SURAT TUGAS
265/B.01/PPPM-NM/IX/2020

Tentang

PENELITIAN YANG DIPUBLIKASIKAN DALAM PROSIDING
Periode September 2020 - Februari 2021

Menulis pada International Conference on Advanced Information Scientific Development
(ICAISD) 2020 6-7 August 2020, West Java, Indonesia

Judul :

Comparison of Text Mining Classification Algorithms in Interbank Money Transfer Application

Menimbang	:	1. Bahwa perlu di adakan pelaksanaan Seminar dalam rangka Seminar.
		2. Untuk keperluan tersebut, pada butir 1 (satu) di atas, maka perlu dibentuk Penulis Seminar.

MEMUTUSKAN

Pertama	:	Menugaskan kepada saudara yang tercantum sebagai Penulis
		Muhamad Ryansyah M.Kom
Kedua	:	Mempunyai tugas sbb:
		Melaksanakan Tugas yang diberikan dengan penuh rasa tanggung jawab.
Ketiga	:	Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila dikemudian hari terdapata kekeliruan akan diubah dan diperbaiki sebagaimana mestinya.

Jakarta,1 September 2020
Ketua PPPM
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri

Norma Yunita, M.Kom

Tembusan

- Ketua Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
- Arsip
- Ybs

www.nusamandiri.ac.id

PAPER • OPEN ACCESS

Comparison of Text Mining Classification Algorithms in Interbank Money Transfer Application

View the article online for updates and enhancements.

Comparison of Text Mining Classification Algorithms in Interbank Money Transfer Application

Siti Masripah[1], Lila Dini Utami[2], Hilda Amalia[3], Dini Nurlaela [4], Muhamad Ryansyah[5], Lestari Yusuf[6]

[1,2]Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika,Bogor, Indonesia

[3,4]Sistem Informasi, Universitas Bina Sarana Informatika,Jakarta, Indonesia

[5,6]Sistem Informasi, Sekolah Tinggi Manajemen dan Ilmu Komputer,Jakarta, Indonesia

E-mail: siti.stm@bsi.ac.id

Abstract. Funds transfer is a series of orders from the sender whose purpose is to move money from the sender to the recipient. The high interbank transaction fees imposed on each bank makes people use an interbank money transfer application, interbank money transfer transactions such as the Flip application are much in demand by the public because there are no administrative fees imposed on users. Opinion of the users of the application is processed using a text mining classification algorithm, namely the Naïve Bayes Algorithm and k-NN, the two algorithms are compared to produce which algorithm has high accuracy in processing the opinion of the flip money transfer application. Based on this matter, researchers conducted a sentiment analysis of the Flip Application,K-Nearest Neighbor (k-NN). After conducting research on sentiment analysis of Flip Applications, the Naïve Bayes classification algorithm has an accuracy of 91.25% and an ROC curve with an AUC value of 0.500. Whereas K-Nearest Neighbor has an accuracy of 85.25% and an ROC curve with an AUC value of 0.937. The Naïve Bayes algorithm can be said to be "good classification" and the public can make the decision to use the Flip Application.

1. Introduction

Funds transfer is a series of orders from the sender whose purpose is to move money from the sender to the recipient [1]. There are many ways to make money transfers [2], either using a Bank or an Indonesian legal entity business entity not a Bank [1]. The high cost of interbank transactions imposed on each bank makes people use an interbank money transfer application, interbank money transfer transactions such as the Flip application are much in demand by the public because there are no administrative fees charged to users. An application of money transfers can be said to be good if the recommendations or opinions of the public are positive. Opinions are used to ensure that services work as expected. This is known as sentiment analysis of an analyst sentiment product or also called opinion mining [3]. User's opinions on an e-commerce website have a significant impact on product reputation [4]. To extract the opinions from a digital product, a text mining method is needed. The ability to mine and analyze data to find new knowledge based on feedback provided by customers will increase the benefits obtained[5]. Analyzing online customer opinion becomes very profitable in business, through analyzing the company's sentiment can make an assessment or measurement of performance[6]. For this reason, a technique is needed to analyze and extract data from text collections provided

1

by the user. One technique for analyzing sentiment is to use text mining. Text mining is a very useful technique for gaining knowledge based on text[7]. Sentiment analysis techniques have been carried out with several data mining methods, one of which is a popular method in text mining, with the text classification algorithm, namely KNN[8]. Another popular method of data mining used for feedback analysis is the Naïve Bayes classification technique, Naïve Bayes can classify feedback accurately[8]. The advantages of the two text mining techniques above can be used to mine text to review the use of the Flip Application. Text mining is used to polarize comments or opinions from users into a number of positive or negative conditions. This mining activity is carried out to determine the reliability of a digital product or application. In this research data mining will be carried out in the form of opinions from Flip Application users. Processing opinions or reviews both positive and negative using sentiment analysis with the Naïve Bayes classification algorithm and the K-NN algorithm, the two algorithms will be compared so that it will show which algorithm is more appropriate to use sentiment analysis of the Flip Application

2. Method
To make this research, the research stages needed are outlined in figure 1

Figure 1. Research Methode.

2.1. Dataset
This research uses review data from the Flip financial application. Only 400 fold reviews were used, consisting of 200 positive review data and 200 negative review data. The data is still in the form of separate texts in the form of separate documents. Positive review data are put together in one folder and given a positive name, while negative review data are put together in one folder and given a negative name.

2.2. Pre-Processing
Pre-processing done include: a. Tokenization In this tokenize process, all the words in each document are collected and the punctuation is removed and removed if there are symbols or anything that is not a letter. b. Stopword Filtering The stopword filtering process is the step of clearing texts of irrelevant words used as an index[10]. c. Stemming Porter The stemming porter algorithm is the most widely used for English stemming[11]. Porter's stemming algorithm looks for basic words by eliminating affixes, the resulting word is sometimes ambiguous or confusing due to inconsistent Indonesian morphological rules[12]

Naive Bayes The Naive Bayes algorithm is a popular method for categorizing text with word frequency as a feature [13]. Naive Bayes is the best algorithm for classification used to classify data based on probability. Naïve Bayes also known as Maximum a Posterior Naïve Bayes has various advantages and disadvantages in various domains and algorithms that are fast and highly scalable[14]

K-Nearest Neighbor The k-nearest neighbor algorithm (k-NN or KNN) is a method for classifying objects based on learning data that is the closest distance to the object[15]. This research was conducted in several stages. The following are some steps in creating a sentiment analysis application using the K - Nearest Neighbor algorithm.

The determinant of classification based on basic examples that do not construct explicit, declarative representations of categories, but relies on category labels attached to training documents similar to test documents is the K-Nearest Neighbor (KNN) algorithm. Application finds k-nearest neighbors between training documents on test documents. The similarity score for each neighboring document nearest the test document is used as the weight of the neighboring document class[16] [17]. Equations for calculating neighboring levels of similarity between two objects are (Nurdiana, Jumadi and Nusantika, 2016).

3. Results and Discussiont
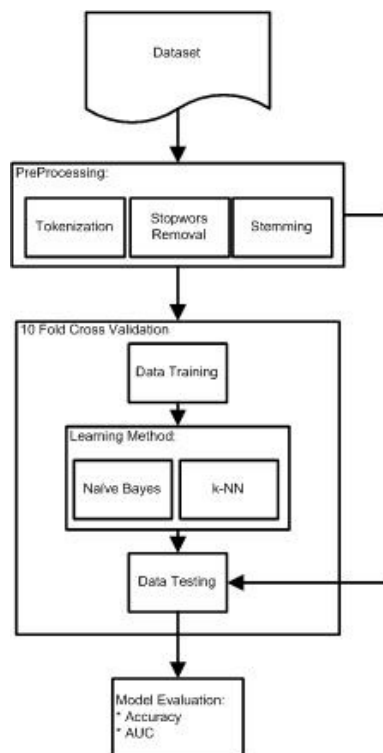Design model for this research outline in figure 2.

Figure 2. Research Methode.

a. Data collection The author collects sample data from Flip App reviews on the Play Store. Reviews consist of 200 negative reviews and 200 positive reviews in Indonesian.

b. Data processing Preprocessing consists of:

1) Tokenization

Table 1. Results of the Tokenization Process

Review	Tokenization
I just tried it, I was very interested, because the work of my country children gave me a 5. star. Please help the transaction process be improved somewhat accelerated .. Everywhere customers must ask for a fast transaction. This is a bit longer than my other T4 subscriptions ... I'm interested ... Just have to be patient a little, .	I'm just trying to be really interested. Because of the work of my children, I love the stars. Please help the transaction process be improved somewhat. Where will the customer ask for a fast transaction? It's a little longer than my other customers.

2) Stopwords

Table 2. Results of the Stopwords Process

Review	Stopwords
I just tried it, I was very interested, because the work of my country children gave me a 5. star. Please help the transaction process be improved somewhat accelerated .. Everywhere customers must ask for a fast transaction. This is a bit longer than my other T4 subscriptions ... I'm interested ... Just have to be patient a little, .	I just tried to be really interested because the work of my children, I love stars, please help the transaction process be improved somewhat, which is where the customers will ask for a fast transaction, this takes a long time compared to my other customers, I'm just interested in being patient a little.

3) Stemming

Table 3. Results of the Stemming Process

Review	Stemming
I just tried it, I was very interested, because the work of my country children gave me a 5. star. Please help the transaction process be improved somewhat accelerated .. Everywhere customers must ask for a fast transaction. This is a bit longer than my other T4 subscriptions ... I'm interested ... Just have to be patient a little,.	I just tried to be really interested because the work of my children, I love stars, please help the transaction process be improved somewhat, which is where the customers will ask for a fast transaction, this takes a long time compared to my other customers, I'm just interested in being patient a little.

Next step after pre-processor is applying the datas to algorithms.

1) Naïve Bayes algorithm From 200 positive reviews and 200 negative reviews, in Table 4 we found that 188 positive prediction reviews were positive, whereas 12 positive predicted reviews were negative. While 23 positive prediction reviews turned negative and 177 negative predicted predictions turned negative. The results of the accuracy of the application of the Rapid Miner using the Naïve Bayes algorithm is 91.25% and the ROC curve with an AUC value of 0.500

Table 4. Confusion Matrix Naïve Bayes Algorithm.

	True Positif	True Negatif	Class Predicition
Pred. Positif	188	23	89,10%
Pred. negative	12	177	93,65%
Class recall	94.00%	88,50%	

Value of accuracy method by Naïve Bayes : 91.25% +/- 3.58% (micro average: 91.25%). The ROC performance curve results from the data test results can be seen in Figure 3 below: AUC: 0.500 +/- 0,000 (micro average: 0.500) (positive class: negative)

2) K-Nearest Neighbor (k-NN) Algorithm From 200 positive reviews and 200 negative reviews, in Table 5 we found that 174 reviews predicted to be positive turned out to be positive, while 26
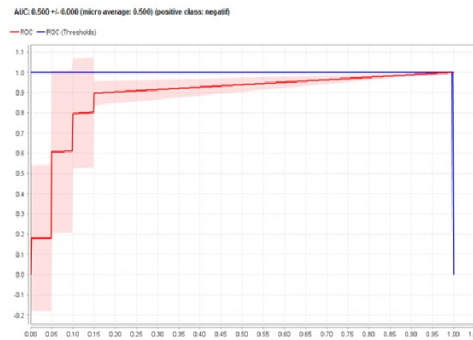
Figure 3. The ROC curve of the
Naïve Bayes algorithm.

reviews predicted to be negative turned out to be positive. While 33 reviews were predicted to be positive and negative were 167 reviews that were negative were negative. The results of the accuracy of the application of the Rapid Miner using the K-Nearest Neighbor Algorithm that is 85.25% and the ROC curve with an AUC value of 0.937.

Table 5. Confusion Matrix K-NN Algorithm.

	True Positif	True Negatif	Class Predicition
Pred. Positif	174	33	84,06%
Pred. negative	26	167	86.53%
Class recall	87.00%	83,50%	

The ROC performance curve results from the test data can be seen in Figure 4 below: AUC: 0.937 +/- 0.041 (micro average: 0.937) (positive class: negative). Naïve Bayes algorithm is
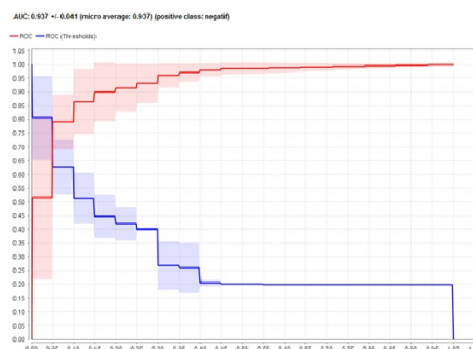

Figure 4. The ROC curve of theK-NN algorithm.

known to have good performance and high accuracy results. It turns out that in the Flip

Application assessment, the Naïve Bayes algotirma still has good performance, the Naïve Bayes algorithm produces an accuracy of 91.25% and an ROC curve with an AUC value of 0.500. Whereas K-Nearest Neighbor produces 85.25% accuracy and ROC curve with AUC value 0.937.

4. Conclusion

The conclusion of processing text mining data on a flip financial application by comparing or comparing the Naïve Bayes algorithm and the k-NN algorithm produces information that the Naïve Bayes algorithm in this research has a higher accuracy value than the K-Nearest Neighbor (k-NN) algorithm.), where the Naïve Bayes algorithm produces an accuracy of 91.25% while the K-Nearest Neighbor produces an accuracy of 85.25% with a difference of 6%. An accuracy can be said to be "good classification" if the accuracy value is in the range 0.80-0.90 so that the Naïve Bayes algorithm can be said to be "good classification".

5. References

[1] Divisi Informasi Hukum dan Direktorat Hukum, 2012 PERATURAN BANK INDONESIA NOMOR 14/ 23 /PBI/2012 TENTANG TRANSFER DANA 53, Desember Jakarta: Bank Indonesia.
[2] Committee on Payment and Settlement system, 2007 General principles for international remittance services January Basel, Switzerland: Bank for International Settlements.
[3] Redhu S, 2018 Sentiment Analysis Using Text Mining: A Review Int. J. Data Sci. Technol. 4, 2 p. 49.
[4] Jack L and Tsai Y D, 2015 Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues Int. Conf. Data Min. DMIN'15 July p. 92–98.
[5] Gallagher C Furey E and Curran K, 2019 The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying Int. J. Data Warehous. Min. 15, 4 p. 21–47.
[6] Al-Saqqa S Al-Naymat G and Awajan A, 2018 A large-scale sentiment data classification for online reviews under apache spark Procedia Comput. Sci. 141 p. 183–189.
[7] Ramesh B and Sathiaseelan J G R, 2015 An Advanced Multi Class Instance Selection based Support Vector Machine for Text Classification Procedia Comput. Sci. 57 p. 1124–1130.
[8] Trstenjak B Mikac S and Donko D, 2014 KNN with TF-IDF based framework for text categorization Procedia Eng. 69 p. 1356–1364.
[9] Maitra S Madan S Kandwal R and Mahajan P, 2018 Mining authentic student feedback for faculty using Naïve Bayes classifier Procedia Comput. Sci. 132 p. 1171–1183.
[10] Rahutomo F and Ririd A R T H, 2019 Evaluasi daftar stopword bahasa indonesia 6, 1 p. st.
[11] Gupta V and Lehal G S, 2013 A Survey of Common Stemming Techniques and Existing Stemmers for Indian Languages 5, 2 p. 157–161.
[12] Verdaningroem N J M and Saifudin A, 2018 UNTUK MENGURANGI KESALAHAN STEMMING BAHASA 10, 2 p. 103–112.
[13] Chandrasekar P and Qian K, 2016 The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier 2016 IEEE 40th Annu. Comput. Softw. Appl. Conf. p. 618–619.
[14] Ahuja R Chug A Kohli S Gupta S and Ahuja P, 2019 ScienceDirect ScienceDirect The Impact of Features Extraction on the Sentiment Analysis Procedia Comput. Sci. 152 p. 341–348.
[15] Yu X G and Yu X P, 2007 Locally adaptive text classification based k-nearest neighbors 2007 Int. Conf. Wirel. Commun. Netw. Mob. Comput. WiCOM 2007 p. 5646–5649.
[16] Sahara S and Wahyudi M, 2015 K-Nearest Neighbors Sebagai Analisis Konf. Nas. Ilmu Pengetah. dan Teknol. 2015 p. 82–87.
[17] Tan S and Zhang J, 2008 An empirical study of sentiment analysis for chinese documents Expert Syst. Appl. 34, 4 p. 2622–2629.