



Author Notification
27 August 2019
Final Revised
28 August 2019
Published
03 September 2019

Analysis Of Warganet Comments On It Services In Mandiri Bank Using K-Nearest Neighbor (K-Nn) Algorithm Based On Itsm Criteria

Febrian Wahyu Ramadhan¹, Husni Teja Sukmana², Luh Kesuma Wardhani³, Kyung Oh Lee⁴

^{1,2,3}Informatics Engineering Study Program, Faculty of Science and Technology, Syarif Hidayatullah State Islamic University Jakarta, ⁴Dept. Computer Engineering, Sun Moon University, Asan Campus, South Korea

To cite this document:

Ramadhan, F., Sukmana, H., Oh, L., & Wardhani, L. (2019). ANALYSIS OF WARGANET COMMENTS ON IT SERVICES IN MANDIRI BANK USING K-NEAREST NEIGHBOR (K-NN) ALGORITHM BASED ON ITSM CRITERIA. *ADI Journal on Recent Innovation (AJRI)*, 1(1), 14-19. <https://doi.org/https://doi.org/10.34306/ajri.v1i1.9>

DOI:

<https://doi.org/10.34306/ajri.v1i1.9>

Abstract

Sentiment analysis is a method for reviewing products or services to determine opinions or feelings about a product. The results of the analysis can be used by companies as evaluation materials and considerations to improve the products or services provided. This study aims to test the level of public sentiment on the quality of Bank Mandiri services that have received ISO 20000-1 with the application of sentiment analysis using the K-NN algorithm based on ITSM criteria. The initial classification in this study uses the lexicon method by detecting words included in sentiment words, the results of which are included as labels on training data and test data. Formation of the classification with the K-NN algorithm by taking into account the results of the training data indexing and weighting of the test data, with the value of k as the decision maker limit. The trial results of 10 scenarios show that the classification using the K-NN algorithm as a sentiment classification is 98% accuracy value of 50 test data to 600 training data, with 24% getting positive sentiment, 22% negative sentiment and 55% neutral sentiment, with f - measure 95.83%. while in testing 100 the test data obtained 79% accuracy value with 21% getting positive sentiment, 42% negative sentiment and 38% neutral with an f-measure value of 68.42%.

Keywords: K-NN Algorithm, Sentiment Analysis, Bank Mandiri, Twitter.

I. INTRODUCTION

The use of the internet is now a part of human needs. Almost all devices are equipped with technology that allows users to surf online. Based on a report released by Tetra Pak Index (2017), there are around 132 million internet users in Indonesia, and 40% of them are social media users. Twitter is one of the most popular microblogging or social media service providers in Indonesia. Based on an article released by Jakarta Globe (2017), the number of Twitter accounts in the world in 2017

Reached 328 million. The number has increased by around 14% compared to 2016, where 24.34 million accounts from Indonesia put Indonesia in the third position of Twitter users.

The high use of social media in Indonesia is widely used by companies or agencies as a means to communicate with the community interactively so that it can directly get feedback from the community (Setyanto, 2016).

Bank Mandiri is one of the State-Owned Enterprises (SOEs) engaged in banking, which has implemented work principles in accordance with IT Service Management (ITSM) criteria, based on ISO 20000: 2011 on IT Service Management. As the role of a bank in general, namely as a data collector, distributor of funds and other bank service providers, Bank Mandiri must always prioritize customer satisfaction by building good corporate governance, one of which is the management of information system services according to criteria ITSM.

Sentiment analysis which is a computational research of opinions, sentiments and emotions is used in determining the sentiments of opinions expressed (Ira & Edi, 2017). The type of sentiment of the opinion can be positive, negative or neutral opinion (Liu, 2012).

The classification process in this sentiment analysis research uses the K-Nearest Neighbor algorithm. The advantage of K-NN algorithm is that it is more resilient to data that has a lot of noise and is more effective against large amounts of data (Alfian et al, 2014) in addition KNN can be used as a classification for various domains (Imandoust et al., 2013). The algorithm was chosen referring to previous research by (Rezwanul et al. 2017) about Sentiment Analysis on Twitter Data Using KNN and SVM. The results of these studies indicate better values based on accuracy, precision, and recall (Rezwanul et al., 2017).

Therefore, this research is important to do, to find out public sentiments towards the quality of Bank Mandiri services with the application of sentiment analysis using K-NN algorithm based on ITSM criteria. So that the results of the analysis can be used by companies as evaluation materials and considerations to improve the products or services provided.

II. METHODE

Method of collecting data

In this study, the data collection process was carried out by the literature study method. Literature study is to collect data, based on theories obtained from reference books, journals, similar thesis and related articles. In addition the author also accesses YouTube, and other sites to learn the concepts of sentiment analysis, natural language processing, lexicon classification and K-NN algorithm

Data Observation

Data observations were taken from the @mandiricare twitter account as a source of data relating to Bank Mandiri IT services through the @mandiricare account to be used as training and test data in the system analysis application. Retrieval of tweets is done from September to October 2018. Twitter data retrieval process uses a library created on the site <https://github.com/abraham/twitteroauthdemo/blob/master/templates/index.html/> using PHP programming language. Successful Twitter data can then be saved to the MySQL database. The data that was successfully obtained and used as a source of information in this study were 700 tweets, with the division of 100 tweets used as test data and 600 tweets used as training data.

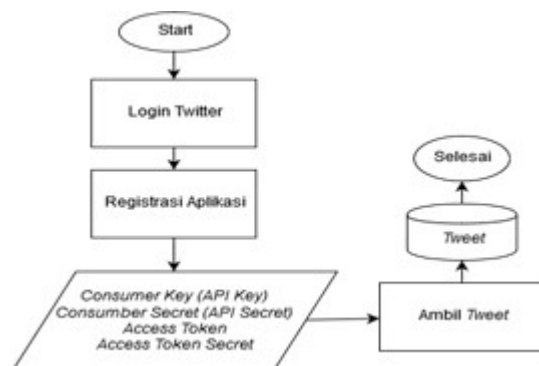


Figure 1. Twitter data retrieval steps

Simulation

Application development in this study uses a simulation method with the following steps:

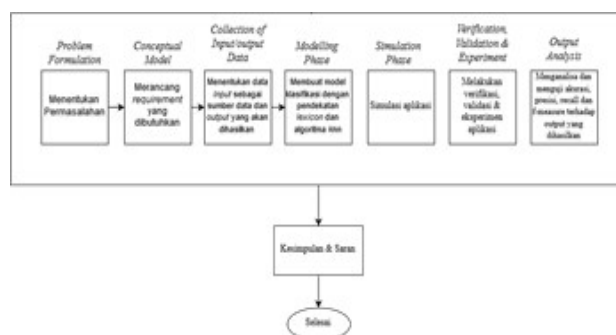


Figure 2. The workflow of the simulation method

Testing

Testing is done based on the value of k , namely $k = 1$ to $k = 10$. Things that were tested include: testing accuracy, precision, recall and f-measure.

III. RESULTS AND DISCUSSION

The results of sentiments with the K-NN algorithm, the authors use 50 test data and 100 test data obtained from crawling Twitter data after passing the lexicon preprocessing and classification process, in conducting testing, the writer uses 5 k values as a comparison of the accuracy of each k .

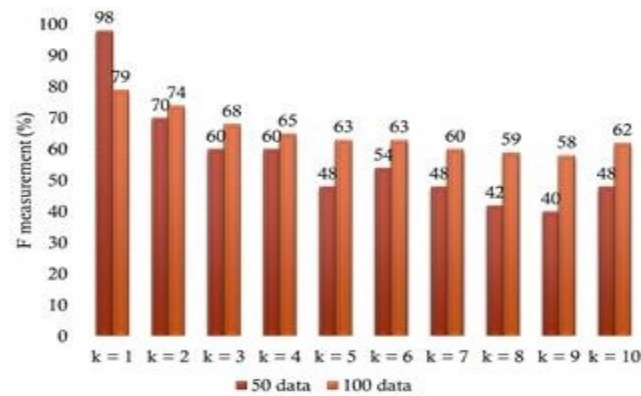


Figure 3. Comparison of K-NN Algorithm Accuracy with 50 test data Based

On figure 3, it is known that the highest accuracy value uses $k = 1$, both with the amount of test data 50 (f measurement = 98%), and the amount of test data 100 (fmeasurement = 79%). Meanwhile, the lowest accuracy value uses $k = 9$, both with the amount of test data 50 (f measurement = 98%), and the amount of test data 100 (f measurement = 79%). Accuracy results tend to decrease because the value of k increases so the greater the range in choosing the biggest weight.

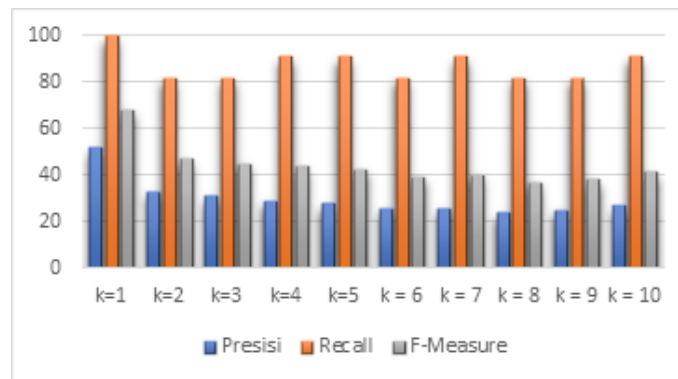


Figure 4. Comparison of Precision, Recall and F-Measure K-NN Algorithms with 100 data

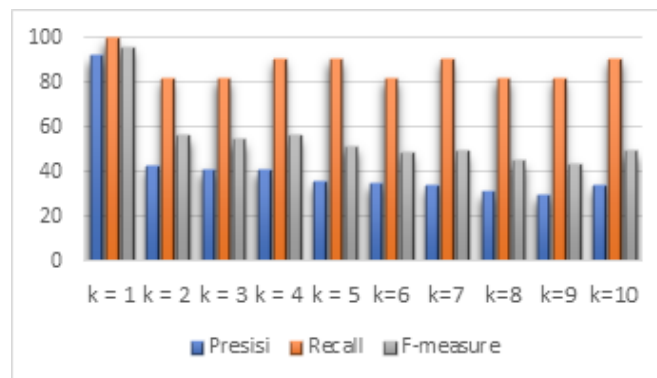


Figure 5. Comparison of Precision, Recall and F-Measure K-NN Algorithms with 50 data

IV. CONCLUSION

The results obtained with each optimal accuracy show positive sentiment results get lower than negative and neutral values, which shows that there are still many users who are still not satisfied with the services of Bank Mandiri. In accordance with the results of the sentiment classification obtained shows that the application of the ISO 20000-1 standard by Bank Mandiri has not yet received a positive response to the services provided by Bank Mandiri. The result of lower positive sentiment towards negative and neutral sentiment is due to the nature of users who are more likely to be quiet if they have received good service and will react to services that are less satisfying. The system is still not optimal in overcoming abbreviated words and / or words not found in the KBBI dictionary.

V. ACKNOWLEDGMENT

Suggestions for Bank Mandiri to improve their services, especially for services related to IT to comply with ISO 20000: 1 standards. As for further research, there should be an increase in the text- processing infrastructure to overcome short or excessive words. Add training data, test data and collection of sentiment words to get better accuracy. Data retrieval is better in the form of a questionnaire than using Twitter because it will be more objective in producing classification. In addition, other algorithms are expected to be used to compare accuracy and performance for better results.

REFERENCES

- [1] Asghar, Muhammad Zubair, Fazal Masud Kundi, Aurangzeb Khan, and Shakeel Ahmad. 2014. "Lexicon-Based Sentimen Analysis in the Social Web." *J. Basic. Appl. Sci. Res* 4(6): 238–48.
- [2] Bassil, Youssef. 2012. "A Comparative Study on the Performance of the Top DBMS Systems." *arXiv preprint arXiv:1205.2889*: 20–31. <http://arxiv.org/abs/1205.2889>.
- [3] Buntoro, Ghulam Asrofi, Teguh Bharata Adji, and Adhistya Erna Purnamasari. 2014. "Sentimen Analysis Twitter Dengan Kombinasi Lexicon Based Dan Double Propagation." *Citee 2014* ISSN: 2085(OCTOBER 2014): 7–8.
- [4] Dwiki, Dyar, and Adriadi Nur. 2015. "PENGEMBANGAN APLIKASI SENTIMEN ANALYSIS MENGGUNAKAN METODE NAÏVE BAYES. (Studi Kasus Sentimen Analysis Dari Media Twitter)." (November): 2–3.
- [5] H, Aris Tri Jaka. 2015. "Preprocessing Text Untuk Meminimalisir Kata Yang Tidak Berarti Dalam Proses Text Mining." *Jurnal Informatika UPGRIS* 1: 1–9.
- [6] Haddi, Emma, Xiaohui Liu, and Yong Shi. 2013. "The Role of Text Pre-Processing in Sentimen Analysis." *Procedia Computer Science* 17: 26–32. <http://dx.doi.org/10.1016/j.procs.2013.05.005>.
- [7] Imandoust, Sadegh Bafandeh, and Mohammad Bolandraftar. 2013. "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background." *International Journal of Engineering Research and Applications* 3(5): 605–10.
- [8] Indriani, Aida. 2014. "Klasifikasi Data Forum Dengan Menggunakan Metode Naïve Bayes Classifier." *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*: 5–10.
- [9] Karyono, Giat, Fandy Setyo Utomo, Arsitektur Sistem, and Temu Balik. 2012. "Temu Balik Informasi Pada Dokumen Teks Berbahasa Indonesia Dengan Metode Vector Space Retrieval Model." *Seminar Nasional Teknologi Informasi dan Terapan 2012* 2012(Semantik): 282–89.
- [10] Kataria, Aman, and M D Singh. 2013. "A Review of Data Classification Using K-Nearest Neighbour Algorithm." *International Journal of Emerging Technology and Advanced Engineering* 3(6): 354–60.

-
- [11] Mario, Hitoshi, Naga Mait, and Mirna Adriani. "Making Stemming Synonym Indonesian Using Algorithm Nazief and Adriani ."
 - [12] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. 2014. "Sentimen Analysis Algorithms and Applications: A Survey." *Ain Shams Engineering Journal* 5(4): 1093–1113. <http://dx.doi.org/10.1016/j.asej.2014.04.011>.
 - [13] Menggunakan, Sekolah, and Algoritma Porter. 2014. "STEMMING BAHASA INDONESIA SEBAGAI MEDIA BELAJAR SISWA."
 - [14] Mujilawati, Siti. 2016. "Pre-Processing Text Mining Pada Data Twitter." *Seminar Nasional Teknologi Informasi dan Komunikasi 2016*(Sentika): 2089–9815.
 - [15] Mustakim, Giantika O. 2016. "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa." *Jurnal Sains dan Teknologi Industri* 13(2): 195–202.
 - [16] Ogc. 2007. *ITIL Library The Official Introduction to the ITIL Service Lifecycle*.
 - [17] Prasath, V. B. Surya, Haneen Arafat Abu Alfeilat, Omar Lasassmeh, and Ahmad B. A. Hassanat. 2017. "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier - A Review." : 1–50. <http://arxiv.org/abs/1708.04321>.
 - [18] Raulji, Jaideepsinh K et al. 2016. "Stop-Word Removal Algorithm and Its Implementation for Sanskrit Language." *International Journal of Computer Applications* 150(2): 975–8887. <http://www.ijcaonline.org/archives/volume150/number2/raulji-2016-ijca-911462.pdf>.
 - [19] Rezwanul, Mohammad, Ahmad Ali, and Anika Rahman. 2017. "Sentimen Analysis on Twitter Data Using KNN and SVM." *International Journal of Advanced Computer Science and Applications* 8(6): 19–25. <http://thesai.org/Publications/ViewPaper?Volume=8&Issue=6&Code=ijacsa&SerialNo=3>.
 - [20] STROE, Ion-Sorin. 2011. "MySQL Databases as Part of the Online Business, Using a Platform Based on Linux." *Database System Journal* 11(3): 3–12.
 - [21] Tommasel, Antonela, and Daniela Godoy. 2018. "A Social-Aware Online Short-Text Feature Selection Technique for Social Media." *Information Fusion* 40: 1–17. <http://dx.doi.org/10.1016/j.inffus.2017.05.003>.
 - [22] Ugm, Fmipa, and Fmipa Ugm. 2014. "Analisis Sentimen Twitter Untuk Teks Berbahasa Indonesia Dengan Maximum Entropy Dan Support Vector Machine." 8(1): 91–100.
 - [23] Vijayarani, S., J. Ilamathi, and Ms. Nithya. 2015. "Preprocessing Techniques for Text Mining - An Overview." *International Journal of Computer Science & Communication Networks* 5(1): 7–16. <http://www.ijcscn.com/Documents/Volumes/vol5issue1/ijcscn2015050102.pdf>.
 - [24] Yazici, Ali L I, Alok Mishra, and Paul Kontogiorgis. 2015. "IT Service Management (ITSM) Education and Research: Global View." *International Journal of Engineering Education* 31(4): 1071–80.