Jurnal Mantik, 5 (4) (2022) 2203-2209

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Sentiment Analysis of Online Investment Applications on Google Play Store using Random Forest Algorithm Method

Wulan Alinda Wahyuni[1], Sudin Saepudin[2], Falentino Sembiring[3]

[1.23]Program Studi Sistem Informasi,
[1.23]Universitas Nusaputra, Sukabumi Jl. Raya Cibatu Cisaat No.21, Cibolang Kaler, Kecamatan. Cisaat, Sukabumi Regency, Jawa Barat 43155, Telp (0266) 210594, Indonesia

E-mail: wulan.alinda_si18@nusaputra.ac.id[1], sudin.saepudin@nusaputra.ac.id[2], falentino.sembiring@nusaputra.ac.id[3]

A R T I C L E I N F O

Article history:

Received: Des 08, 2021
Revised: Jan 10, 2022
Accepted: Feb 30, 2022

Keywords:

Sentiment Analysis,
Investment,
Google Playstore,
Random Forest

A B S T R A C T

Online investment is an investment to get a long profit. Where this investment has a function to store funds long-term and in order to have a higher value than annual inflation. Currently online and offline investments are available in the market to reach the increasing interest of beginner investors. In addition, to open an online investment account can also now be very easy, fast and flexible. There are many online investment apps that can be downloaded through the Google Play Store. However, with the number of stock investment applications sometimes makes one have to choose which application is the best that can be used. Therefore, the author conducted a study to find out the best stock investment application to be recommended to investors who will start investing. The purpose of this study is to analyze the sentiment of online investment application reviews, namely stockbit, Hsb Investment, and Seedlings. In this study, the method to be used is the Random Forest method. Based on the analysis and testing conducted, the conclusion that can be drawn is the result of the implementation of random forest algorithms for Stockbit applications in this study resulted in an accuracy value of 62.50%, for seed applications in this study resulted in an accuracy value of 63.39%. And for HSB applications in this study resulted in an accuracy value of 96.25%.

1. Introduction

As computerized innovation advances today, many individuals are creating a variety of applications that can be utilized by the general public to obtain or deliver data effectively and quickly. Aspects that affect the era of digital technology is one of them is the economic aspect. Where economic activities such as buying and selling, transacting, to investing are mostly done online. At this time many online investment applications that offer great convenience and benefits to its users. The application can be accessed on android smartphones and iphones by downloading on google play store or appstore. With the online investment application, many people began to make investments to raise funds and develop assets to meet all needs in the future.

There are many investment apps in the Google Play Store, but the online investment application so far there are advantages and disadvantages of each. For shortcomings such as late purchases take days, take a long time to access and it is difficult to withdraw funds. This problem is very influential on the satisfaction of users who should be a benchmark of service quality of the online investment application.

But to choose a trusted online investment application requires a collection of information obtained from the experience of previous application users. To collect this information, you must take a review data on an online investment application in the Google Play Store. In order to be able to know the reviews of application

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

users and categorize the reviews are negative or positive, which will later get accuracy and compared values from each of these online investment applications.

The previous study conducted a sentiment analysis of seedling and bareksa applications with a total of 998 reviews, including 484 positive and 514 negative for Bareksa applications, while for Seed applications, using 1063 data sets including 541 positive and 522 negative. The data also goes through the stages of preprocessing and modeling. The study used the CRISPDM (Cross Industry Standard Process for Data Mining) model and the algorithm used in the study was Knearest Neighbors. Based on results obtained during the modeling phase using the Knearest Neighbors algorithm and a 60:40 ratio for training and test data, the recall and precision values generated for each application were 85.14%, 91.91% and 76.44% while with 81.70% bareksa, 87.15%, 75.73%.

Based on that background, the authors submitted the research topic "Sentiment Analysis of Online Investment Applications in the Google Play Store Using Random Forest Algorithm Method". In this study will be taken 3 samples of online investment applications, namely HSB Investment, Seedlings, and Stockbit. Through this research, the author hopes to be a recommendation to the public who will start investing online and can provide information to application owners in order to improve the quality of service of the online application.

2. Method

2.1 Research Flow Scheme

To complete a consistent study there are several stages to be used as a guideline or reference in this study, the author has created a frame of mind to complete a study including:

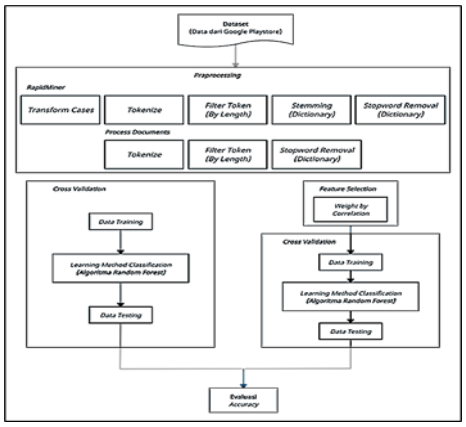


Fig 2. Frame of mind

2.2 Data Collection

The process of collecting data is done by scraping using Webharvy tools by collecting 239 comments from users of online investment applications on Google Playstore. Peeneliti also used google form as a questionnaire distributed to users of about 31 users. With the testing intervals include satisfied, dissatisfied, quite satisfied and very satisfied. In the study the authors also created a questionnaire form.

Here is a sample of data obtained through scraping data with webharvy:

Table 1.
Sample Data

No.	Text
1.	let's learn a very easy and safe investment, for the foreseeable future, immediately download the seed application.
2.	Ka, the seedlings are already good and ok, recomended really, but less of one in the telegram group that the seed admin is still a beginner, how to respond to investors less understanding and intelligent. Please fix it, brother.
3.	The comments of the seed application are not good, complicated, it's a big mistake, the seeds are ordinary luaar, which must have joined more profit than in a regular bank, I love 5 stars, hopefully more advanced again, please make the application entry feature pakek pin or paswoord for the application, so that it can not be opened people
4.	It's been a long time in diem in his savings, just want to join again. Precisely the application g can be opened for reasons the network does not exist. Even though my internet is covered. Please make it
5.	I had a problem, when I tried to try to make a transaction of 100,000 continued I resold the next day 98,000 and

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

No.	Text
	the money did not enter my account, from March 8 to now March 26. It's not possible to work that long.

Fig 3. Application user questionnaire

2.3 Text Processing

Preprocessing is a stage in the construction of a sentiment analysis system. At this stage the data is first prepared so that the next process such as classification is better when used. Preprocessing has several stages, in this research the stages carried out are case folding, cleaning, toke nization, stopword removal, stemming and, remove unknown word. . Case folding aims to convert all letters into lowercase letters[12], cleaning aims to remove punctuation marks or symbols[13], tokenization which aims to cut a k alimat by separating each word contained in the The sentence[12], stopword removal aims to eliminate words that are not important and in the document[12], stemming is the process of converting the word into a basic word[12], and remove unknown word It aims to eliminate words that have no implied meaning in the English language such as br.

Text preprocessing is the initial stage in a study of a collection of many texts that are prepared to become a data that will be processed further. A text cannot be processed directly by a search algorithm, therefore it takes preprocessing text to convert text into numeric data[14].

This process consists of several stages of document cleaning such as the following:

a. Tokenize

Tokenize is to remove punctuation marks and certain characters according to the number of text.

b. Stopword Removal or Filtering

Stopwords removal is a process of removing words that are considered unimportant and have no connection with research that is feared will affect the results of the study by checking the words of the parsing results.

c. Stemming

Stemming is the mapping or process of subtraction in the basic word. The purpose of the stemming process is to eliminate both in the form of prefixes, suffixes and confixes that exist in each word.[15]

3. Results and Analysis

For these results and discussions, the author will explain the process that was lived during the study. Here the author uses webharvy and repidminer tools in processing data using *the Random Forest algorithm method.* Here are some research processes:

3.1 Data Capture

In this study, the data obtained was a user review of several online investment applications in the Google Play store, namely HSB Investment, Seeds, and Stockbit

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

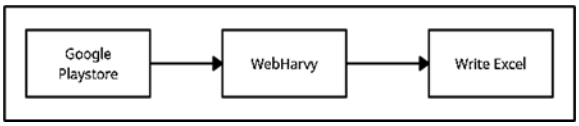


Fig 4. Data collection process

After getting the top three apps, the next step is to collect data using Webharvy tools. The total number of data obtained was 239 comments. Data consists of raw data collected in the form of tables in Microsoft Excel. As follows:

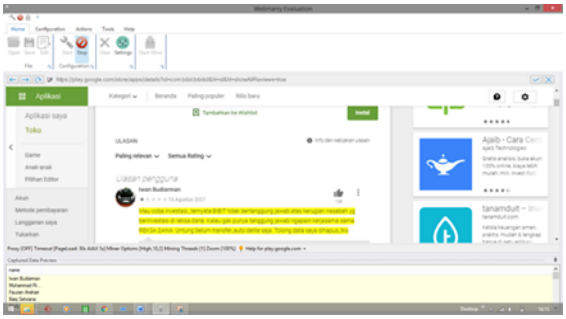


Fig 5. Scraping data with Webharvy

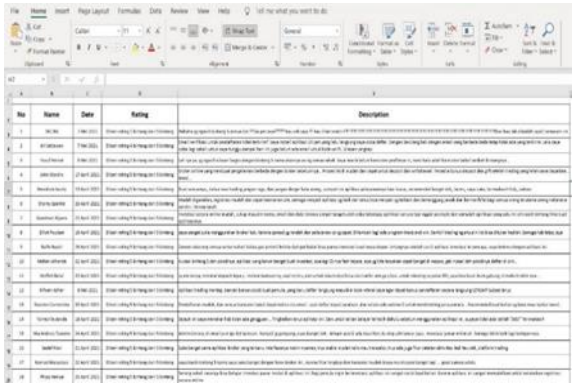


Fig 6. Excel data results

3.2 Pre-processing data

The pre-processing stage performed on the RapidMiner application :

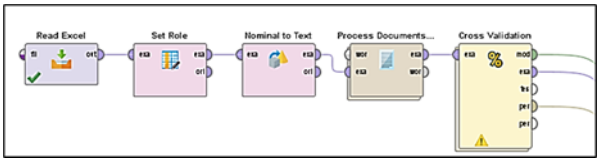


Fig 7. Pre-processing stage

a. Read Excel

In this process enter the excel data that has been obtained on WebHarvy. The data is user review data in the form of name, time, rating and comments.

Fig 8. Read Excel

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

b. Set Role

Acting as an attribute, this operator is used to change the role of one or more Attributes.

c. Nominal to Text

This operator then converts the selected nominal attribute type into text. It also maps all the values of these attributes to the corresponding string values.

d. Process Documents

This stage includes several advanced processes, including:

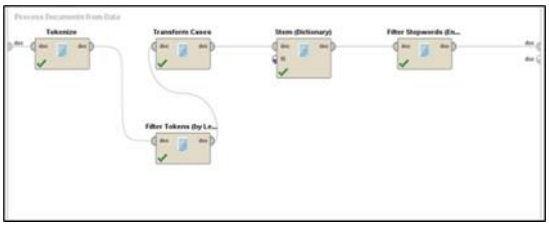


Fig 9. Text Processing

1) Tokenize

Continue the process by using the tokenize function in RapidMiner to eliminate such non-letter characters, such as symbols, spaces, numbers, and other characters. Although a written comment at all times includes a number in its first or final sentence to indicate that the sentence was repeated, so did a study, if you find that a word using an additional digit should remove it.

2) Case Folding / Transform Cases

The use of the "TF-IDF" vector to resume the pre-processing stage so that the data is ready for processing. Then the letter conversion is done with the aim of converting all the text into lowercase letters.

3) Stemming

The process continues by eliminating the appendings of each word, making it a basic word, and at this stage also attempts to eliminate misspelled words.

4) Clean One Character or More

The next process is to eliminate a word with a certain number of letters through the token filter function (by length) parameters min chars 4 and max chars 25, and the number of letters in the word is limited to a minimum of 4 and a maximum of 25 texts.

5) Remove Stopword

Words that are considered always appear or are not important, such as tense, connector, and so on. Then stopwords are processed in sentences. To do the process of removing the word needs to be donen the removal of a data or vocabulary.

3.3 Data Processing

At this stage, random forest algorithms are used to perform cross-validation processes that include training and testing processes to get accuracy values based on (random forest application performance and models).

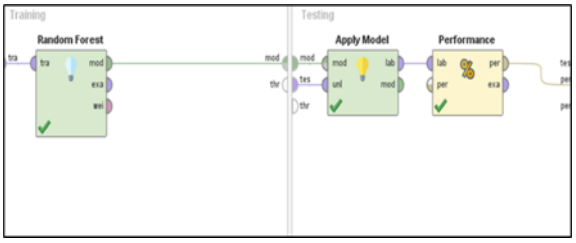
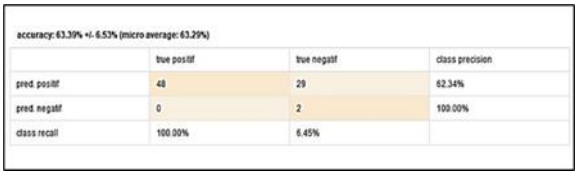


Fig 10. Cross validation process stage- random forest

After doing the process of training data using random forest algorithms. Then automatically the system reads all the data input, then with some calculations and formulas the system obtained the accuracy value of each application.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

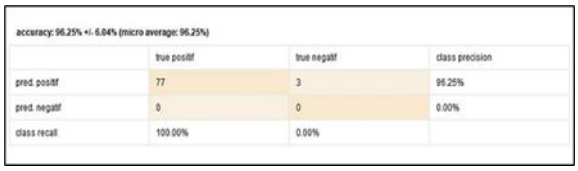


accuracy: 63.39% +/- 6.53% (micro average: 63.29%)

	true positif	true negatif	class precision
pred. positif	48	29	62.34%
pred. negatif	0	2	100.00%
class recall	100.00%	6.45%	

Fig 11. Seedling application accuracy value

The accuracy value obtained by the random forest algorithm of the Seed application is 63.39% with a margin of +/- 6.53% with a micro average value of 63.29%. Thus the maximum accuracy value is 63.39%. So, from the cross validation process of training and testing, the random forest algorithm for seedling applications produces an accuracy value of 63.39%.

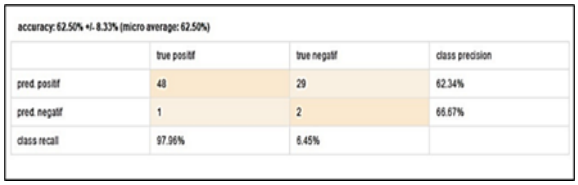


accuracy: 96.25% +/- 6.04% (micro average: 96.25%)

	true positif	true negatif	class precision
pred. positif	77	3	96.25%
pred. negatif	0	0	0.00%
class recall	100.00%	0.00%	

Fig 12. Accuracy value of HSB Investment application

The accuracy value obtained by the HSB Investment Random Forest algorithm is 96.25, the margin is +/- 6.04, and the micro-mean value is 96.25%. Thus, the maximum accuracy value is 96.25% and the minimum accuracy value is 96.25%. Therefore, from the training and testing process in HSB applications produce an accuracy value of 96.25%.



accuracy: 62.50% +/- 8.33% (micro average: 62.50%)

	true positif	true negatif	class precision
pred. positif	48	29	62.34%
pred. negatif	1	2	66.67%
class recall	97.96%	6.45%	

Fig 13. Stockbit app accuracy value

The random forest algorithm applied by Stockbit Trade New obtained a precision value of 62.50, a margin of +/- 8.33 and a micro average value of 62.50%. Therefore the maximum precision value is 62.50%, while for training and testing random forest algorithms Stockbit applications get a precision value of 62.50%.

3.4 Apikasi User Test

Fig 14. Percentage of Online Investment Application Usage Results

Based on the questionnaire that has been made, there is a conclusion in the form of application presentation results with the use of seed applications 70%, HSB 13.3%, and Stockbit 16.7%.

4. Conclusions

From the tests that have been done using the random forest algorithm method, the accuracy value of the Seed investment application is 63.39%, the HSB investment application is 96.25%, and the Stockbit Application is 62.50%. In the results of the questionnaire on the use of online investment applications, out of 30 respondents to the satisfaction of application users, respondents chose to use the Seed application to invest with a presentation value of 70%.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

References

[1] D. Adhi Putra, "Sentiment Analysis on User Reviews of Seedling And Bareksa Applications with KNN Algorithm," JATISI (Jurnal Tek. Inform. and Sist. Information), vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.

[2] F. N. Zamzami, A. Adiwijaya, and others, "Sentiment Analysis of Film Reviews Using Modified Balanced Random Forest methods and Mutual Information," J. MEDIA Inform. BUDIDARMA, vol. 5, no. 2, pp. 415–421, 2021.

[3] Ulfa, M.A., Irmawati, B., Husodo, A.Y. (2018). "Twitter Sentiment Analysis using Naïve Bayes Classifier with Mutual Information Feature Selection". Journal of Computer Science and Informatics Engineering. 2(2). 106-111

[4] E.M. Sipayung, H. Maharani, and I. Zefanya, "Designing Customer Comment Sentiment Analysis System Using Naive Bayes Classifier Method," J. Sist. Inf., vol. 8, no. 1, pp. 958–965, 2016, [Online]. Available: https://ejournal.unsri.ac.id/index.php/jsi/article/view/3250/1907

[5] Afriyeni and D. Marlius, "Analysis of Investment Returns and Risks (Study on Indonesia Stock Exchange-Listed Manufacturing Industry)," pp. 1–14, 2019, doi: 10.31219/osf.io/cfb92.

[6] E.B. Santoso, A. Nugroho, Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook, Jurnal Eksplora Informatika 9, 60-69 (2019)

[7] R. Feldman, J. Sanger, The Text Mining Handbook. New York: Cambridge University Press (2007) 9

[8] E3S Web of Conferences 202, 16006 (2020) https://doi.org/10.1051/e3sconf/202020216006 ICENIS 2020

[9] Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," Indones. J. Artif. Intell. Data Min., vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.

[10] L. Ratnawati and D. R. Sulistyaningrum, "Application of Random Forest to Measure The Severity of Disease in Apple Leaves," vol. 8, no. 2, 2019

[11] L. Breiman, Random Forests. Machine Learning 45, 5-32 (2001)

[12] L. Breiman, A. Cutler, Manual on Setting Up, Using, and U nderstanding RandomForest 4.0 (2003)

[13] Mubarok, M.S., Adiwijaya, Aldhi, M.D. (2017). "Aspect-based sentiment analysis to review products using Naïve Bayes". International Conference on Mathematics: Pure, Applied and Computation: Empowering Engineering using Mathematics. 1-8.

[14] Bakar, M.Y.A., Adiwijaya, Faraby, S.A. (2018). "Multi-Label Topic Classification of Hadith of Bukhari (Indonesian Language Translation)Using Information Gain and Backpropagation Neural Network". 2018 International Conference on Asian Language Processing (IALP). 344-350

[15] W. Gata, "Accuracy of Text Mining Using K-Nearest Neighbour Algorithm on SMS News Content Data,"vol. 6, pp. 1–13, 2017.

[16] Sari, Retno, and Ratih Yulia Hayuningtyas. "Application of Naive Bayes Algorithm for Sentiment Analysis on Website-Based TMII Tourism." Indonesian Journal on Software Engineering (IJSE) 5.2 (2019): 51-60.