

A Naïve Bayes Sentiment Analysis for Fintech Mobile Application User Review in Indonesia



Rein Rachman Putra¹, Monika Evelin Johan², Emil Robert Kaburuan³

¹Information System Management, Binus University, Indonesia, rein.putra@binus.ac.id

²Information System Management, Binus University, Indonesia, monika.johan@binus.ac.id

³Information System Management, Binus University, Indonesia, emil.kaburuan@binus.ac.id

ABSTRACT

The growth of Fintech industries in Indonesia, as the 5th most internet users in the world, is tremendous, with a predicted of 16,3% year on year user growth and a total of \$176,75 Million USD in investment on Fintech Startup it has become one of the biggest potential market in the world. Therefore, with this fast-growing market, Fintech Companies need to know their user opinion in Realtime in order to face their competitor on the market. Existing user review on Application Store already existed but there are no classification based on the subject of the review, because it only focus on the application not on the business level, therefore there are in need of a sentiment analysis to classify a user review based on the subject, in this paper we do a sentiment analysis using Naïve Bayes algorithm in Bahasa Indonesia and in English, while also seeing the significance of data cleansing on the data, the resulted accuracy of the test in this paper between the analysis is 78% for English and Bahasa, and 75% for Bahasa Only.

Key words- Fintech, mobile application, Naïve Bayes, machine learning, sentiment analysis, text mining.

1. INTRODUCTION

Fintech is a new technology that aims to compete to traditional Financial services market, with a year on year growth of 16,3% and as predicted by Statista it will keep increasing by 42% until 2023 [1] is one of the most potential startup model in Indonesia, in just 3 years there are already 113 registered Fintech company in Indonesia [2] with now coming from various investor, even conventional Financial Service Industries companies also joins [2]. With most Fintech user dominated by mobile, the companies also needs to invest heavily in application development, therefore user review is also important not only on how they develop the application but also on how well do they perform on the market. There are already user review tools provided by Platform provider, it gives rating based on user opinion whether it is a positive or negative review, however the user review tools did not make a

distinction based on the subject of the review whether is intended for the Application, or for the business model of the companies. Machine learning was introduced in 1959 [3] as a way for a computer system to perform a task without using an explicit commands because it able to adapt based on the data provided by the user, it uses math and statistics in order to perform the tasks [4]. Machine Learning has been tried for the usage of sentiment review before [5], the analysis compared 3 classification Model Naïve Bayes, Decision Tree, and Support Vector Model [5], however there is not yet a study case for a Fintech application in Indonesia, also previously the research is conducted using English, in Indonesia there is a need to do a classification based on Bahasa, and also in this research we will do a comparison between un processed and preprocessed data to see the significance of the process.

2. LITERATURE STUDY

2.1. Fintech (Financial Technology)

Financial technology, often shortened to fintech, is the technology and innovation that aims to compete with traditional financial methods in the delivery of financial services [6]. Using smartphones for mobile banking, investment services, and cryptocurrency are examples of technologies aimed at making financial services more accessible to the public. Companies engaged in the field of fintech consist of startup companies and also existing financial companies, which are trying to replace or improve their financial services. Many want to manage solutions and technology to improve and develop their services, as well as improve their competitiveness. Financial Services Authority or OJK in Indonesia categorized Fintech into 4 categories:

- Payment, clearing, and settlement. This is fintech, which provides payment system services, both operated by the banking industry and those carried out by Bank Indonesia.
- E-aggregator. This Fintech collects and processes data that can be used by consumers to help decision making, by providing product comparisons ranging from price, features, to benefits.
- Risk and investment management. Fintech provides services such as advisors (software that

provides financial planning services and e-trading and e-insurance platforms.

- Peer to peer lending (P2P). Fintech brings together lenders (investors) with loan seekers in one platform.

2.2. Machine Learning

Machine learning (ML) is the study of algorithms and statistical models that are used in a computer systems to perform a specific task without having to use a specific instruction for it in order to perform [3], it is a part of Artificial Intelligence technology. The Machine learning built mathematical model called "training data", in order to make predictions or decisions without having a specific command to do it [7]. Machine learning algorithms are used in various scenario and technology such as email filtering and computer vision, where manual work is not possible.

Machine learning is a part of computational statistics, it focuses on probability calculation using computers. It also related to the study of mathematical optimization, where it contributes to the methods, theory and application domains to the machine learning theory. Data mining is a field of study within machine learning and focuses on exploratory data analysis through unsupervised learning [8].

2.3. Previous Research

A review on text based emotion recognition has been done previously by Priyanka Thakur, Dr. Rajiv Shrivastava shows that Text Based Review contribute greatly for a company in making a business decision [9]. In this study, methods of text recognition for extracting text are compared, it also shows that English text Emotion has a very good future. By using SVM, . Hidden Markov Model (HMM) and K Nearest Neighbor (KNN), his study compare all three model to determine each of the model accuracy

Sentiment Analysis review for online product review has been previously conducted by Oumayma Oueslati, Ahmed Ibrahim S. Khalil, Habib Ounelli, the study shows that SVM Algorithm achieved 97.95%. accuracy [10] it also emphasized on the negative review of the customer is a decisive factor in helping to determine whether the review is negative or positive compared to the statistical approach.

A study by Ece Calikus was conducted using the Databox Machine Learning Framework, comparing the SVM algorithm, Naive Bayes, and Maximum Entropy for the classification of sentiment analysis from customer reviews of a mobile application [11]. The most important contribution of this research is to prove that the supervised-learning approach is successful and can be applied to sentiment analysis of mobile application reviews. Nearly 90% accuracy results prove that this approach can be useful for real world applications. In addition, this research experiment determines that the Naive Bayes Multinomial algorithm provides the best results in terms of accuracy and speed.

Previous research on sentiment classification has been done on Facebook Status Updates, comparing two machine learning algorithms SVM and Naive Bayes [12]. The results found that both Naive Bayes and SVM algorithms can achieve high accuracy to classify sentiments by combining different features. Although the user's Facebook status has unique characteristics when compared to other corpus such as news, blog reviews, etc., the results of the study show. That machine learning algorithm can classify these statuses with similar steps and processes. So, in this study, we will use a dataset from user reviews on the Google Play Store that is found to also contain unique characteristics and languages that are mixed between Indonesian and English.

3. METHODOLOGY

The methodology that is going to be used in this research is Text Mining Text mining process is similar but not identic data mining, the differences are the data mining tools are created for structured data, whereas for text mining it is able to process unstructured data, [13] this process is explained in Figure 1.

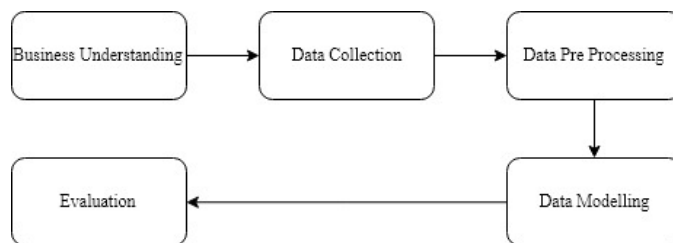


Figure 1: Research Framework

3.1. Business Understanding

Business Understanding is learned by understanding the fintech industries in Indonesia landscape, and by understanding the importance of user review on the application [14].

3.2. Data Collection

The data collection is performed on Google Android Play Store, using Python programming language, there are 2 sets of data, one is by mixed review the other one is in Bahasa only, below is the table to explain the data collected

3.3. Data Pre-Processing

In the preprocessing stage there are two steps used to prepare the collected text data in order to be modelled.

a) Data Labeling

Data labeling is process of labeling a sets of data to define the meaning of the data, in this Research the label that we use are "Application" to label the review that talks more about the application and "Company" to label the review that talks more about the company in general.

b) Data Cleansing

Data cleansing is the process of correcting records or data from a record by replacing, modifying, or deleting the data but also by preserving the meaning of the data presented. In This research we are doing stop word Removal and stemming for data cleansing.

3.4. Data Modeling

Data modeling is the process of creating a data model for the purpose of Information processing using a certain technique, as already performed in a previous research regarding a sentiment analysis Indonesian language based movie review [15] we are using a same Data modeling technique Naïve Bayes.

Making the simplifying assumption that the values of f_i for $i = 1...m$ are conditionally independent given the correct classification of d [15]. Bayes' rule states:

$$P(s|d) = \frac{P(s)P(d|s)}{P(d)} \quad (1)$$

Using equation (1) conditional independence assumption, naive Bayes classification makes the following approximation:

$$P(d|s) \approx \prod_{i=1}^m P(f_i|s) \quad (2)$$

A slight variation is multinomial Naive Bayes, where different occurrences of the same word in a document are treated as separate events, e.g. [15].

$$P(d|s) \approx \prod_{i=1}^m P(f_i|s)^{n_i^{(d)}} \quad (3)$$

The $P(f_i|s)$ probabilities in equation (3) are obtained by counting the frequency of occurrences of the unigrams within a document. Although a very simple approach, with assumptions that disregard word order, syntax, and discourse coherence, naïve Bayes has been shown to be effective in similar applications such as text classification [15].

3.5. Evaluation

Model Evaluation is an important part in determining a suitable model in analyzing a data. In this research we are going to use confusion matrix as method to evaluate Naïve Bayes as our model.

a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm,

there are Four Basic Terms in Confusion Matrix

a) true positives (TP)

Predicted Positive Outcomes and it is true

a) true negatives (TN)

Predicted Negative Outcomes and it is True

b) false positives (FP)

Predicted Positive Outcomes but it is False

c) false negatives (FN)

Predicted Negative Outcomes but it is False

Whereas from the information above [16] we can derive the following Equation where (P) is a positive condition and (N) is a negative condition:

$$Recall = \frac{TP}{P} = \frac{TP}{TP - FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1\ Score = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$$

4. RESULT AND DISCUSSION

4.1. Data Collection

The fintech mobile application used in this study is a digital / online loan application, where users can make loans or make payments at several merchants that already have an agreement, without the need to have a bank account and without a card. Users only need to register using a resident identity card (KTP). In this study, data were obtained from user reviews of Android-based applications on the Google Play Store. Based on our observations, reviews written by users are not only positive or negative, but also have subject whether the review is for the application or the company as the provider / application developer.

Example of a review in Bahasa on the company:

“untuk persyaratan rekening koran dan pajak tidak punya persyaratan nya cukup ribet; sedangkan express kredit hanya 30 hari.”

Example of a review in Bahasa on the application

“Proses upload selfienya kenapa lama sekali ya? ngulang berkali-kali gagal terus. padahal sudah sesuai dengan permintaan.”

The data used in this study are 10,000 user reviews of an Indonesian fintech Android application obtained by doing data scrapping using the Python programming language on the Google Play Store. The data is then stored in a Table format for further processing

4.2. Data Pre-Processing

Before processing the algorithm in machine learning, data preprocessing needs to be done. The first step is to do data labeling. Each review is labeled "Company" or "Application" based on the subject of the review. Labeling is also carried out to mark reviews in English for later processing separately between original data (Indonesian + English) and data that has been cleared from an English-language review. The result is now it has two datasets as follows:

Table 1: Datasets

Dataset A	Bahasa Indonesia + English
Dataset B	Bahasa Indonesia

As explained in Table 1 The data cleansing process carried out is stop word removal, stemming, and tokenization. Previously, the entire sentence was changed to the lower case so as not to distinguish the meanings for the same words with capital letters. After that also cleaning data from punctuation such as <"',>. Stop removal is done to delete or eliminate content / words that do not give importance to the dataset, such as pronouns, prepositions, conjunctions, etc. [12]. Then stemming is done, which is to change words into their basic words, by removing prefixes and / or suffixes so that they become basic words [12]. After that, tokenization is done. The tokenization is in a unigram pattern. The whole process is done in the Python programming language. We use a library for preprocessing Indonesian language data, namely Literature. Data that has been processed is processed into a dataset for the next stage stored in form of a table.

4.3. Data Modelling

Data modeling was performed using the Naïve Bayes algorithm. First, we process data with a dataset that is not cleansed, i.e. data that is only labeled without a stop word removal and stemming process. Then, we process the dataset that has been cleansed, stop word removal and stemming. This was done on the two datasets namely Dataset A and Dataset B in.

4.3.1 Dataset A

Table 2: Dataset A

Data Type	Data Number	Counter	
		Application	Company
Total	10000	6424	3576
Training	8000	5145	2855
Testing	2000	1279	721

4.3.2 Dataset B

Table 3: Dataset B

Data Type	Data Number	Counter	
		Application	Company
Total	6346	3321	3025
Training	5076	2673	2403
Testing	1270	648	622

4.4. Evaluation

To measure the accuracy of the Naïve Bayes model we use ini 2 dataset in Table 2 and Table 3 we use confusion matrix to measure the results:

4.4.1 Dataset A

4.4.1.1 Without Data Cleansing

Table 4: Result of Dataset A without Data Cleansing

	Application	Company
Application	TP - 1117	FP – 162
Company	FN - 278	TN – 443

	precision	recall	f1-score	support
Application	0.80	0.87	0.84	1279
Company	0.73	0.61	0.67	721
accuracy			0.78	2000
macro avg	0.77	0.74	0.75	2000
weighted avg	0.78	0.78	0.78	2000

4.4.1.2 With Data Cleansing

Table 5: Result of Dataset A with Data Cleansing

	Application	Company
Application	TP - 1125	FP – 154
Company	FN - 282	TN – 439

2.

	precision	recall	f1-score	support
Application	0.80	0.8	0.84	1279
Company	0.74	0.6	0.67	721
accuracy			0.78	2000
macro avg	0.77	0.7	0.75	2000
weighted avg	0.78	0.7	0.78	2000

4.4.2 Dataset B

4.4.2.1 Without Data Cleansing

Table 6: Result of Dataset B without Data Cleansing

	Application	Company
Application	TP - 453	FP - 195
Company	FN - 127	TN - 495

	precision	recall	f1-score	support
Application	0.78	0.70	0.74	648
Company	0.72	0.80	0.75	622
accuracy			0.75	1270
macro avg	0.75	0.75	0.75	1270
weighted avg	0.75	0.75	0.75	1270

4.4.2.2 With Data Cleansing

Table 7: Result of Dataset B with Data Cleansing

	Application	Company
Application	TP – 440	FP – 208
Company	FN - 126	TN – 496

	precision	recall	f1-score	support
Application	0.78	0.68	0.72	648
Company	0.70	0.80	0.75	622
accuracy			0.74	1270
macro avg	0.74	0.74	0.74	1270
weighted avg	0.74	0.74	0.74	1270

5. CONCLUSION

Based on the research conducted in 10,000 User Review with two different data set with uses Dataset A in Table 4 and Table 5 and Dataset B in Table 6 and 7 which sets using Naïve Bayes here are the accuracy results of each test conducted:

Table 8: Summary based on the Accuracy

Dataset	Data Cleansing	Accuracy
A	Yes	78%
A	No	78%
B	Yes	74%
B	No	75%

With the accuracy percentage of Naïve Bayes Model with dataset A and B showed in Table 8 used in this research remains the same around 78% for Dataset A and 74% for Dataset B, it can be concluded that data Cleansing process on Fintech Application user review in Indonesia with does not contribute significantly in the analysis, however we found that there is around 3% margin of accuracy with Bahasa Indonesia and English compared to Bahasa Only, it suggested that Naïve Bayes model is also better used in English language review rather than Bahasa.

Further research regarding this Paper can also be conducted with different Model, one of the could be using Support Vector Machine or Maximum Entropy while also focusing the research more on the language difference impact on a sentiment analysis research.

REFERENCES

1. Statista, "Fintech In Indonesia," Statista, 2019.
2. Otoritas Jasa Keuangan, **Penyelenggara Fintech**

Terdaftar dan Berizin di OJK per 31 Mei 2019, Otoritas Jasa Keuangan, May 2019. [Online]. Available: <https://www.ojk.go.id>.

3. IBM, **"Some Studies in Machine Learning Using the Game of Checkers"**, IBM, 1959.
4. C. M. Bishop, "Pattern Recognition and Machine Learning," Springer, 2006.
5. Z. Singla, S. Randhawa and S. Jain, "Sentiment analysis of customer product reviews using machine learning," *International Conference on Intelligent Computing and Control (I2C2)*, 2017. <https://doi.org/10.1109/I2C2.2017.8321910>
6. T. C. W. Lin, "Infinite Financial Intermediation," *50 Wake Forest Law Review* 643, 2015.
7. A. Samuel, "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming," *Artificial Intelligence in Design*, 1996.
8. J. H. Friedman, "Data Mining and Statistics: What's the connection?," *Computing Science and Statistics*.
9. D. R. S. Priyanka Thakur, "A Review on Text Based Emotion Recognition System," *International Journal of Advanced Trends in Computer Science and Engineerin*, vol. 7, pp. 67-71, 2018. <https://doi.org/10.30534/ijatcse/2018/01752018>
- 10 A. I. S. K. H. O. Oumayma Oueslati, "Sentiment Analysis for Helpful Reviews Prediction," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 7, pp. 34-40, 2018. <https://doi.org/10.30534/ijatcse/2018/02732018>
- 11 E. Calikus, "Mobile App Analytics & Sentiment Analysis of Customer Reviews," 2015.
- 12 J. D. Z. & P. M. J. L. H. Akaichi, "Text mining facebook status updates for sentiment classification," in *17th International conference on system theory, control and computing (ICSTCC)*, 2013.
- 13 S. I. M. J. & N. M. Vijayarani, "Preprocessing techniques for text mining-an overview," *International Journal of Computer Science & Communication Networks*, 2015.
- 14 Fintechnews Singapore, "Fintech Indonesia Report 2018 – The State of Play for Fintech Indonesia," 2018. [Online]. Available: <http://fintechnews.sg/20712/indonesia/fintech-indonesia-report-2018/>.
- 15 F. Ruli Manurung, "Machine Learning-based Sentiment Analysis of Automatic Indonesian," 2008.
- 16 D. M. W. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, 2011.