Journal of Information Technology and Computer Science (INTECOMS)
Volume 6 Nomor 2, Desember 2022
e-ISSN : 2614-1574
p-ISSN : 2621-3249

SENTIMENT ANALYSIS OF BNI MOBILE APPLICATION USING THE K-NEAREST NEIGHBOR ALGORITHM (KNN) WITH PARTICLE SWARM OPTIMIZATION (PSO) FEATURE SELECTION

Dewi Yustika Lakoro[1], Ema Utami[2], Dhani Ariatmanto[3]
[1,2,3]Universitas Amikom Yogyakarta
Dewiyustikalakoro@students.amikom.ac.id[1]

ABSTRACT

Analysis sentiment is field studies analyze opinions, sentiments, evaluations, attitudes and emotions to entity like products, services, organizations, individuals, issues, events, films and topics. Sentiment analysis first succeed used previously in various fields, such as movie ratings, service ratings, product ratings, etc. and recently This get popularity in the field economy especially on e- commerce applications. Moment This industry banking do various innovation finance that is shift their focus from banking traditional to banking based technology for fulfil need customers as well as For increase Power competitive. Mobile banking is one of them from innovation that. Response in use of this mobile banking application enter every the day with amount response as much hundreds so that response the difficult For sorted become responses included positive or including negative response. Study This using private data which is comment BNI Mobile banking application in the application laystore. Existing data Then will be done preprocessing with do a number of stages, ie started from tokenization, stemming, stopword removal and Term Frequency Inverse Document Frequency (TF-IDF) was carried out. Initial process carried out in classification is accept input comment data then the preprocessing process is carried out, then included in the classification model, method classification used is K-Nearest Neighbor with particle swarm optimization Optimization and finally is the results issued is accuracy from method used. After That researcher do optimization use one algorithm optimization namely PSO with combine with KNN or called KNN-PSO gain accuracy of 92.33% or only has an error of 7.67%. If seen from the amount of data is successful do classification as many as 277 data, in meaning only has 23 errors or missing data succeed classified with good.

***Keyword**: Sentimen Analysis, BNI Mobile, TF-IDF, KNN, PSO*

INTRODUCTION

Language is powerful tool for communicate and convey information. This is alsoa means for express emotions and sentiments. Analysis sentiment is field studies analyze opinions, sentiments, evaluations, attitudes and emotions to entity like products, services, organizations, individuals, problems, events, films and topics (Ridwansyah, 2022).

Sentiment analysis is carried out at levels different details with different levels. Sentiment Analysis used for count polarity sentiment text on level sentence. Sentiment Analysis aim For mining opinion or analysis sentiment on content textual (Sharma and Ghose 2020).

Moment This industry banking do various innovation finance that is shift their focus from banking traditional to banking based technology for fulfil need customers as well as for increase Power competitive. Mobile banking is one of them from innovation that. Mobile banking can defined as do various transaction banking such as fund transfers, checks balance, investment, payment billing made through use cell phone. Facility this remove limitation space and time from various transaction banking (Nursiah et al., 2022).

Analysis sentiment is field very popular research in text mining you can made as method for overcome matter that. Basic ideas in analysis sentiment this that is find polarity from documents and classify them become positive or negative on the model that can be used with one machine learning algorithm. Classification process there is Lots algorithm possible classification used (Ainurrohma, 2021).

On research analysis sentiment previously about Sentiment Analysis on Twitter using Information Gain and Naïve Bayes which was researched by Sari, Insan

and Rini that propose algorithm Naïve Bayes classification with Information gain optimization. The Naïve Bayes algorithm was chosen because strength accuracy (Widya Sihwi, Prasetya Jati, and Anggrainingsih 2018).

Study other about Optimization feature with use Particle swarm Optimization (PSO) with combining with naïve Bayes for recommendation clothes (Ismail et al. 2020) with propose PSO based Naïve Bayes algorithm obtains mark accuracy with good in comparison with Naïve Bayes algorithm without use optimization features.

Study other related with Analysis Entitled sentiment opinion in online media using Particle Swarm Optimization and the Naïve Bayes algorithm to get more value Good compared to with SVM and SVM+PSO algorithms. So you can concluded that application optimization can increase accuracy. Models in Naïve Bayes Classifier and Particle Swarm Optimization can give solution for classification Sentiment Analysis (Idrus, Brawijaya, and Maruloh 2018).

Study other about Implementation of Text Mining Analysis Sentiment is also in the Twitter application using K-Fold Cross Validation and Naïve Bayes Classifier. Data used in study This is collection of tweets regarding tourist Medan city from 1 December 2019 to 8 December 2019. The word count results are results from analysis sentiment using RStudio software in tourism Medan city is 1803 positive tweets or 80.2% while sentiment tweets negative namely 137 or amounting to 19.8% (Ridwansyah , 2022).

Study other about K-Nearest Neighbor Based on Particle Swarm Optimization for Analysis Sentiment Against Tokopedia. Objective from study this is for grouping perception customer based on class positive, neutral, and negative. Produce accuracy evaluation best of 97.9%, precision of 96.17%, recall of 96.62%, and f-measure of 96.39% (Pajri et al., 2020).

Study this expected can do analysis sentiment about review mobile banking applications especially BNI Mobile application so can is known accuracy as well as influence from application from Optimization Particle Swarm Optimization feature inside analysis sentiment with apply method for classification namely K-Nearest Neighbor

METHOD

1. **Method Data Collection**
   Study This using private data which is comment BNI Mobile banking application in the application playstore. Data taken from comment user application. Existing data then labeled with do validation by translation with two labels, viz comment labeled positive and negative labels. The amount of data taken totaling 300 comments using Indonesian. Existing data Then will be done preprocessing with do a number of stages, ie started from tokenization, stemming, stopword removal and Term Frequency Inverse Document Frequency (TF-IDF) was carried out.
2. **Method Data Analysis**
   Initial process carried out in classification is accept input comment data then the preprocessing process is carried out, then included in the classification model, method classification used is K-Nearest Neighbor with particle swarm optimization Optimization and finally is the results issued is accuracy from method used .

3. Research Flow

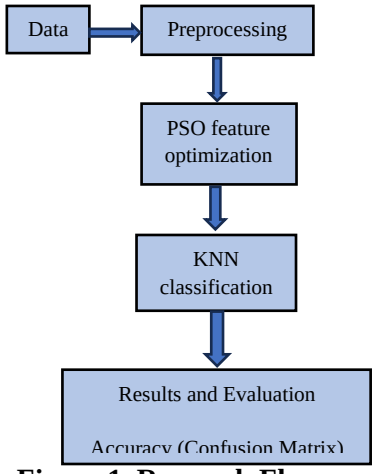


Figure 1. Research Flow

a. Stage 1
Dataset: Dataset used is Comment data user BNI Mobile banking application available on Playstore. Data used totaling 300 with own two labels viz Positive and Negative. After the data is collected then do it preprocessing.

b. Stage 2
Preprocessing: Data used will be done preprocessing with do tokenization, stemming, stopword removal and Term Frequency Inverse Document Frequency (TF-IDF) was carried out. After doing it preprocessing enter optimization features.

c. Stage 3
Optimization Features: after doing it preprocessing so will be done selection feature with using Particle Swarm Optimization. On optimization the with enter population size value, as well determine maximum value of number generation.

d. Stage 4
Classification process: At stage This after doing it selection feature then the classification process is carried out with use K-Nearest Neighbor method.

e. Stage 5
Results and Evaluation: Results and evaluation use the confusion matrix for see mark accuracy, precision and recall

RESULT AND DISCUSSION

1. Data Collection

Stages early in the research **t**his ie data collection. Data taken ie comments on the BNI Mobile banking application on Playstore as many as 300 data. Existing data Then given a sentiment label positive and negative. The amount of data taken totaling 300 comments using Indonesian. the data consists of 215 sentiments positive and 85sentiment negative. Existing data Then will done preprocessing with do a number of stages, ie started from tokenization as many as 2 stages, stage First For symbols or interpreted regular expression ([!"#$%&'()*+,./:;<=>?@\[\\\]_`{|}~]), stage 2nd non-letter tokenizing. After That Transform cases to lower case, next filter tokens by length min chars 4-25 char, after that filter stopwords (dictionary) and finally do steaming (dictionary). Whole stages the enter TF-IDF. The dataset table can be seen in table 1 below this:

Table 1. Dataset table

No	Comment User	Sentiment
1	Application this is very helpful in transaction origin network stable sometimes wear wifi looks ok ok aj whereas again in trouble So blamed the application whereas the application Cool very added the features are easy and so make easy in transaction. Accept love Mbanking BNI yang always make So easy in transaction. Keep it up yes the security.	Positive
2	Application sometimes like error arrives", want transaction, so to inhibited. I don't know if I want to update, I don't know. Rather annoying indeed, then The application is also	Negative

	slow, you know I want to update Jan half ", so it's not complicated klo mo Transactions must wait for updates.

The process of tokenizing, transform case, filter token by length, stopword and steaming as following:

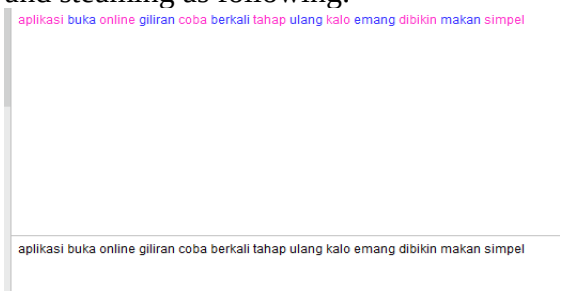


Figure 2. Tokenizing process

Following display of output data from processed TF-IDF use RapidMiner app:

Table 2. Output Table From TF-IDF Using Rapidminer

Word	Attribute Name	Total Occirrences	Document Occurrences	Negative	Positive
Amen	Amen	1	1	0	1
Abdet	Abdet	1	1	0	1
Program	Program	1	1	1	0

No	Word	In Documents	Total	In Class (Negative)	In Class (Positive)
1	Amen	1	1	0	1
2	Abdet	1	1	0	1
3	Program	1	1	1	0

After That use the SMOTE UP sampling operator for balance the data between two classes for performance more good. The two classes in question are sentiment negative and positive. The SMOTE process can seen in the image below this:

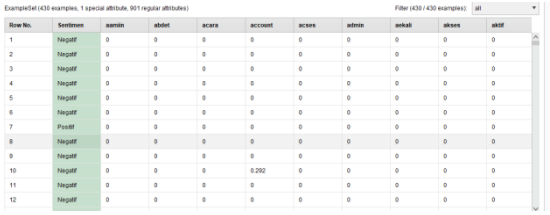


Figure 3. SMOTE process

2. Data Processing Process Uses KNN

After that's the next process ie use KNN algorithm, but before That researcher do evaluation between training and testing datasets using Cross Validation for do validation of training and testing data on 300 existing data obtained.

a. KNN Algorithm

Next process ie do analysis use KNN algorithm. Researcher's KNN analysis process break it down as following:

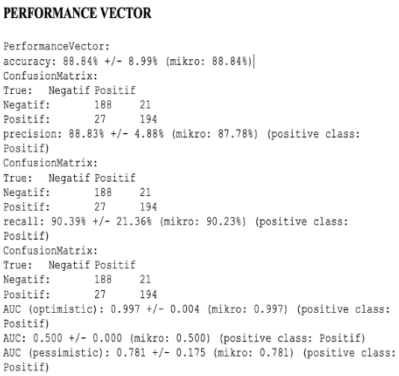


Figure 4. KNN Analysis Process

From the results analysis found that KNN algorithm already Good in do predictions proven from table 3 confusion matrix KNN get results accuracy 88.84% like seen below this:

Table 3. KNN Confusion Matrix

	True negative	True Positive	Class Precision
Pred Negative	188	21	89.95%
Pred Positive	27	194	87.78%
Class Recall	87.44%	90.23%	

From table 3 above explain that results processing using KNN get accuracy amounting to 88.84%. For predictions negative and true negative as many as 188 data. Whereas predictions Negative and true Positive or incorrect data 21 data with class precision 89.95%. For predictions but true Negative 27 data whereas For predictions Positive and true Positive as many as 194 with class precision of 87.78%.

b. PSO-KNN Algorithm

Example SMOTE UP sampling for 20 data taken:

Table 4. SMOTE UP Example

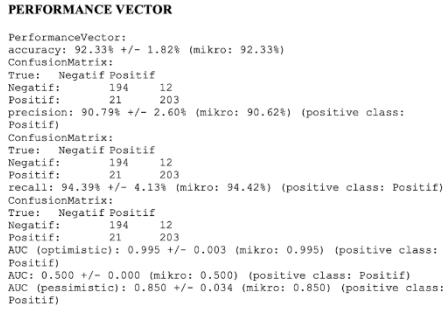
NO	SENTIMENT	TEXT
1	Positive	application help transaction network stable sometimes wear wifi disturbance blamed the application the application Cool very added its features easy easy transaction Sorry especially now full wifi network the application wronged accept love mbanking easy keep your transactions safe
2	Negative	application sometimes like transaction errors hampered by updating likes sure annoying indeed then slow updating the application is complicated Transactions must wait for updates

c. PSO Parameters
Population Size: 5
Maximum number of generations: 30
The result weight attribute PSO 20 data processing

Table 5. PSO parameters

Attributes	Weight
Amen	1
Abdet	0
Program	0.712599
Accounts	1
Access	0.053995
Admin	0.471494
Sometimes	0
Access	1
Active	0
Activity	0.335457
Activation	0
Its Activation	0
Activity	0.570531
Activity	1
Account	0.670953
Accurate	0.260083
Alert	1
Alhamdulillah	1
Thank God	0.046895
Safe	0.165989

d. Confusion Matrix


PERFORMANCE VECTOR

PerformanceVector:
accuracy: 92.33% +/- 1.82% (mikro: 92.33%)
ConfusionMatrix:
True: Negatif Positif
Negatif: 194 12
Positif: 21 203
precision: 90.79% +/- 2.60% (mikro: 90.62%) (positive class: Positif)
ConfusionMatrix:
True: Negatif Positif
Negatif: 194 12
Positif: 21 203
recall: 94.39% +/- 4.13% (mikro: 94.42%) (positive class: Positif)
ConfusionMatrix:
True: Negatif Positif
Negatif: 194 12
Positif: 21 203
AUC (optimistic): 0.995 +/- 0.003 (mikro: 0.995) (positive class: Positif)
AUC: 0.500 +/- 0.000 (mikro: 0.500) (positive class: Positif)
AUC (pessimistic): 0.850 +/- 0.034 (mikro: 0.850) (positive class: Positif)

Figure 4. KNN-PSO Analysis Process

Table 5. KNN-PSO confusion matrix

	True Negative	True Positive	Class Precision
Pred. Negative	194	12	94.17%
Pred. Positive	21	203	90.62%
Class Recall	90.23%	94.42%	

From table confusion matrix above seen that PSO KNN experienced enhancement initial accuracy only using KNN with accuracy 88.84% after added algorithm PSO optimization to 92.33% occurred enhancement accuracy of 3.49% after use algorithm PSO optimization. From the table above describe that predictions Negative is also true Negative amounting to 194 total data and predictions Negative and true Positive or data that is not succeed predicted only 12 data with class precision 94.17%. Furthermore For predictions Positive but true Negative or misclassification 21 amount of data and predictions Positive is also true Positive a total of 203 data with class precision of 90.62%. Proven that KNN-PSO can increase accuracy of 3.49% with do combination both. KNN-PSO got accuracy highest from results comment dataset processing via the selected BNI Playstore in a way random as many as 300 comment data.

e. SVM and Logistic Regression (LR)
Researchers also do a number of analysis use SVM algorithm, and also Logistic Regression (LR) for compared to with KNN-PSO research created. Researcher get results seen in table 6 as following:

Table 6. Comparison algorithm

Algorithm	Accuracy
VECTOR MACHINE SUPPORT	80.39%
LOGISTIC REGRESSION	73.33%
K-NEAREST NEIGHBORS	88.84%
KNN-PSO	**92.33%**

CONCLUSION

From the results study this can concluded that KNN can also do classification with Good compared to a number of algorithm other such as Naïve Bayes, SVM, and so on researched previously in the journals found were 88.84% meaning KNN algorithm only has

11.16 classification errors. If averaged in a dataset of 300 data, KNN was successful do classification as many as 266 amounts of meaningful data only only has 34 data errors. After That researcher do optimization use one algorithm optimization namely PSO with combine with KNN or called KNN-PSO gain accuracy of 92.33% or only has an error of 7.67%. If seen from the amount of data is successful do classification as many as 277 data, in meaning only has 23 errors or missing data succeed classified with good

REFERENCES

Ainurrohma. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. *PRISMA, Prosiding Seminar Nasional Matematika*, *4*, 493–499. https://journal.unnes.ac.id/sju/index.php/prisma/

Idrus, Ali, Herlambang Brawijaya, and Maruloh. (2018). Sentiment Analysis Of State Officials News On Online Media Based On Public Opinion Using Naive Bayes Classifier Algorithm And Particle Swarm Optimization. In *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–7. Parapat, Indonesia: IEEE. https://doi.org/10.1109/CITSM.2018.8674331.

Ismail, Abdul Rahman, Ahmad Zainul Fanani, Guruh Fajar Shidik, and Muljono. (2020). Implementation Of Naive Bayes Algorithm With Particle Swarm Optimization In Classification Of Dress Recommendation. In *2020 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 174–78. Semarang, Indonesia: IEEE. https://doi.org/10.1109/iSemantic50169.2020.9234293.

Nursiah, N., Ferils, M., & Kamarudin, J. (2022). Analisis minat menggunakan mobile banking. *Akuntabel*, *19*(1), 91–100. https://doi.org/10.30872/jakt.v19i1.10711

Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *2*(5), 178–185. https://doi.org/10.30865/klik.v2i5.362

Sharma, Ankita, and Udayan Ghose. (2020). Sentimental Analysis of Twitter Data with Respect to General Elections in India. *Procedia Computer Science* 173: 325–34. https://doi.org/10.1016/j.procs.2020.06.038.

Pajri, D., Umaidah, Y., & Padilah, T. N. (2020). K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(2). https://doi.org/10.28932/jutisi.v6i2.2658

Widya Sihwi, Sari, Insan Prasetya Jati, and Rini Anggrainingsih. (2018). Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier. In *2018 International Seminar on Application for Technology of Information and Communication*, 190–95. Semarang: IEEE. https://doi.org/10.1109/ISEMANTIC.2018.8549757.