Evaluation of the Performance of Naïve Bayes and Logistic Regression in Data Analysis Sentiment Users Digital Wallet

Puspita Wanny, Muhammad Syahputra Novelan

Abstract

DANA digital wallet is application service digital finance available on Google Play Store and used For make it easier users do various transaction non - cash . Application This provide feature digital payments , money transfers, purchases credit , payment bills , as well as various transaction other through device mobile . As one of the application many digital wallets used Indonesian society , DANA obtains diverse reviews and ratings from users on the Google Play Store. Analysis sentiment done For identify opinion users to in three category , namely positive , neutral , and negative . Research data collected through the web scraping process with amount initial 1,248 reviews , then processed through case folding, cleaning, normalization stages text , tokenizing, stopword removal, and stemming before used in classification . Model evaluation is carried out against 958 reviews use accuracy, precision, recall, F1-score, and confusion matrix metrics . Research results show that Naïve Bayes obtains The accuracy value is 79.65%, macro precision 84.04 %, macro recall 63.47 %, and macro F1-score 64.61 %. Meanwhile, In addition , Logistic Regression obtained an accuracy of 88.94%, macro precision of 86.35 %, macro recall of 82.75 %, and macro F1-score of 84.19 %. Logistic Regression also produced 852 predictions . Correct of 958 reviews , more tall compared to Naïve Bayes which produced 763 predictions true . Based on results said , Logistic Regression shows better performance better and better balanced in classify sentiment users in the analyzed dataset . Research This expected can become reference in application of text mining and machine learning for understand opinion users as well as support evaluation quality service application digital wallet .

Keywords *: Analysis Sentiment , Naïve Bayes, Logistic Regression, Digital Wallet , DANA*

Puspita Wanny
of Technology Study Program Information , Faculty Postgraduate Program of Panca Budi Development University , Medan, Indonesia
e-mail: puspitawanny142@gmail.com

Muhammad Syahputra Novelan [2]
e-mail: putranovelan@dosen.pancabudi.ac.id
3rd International Conference Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Cross-Sector and Cross-Disciplinary Strategies for Achieving the SDGS
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Development technology information and communication has bring change significant to various aspect life society , including in field transaction finance . One of the form development the is emergence service digital finance in the form of digital wallet (e-wallet) that allows users do transaction in a way electronic with easy , fast , and practical . Digital wallets are utilized For various needs , such as payment online shopping , in- store transactions , purchases credit , payment bills , to fund transfers[1]. The presence of service This give alternative for public in do transaction without must always using cash . Along with increasing use device mobile and access towards the internet, digital wallets are increasingly become part from activity economy modern society . Conditions the push provider service For Keep going increase quality applications , security transactions , convenience usage , as well as satisfaction users to be able to compete in industry increasingly digital finance develop .

In context digital wallet , analysis sentiment can used For know trend response users to services available , both in category positive , negative , and neutral . Implementation analysis sentiment allows previous review data shaped text No structured processed become more information easy understood and analyzed . Information the can help identify aspect services that obtain appreciation as well as common problems complained by users[2].

One of algorithm that can implemented in classification sentiment is Naïve Bayes. Algorithm This is method classification probabilistic which uses Bayes' Theorem with assumptions that features used nature independent in a way conditional to predicted class[3]. In analysis text , Naïve Bayes many used because of the computing process relatively simple , needs source Power the computation relatively low , and his abilities in handling text data dimensionless high . Algorithm This estimate probability something review including to in category sentiment certain based on characteristics of the words contained in it . Thus , Naïve Bayes can used For classify review users digital wallet to in category sentiments that have been determined . However , the assumption independence interwords used in algorithm This No always reflect word relationships in Language natural[4]. Because of this that , the performance of Naïve Bayes needs to be evaluated using appropriate test data For know his abilities in identify sentiment users in a way accurate .

Besides Naïve Bayes, Logistic Regression is also algorithm classification that can be used For do analysis sentiment on text data . Although use regression terms , algorithms This can implemented For predict class categorical with estimate probability a data includes to in class certain[5]. In classification text , Logistic Regression in general can combined with representation feature such as Bag of Words or Term Frequency–Inverse Document Frequency (TF-IDF) for change text become form numerical that can processed by the model. Algorithm This learn connection between word and category features sentiment through formation function decision based on weight features obtained during the training[6]. His ability in learn contribution feature to predicted class make Logistic Regression one of the relevant methods For classification sentiment . Although Thus , its performance can influenced by data quality , selection features , model parameters, imbalance distribution class , as well as pre-processing processes text . Therefore that , evaluation required For know to what extent the algorithm This capable classify sentiment users digital wallet consistent[7].

Difference characteristics between Naïve Bayes and Logistic Regression makes comparison second algorithm the important For done . Naïve Bayes relies on calculation probability based on distribution features on each class , while Logistic Regression studies connection between features and probabilities class through function logistics[8]. Second algorithm the own different approaches in processing information so that can produce level success different classifications on the same dataset . Evaluation that only based on one size performance potential give less picture comprehensive , especially when interclass data distribution No balanced[9]. Therefore that , research This use a number of metric evaluation , including accuracy, precision, recall, and F1-score, for evaluate performance of each algorithm . In addition , the confusion matrix can used For identify amount correct prediction and error

classification on each category sentiment[10]. By using metric evaluation mentioned , comparison second algorithm expected can give a clearer picture objective about ability classification sentiment based on the dataset used.

Study This expected can give contribution academic in application of text mining and learning machine learning, especially in classification sentiment in Indonesian[11]. From the side practical , results study potential become material consideration for developers and providers service digital wallet in understand trend opinion users as well as identify aspect services that require improvements . In addition , the results comparison can become reference in determine more algorithms in accordance For need analysis sentiment on the characteristics of the dataset studied[12].

Literature Review

Analysis sentiment is one of the application of text mining and processing Language Natural Language Processing (NLP) which aims identify polarity opinion in text to in category positive , negative , or neutral . In context application DANA digital wallet , analysis sentiment to review users on the distribution platform application can utilized For understand level satisfaction , complaints , and perceptions users to aspect services , such as convenience transactions , security , stability applications , as well as quality service customer . Analysis process generally covers review data collection , labeling sentiment , preprocessing text through case folding, cleansing, tokenization , deletion stopwords , and stemming, then to be continued with extraction feature use method such as Term Frequency–Inverse Document Frequency (TF-IDF)[13]. Stage classification can done with compare Naïve Bayes and Logistic Regression algorithms . Naïve Bayes is algorithm probabilistic which applies Bayes' Theorem with assumptions independence conditional between features , so that relatively efficient For classification text dimensionless high . Meanwhile that , Logistic Regression is algorithm classification that models probability something class based on combination weighty from feature text and function logistics , so that can become strong comparator in task classification sentiment . Although both of them Lots used in classification text , performance of each algorithm depends on the characteristics of the data, distribution class , representation features , as well as quality preprocessing[14]. Therefore that , evaluation second algorithm need done use consistent division of training data and test data as well as metric evaluation in the form of accuracy, precision, recall, and F1-score, accompanied by a confusion matrix for identify pattern error classification . The use of stratified split or validation cross can also help produce further evaluation representative , especially when amount reviews on each class No balanced .

Research Methodology

Research data obtained from review users DANA application on Google Play Store through a data collection process using web scraping techniques or tool retrieval of appropriate review data , with attribute in the form of text review , date publications and information relevant supporting data. Collected data furthermore through stage selection For delete duplicate data , review empty , and data that is not relevant , then done labeling sentiment to in category positive and negative or category positive , negative , and neutral in accordance with objective research[15]. Stage preprocessing text includes a folding case for standardize letter become letter small , cleaning for delete characters that are not required , tokenization For separate text into words, deletion stopwords , as well as stemming for change the word to form basically . After that , text changed become representation numeric use Term Frequency–Inverse Document Frequency (TF-IDF) method to be able to processed by algorithm learning machine . The dataset is then shared into training data and test data with 80:20 proportion using stratified split for maintain proportion class sentiment on both data section . Next , Naïve Bayes models, specifically Multinomial Naïve Bayes, and Logistic Regression were trained using training data with stages preprocessing , weighting features , and consistent data sharing for comparison ongoing in a way fair . Both models then tested using the same test data , and its

performance evaluated based on accuracy, precision, recall, and F1-score metrics , as well as a confusion matrix for know amount prediction right and wrong classification on each class sentiment .

Results

4.1 Dataset

Dataset obtained through *scrapping* around 1,248 reviews users DANA application on the *Google Play Store platform* . Each row represents One review equipped users with a number of attribute supporters , such as identity reviews , username , score assessment , time publication reviews , content comments , number sign likes (thumbs up count), and version applications used . Attributes the give information about evaluation users as well as context reviews that can utilized in the analysis process sentiment . As for the example of the results data *scraping* can seen in Table 1.

Table 1. Comment Scraping Result Dataset Play Store DANA Application

No	Username	Comment	Number of Likes
1	@Jae Cool	I Like with This fund application is very helpful my	0
2	@Tomi Sumadi	Excellent	0
3	@Yossi Soenatha	fast response safe	0
4	@Rahman Banjar	After updating , it gives an error and can't be used opened the fund apk .. Help repaired Again yes	0
5	@Januar hafsya	new Got a TF, you know the money his/her directly lost , unable to run where 🤦‍♀️ 💰 🤦‍♀️ 💰 🤦‍♀️ 💰 500,000 missing so no , ask Going to CS for free , it won't happen. can come back too money	0
6	@Ulasmo 7194	Want to I haven't been able to enter Visa debit since yesterday , this disturbance his until when ?	0
7	@IRVAN Hidayat	No There is service help that can be finish customer issues	0
8	@Dwi putri Utami	installment fund balance No can be used qiuris	0
9	@Bastiar gaming STUPID	method active in instant funds	
10	@Alvayed Cudamaja	Help login	
11	@raska aditia	can not enter to in your mind	
12	@Bayu Bayu	My DANA bro bug yeah, why ca n't it be opened? urgent this Want to pull cash Thanks	
…	……..	……..	……
1248	@Syraah Pondan	very useful 👍👍	0

4.2 Preprocessing

Preprocessing stage is the initial data processing process which aims to For clean and simplify text review users DANA application from Google Play Store before used in the analysis process sentiment . In the study here , preprocessing is done through a number of stages , namely case folding for change all over letter become letter small , cleaning for delete characters that are not relevant such as URLs, symbols , signs read , and other elements that are not required , as well as normalization text For correct the word no standard , abbreviation , or Language everyday found in review user . Next , it is done tokenization For break sentence into words, stopword removal for eliminate less common words give information important in classification with still maintain negation such as “ not ” and “ not ”, as well as stemming for changing affixed words become a basic word . The following This is results from Cleaning and case folding are presented in table 2.

Table 2. Cleaning and Case Folding

No	Cleaning	Folding Case
1	I Like with This fund application is very helpful my	I Like with This fund application is very helpful my

2	Excellent	Excellent
3	fast response safe	fast response safe
4	After updating , it gives an error and doesn't work opened the fund apk Help repaired Again yes	After updating , it gives an error and can't be used opened the fund apk .. Help repaired Again yes
5	new get if you know, you know money his/her directly lost, unable to run where lost so just ask Going to CS for free is not possible. can come back too money	new get tf an, you know money his/her directly lost , unable to run where 🤦‍♀️ 💰 🤦‍♀️ 💰 🤦‍♀️ 💰 500,000 missing so no , ask Going to CS for free , it won't happen. can come back too money
6	Want to I haven't been able to enter Visa debit since yesterday This disturbance his until when	Want to I haven't been able to enter Visa debit since yesterday , this disturbance his until when ?
7	No There is service help that can be finish customer issues	No There is service help that can be finish customer issues
8	installment fund balance No can be used qiuris	installment fund balance No can be used qiuris
9	method activate instant funds	method active in instant funds
10	help login	help login
11	Ca n't enter to in your mind	ca n't enter to in the mental mind
12	my sister, my money to beat Yes, you can open it. urgent this want pull cash thank you	my sister, my money to beat Yes, you can open it. urgent this want pull cash thank you
…..	……	…….
1248	very useful	very useful 👍👍

At the stage *preprocessing* , is done normalization and tokenizing for prepare review data users DANA application on *the Google Play Store platform* so that you can used in the classification process sentiment . Normalization done For change the word no standard , abbreviations , and language daily in review users DANA application becomes more word forms standard and consistent for easy analyzed . Next , the next stage tokenization done with break text review into word units so that every word can processed as feature in analysis sentiment . Second stages This aim For increase consistency of text data and support the classification process sentiment use Naïve Bayes and Logistic Regression algorithms . The results of second stages the presented in table 3.

Table 3. Normalization and Tokenizing

NO	Normalization	Tokenizing
1	I Like with This fund application is very helpful my	I | like | with | this | fund | application | it | really | helps | me
2	Excellent	Excellent
3	fast response safe	fast | response | safe
4	after updating , there is no error Can opened the fund application Help repaired Again yes	after | in | update | instead | error | can't | open | the | dana | application | please | fix | again | yes
5	new get a transfer and then you know it money his direct lost don't know how to run Where to lost so just ask to CS is free No going to be Can come back too money	just | got | transfer | an | know | know | the money | immediately | lost | don't | know | ran | where | lost | just like that | ask | to | cs | useless | no | will | be able | to | come back | also | that | money

6	Want to enter visa debit no Can from Yesterday This disturbance his until when	want to | enter | debit | visa | ca n't | from | yesterday | this | disruption | until | when
7	No There is service help that can be finish customer issues	no | service | help | that | can | solve | customer | problems
8	installment fund balance No can be used qiuris	balance | funds | installments | cannot | be | used | qiuris
9	method active in instant funds	how to | active | in | funds | instant
10	help login	help | login
11	No Can enter to in your mind	can't | get | into | into | mentally | always
12	my sister, my money to beat Yes, you can open it. urgent this want pull cash accept love	sis | funds | my | bug | yes | you | can | open | urgent | this | want | withdraw | cash | thank you | thanks
……	……	…….
1248	very useful	very | useful

After stage *tokenizing* , the Stopword Removal and Stemming processes are carried out to reduce unnecessary words necessary and simplify word form . Stage Stopword Removal is performed with remove less common words give information important in analysis sentiments , such as “the”, “and”, “in”, and “ to ”, with still maintain negation such as “ no ” and “ not ” for meaning sentiment No changed . Next , the Stemming stage is carried out For changing affixed words being the basic word , so variations of words that have meaning similar can standardized . Second stages This aim For reduce number of words that are not relevant and improving effectiveness of the classification process sentiment use Naïve Bayes and Logistic Regression algorithms . The results of second stages the shown in table 4.

Table 4. Stopword Removal and Stemming

No	Stopword Removal	Stemming
1	I | like | the application | funds | very | helpful | meh	I Like the fund application is very helpful my
2	Excellent	Excellent
3	fast | response | safe	fast response safe
4	update | even | error | not | opened | application | funds | fixed	update instead error no open good fund application
5	new | get | transfer | an | know | know | money | immediately | lost | no | know | run | lost | ask | cs | useless | no | will | that | money	new get a transfer and then you know it money direct lost don't know how to run lost ask CS for free No will that money
6	want | enter | debit | visa | no | yesterday | disruption | arrive | when	Want to enter visa debit no Yesterday disturb until when
7	no | service | help | solve | problem | customer	No service help finished customer issues
8	balance | funds | installments | not | use | qiuris	installment fund balance useless qiuris
9	how to | active | in | funds | instant	method active in instant funds
10	help | login	help login
11	not | enter | mental | always	No always mentally enter
12	sis | funds | ku | bugging | ko | g | open | urgent | nie | want | pull | cash | accept | love	my funds bro bug why won't it open urgent this Want to pull cash accept love
……	……..	…….

1248	very | useful	very useful

4.3 Visualization of WordCloud Data Processing Results

Data visualization using WordCloud was performed to display the most frequently occurring words in DANA app user reviews after going through preprocessing stages, which include cleaning, normalization, tokenization, stopword removal, and stemming. In this visualization, the size of each word indicates its frequency of occurrence in the collection of review texts, so that words displayed in larger sizes have a higher frequency of occurrence than others. The WordCloud visualization aims to provide an overview of the topics, experiences, and tendencies of user responses to the DANA app based on the dominant words in the dataset. Frequently appearing words can reflect aspects of the service that are frequently discussed by users, such as ease of transactions, application usage, balance, payments, or problems experienced. The results of this visualization were then used as an exploratory analysis to help understand the characteristics of the data before sentiment classification using the Naïve Bayes and Logistic Regression algorithms. Visualization results WordCloud from all comment data shown in Figure 1.



Figure 1. WordCloud

4.4 Data Labeling

Labeling process aim For grouping review to in three category , namely positive , neutral , and negative , with consider words that contain sentiment in text reviews . Based on results labeling on files, obtained as many as 574 reviews labeled positive , 531 reviews labeled neutral , and 142 reviews labeled negative . Labeling This supported by columns score_sentiment showing mark results calculation sentiment as well as column positive_words and negative_words which record the identified words own trend sentiment certain . Reviews containing words with trend positive , such as “ like ”, “ help ”, and “ satisfied ”, can categorized as sentiment positive , whereas reviews that show complaint or dissatisfaction can leading to sentiment negative . Meanwhile that , the review that got score sentiment neutral or No show trend enough sentiment strong entered to in category neutral . Labeling results the furthermore used as a class label in the classification process use Naïve Bayes and Logistic Regression algorithms .

Table 3. Data Labeling

Comment	**Sentiment score**	**label**
I Like the fund application is very helpful my	2	Positive
Excellent	3	Positive
fast response safe	2	Positive
new get a transfer and then you know it money direct lost don't know how to run lost ask CS for free No will that money	-1	Negative
installment fund balance useless qiuris	0	Neutral
thank God it's good	2	Positive

more program funds practical fast safe	3	Positive
transaction So more easy safe and good	2	Positive
arrive Dana account is difficult to access tonight still have some money left for lunch open mlh cm rupiah rb ny lost update , just look again It's hard to open , uninstall , then install again using it sm phone number usually it can, but sometimes it ca n't very How	-3	Negative
very good transaction comfortable fluent	4	Positive
method active in instant funds	0	Neutral
Want to enter visa debit no Yesterday disturb until when	-1	Negative
help login	0	Neutral
Dana says cut thousand	0	Neutral
Good believe	2	Positive
Excellent easy once for all transactions	2	Positive
problem very bad service very difficult ask for help ask customer care with the bot, it just keeps going round and round answer of course application No clear	-4	Negative
I no ever use installment funds still pay installments please Good	0	Neutral
update instead error no open good fund application	-1	Negative
application is very difficult for complain chat ana bot suru ask then email customer care reply customer care instead ask me to chat or not road installment fund withdrawal frozen instant No very activity for disappointed annoyed service bad	-4	Negative
That's it.	0	Neutral

4.5 Classification Sentiment

In the research this is the dataset that has been through stage *preprocessing* and labeling sentiment Then shared into two parts , namely training data and testing data. Data division is carried out use 80:20 scheme , where 998 data (80%) are used as training data for training a classification model sentiment , while 250 data (20%) is used as testing data for test model capabilities in classify data that has not been Once studied previously . Thus , the total data used in study This is 1248 data.

Table 4. *Training* and *Testing Data*

Data Types	Amount of Data	Percentage
Training	998	79.99%
Testing	2250	20.01%
Total	**1248**	**100%**

4.6 Application of Naïve Bayes

Naïve Bayes algorithm is applied For classify sentiment users DANA application based on reviews obtained from Google Play Store. Data used in this stage classification consists of over 958 reviews that have been through preprocessing and labeling processes sentiment to in three category , namely positive , neutral , and negative . The preprocessing stage produces more text structured through the process of case folding, cleaning, normalization , tokenization , stopword

removal, and stemming. Next , the text preprocessing results are used as base in the classification process For predict category sentiment every review . The resulting predictions Then compared to with actual label For know level model accuracy in recognize sentiment users to DANA application .

Evaluation results , the Naïve Bayes algorithm obtained accuracy value of 79.65%, which shows that the model is successful classify around 79.65% of the total 958 reviews in a way True . The macro precision value reached 84.04%, while the macro recall was 63.47% and the macro F1-score was 64.61%. That is , the weighted precision value is 81.02%, the weighted recall is 79.65%, and the weighted F1-score is 77.16%. These results show that the model has level accuracy enough prediction Good in a way overall , but his abilities in recognize all over class sentiment in a way evenly Still need improved . Difference between The precision macro and recall macro values also show that accuracy model predictions do not always followed by the same ability the good in find all over reviews on each category sentiment .

Based on the classification report, the model performance differs in each category sentiment . Class positive get precision value of 84.03%, recall of 93.05%, and F1-score of 88.31%, so that become class with performance best . Class neutral achieved a precision of 73.37%, a recall of 80.37%, and an F1-score of 76.71 %. that , class negative obtained a precision of 94.74%, but recall value only reached 16.98% and F1-score of 28.80%. This show that when the model predicts something review as negative , prediction the tend right , but the model is still difficulty identify part big review the actual negative . Thus , the ability Naïve Bayes classification on class negative Still become weakness main thing that is needed noticed.

The results of the confusion matrix clarify pattern error classification that occurred in 958 reviews testing . Of the 106 reviews that were current labeled negative , only 18 reviews were successful predicted as negative , while 77 reviews were misclassified as neutral and 11 reviews as positive . In class neutral , from 377 reviews actual , as many as 303 reviews succeed predicted with Correct as neutral , while 1 review predicted as negative and 73 reviews as positive . As for the class positive , from 475 reviews actual , as many as 442 reviews succeed classified with true , while 33 reviews predicted as neutral and not There is predicted reviews as negative . This result show that error most prominent classification happened in the review negative tendencies recognized as neutral . Condition the can indicates that characteristics of words in review negative Not yet Enough differentiated by model or there is mismatch between content text and its sentiment label .

In a way overall , results testing show that Naïve Bayes algorithm has sufficient ability Good in classify sentiment positive and neutral users DANA application , but not optimal in recognize sentiment negative . The accuracy value of 79.65% indicates that the model is capable give correct predictions in part large test data . However , the macro F1-score value of 64.61% indicates that model performance yet evenly distributed across the three category sentiment , especially Because low recall in class negative . Therefore that , model evaluation is not Enough only based on accuracy values , but also necessary considering precision, recall, F1-score, and confusion matrix. The following shown in Figure 2 Naïve Bayes Confusion Matrix.

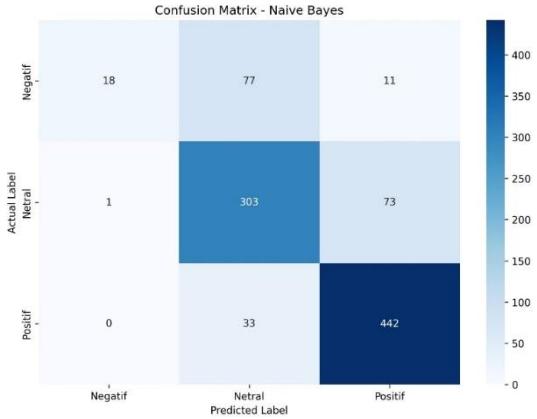


Figure 2. *Confusion Matrix Naive Bayes*

4.7 Implementation of SVM

Data analyzed totaling 958 reviews , with actual label distribution consists of over 475 reviews positive , 377 reviews neutral , and 106 reviews negative . Before done classification , text review through stages Preprocessing includes case folding, cleaning, word normalization , tokenizing, stopword removal, and stemming . the aim For clean as well as standardize text so that it can processed by the classification model.

The Logistic Regression Model obtains accuracy value of 88.94%, which shows that part big review succeed classified in accordance with its actual label . The macro precision value reached 86.35%, macro recall 82.75 %, and macro F1-score 84.19 %. Meanwhile, In that case , the weighted precision value reached 89.03%, the weighted recall was 88.94%, and the weighted F1-score was 88.80%. Based on the classification report, the category positive own performance best with precision value of 94.99%, recall of 91.79%, and F1-score of 93.36%. Category neutral obtained a precision of 84.06%, a recall of 92.31%, and an F1-score of 87.99%. The categories negative obtained a precision of 80.00%, a recall of 64.15%, and an F1-score of 71.20%.

Evaluation results through the confusion matrix shows that of 106 reviews that are current categorized negative , as many as 68 reviews succeed predicted as negative , while 32 reviews were misclassified as neutral and 6 reviews as positive . From 377 reviews neutral , as many as 348 reviews classified with true , while 12 reviews predicted as negative and 17 reviews as positive . Meanwhile that , from 475 reviews positive , as many as 436 reviews succeed predicted with true , while 5 reviews classified as negative and 34 reviews as neutral . Thus , the model produced 852 correct predictions . from a total of 958 reviews , according to with level accuracy around 88.94%. The following shown in Figure 3 Confusion Matrix Logistic Regression.

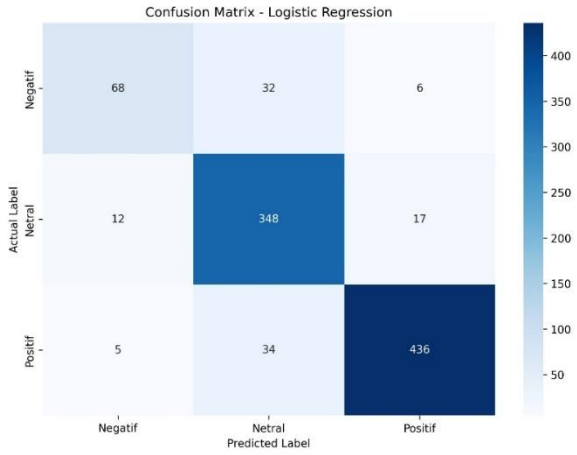


Figure 3. *Confusion Matrix* Logistic Regression

4.8 Comparison Naïve Bayes Algorithm and Logistic Regression

Comparison Naive Bayes and Logistic Regression algorithms on 958 review data , the Logistic Regression model shows better performance Good compared to Naive Bayes. Logistic Regression obtains mark accuracy of 88.94%, macro precision 86.35 %, macro recall 82.75 %, and macro F1-score 84.19 %, while Naive Bayes obtained accuracy of 79.65%, macro precision of 84.04 %, macro recall of 63.47 %, and macro F1-score of 64.61 %. Based on the confusion

matrix, Logistic Regression is able to classify 852 data with true , while Naive Bayes classified 763 data with Correct .

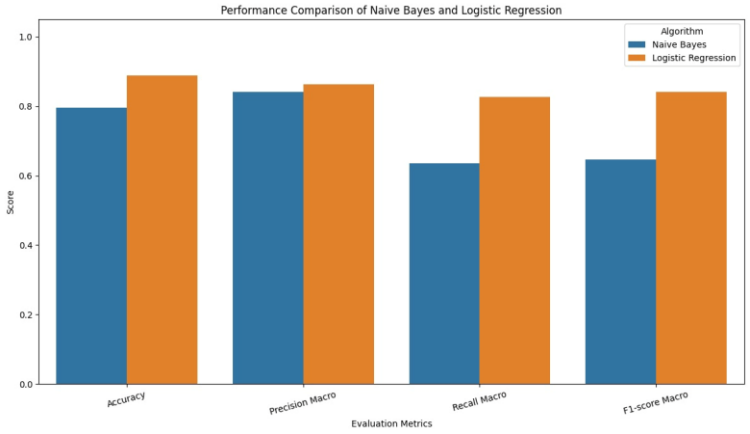


Figure 4. Comparison diagram

Discussion

Study This analyze sentiment users application DANA digital wallet based on reviews obtained from Google Play Store with compare Naïve Bayes and Logistic Regression algorithms . Based on document research , the data collection process resulted in 1,248 reviews user who loads information about experience , judgment , as well complaint to service application . The data Then through the selection , pre-processing and labeling processes sentiment For identify trend opinion users to in three category , namely positive , neutral , and negative . Labeling results beginning show as many as 574 reviews categorized positive , 531 reviews categorized neutral , and 142 reviews categorized negative . Next , evaluate classification done against 958 reviews , with actual label distribution as many as 475 reviews positive , 377 reviews neutral , and 106 reviews negative . Difference amount of data between stage collection and testing show that data used at stage classification No covers all initial data , so that consistency recording amount of data on each stages need noticed in reporting results study .

Before the classification process conducted , review data processed through preprocessing stages consisting of top case folding, cleaning, normalization text , tokenizing, stopword removal, and stemming. Case folding is used For standardize letter become letter small , while cleaning aims delete characters that are not required . Normalization done For standardize the word no standard , abbreviations , and language everyday that often found in review users . Next , tokenizing breaks down sentence into word units , stopword removal removes less common words give information important with still retaining the negated word , and stemming changes the affixed word become form basically . Text data that has been cleaned Then used in the classification process sentiment . In addition , visualization WordCloud used For describe frequently used words appear in review so that can help understand many topics discussed users , such as convenience transactions , balances , payments , and constraint applications and services customer .

Test results Naïve Bayes algorithm shows The accuracy value is 79.65%, the macro precision is 84.04%, the macro recall is 63.47%, and the macro F1-score is 64.61%. Based on category sentiment , class positive get precision value of 84.03%, recall of 93.05%, and F1-score of 88.31%, while class neutral obtained a precision of 73.37%, a recall of 80.37%, and an F1-score of 76.71%. In the class negative , the precision value reached 94.74%, but the recall was only of 16.98% and F1-score of 28.80%. The confusion matrix results show that from 106 reviews negative , only 18 were successful classified in a way Correct as negative , while 77 reviews were miscategorized as neutral and 11 reviews as positive . Findings This show that Naïve Bayes is sufficient Good in recognize sentiment positive , but Still own limitations in detect review negative . High precision in class negative show that prediction negative

generated tend right , but low recall indicates that part big review the actual negative Not yet succeed identified by the model.

Different with Naïve Bayes, Logistic Regression produces performance more classification tall based on results evaluation in document . This model obtained an accuracy of 88.94%, a macro precision of 86.35%, a macro recall of 82.75%, and a macro F1-score of 84.19%. Based on category sentiment , class positive obtained a precision of 94.99%, a recall of 91.79%, and an F1-score of 93.36%. Class neutral obtained a precision of 84.06%, a recall of 92.31%, and an F1-score of 87.99%, while class negative achieved a precision of 80.00%, a recall of 64.15%, and an F1-score of 71.20%. Based on the confusion matrix, 68 of the 106 reviews negative succeed classified with true , while 32 reviews predicted as neutral and 6 reviews as positive . From 377 reviews neutral , as many as 348 were successful classified with true , whereas from 475 reviews positive , as many as 436 were successful predicted with true . In Overall , Logistic Regression produced 852 predictions . Correct of 958 reviews tested . These results show that Logistic Regression has ability more Good in recognize third category sentiment , especially Because class recall value negative Far more tall than that obtained by Naïve Bayes.

Comparison second algorithm show that Logistic Regression is more superior than Naïve Bayes in classification sentiment review users DANA application based on accuracy value and F1-score macro. Logistic Regression obtains accuracy of 88.94%, while Naïve Bayes obtained 79.65%, so that there is difference by 9.29 points percentage . In addition , the F1-score of macro Logistic Regression reached 84.19%, more than tall compared to Naïve Bayes which obtained 64.61%. Difference This show that Logistic Regression has better performance balanced in classify sentiment positive , neutral , and negative on the test data used . Although Thus , Logistic Regression is still own room repairs , especially in recognize review negative Because its recall value 64.15 % has not yet reach same level with class positive and neutral . Research results This indicates that election algorithm need consider more from just accuracy , including ability detect every class sentiment . Therefore that , Logistic Regression can chosen as a better model in accordance for research datasets this , with notes that consistency amount of data, quality labeling , and data sharing is necessary ensure that the results evaluation more can reliable .

CONCLUSION

Based on results research , can concluded that Logistic Regression algorithm has performance more Good compared to Naïve Bayes in do analysis sentiment to review users application DANA digital wallet on the tested dataset . Logistic Regression obtained The accuracy value was 88.94%, macro precision 86.35 %, macro recall 82.75 %, and macro F1-score 84.19 %, while Naïve Bayes obtained an accuracy of 79.65%, macro precision 84.04 %, macro recall 63.47 %, and macro F1-score 64.61 %. The confusion matrix results show that Logistic Regression is capable classified 852 of 958 reviews with true , while Naïve Bayes classified 763 reviews with true . The advantages of Logistic Regression are mainly seen in ability recognize sentiment negative , with recall value of 64.15%, compared to Naïve Bayes which only reached 16.98%. Thus , Logistic Regression is a more accurate model. in accordance For classification sentiment in this dataset . However , research furthermore need notice consistency amount of data, validity labeling and distribution class as well as consider parameter optimization or handling strategy imbalance class so that the model's capabilities in recognize sentiment negative can improved .

Reference

[1] L. R. Handoko, A. Faza, M. A. Barata, and A. E. Pajri, “Comparison of the Performance of K-Nearest Neighbor and Naive Bayes Algorithms for Sentiment Analysis of PinjamYuk Application User Reviews Using SMOTE and TF-IDF,” vol. 10, no. 4, pp. 3991–4000, 2026.

[2] P. A. Aziz, S. Barokah, N. Ilahi, S. Moka, and A. M. Sajiah, "Penerapan Hadoop untuk Analisis Sentimen Berbasis Big Data pada Ulasan Aplikasi Transportasi Online," vol. 5, no. 1, pp. 51–60, 2025, doi: 10.54259/satesi.v5i1.4051.

[3] J. Nasional, S. Informasi, N. Cahaya, I. Tahyudin, and A. Shouni, "Perbandingan Algoritma Support Vector Machine , Decision Tree , dan Logistic Regresion Pada Analisis Sentimen Ulasan Aplikasi Netflix," vol. 02, pp. 110–117, 2024.

[4] A. R. Firdaus *et al.*, "PERBANDINGAN KINERJA NAÏVE BAYES DAN INDOBERT DALAM," vol. 10, no. 5, pp. 9418–9424, 2026.

[5] F. Farahdinna *et al.*, "PEMANFAATAN MACHINE LEARNING DALAM ANALISIS SENTIMEN PENGGUNA E-WALLET MENGGUNAKAN SVM," vol. 9, no. 2, pp. 3522–3526, 2025.

[6] N. Ayu, S. Hardianti, and D. Abdullah, "A Comparative Analysis Of User Sentiment And Review Patterns On The Dana And Seabank Apps Using The Naive Bayes , Support Vector Machine And Logistic Regression Algorithms Analisis Komparatif Sentimen Dan Pola Ulasan Pengguna Aplikasi Dana Dan Seabank Menggunakan Algoritma Naive Bayes , Support Vector Machine , Dan Logistic Regression," vol. 5, no. 3, pp. 1181–1196, 2026.

[7] C. Dometian and S. Jatmiko, "Analisis Sentimen Ulasan Google Play Store : Studi Komparatif Algoritma SVM , Naïve Bayes , dan Logistic Regression," vol. 15, no. 3, pp. 610–620, 2025.

[8] R. A. Prasetyo, W. T. Saputro, and D. Chirzah, "Jurnal Informatika : Jurnal pengembangan IT Perbandingan Logistic Regression , SVM , dan Random Forest untuk Analisis Sentimen Ulasan Aplikasi Gopay," vol. 10, no. 4, pp. 1176–1188, 2025, doi: 10.30591/jpit.v10i4.8796.

[9] M. F. Zaenudin and Y. Sibaroni, "COMBINATION OF LOGISTIC REGRESSION AND NAÏVE BAYES," vol. 10, no. 2, pp. 1286–1298, 2025.

[10] M. A. Hermansyah, M. Arifin, and P. Setiaji, "Sentiment Classification of Indonesian PayLater App Reviews Using Random Forest and Logistic Regression," vol. 9, pp. 193–203, 2026.

[11] L. R. Fitriati, R. G. Guntara, and B. M. Purwaamijaya, "Analisis Sentimen Ulasan Pengguna GoPay di Google Play Store menggunakan Model IndoELECTRA," pp. 2295–2304, 2025, doi: 10.33364/algoritma/v.22-2.3053.

[12] A. Rahman *et al.*, "Analisis Sentimen Pengguna terhadap Aplikasi Shopee Menggunakan Metode Naive Bayes sebagai Pendukung Pengambilan Keputusan dapat digunakan untuk," vol. 3, no. 4, pp. 951–960, 2026.

[13] E. Nurrohman *et al.*, "Analisis sentimen pengguna mypertamina pada google playstore menggunakan naïve bayes classifier," vol. 3, no. 04, pp. 1–8, 2026.

[14] N. A. Rahmi and E. Sawitri, "Peningkatan Akurasi Klasifikasi Sentimen Pengguna Dompet Digital Menggunakan Stacking Ensemble Machine Learning," vol. 4, pp. 97–109, 2026.

[15] F. Wijaya, "KLASIFIKASI SENTIMEN ULASAN APLIKASI OVO MENGGUNAKAN SVM , LOGISTIC REGRESSION , BI-LSTM , DAN," vol. 21, no. 1, pp. 84–93, 2026.