JOURNAL OF TECHNOLOGY AND OPEN SOURCE
Vol. 8, no. 2, December 2024, pp. 277~288
e-ISSN: 2622- 1659, accredited Four Grade by Kemenristekdikti , Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v7i2.4972

Sentiment Analysis of Marketplace Application Reviews Using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN)

Arief Ichwani [1] , Munawar [2] , Rilla Gantino [3]
[1.2] Informatics Engineering , Esa University Superior
[3] Master of Accounting , Esa University Superior

Article Info

Article history:

Received 10 13, 2024
Revised 12 02, 2024
Accepted 12 25, 2024

Keywords:

Shopee
Sentiment Analysis
Machine Learning
Support Vector Machine (SVM)
K-Nearest Neighbors

ABSTRACT

Shopee is one of the most popular online marketplaces in Indonesia, with more than 103 million users in 2023. Most users consider factors such as customer reviews, ratings, prices, and free shipping promotions before making a purchase. Analyzing user reviews is essential to understand consumer perceptions of services, identify satisfaction or dissatisfaction, and detect potential issues that need to be addressed. However, sentiment analysis faces challenges in processing text with diverse language styles, structures, and informal expressions. To overcome these challenges, this study applies machine learning algorithms—Support Vector Machine (SVM) and K-Nearest Neighbors (KNN)—for classifying sentiment in Shopee user reviews. Data labeling using the Lexicon InSet method produced 9,509 positive reviews (47.55%), 7,686 negative reviews (38.43%), and 2,805 neutral reviews (14.03%). Based on the Confusion Matrix results, SVM outperformed KNN, particularly in classifying negative and neutral sentiments with higher accuracy. These findings indicate that SVM is a more effective and efficient model for sentiment analysis of user reviews on the Shopee platform[1].

Corresponding Author:

Arief Ichwani
Faculty of Computer Science
Esa University Superior
West Jakarta, Indonesia
Email: arief.ichwani@esaunggul.ac.id

1. Introduction

With development rapid growth of e-commerce, especially in the Southeast Asia region , reviews and responses users be one of very important data source for company in understand quality service as well as satisfaction users to application . Reviews users can reflect experience shop in a way directly , giving outlook about excess and weakness from features offered by marketplace platforms, such as Shopee.

Shopee is one of the the most popular marketplace application in Indonesia. In 2023 , the number of Shopee users in Indonesia reached 103 million , making it the platform with users the most in the country [1] . With height amount users , the majority buyers on Shopee tend to consider a number of factor important before do purchases , such as evaluation customers (online customer rating), reviews customers (online customer reviews), prices , and free shipping offers send . Factors This become gauge measuring main in choose shop , okay shop normal and star sellers. Although online stores on Shopee have registered and trusted , consumers still do evaluation Alone based on various consideration For ensure quality and trust moment shopping .

Analysis to review users can help company identify contributing factors to success service or reflect potential problems that need to be addressed quick overcome [3]This data can utilized For increase performance application , repair services , and adjust marketing strategies . However , for manage and extract information valuable from a very large and varied number of reviews is required method precise and efficient analysis [4][5].

One of common methods used is analysis sentiment , which aims For identify emotion or opinion behind every review user . Analysis This give useful mapping For understand perception users to product or service , whether they feel satisfaction , dissatisfaction , or give neutral response [6]. Classification sentiment This divided in three category main , namely positive , negative , and neutral . Reviews positive usually reflect satisfaction users to applications , often related with availability products , convenience transactions , or speed shipping . Reviews negative , on the other hand , indicates existence problems faced users , such as disturbance technical , delay delivery , or mismatch product [5][6]. Meanwhile that , review neutral tend contain view objective that is not convey emotion strong , but still relevant in give input about experience users in a way general [3] [2] .

However , the challenge main in analysis sentiment is How processing highly varied text data in language , structure , and style writing . User often use informal language , slang, or abbreviations , which make it difficult interpretation automatically . Therefore that , is necessary powerful Natural Language Processing (NLP) techniques as well as machine learning algorithm capable handle variation the with good . Use NLP techniques such as Word2Vec can help in increase accuracy processing Language experience [3] . The importance of selection feature in dealing with inconsistent data balanced , where the analysis more sentiment appropriate can achieved with combination technique election appropriate features and powerful algorithms [6] .

Two common algorithms used in analysis sentiment are Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). SVM is algorithm learning purpose machine For find the best separating hyperplane existing classes in data, such as sentiment positive and negative . SVM is very effective For handle data with dimensions high , such as review text , which can contain thousands feature in word forms [3], [4] . Algorithm This own ability For overcome overfitting problem , especially on complex datasets [2] .

Temporary that is , KNN is algorithm based neighbor the closest one determines class from a new data based on class majority from neighbors closest . Although simple , KNN can give competitive results , especially when the data is analyzed Enough size and distribution No too varies [8] [5] . Algorithm This Work with compare word patterns in review new with reviews that have been known sentiment , and classify review new the based on similarities . In context analysis review users on Shopee, SVM and KNN often used For determine sentiment and evaluate effectiveness of improvement strategies services on the application shopee .

Study This aim For compare accuracy and performance second algorithm the in do analysis sentiment on reviews Shopee marketplace application . With do comparison this research can give outlook about which method is better effective in categorize review users , so that can give more recommendations appropriate for developer application in repair experience shop user . Study this is also expected can contribute to the improvement quality service through utilization of review data in a way more optimal, which in the end will support growth e-commerce businesses like Shopee [9] [9] .

2. Research Method

Method data collection in research This there with secondary data collection from Google Play Store app shopee . As for steps to take do analysis sentiment using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) . as following :

1) Data collection
   Step :
   a) Shopee review data can obtained through scraping from the Shopee site or using available datasets in a way public .
   b) Data includes text reviews and sentiment labels (positive , negative , neutral).
2) Data *Preprocessing*
   Step:
   a) Cleanup : Delete character special , numbers , and signs read what is not relevant .
   b) Case Normalization: Change text review become letter small (lowercase) to make it more consistent .
   c) Stop Word Removal: Remove unnecessary words own meaning significant such as "and", " or ", " to ".

d) Stemming/Lemmatization: Change words into form base For consistency word analysis (for example , " main " becomes "main").

Feature Extraction

Step:

a) Change text review become representation numerical that can used For machine learning modeling .
b) Method commonly used :
c) Bag of Words : Build representation based word frequency .
d) TF-IDF (Term Frequency-Inverse Document Frequency): Calculating weight the importance of words based on frequency its emergence .

3) Split (Train-Test Split)

Step:

a) Divide the dataset into two parts : training data (training set) and test data (testing set) so that it can evaluate model performance with Good .
b) Typically , training data taken as much as 80%, while the test data is as much as 20%.

4) Model Development (Training)

Steps (SVM):

a) Apply Support Vector Machine (SVM) to train the model using training data .
b) Select the appropriate kernel (linear, RBF, polynomial), adjust with data and goals analysis .

Steps (KNN):

a) Apply K-Nearest Neighbors (KNN) algorithm for train the model on training data .
b) Determine optimal k value (number neighbor nearest) to more classification accurate .

Tools:

a) For SVM, use ` svm.SVC ( )` from Scikit-learn.
b) For KNN, use ` KNeighborsClassifier ( )` from Scikit-learn.

5) Model Testing

Step:

a) After the model is trained , do testing on test data for measure how much good model can predict sentiment that has not been Once seen previously .
b) Compare model prediction with original labels on test data.

Tools:

a) Use the ` predict( )` method of the trained model For predict results sentiment on test data.

6) Model Evaluation

Step:

b) Measuring model performance using metric evaluation like :
c) Confusion Matrix: For see classification right and wrong of the model.
d) Accuracy: Percentage correct prediction .
e) Precision, Recall, and F1-Score: Metrics important additions in evaluation classification .

7) Model Optimization

Step:

1) Do model optimization for increase accuracy and performance .
2) A number of method optimization includes :
   a) Grid Search or Random Search for find the best parameters on SVM and KNN.
   b) Cross-validation for ensure model stability and avoid overfitting.

Tools:

a) Use ` GridSearchCV ` from Scikit-learn to best hyperparameter search .

8) Visualization of Results

Step:

b) Visualize results analysis sentiment For more easy understood .
c) Use visualization like WordCloud For display frequently used words appear in reviews .
d) Use Confusion Matrix and ROC Curve to illustrate model performance .

Tools:

a) Use Matplotlib, Seaborn, or WordCloud For visualization .

9) Conclusion and Documentation of Results

Step:

b) For conclusion from results analysis sentiments that are carried out .
c) Document it the whole process of data collection up to model evaluation .

d) Tell me which model is better effective in analysis sentiment between SVM and KNN and recommendation For future improvements .

3. Result and Discussion

3.1 Data Collection

This study utilizes review data collected from the Shopee application available on the Google Play Store. The retrieved data were then converted into Microsoft Excel format for further processing and analysis. From the data collection process, a total of 20,000 user reviews were initially obtained. After performing data cleaning and duplicate removal, the final dataset consisted of 15,209 unique records that were ready for analysis. The review data were collected using the web scraping method with the assistance of the “google-play-scraper” library, which automatically extracts user reviews from the Google Play Store based on the latest review order (sort = Sort.NEWEST)[11].

This approach ensured that the dataset reflected the most recent user opinions and experiences, allowing the sentiment analysis to capture current trends in user satisfaction and feedback dynamics. In addition, the automated scraping process minimized manual collection errors and improved the efficiency and accuracy of data acquisition.

3.2 Preprocessing

Stages This is stages crucial aiming For ensuring clean , consistent , and ready data For analysis more deep . Collected data need processed through a number of stages *data preprocessing* that is *cleansing* , *case folding* , *tokenizing* , *normalizing* , *stopwords* , and *stemming* [12].

a) Cleansing, case folding , and normalization

Cleansing stages is For remove elements that are not relevant so that text become more clean and easy analyzed . The results from stages *cleaning* .

Table 1. Results *of Cleaning* , Case Folding, and Data Normalization

Review Text	cleaning	case_folding	normalisasi
Now you don't have to worry about shopping online at Shopee anymore. It's always safe and secure. Because there is a money back guarantee if the item is not as expected. There is a free return policy. Thank you Shopee ðŸ™□ðŸ™□	Now you don't have to worry about shopping online at Shopee anymore, it's always safe and controlled because there is a money back guarantee if the item is not as expected, there is free return of goods, thank you Shopee	Now you don't have to worry about shopping online at Shopee anymore, it's always safe and controlled because there is a money back guarantee if the item is not as expected, there is free return of goods, thank you Shopee	Now you don't have to worry about shopping online at Shopee anymore, it's always safe and controlled because there is a money back guarantee if the item is not suitable, there is free return of goods, thank you Shopee
I've only tried using Shopee once, but it's tiring. Orders don't match estimates, and shipping takes a long time. I'm not comparing it to the blue shop next door. Even though they're overloaded, their delivery service is always on time.	I've only tried using Shopee once, but it turns out I'm tired. The order doesn't match the estimated delivery time. I don't want to compare it with the blue shop next door. Even though it's a load oper, the delivery service is always on time.	I've only tried using Shopee once, but it turns out that the order doesn't match the estimated delivery time. I don't want to compare it with the blue shop next door, even though the delivery service is always on time.	I've only tried using Shopee once, but it turns out that the order doesn't match the estimated delivery time, so I don't want to compare it with the blue shop next door, even though they use the shipping service, it's always on time.

really makes all purchasing and payment processes easier ðŸ'□	greatly simplifies all purchasing and payment processes	greatly simplifies all purchasing and payment processes	greatly simplifies all purchasing and payment processes
Shopee is getting slower..	Shopee is getting slower	Shopee is getting slower	Shopee is getting slower
I am very satisfied shopping at Shopee, thank you ðŸ™	I am very satisfied shopping at Shopee, thank you	I am very satisfied shopping at Shopee, thank you	I am very satisfied shopping at Shopee, thank you.
Can you pay on the spot or pay when the goods arrive	You can pay on the spot instead of paying when the goods arrive	You can pay on the spot, you don't have to pay when the goods arrive	can pay on the spot or not pay when the goods arrive

b) *Stokenizing* , *stopword removal* ,

Table 2. Stopword Results

tokenize	stopword removal
['service', 'courier', 'very', 'bad', 'to', 'seller']	['service', 'courier', 'bad', 'seller']
['new', 'try', 'once', 'use', 'shope', 'apparently', 'rewarded', 'order', 'no', 'according to', 'long estimate', 'delivery', 'yes', 'not', 'want', 'compare', 'same', 'shop', 'blue', 'next door', 'even though', 'operate', 'load', 'service', 'send', 'yes', 'always', 'right', 'time']	['try', 'use', 'shopee', 'kapuk', 'order', 'according to', 'estimated time', 'delivery', 'yes', 'compare', 'shop', 'blue', 'next door', 'oper', 'load', 'service', 'send', 'yes']
['very', 'make it easier', 'all', 'process', 'purchase', 'and', 'payment']	['make it easier', 'process', 'purchase', 'payment']
['shopee', 'more', 'slow']	['shopee', 'slow']
['I', 'very', 'satisfied', 'shopping', 'at', 'shopee', 'thank', 'thank you']	['satisfied', 'shopping', 'shopee', 'receive', 'thank you']
['can', 'pay', 'at', 'place', 'no', 'pay', 'yes', 'pass', 'goods', 'already', 'arrived']	['pay', 'pay', 'yes', 'pass', 'goods']

c) and *stemming* .

In stages This done *s temming* namely the process of changing words in A text become form basically (*root word*). This process aim For simplify words that have form derivatives[13] . The results are is as following :

Table 3. Stemming Results

stopword removal	steming_data
['service', 'courier', 'bad', 'seller']	bad courier service sell
['try', 'use', 'shopee', 'kapuk', 'order', 'according to', 'estimated time', 'delivery', 'yes', 'compare', 'shop', 'blue', 'next door', 'oper', 'load', 'service', 'send', 'yes']	Try using Shopee Kapuk, order according to the estimated delivery time compared to the Blue Shop, which operates the delivery service.
['make it easier', 'process', 'purchase', 'payment']	easy buy pay process
['shopee', 'slow']	Shopee Lot
['satisfied', 'shopping', 'shopee', 'receive', 'thank you']	satisfied shopping at Shopee, thank you
['pay', 'pay', 'yes', 'pass', 'goods']	pay when you receive the goods
['sis', 'please', 'men', 'non', 'activate', 'sea', 'bank', 'wife', 'buy', 'pridok', 'assessment with', 'sticker', 'shopee', 'receive', 'ordered', 'activate', 'sea', 'bank',	Sis, please deactivate Sea Bank, wife buys Pridok, assessment with Shopee sticker, accepts, told to activate Sea Bank, told to activate ES loan,

'ordered', 'active', 'espinjam', 'profit', 'husband', 'go home', 'say', 'fraud', 'code', 'otp', 'sent', 'fraudster']	fortunately husband comes home, says fraud, OTP code, sends fraud
['service', 'shipping', 'cheap']	cheap shipping service
['application', 'good', 'sometimes', 'like', 'free', 'shipping']	good app sometimes likes free shipping
['shopping', 'online', 'shopee', 'scared', 'safe', 'controlled', 'guarantee', 'money', 'goods', 'according to conditions', 'free', 'returns', 'goods', 'thank you', 'shopee']	Shopee online shopping is safe, guaranteed money, goods are as is, free to return goods, thank you Shopee

d) Data Labeling

Data labeling is a crucial process in data analysis . At this stage this , data labeling will be use *Lexicon InSet* For grouping sentiment- related data in *the marketplace* Shopee to in three category that is positive , negative and neutral . Furthermore , the results data labeling by *Lexicon InSet* will validated by experts Language For ensure its accuracy[14] .

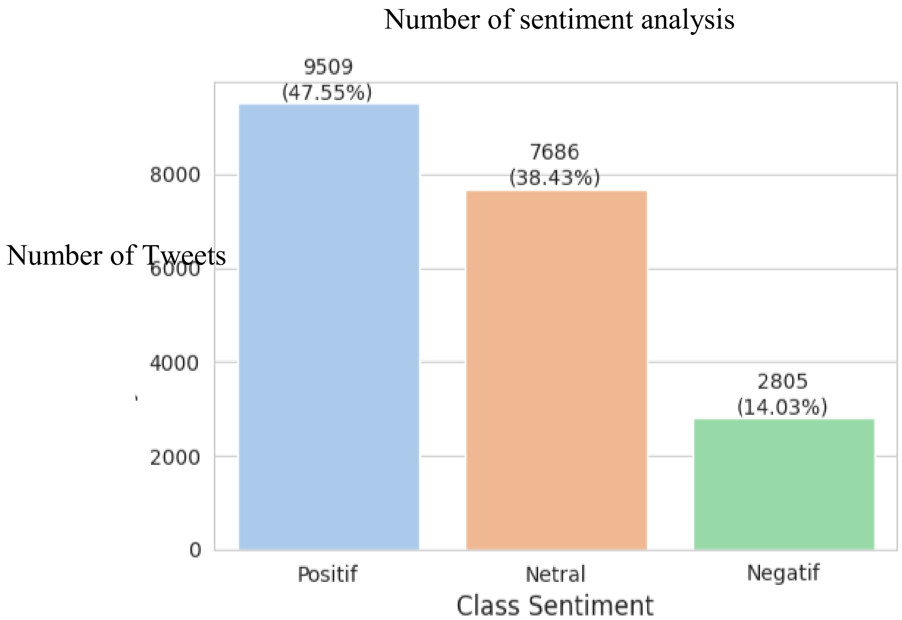

Figure 1. Number Analysis Sentiment from Shopee

Figure 1 shows distribution sentiment based on results labeling use *Lexicon* in diagram form . The majority *Tweet* has sentiment positive as big as 47.55 %, which shows that review related Shopee on Google Play Store dominated by positive sentiment . Meanwhile that , sentiment negative (38 , 43 %) and sentiment neutral (14.03 %) reflects No show clear feeling[15].

e) Training Data and Data Testing

After the labeling process finished , the data will be shared become two part namely training data and test data, before enter stage training using SVM, KNN, and Naïve Bayes models . Training data functioning as data set used For training *machine learning* models , enabling algorithm learn patterns , features , and relationships between variable . While that , test data is used For measure model performance after the training process finished , without follow as well as in stage learning . In study here , the data is shared with proportion of 80:20, as explained following This :

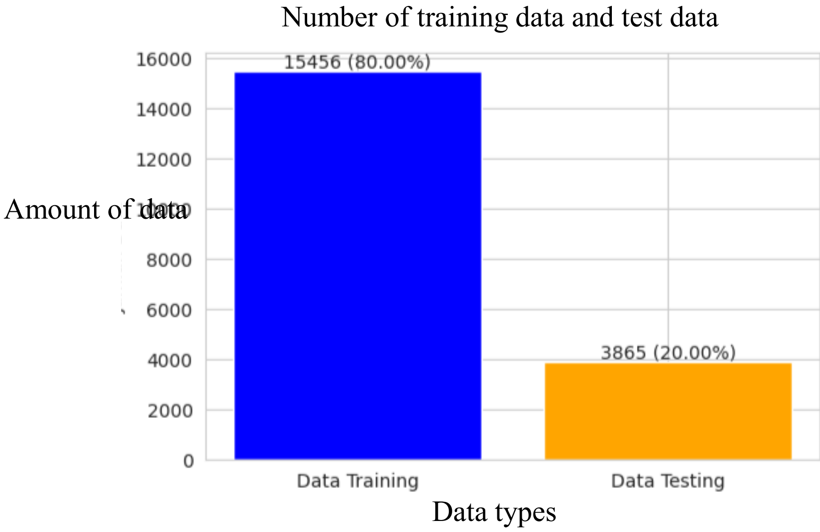

Figure 2. Number of Training Data and Test Data

In Figure 2 above , the *review data* is the training data . is 15,456 reviews and the *review data* that becomes the test data is 3,865 *reviews* .

f) Model Evaluation

At the stage previous data testing , the code used also includes visualization *Confusion Matrix* for show distribution error predictions . Here This is results SVM model evaluation in do classification , which is displayed through *Confusion Matrix* [16].

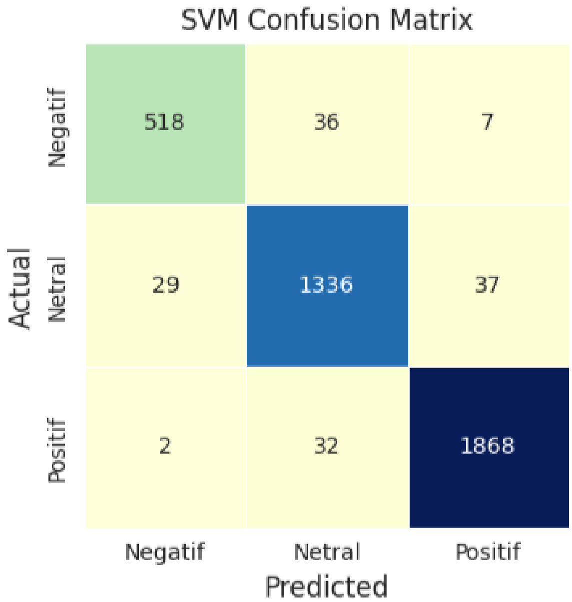

Figure 3. SVM Confusion Matrix

Confusion Matrix above show performance of the SVM model in classify sentiment become three category , namely positive , negative , and neutral . Of the total predictions , the model was successful classify 518 samples with Correct as negative , but there are 36 samples falsely predicted negative as neutral and 7 samples falsely predicted negative as positive[17] . For category neutral , capable model predict with Correct as many as 1,336 samples , however there are 29 samples wrongly predicted neutral as negative and 37 samples wrongly predicted neutral as positive . Meanwhile that , in the category positive , the model shows performance the best [18].

Total 20,000 samples the predicted with true , but Still there are 29 samples false positive prediction as negative and 37 samples false positive prediction as neutral [19]. Visually , the color in This *Confusion Matrix* represent amount sample in every category predictions . The more dark color blue , more and more tall amount classified samples in cell said . Cell with color darkest blue show the most dominant prediction , namely in the category predicted positive with true (1,868 samples). On the other hand , the more bright show amount more samples a little , as in a mistake classification category negative to positive (7 samples). Colors This help in identify pattern error classification , where the error the biggest appears in the classification category neutral to positive and vice versa , which shows possibility existence similarity pattern between second category This Confusion Matrix give clear picture about the areas where the model can improved For reduce error classification , especially in differentiate sentiment neutral and positive[20] .

From the picture The *Confusion Matrix* We can count accuracy and metrics other with use values from The *Confusion Matrix* .

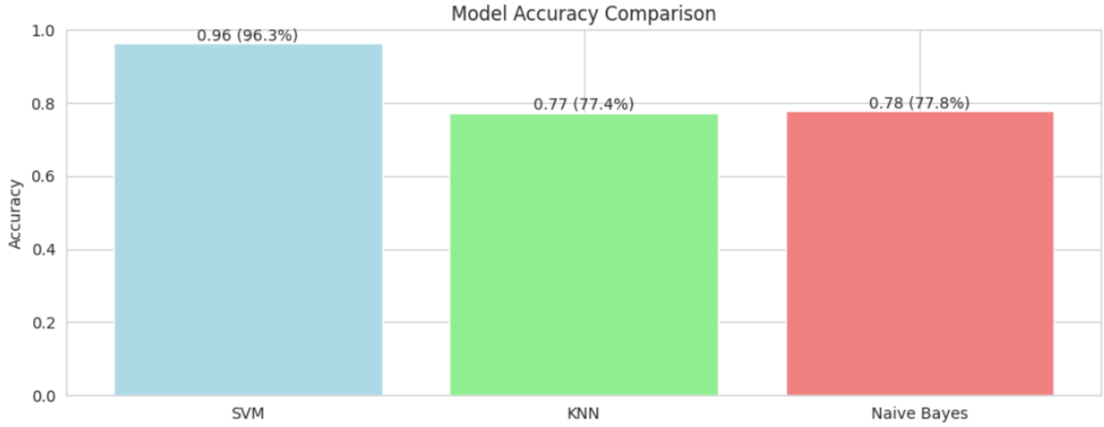

Figure 4. Comparison of Accuracy models

Based on results comparison from Figure 6.16 shows that the SVM model has greater accuracy Good with value of 96.3% compared previously .

	precision	recall	f1-score	support
Negatif	0.944	0.923	0.933	561.000
Netral	0.952	0.953	0.952	1402.000
Positif	0.977	0.982	0.980	1902.000
accuracy	0.963	0.963	0.963	0.963
macro avg	0.957	0.953	0.955	3865.000
weighted avg	0.963	0.963	0.963	3865.000

Figure 5. Classification *Report* for SVM

Report the above classification serve metric evaluation for classification models sentiment SVM model, which was applied to three category sentiment that is negative , neutral and positive [21]. This model get accuracy overall by 96% of all test data. Positive class show performance best with *F1 - Score* of 0.98, which indicates optimal balance between *Precision* (98%) and *Recall* (98%), so the model is capable of recognize sentiment positive with level high accuracy as well as only A little experience error classification . However , in class negative and neutral , *F1 - Score* is 0.93 and 0.95 respectively , which indicates existence challenge in differentiate second category This with precise . More *precision* low in class neutral (0.95%) and More *recall* low in class negative (0.92%) indicates that the model is still often make error in classify sentiment neutral and not yet fully capable recognize all negative data with true . In overall , results evaluation This show that the model has good performance in identify sentiment positive , but Still need repair in recognize sentiment negative and neutral , especially in increase accuracy classification into categories neutral which is still often miscategorized as sentiment negative or positive[22] .

Besides results from *Confusion Matrix* of the SVM model, in study this is also available results visualization *Confusion Matrix* from use of the KNN model. The following is results from The *Confusion Matrix* .

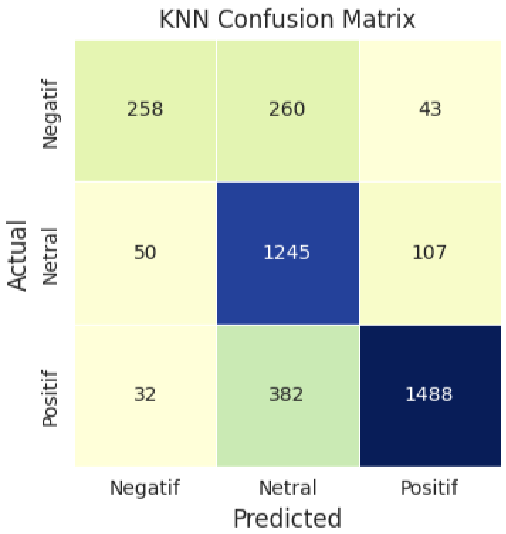

Figure 6. KNN Confusion Matrix

The Confusion Matrix in Figure 6.17 shows performance of the KNN model in classify sentiment become three category , namely positive , negative , and neutral . Of the total predictions , the model was successful classify 258 samples with Correct as negative , but there are 260 samples falsely predicted negative as neutral and 43 samples falsely predicted negative as positive . For category neutral , capable model predict with Correct as many as 1245 samples , but there are 50 samples wrongly predicted neutral as negative and 10 samples wrongly predicted neutral as positive [23]. Meanwhile that , in the category positive , the model shows performance at its best , with 1488 samples predicted with true , but Still there are 32 samples false positive prediction as negative and 382 samples false positive prediction as neutral . Visually , the color in This *Confusion Matrix* represent amount sample in every category predictions[24] . The more dark color blue , more and more tall amount classified samples in cell said . Cell with color darkest blue show the most dominant prediction , namely in the category predicted positive with true (1,488 samples). On the other hand , the more colors bright show amount more samples a little , as in a mistake classification category negative to positive (43 samples). Colors This help in identify pattern error classification , where the error the biggest appears in the classification category neutral to positive and vice versa , which shows possibility existence similarity pattern between second category This Confusion Matrix give clear picture about the areas where the model can improved For reduce error classification , especially in differentiate sentiment neutral and positive [25].

Classification Report for KNN:

	precision	recall	f1-score	support
Negatif	0.759	0.460	0.573	561.000
Netral	0.660	0.888	0.757	1402.000
Positif	0.908	0.782	0.841	1902.000
accuracy	0.774	0.774	0.774	0.774
macro avg	0.776	0.710	0.723	3865.000
weighted avg	0.797	0.774	0.771	3865.000

Figure 7. Classification *Report* for KNN

Then Report The classification in Figure 7 presents metric evaluation for classification models KNN model sentiment , which is applied to three category sentiment that is negative , neutral and positive . This model get accuracy overall by 77% of all test data. Positive class show performance best with *F1 - Score* of 0.84, which indicates optimal balance between *Precision* (90%) and *Recall* (78%), so the model is capable of recognize sentiment positive with level sufficient accuracy tall as well as only A little experience error classification . However , in class negative and neutral , *F1 - Score* is 0.57 and 0.75 respectively , which indicates existence challenge in differentiate second category This with precise . More *precision* low in class neutral (66%) and More *recall* low in class negative (75%) indicates that the model is still often make error in classify sentiment neutral and not yet fully capable recognize all negative data with true[26] . In overall , results evaluation This show that the model has poor performance Good in identify sentiment positive , and also requires repair in recognize sentiment negative and neutral , especially in increase accuracy classification into categories neutral which is still often miscategorized as sentiment negative or positive .

4. Conclusion

Based on Based on results data processing and results analysis , then conclusion from study This is

a) Data labeling is done use method *Lexicon InSet* show that 9,509 reviews (47.55 %) sentiment positive , 7,686 reviews (38.43 %) negative , and 2,805 reviews (1 4.03 %) neutral .

b) Based on results *Confusion Matrix* , SVM more superior in catch pattern sentiment compared to KNN , especially in classify sentiment negative and neutral with more accurate . This is show that the SVM model is a better algorithm accurate and efficient in categorize sentiment review users on Shopee.

Acknowledgement

Writer say accept love to Universitas Esa Unggul Superior on support funding through the Grant Program Esa University Internal Research Superior Year 2024. Greetings accept love is also conveyed to Fakultas Ilmu Komputer on help facilities and guidance technical during the collection process as well as data analysis . Awards special given to colleagues researchers and students who participated help in the research process This until finished .

References

[1] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J. Big Data*, 2024, doi: 10.1186/s40537-024-00973-y.

[2] A. Muslikhun, "Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian Online di

Marketplace Shopee Alfin Muslikhun semakin meningkat . Hasil survey dari (Similarweb , 2023) menyatakan di tahun 2023 , pembelian online . Selain itu juga penelitian dari (Oktaviana,” no. 4, 2024.
[3] M. B. Sitepu, I. R. Munthe, and S. Z. Harahap, “Implementation of Support Vector Machine Algorithm for Shopee Customer Sentiment Anlysis,” *Sinkron*, vol. 7, no. 2, pp. 619–627, 2022, doi: 10.33395/sinkron.v7i2.11408.
[4] N. Rezki, S. A. Thamrin, and S. Siswanto, “Sentiment Analysis of Merdeka Belajar Kampus Merdeka Policy Using Support Vector Machine With Word2Vec,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 1, pp. 0481–0486, 2023, doi: 10.30598/barekengvol17iss1pp0481-0486.
[5] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, “Corrigendum: Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 2, p. 169, 2018, doi: 10.20473/jisebi.4.2.169.
[6] A. W. Sari, T. I. Hermanto, and M. Defriani, “Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine,” *Sinkron*, vol. 8, no. 3, pp. 1366–1378, 2023, doi: 10.33395/sinkron.v8i3.12447.
[7] S. Kübler, C. Liu, and Z. A. Sayyed, *To use or not to use: Feature selection for sentiment analysis of highly imbalanced data*, vol. 24, no. 1. 2018. doi: 10.1017/S1351324917000298.
[8] Junadhi, Agustin, M. Rifqi, and M. K. Anam, “Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection,” *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 216–225, 2022, doi: 10.23887/janapati.v11i3.51531.
[9] Shahmirul Hafizullah Imanuddin, Kusworo Adi, and Rahmat Gernowo, “Sentiment Analysis on Satusehat Application Using Support Vector Machine Method,” *J. Electron. Electromed. Eng. Med. Informatics*, vol. 5, no. 3, pp. 143–149, 2023, doi: 10.35882/jeemi.v5i3.304.
[10] Alqis Rausanfita, A. Ramadan, M. Dzulfikar Fauzi, Q. E. P. Mafidah, E. Ramona, and Y. Mahardika Putra, “Sentiment Analysis of Comments on Instagram Posts of Indonesia’s 2024 Presidential Candidates Using The Support Vector Machine Method,” *Tech-E*, vol. 7, no. 1, pp. 21–31, 2023, doi: 10.31253/te.v7i1.2289.
[11] P. Cunningham and S. J. Delany, “K-nearest neighbour classifiers-a tutorial,” *ACM Comput. Surv.*, 2021, doi: 10.1145/3459665.
[12] M. Suyal and P. Goyal, “A review on analysis of k-nearest neighbor classification machine learning algorithms based on supervised learning,” ... *J. Eng. Trends ...*, 2022, [Online]. Available: https://www.researchgate.net/profile/Parul-Goyal-5/publication/362383623_A_Review_on_Analysis_of_K-Nearest_Neighbor_Classification_Machine_Learning_Algorithms_based_on_Supervised_Learning/links/65533e1cb1398a779d88c6a4/A-Review-on-Analysis-of-K-Nearest-Neighbor-Classification-Machine-Learning-Algorithms-based-on-Supervised-Learning.pdf
[13] A. R. Lubis and M. Lubis, “Optimization of distance formula in K-Nearest Neighbor method,” *Bull. Electr. Eng. ...*, 2020, [Online]. Available: https://mail.beei.org/index.php/EEI/article/view/1464
[14] S. Kang, “k-Nearest Neighbor Learning with Graph Neural Networks,” *Mathematics*, 2021, [Online]. Available: https://www.mdpi.com/2227-7390/9/8/830
[15] H. Henderi, T. Wahyuningsih, and E. Rahwanto, “Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer,” ... *J. Informatics ...*, 2021, [Online]. Available: http://ijiis.org/index.php/IJIIS/article/view/73
[16] Z. Pan, Y. Wang, and Y. Pan, “A new locally adaptive k-nearest neighbor algorithm based on discrimination class,” *Knowledge-Based Syst.*, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705120304147
[17] M. M. Kumbure, P. Luukka, and M. Collan, “A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean,” *Pattern Recognit. Lett.*, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865520303792
[18] A. X. Wang, S. S. Chukova, and B. P. Nguyen, “Ensemble k-nearest neighbors based on centroid displacement,” *Inf. Sci. (Ny).*, 2023, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025523001731
[19] M. Bansal, A. Goyal, and A. Choudhary, “A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning,” *Decis. Anal. J.*, 2022, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772662222000261
[20] Z. Mushtaq, A. Yaqub, S. Sani, and ..., “Effective K-nearest neighbor classifications for Wisconsin

breast cancer data sets," *J. Chinese* ..., 2020, doi: 10.1080/02533839.2019.1676658.
[21] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review," ... *Data Anal.* ..., 2020.
[22] N. O. F. Daeli and A. Adiwijaya, "Sentiment analysis on movie reviews using Information gain and K-nearest neighbor," *J. Data Sci. Its Appl.*, 2020.
[23] P. Cunningham and S. J. Delany, "k-Nearest neighbour classifiers: (with Python examples)," *arXiv Prepr. arXiv2004.04523*, 2020, [Online]. Available: https://arxiv.org/abs/2004.04523
[24] Y. Tang, Y. Chang, and K. Li, "Applications of K-nearest neighbor algorithm in intelligent diagnosis of wind turbine blades damage," *Renew. Energy*, 2023, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148123007164
[25] M. Arunadevi, M. Rani, R. Sibinraj, M. K. Chandru, and ..., "Comparison of k-nearest neighbor \& artificial neural network prediction in the mechanical properties of aluminum alloys," *Mater. Today* ..., 2023, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221478532304782X
[26] G. Lin, A. Lin, and D. Gu, "Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient," *Inf. Sci. (Ny).*, 2022, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025522006843