ 100\% \quad (7)$$

While the precision value is the number of positive category data that is divided correctly and divided by the total data that is positively divided, precision can be obtained in the equation

$$Presisi = TPFP + TP\ X\ 100\% \quad (8)$$

Meanwhile, recall reflects the percentage of positive category data correctly grouped by the system.

$$Recall = TPFN + TP\ X\ 100\% \quad (9)$$

RESULT

Model Accuracy

In measuring the accuracy of the classification, it can be measured by using the accuracy value when testing the data using all the models carried out by researchers in this study. The results of the comparison of the accuracy of each model can be seen in Table 2.

Table 2. Model Performance Accuracy Results

Model	Testing Accuracy (%)
Random Forest	53
Support Vector Machine	55
Multinomial Naïve Bayes	52
Decision Tree	44
K-Nearest Neighbor	42

The results of performance testing of classification accuracy are still very low. This is due to the condition of the class after labeling and preprocessing encounter unbalanced class results or imbalance (Fandi & Sephia Dwi Arma Putri, 2023). Where there are more Neutral classes than Negative and Positive, it is necessary to conduct experiments on the dataset. The experiment that will be carried out is by technique Oversampling and *Undersampling*. Where the goal is to balance class data by multiplying classes from the minority to the majority class (*Oversampling*) (Ciptady et al., 2022), or reducing the majority class to the minority class (*Undersampling*) (Informatics et al., 2023). The application of the classification method to sentiment analysis is carried out by dividing the dataset where 80% is declared as train data and 20% is stated as testing data. The following are the results of the accuracy comparison after the *Undersampling* and *Oversampling* techniques are applied which can be seen in Table 3.

Table 3. Performance Evaluation

Model	Test Accuracy Normal Data (%)	Test Accuracy Undersampling (%)	Test Accuracy Oversampng (%)
Random Forest	53	40	85
Support Vector Machine	55	45	77
Multinomial Naïve Bayes	52	49	76
Decision Tree	44	43	82
K-Nearest Neighbor	42	30	90

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12907 p-ISSN : 2541-044X

The results of the comparison of technical experiments to this dataset can be seen by technique oversampling resulting in the most significantly high classification accuracy performance at the time testing data. Therefore, the oversampling technique was chosen to be applied in this case in order to be able to support the research process and produce better performance.

Classification Report

The previous results show the overall performance of the classification model across all experiments to the applied dataset. This subsection shows the detailed performance of each label by presenting a classification report. The rating report consists of several matrices: precision, recall, and F1-Score. Tables 4, 5, and 6 show the performance of negative, neutral, and positive signs, respectively.

Table 4. Classification Report for Positive Label

Model	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	81	97	88
Support Vector Machine	78	81	79
Multinomial Naïve Bayes	81	84	83
Decision Tree	74	94	83
K-Nearest Neighbor	86	99	93

Table 5. Classification Report for Negative Label

Model	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	100	62	76
Support Vector Machine	74	68	71
Multinomial Naïve Bayes	77	71	74
Decision Tree	99	56	72
K-Nearest Neighbor	99	74	85

Table 6. Classification Report for Neutral Label

Model	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	80	97	88
Support Vector Machine	77	82	79
Multinomial Naïve Bayes	69	73	71
Decision Tree	80	87	88
K-Nearest Neighbor	86	97	91

DISCUSSIONS

In the research that has been done, the application of experimental techniques to the dataset has been proven to have a very significant effect. The results show that the oversampling technique in this research case can produce very good accuracy performance measurements from classification. This is because by applying *oversampling*, data information on the minority class is improved so that it can produce more information to examine and produce better decisions as well. Whereas *undersampling* removes many majority samples to approach minority samples to achieve class balance. This can result in the loss of important information from the data set. This loss of information can affect the model's ability to learn meaningful patterns from the data. Thus, the *oversampling* technique in this case is the most effective for use on the dataset and then the classification process is applied to the 5 classification method models, all of which can improve classification accuracy properly. In this study, the condition of the dataset after the labeling process is important and needs more attention. Because this proves that if the data class conditions are unbalanced it will greatly affect the performance of the model for classifying. With unbalanced data conditions, the results of the model performance decrease. Then after experimenting with the dataset to deal with unbalanced class conditions, this yielded positive results, where all classification models were able to improve properly.

*name of corresponding author

CONCLUSION

Sentiment analysis research on the satisfaction level of PDDikti operators on the increase in the PDDikti Neo feeder application was successfully carried out. The method used is the Decision tree algorithm. K-Nearest Neighbor, Multinomial Naïve Bayes, Random Forest, and Support Vector Machine can classify operator sentiment data well. The accuracy of classification success during the best data testing process is the K-Nearest Neighbor algorithm which reaches 90%. Then the lowest accuracy during the data testing process is the Multinomial Naïve Bayes algorithm with an accuracy of 76%. This proves that the performance of the K-NN algorithm is better than the other 4 algorithms in the case of this study.

REFERENCES

Asro'i, A., & Februariyanti, H. (2022). Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan Ppkm Menggunakan Metode K-Nearest Neighbor. *Jurnal Khatulistiwa Informatika*, *10*(1), 17–24. https://doi.org/10.31294/jki.v10i1.12624

Ciptady, K., Harahap, M., Jonvin, J., Ndruru, Y., & Ibadurrahman, I. (2022). Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science. *Data Sciences Indonesia (DSI)*, *2*(1). https://doi.org/10.47709/dsi.v2i1.1708

Desiani, A., Akbar, M., Irmeilyana, & Amran, A. (2022). Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular. *Jurnal Teknik Elektro Dan Komputasi (ELKOM)*, *4*(2), 207–214.

Fandi, F. Y. P., & Sephia Dwi Arma Putri. (2023). Komparasi Metode Smote Dan Adasyn Untuk Penanganan Data Tidak Seimbang Multiclass. *Jurnal Informatika Polinema*, *9*(3), 331–338. https://doi.org/10.33795/jip.v9i3.1330

Hendriyanto, M. D., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine. *INTECOMS: Journal of Information Technology and Computer Science*, *5*(1), 1–7. https://doi.org/10.31539/intecoms.v5i1.3708

Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, *10*(1), 18–26. https://doi.org/10.34148/teknika.v10i1.311

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121. https://doi.org/10.22146/ijccs.65176

Kewsuwun, N., & Kajornkasirat, S. (2022). A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier. *International Journal of Electrical and Computer Engineering*, *12*(3), 2829–2838. https://doi.org/10.11591/ijece.v12i3.pp2829-2838

Khotimah, A. C., Utami, E., Universitas, I., Yogyakarta, A., Teknik, M., Universitas, I., & Yogyakarta, A. (2022). *COMPARISON NAÏVE BAYES CLASSIFIER , K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN THE CLASSIFICATION OF INDIVIDUAL ON PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER , K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI*. *3*(3), 673–680.

Laurensz, B., & Eko Sediyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *10*(2), 118–123. https://doi.org/10.22146/jnteti.v10i2.1421

Mardiana, L., Kusnandar, D., & Satyahadewi, N. (2022). Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak. *Bimaster*, *11*(1), 97–102.

Morama, H. C., Ratnawati, D. E., & Arwani, I. (2022). *Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier*. *6*(4), 1702–1708.

Mukarramah, R., Atmajaya, D., & Ilmawan, L. B. (2021). Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter. *ILKOM Jurnal Ilmiah*, *13*(2), 168–174. https://doi.org/10.33096/ilkom.v13i2.851.168-174

*name of corresponding author

Nurmawiya, & Harvian, K. A. (2021). Public sentiment towards face-to-face activities during the COVID-19 pandemic in Indonesia. *Procedia Computer Science*, *197*(2021), 529–537. https://doi.org/10.1016/j.procs.2021.12.170

Oktafani, M., & Prasetyaningrum, P. T. (2022). Implementasi Support Vector Machine Untuk Analisis Sentimen Komentar Aplikasi Tanda Tangan Digital. *Jurnal Sistem Informasi Dan Bisnis Cerdas*, *15*(1), 10–19. https://doi.org/10.33005/sibc.v15i1.2697

Prasetyo, K. G., & Pahlevi, S. M. (2019). Analisis Perbandingan Algoritma Decision Treedengan Support Vector Machineuntuk Mendeteksi Kompetensi Mahasiswakonsentrasi Informatika Komputerstudi Kasus : Politeknik Lp3I Jakarta, Kampus Depok. *Jurna Lentera ICT*, *5*(2), 11–26.

Ramon, E., Nazir, A., Novriyanto, N., Yusra, Y., & Oktavia, L. (2022). Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, *5*(2), 143–150. https://doi.org/10.47080/simika.v5i2.2185

Sastypratiwi, H., Muhardi, H., & Noveanto, M. (2022). *Klasifikasi Emosi Pada Lirik Lagu Menggunakan Algoritma Multiclass SVM dengan Tuning Hyperparameter PSO*. *6*, 2279–2286. https://doi.org/10.30865/mib.v6i4.4609

Shah, P., Swaminarayan, P., & Patel, M. (2022). Sentiment analysis on film review in Gujarati language using machine learning. *International Journal of Electrical and Computer Engineering*, *12*(1), 1030–1039. https://doi.org/10.11591/ijece.v12i1.pp1030-1039

Sholeha, E. W., Yunita, S., Hammad, R., Hardita, V. C., & Kaharuddin, K. (2022). Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *3*(4), 203–208. https://doi.org/10.35746/jtim.v3i4.178

Simanjuntak, T. H., Mahmudy, W. F., & Sutrisno Sutrisno. (2017). Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*, *No.2*(2), 75–79. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15/21

Siringoringo, R. (2018). Klasifikasi data tidak Seimbang menggunakan algoritma SMOTE dan k-nearest neighbor. *Jurnal ISD*, *3*(1), 44–49.

Sulastomo, H., Gibran, K., Maryansyah, E., & Tegar, A. (2022). Analisis Sentimen Pada Twitter @Ovo_Id dengan Metode Support Vectore Machine (SVM). *Jurnal Sains Komputer & Informatika (J-SAKTI*, *6*(2), 1050–1056.

Turjaman, R. M., & Budi, I. (2022). Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter). *Jurnal Darma Agung*, *30*(2), 266. https://doi.org/10.46930/ojsuda.v30i2.1672

Verawati, I., & Audit, B. S. (2022). Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u. *Jurnal Media Informatika Budidarma*, *6*(3), 1411. https://doi.org/10.30865/mib.v6i3.4132

Wijaya, A. (2022). Analisis Sentimen Kepuasan Pelanggan pada Penyedia Layanan Pengantaran Makanan. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(3), 1699–1708. https://doi.org/10.35957/jatisi.v9i3.2196

*name of corresponding author