

USER ANALYSIS OF INFOBMKG APPLICATION IN THE PERSPECTIVE OF HUMAN COMPUTER INTERACTION USING SUPPORT VECTOR MACHINE ALGORITHM

Ilham Fannani^{1*}, Enggar Novianto², Alfin Syarifuddin Syahab³
Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia¹²³
E-mail address: 6220211005.fannani@student.uty.ac.id¹

Received: 18, May, 2023

Revised: 25, May, 2023

Accepted: 30, May, 2023

ABSTRACT

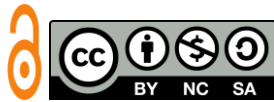
On the Google Play Store, users often read other users' app reviews and reputations, before downloading an app. This makes the analysis of user reviews very interesting for app owners to make future decisions. This study aims to analyze user reviews of the INFOBMKG application on the Google Play Store, using sentiment analysis. This user review analysis uses the Support Vector Machine (SVM) method. The evaluation proposal was made from more than 3,000 user reviews collected from the INFOBMKG application on the Google Play Store. The results of the analysis using the Support Vector Machine produce an accuracy of 85.54% and the most frequently reviewed positive review results are "Good", while the most frequently reviewed negative reviews are "Error". Which indicates a complaint against INFOBMKG users, and from the negative words that appear most often, there are two combinations of the two words that appear most often together, namely the word "very helpful" and the word "less accurate", which indicates that user often complain about problems related to application performance. The results of the sentiment analysis process of testing 3000 review data using the fold = 5 test value in the Support Vector Machine (SVM) method obtained an accuracy of 85.54% which produces predictions on data testing, namely 1500 positive reviews and 1500 negative reviews 1500 reviews.

Keywords: Text Mining, Sentiment Analysis, Support Vector Machine.

1. INTRODUCTION

Information technology has largely impacted all aspect of human life at this point. The rapid development of information technology has affected every aspect of the general public's quality of life. Bidang human-computer interaction theory is a body of knowledge that explores how computer technology affects human activity and work. The primary goal of any human-computer interaction is to provide human beings with ease of use when using computers and the ability to obtain needed goods. Keep your Introduction to be very concise, well structured, and inclusive of all the information needed to follow the development of your findings. Do not over-burden the reader by making the introduction too long. Get to the key parts other paper sooner rather than later (Hamidah, Nugroho, & Surejo, 2023).

In addition, interactions that are untapped resources in human life serve as the single best way to learn new languages, whether they be modern ones or ancient ones, or media technologies. The primary area of computer science that examines a computer-human interaction is how such interaction is conducted (Kaffah & Anshori, 2021). Humans interact



with computers in many ways, so that the interface design between the two is very interesting to be used as research material. In designing an effective interface, of course, you have to go through ideas, ideas and even research involving various disciplines, including: computer science, cognitive and behavioral psychology, anthropology, sociology, ergonomics, industrial design (Armanda, Amin, & Wardhani, 2022).

The Meteorology, Climatology and Geophysics Agency (BMKG) is a Non-Departmental Government Institution that provides various services to provide information to the general public, with the main use of a mobile application named INFOBMKG. The INFOBMKG application is a tool that provides information including early warnings, earthquakes, weather, climate, and other topics that are in line with the BMKG meeting room (Primanda, Burga, Juliansyah, & Nurfauziah, 2022).

2. THEORY

2.1. Human Computer Interaction

Human Computer Interaction is a field of study that focuses on designing, evaluating, and implementing interactive computer systems for use by people as well as researching widespread phenomena in their surroundings (Mariam, 2023). Human Computer Interaction's goal is to improve and fix conditions related to security, usefulness, effectiveness, efficiency, usability, and appeal (Vinodhini & Chandrasekaran, 2012); (Aman et al., 2023).

2.2. Information System

A system of information is a piece of equipment with a single, common connection that has the ability to gather, transmit, store, and display data in order to support the creation of keputusan and pengawasan within any institution or organization (Amiruddin, 2022). With such a system, information about people, places, and objects can be found in authorities or other organizations that are in charge of it (Primanda et al., 2022).

2.3. Application

Application is a tool that is used for the specific task at hand, such as resizing documents, customizing windows, or playing games. Another point about the application is that it's a particular group of files that's meant to be used for specific activities that are related to each other (Bagus Tri, 2020).

3. METHOD

3.1 Support Vector Machine

This study analyzes reviews users on the INFOBMKG application on the Google website Play uses the Support Vector Machines algorithm with stages of research that can shown in figure 1.

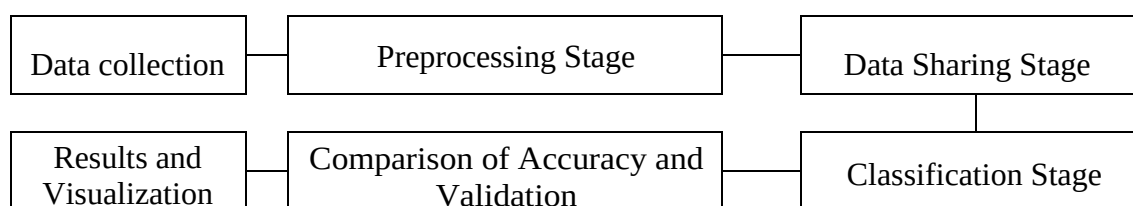


Figure 1. Research Flow

a. Data collection

The process of collecting data in this study namely taking the primary data obtained directly from the original source (not through intermediary media) in particular collected by researchers to be answered research problem (Johnson & Gupta, 2012). Data is taken from 1000 user reviews with the most relevant review category, because Relevant categories display review data that is most related to the app and content from clear user reviews about complaints or satisfaction from user reviews INFOBMKG the app on the google play site with scraping data collection method.

For scraping in this study using extensions in Google Chrome scraper. Then obtained review data saved to a file of type .csv so that the data which is used becomes Jupyter when read csv is readable. Data collection is done by retrieve user review data for the INFOBMKG app on the Indonesian language Google Play site with URL address: shorturl.at/vxEH0 with using web scraping techniques. Web engineering scraping is the process of taking a semi-structured documents from the internet, generally in the form of web pages that are built with a markup language like HTML or XHTML that aims to fetch the information from that page is good wholly or partly for use for other purposes (Tripathy, Agrawal, & Rath, 2015).

b. Preprocessing Stage

At the preprocessing stage of data processing that has been obtained will be processed for find the necessary data and get rid of useless data analysis stage. Data Transformation with stages (1) Reading Data, namely initial data saved in .csv format will be converted into the Jupyter database, so the data well read Jupyter. (2) Case Folding, namely data will be changed to equal (lowercase). Data Cleaning with stages (1) Tokenizing, namely the text on the data that has been obtained will be separated for each word. (2) Screening, i.e. remove the words that don't have influence or not informative. (3) Spelling, namely correcting the word to have errors in spelling or word order which is abbreviated. (4) N-gram, namely connecting related words.

c. Data Sharing Storage

A collection of words that can express positive or negative sentiments called sentiment lection (Bhakuni, Kumar, Iwendi, Singh, & others, 2022). According to (Christina & Yasa, 2021), A sentiment lexicon can be formed in a number of ways manually by searching for words sentiment and classify the word sentiment be a positive or negative sentiment manual by the researchers. The data is divided into two parts in a ratio of 9:1, that is 9 parts for training data and 1 part for data testing. Data Training, namely the data that will be processed by the method to be used, whose results will be used as predictions for test data. Data Testing, that is data to be tested and predicted.

d. Results and Visualization

Classification stage by using SVM method, looking for hyperline (limit decision) best separates each There are two types of reviews, namely positive reviews and negative reviews by using Supports the Vector Machine algorithm. And then the data will be used for prediction test data.

e. Comparison of Accuracy and Validation

Process accuracy and comparison evaluation in this study using k fold cross validation, using value fold = 2 to 10, this test is done for get the best accuracy in the process exam. And then evaluation use the confusion matrix for know the right and wrong estimates of classification results.

f. Classification Stage

At this stage the withdrawal is made conclusions drawn based on the results classification by using the method SVM. As well as the results of the visualization process. Use of the Support Vector Machine method in research to review sentiment analysis in the INFOBMKG app based on research who have drawn conclusions that SVM shows the best performance for sentiment analysis classification (Hartmann, Huppertz, Schamp, & Heitmann, 2019).

4. RESULTS AND DISCUSSION

4.1 Scraping Process Machine Support Vector Method

The use of the SVM algorithm model can be: done using the operator "Support Vector Machine" on Jupyter. Based on previous research (Sharma & Dey, 2012) SVM is the most successful method for classification by dataset review. SVM's ability to minimize errors training data and minimize errors influenced by dimensions. SVM is also a one of the right ways to use in solving high-dimensional problems however the number of data samples is limited and SVM easy to implement on data already have a library.

The value in the parameter selected for The SVM operator is the default value provided by Jupyter. In application classification process using models SVM, the results obtained from the training data then applied to the test data. Then the accuracy of the data is recalculated for test data. Best accuracy results by using this SVM method generates a weight for each word There is. From the weighting of the words used on the classification process to predict sentiment of the data to be tested. Result of SVM for word weighting.

Weighting these words have different values, if weighting numbers produce numbers that minus (-), then the word is categorized as a negative word, as an example word "problem_active" is categorized as a negative word because it has a weight value of -0.003. but if generate a number greater than 0 or not minus, then the word is categorized as a positive word, as an example "account_safe" which is categorized as a word positive because it has a weight of 0.009. Specifically this word is used for predict sentiment on testing data, the results obtained are as shown in Figure 2.



Figure 2. Prediction Results of SVM Testing Data

In Figure 2. is the result of the prediction from testing data. Of the 3000 reviews predicted using the SVM classification algorithm, 2639 positive reviews and 361 negative reviews were predicted.

4.2 Comparison of Accuracy and Validation

4.2.1 Comparison of Accuracy

At the validation stage, testing is carried out using both classification algorithms using cross validation, testing is carried out using fold = 2-10. This test aims to find the best accuracy of each algorithm. The results of testing each fold value are in table 1.

Table 1. Validation Test Results from the Algorithm

SVM		
Accuracy Fold	SVM	Mark Ascension
2	84,20%	8,23%
3	84,87%	8,45%
4	84,21%	7,9%
5	85,54%	9,34%
6	84,32%	7,79%
7	85,43%	8,79%
8	83,76%	7,34%
9	84,65%	8,12%
10	84,99%	8,57%

Based on Table 1. the process of the experiments that have been carried out using the SVM method, the SVM method with a fold = 5 value will be used in the validation testing model, because the test with a fold = 5 value shows the best accuracy results compared to other fold values. The results of the fold=5 value will be used for validation testing for the SVM method.

4.2.2 Validation

At the validation test stage, this research uses a confusion matrix in the process of calculating accuracy by knowing the amount of test data that is classified correctly and the

amount of test data that is wrong in its classification. A comparison of the confusion matrix with testing using fold=5 for SVM can be seen in Table 2.

Table 2. Confusion Matrix

Prediction	SVM	
	Negative	Positive
Negative	1204	927
Positive	253	616
Akurasi 85,54%		

Based on Table 2, it explains how big the percentage of data is predicted by the SVM method. True negative is the amount of observational data with negative categories that can be predicted negatively (correctly predicted) by each method. True positive, namely the number of observation data with positive categories that can be predicted positively (predicted correct) by each method. False negatives are the large number of observational data that are in the negative category but there are prediction errors. Meanwhile, False positives are the number of observational data that are in the positive category but there are prediction errors.

SVM obtained prediction results in the positive class as many as 1543 positive reviews were tested, there were 616 reviews which were classified as correct and there were prediction errors of 951 reviews which were included in the negative reviews. Then in the negative reviews with a total of 1457 reviews tested, there were 1204 reviews that were correctly classified as negative reviews and there were prediction errors of 253 reviews which were included in the positive reviews. then from the Confusion Matrix value obtained an accuracy of 85.54%, there are 3000 reviews that are classified as correct with the SVM model. The process of calculating the accuracy value is carried out using a calculation formula (Gorunescu, 2019). So to calculate the value of accuracy in the SVM method as follows:

$$\text{Accuracy} = \frac{TP + TN + IN}{TP + FN + FP + TN} \times 100\%$$

$$\text{Accuracy} = \frac{616 + 1204 + 745}{616 + 927 + 253 + 1204} \times 100\%$$

$$\text{Accuracy} = \frac{2565}{3000} \times 100\%$$

$$\text{Accuracy} = 85,54\%$$

The results of the classification process carried out using 3,000 training data reviews and 100 testing data reviews using fold=5 in the Support Vector Machine (SVM) method obtained an accuracy of 85.54% which resulted in correct predictions with 616 positive reviews and 1204 negative reviews.

4.3 Visualization

Visualization is carried out on the sentiment class in each classification. Visualization is used with the aim of viewing positive and negative sentiment information, as well as extracting information in the form of topics that are most frequently discussed by INFOBMKG users. Of

the many existing review texts, the information that is considered the most important can later be retrieved. As for the visualization for all the data that appears most frequently in the image. In the classification process that has been carried out, the results of the classification review of the entire data are 3,000 reviews. The word that appears most often is the word "Excellent" 239 times, the word "Very Helpful" 211 times, and the word "Error" 206 times and so on.

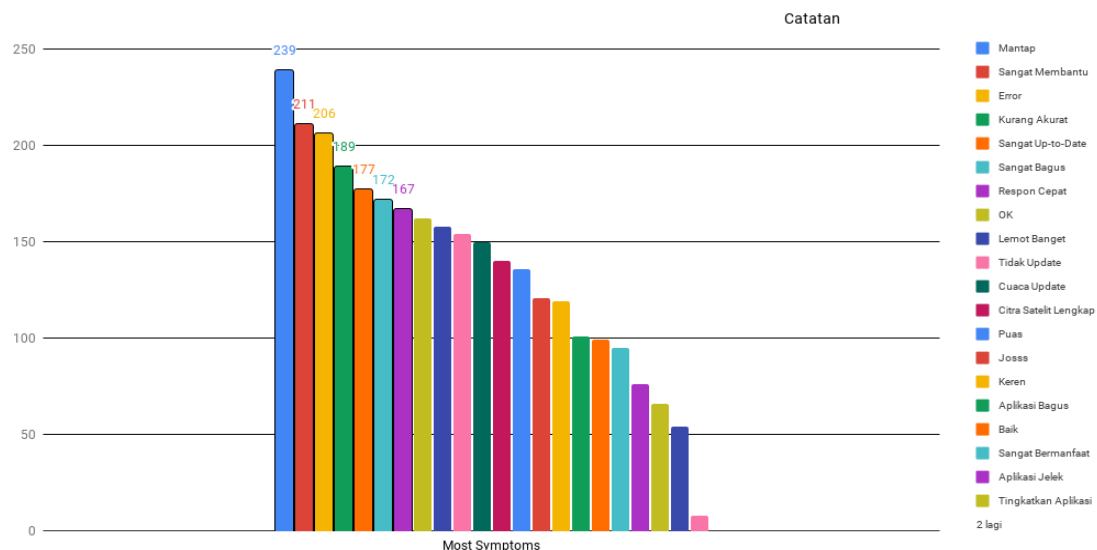


Figure 3. The most frequently occurring words from the entire data

4.4 Positive Reviews

The positive review data used was manually selected labeled data with the most frequently occurring "kernel model" vector values. Where the process of extracting positive review information is carried out repeatedly so that positive review information is obtained from INFOBMKG users who are most frequently discussed. Positive reviews are identified based on the frequency of words in the review data. Visualization for positive review data can be seen in Figure 4.

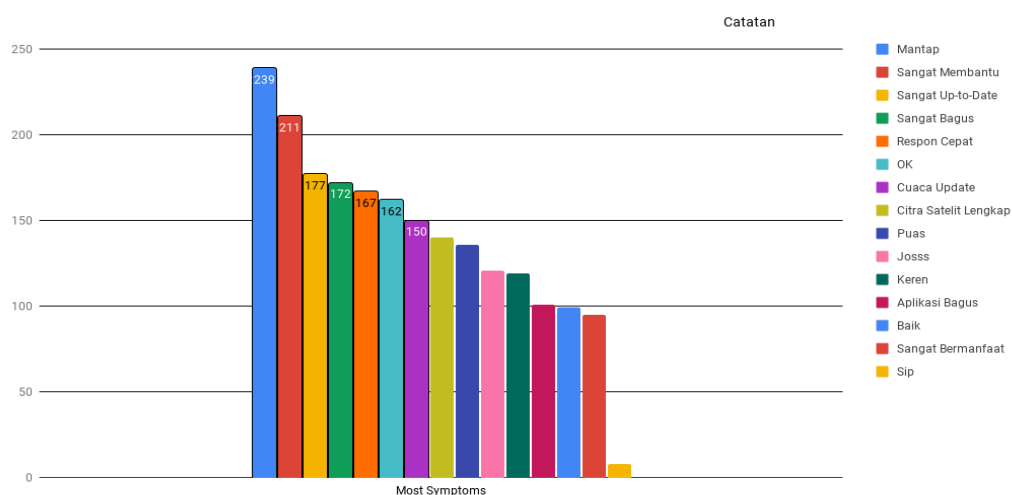


Figure 4. The Most Frequently Appearing Positive Words

In Figure 4 it can be seen that the results of the classification of 3000 reviews show that the most frequent positive words are the words "Great" 239 times, the words "Very Helpful" 211 times, and the words "Very Up- to-Date". 177 times and so on. The series of words that

appear the most often as shown in Figure 4 are positive conversations most frequently discussed by INFOBMKG users.

4.5 Negative Reviews

The negative review data used was manually selected labeled data and with vector values in the most frequently occurring "kernel model". The process of extracting negative review information was carried out repeatedly to obtain information from the negative reviews of INFOBMKG users who were the most discussed. These negative reviews are identified based on the frequency of words in the review data. The following is a negative review data visualization as shown in Figure 5.

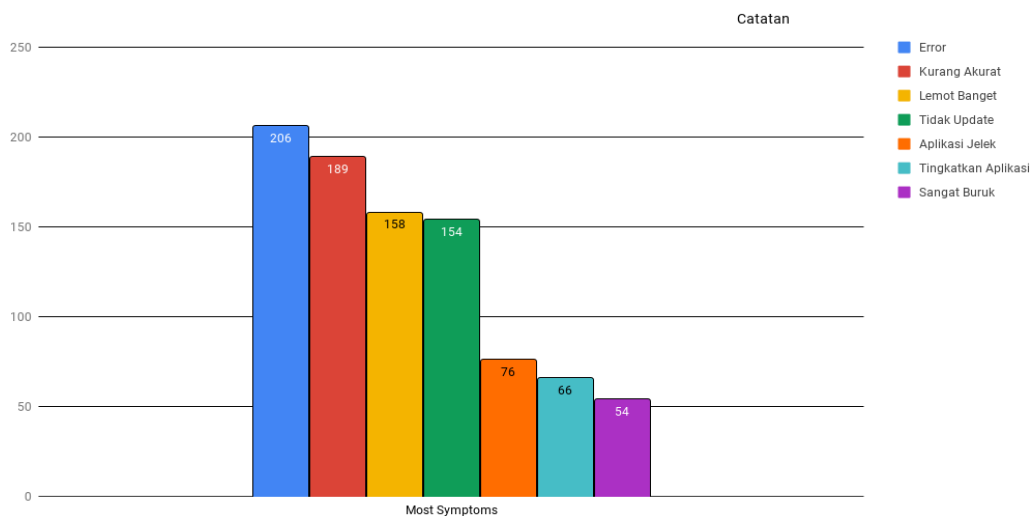


Figure 5. The Most Frequently Appearing Negative Words

In Figure 5 it can be seen that from the results of the review classification of 3000 negative words the most frequently appearing were the word "Error" 206 times, the word " Less Accurate " 189 times, and the word "Very Slow" 158 times. times and so on. The most frequently occurring set of words as shown in Figure 5 is the negative discussion most frequently discussed by INFOBMKG users.

4.6 Generate n-Grams

The use of generate n-grams is to associate or connect words. The use of generate n-grams in this study is with length = 2, meaning that the words to be connected are 2 words. The results obtained from connecting the 2 words that appear most often are as shown in Figure 5.

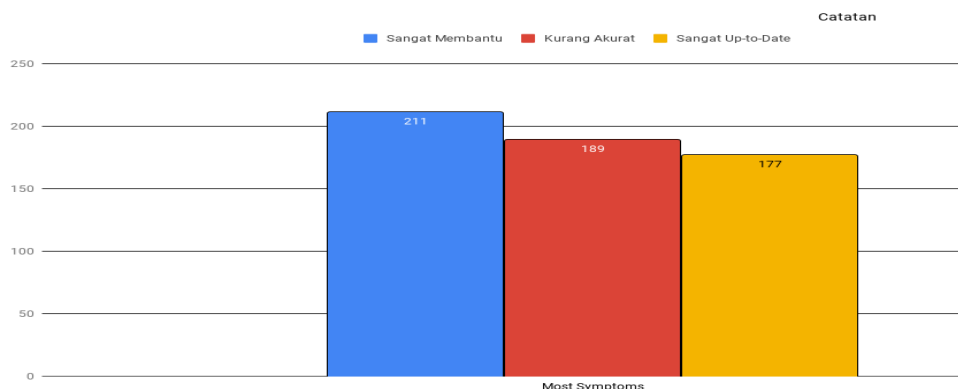


Figure 6. Generate n-Gram Most Appears

In Figure 6. it can be seen that the results of generating n--grams from 3000 reviews, the words that appear most frequently are “Very Helpful” 211 times, “Less Accurate” 189 times, “Sangat_Up-to-Date”, 177 times etc. Meanwhile, to generate n-grams with a positive category, as shown in Figure 7.

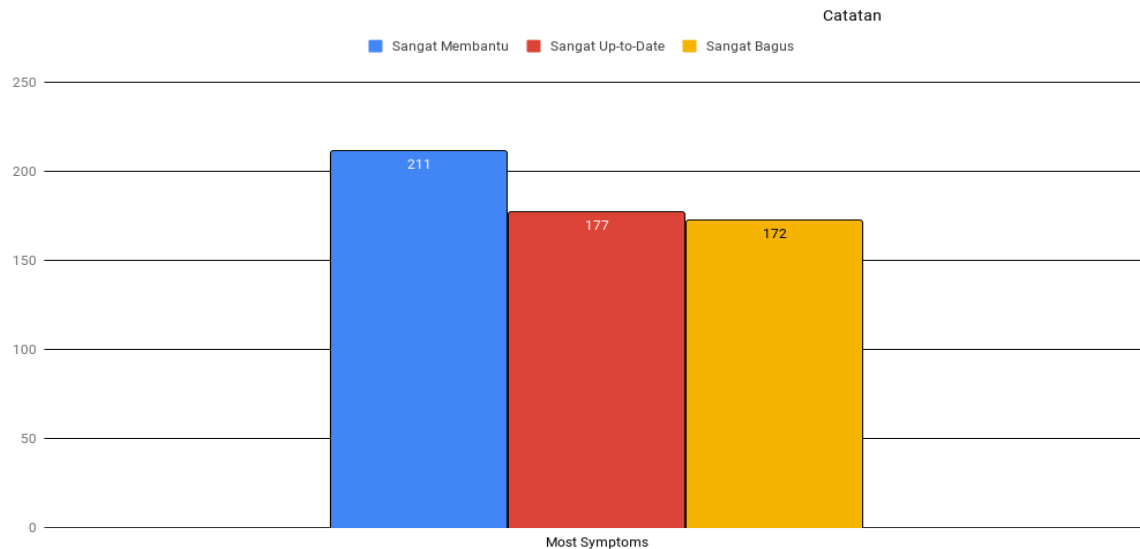


Figure 7. Generate n-Gram Positive Words

In Figure 7. it can be seen that the results of generating n--grams from 3000 reviews, the words that appear most frequently are “Very helpful” 211 times, “Very_Up-to-Date” 177 times, “Very Good”, 172 times etc. From these 2 connecting words it can be concluded that the INFOBMKG application is very helpful and users are very grateful for the INFOBMKG application, they are also happy with the updated information and the application performance is very good and user friendly. Meanwhile, to generate n-grams with a negative category, as shown in Figure 8.

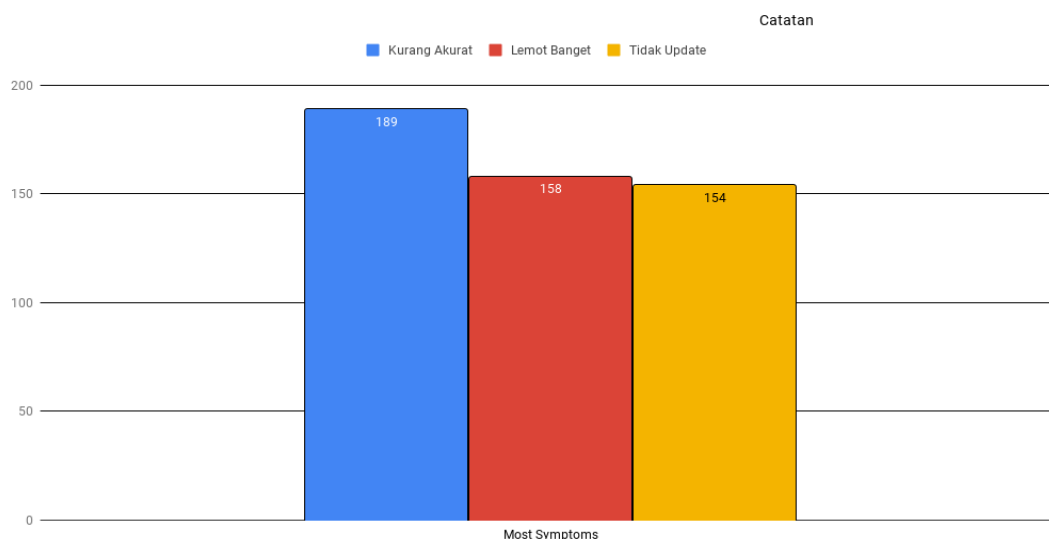
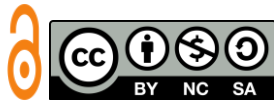


Figure 8. Generate n-Gram Negative Words

From the results above, the 2 words that appear most frequently in the negative category out of 3000 data reviews are "Kurang_Akurat" 189 times, "Lemot_Banget" 158 times, "Tidak_Update" 154 times and so on. From the 2 linked words it can be concluded that users



often discuss inaccurate information in applications such as not being timely and users complain about the long application process. INFOBMKG application users also complain that application information is not updated.

The results of the visualization process provide information in the form of words that appear most often, positive review words that often appear, negative review words that often appear, and two combinations of words that appear most often. The word “Mantap” that appears most often is the word “Sangat_Membantu” 211 times, while the word “Error” that appears most often is the word “Kurang_Akurat” 189 times. And from the words that appear most often, there are two combinations of the two words that often appear, namely "Sangat_Membantu" for two word combinations that are included in the positive category and "Kurang_Akurat" for two combinations of words that are included in the positive category. included in the positive category and negative category.

5. Conclusion

Based on the experimental results shown in Figure 7. it was concluded that the words with positive values that were most often reviewed by users were about "Excellent" which appeared 239 times, while the words with negative values could be seen in Figure 8. The most frequently reviewed by users is regarding "Error" which appears 206 times, which indicates a complaint against INFOBMKG users. And from the negative words that appear most often, there are two combinations of the two words that appear most often together, namely the word "Very Helpful" and the word "Less Accurate", which indicates that user often complain about problems related to application performance. The results of the sentiment analysis process of testing 3000 review data using the fold = 5 test value in the Support VectorMachine (SVM) method obtained an accuracy of 85.54% which produces predictions on data testing, namely 1500 positive reviews and 1500 negative reviews. 1500 reviews.

In this study, the review data used was still limited to the most relevant review categories, so it needs to be developed in further research using other categories of reviews or using review data from a certain time span. For further research, the emoticon detection approach can be used with the Lexicon Based Approach method.

REFERENCES

- Aman, A., Kasmi, M., Iskandar, A., Zam, W., Mustika, N., Laswi, A. S., ... others. (2023). The Virtual Tour Panorama as a Guide and Education Media of the Historic Objects at Datu Luwu Palace. *Ingenierie Des Systemes d'Information*, 28(2), 425.
- Amiruddin, E. G. (2022). Web Based Aviation Communication Tool Information System. *Ceddi Journal of Information System and Technology (JST)*, 1(1).
- Armanda, T. A., Amin, A. S., & Wardhani, I. P. (2022). Menu Dinamis pada Aplikasi Dalam Perspektif Human-Computer Interaction.
- Bagus Tri, M. (2020). Perancangan Sistem Informasi Management Siswa Berprestasi Berbasis Android Pada Smk Pgri Rawalumbu. *Jurnal Sains & Teknologi Fakultas Teknik*, 10(2), 30–39.
- Bhakuni, M., Kumar, K., Iwendi, C., Singh, A., & others. (2022). Evolution and Evaluation: Sarcasm Analysis for Twitter Data Using Sentiment Analysis. *Journal of Sensors*, 2022.
- Christina, Y., & Yasa, N. (2021). Application of theory of planned behavior to study online

- booking behavior. *International Journal of Data and Network Science*, 5(3), 331–340.
- Hamidah, I., Nugroho, B. I., & Surejo, S. (2023). Penerapan Interaksi Manusia Dan Komputer Pada Antarmuka Sistem Informasi Akademik. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 5(1), 111–120.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38.
- Johnson, F., & Gupta, S. K. (2012). Web content mining techniques: a survey. *International Journal of Computer Applications*, 47(11).
- Kaffah, S. A., & Anshori, I. F. (2021). Analisa Aplikasi Cake Berdasarkan Prinsip Dan Paradigma Interaksi Manusia Dan Komputer Menggunakan Evaluasi Heuristic. *JTIK (Jurnal Teknik Informatika Kaputama)*, 5(2), 291–299.
- Mariam, M. (2023). Application of Customer Relationship Management in Maintaining Customer Loyalty (Case Study Hotel Melati). *Ceddi Journal of Information System and Technology (JST)*, 2(1), 1–8.
- Primanda, D., Burga, M. N., Juliansyah, A., & Nurfauziah, N. (2022). EFEKTIVITAS APLIKASI INFO BMKG DALAM MEMBERIKAN INFORMASI CUACA DAN BENCANA TERHADAP MASYARAKAT KOTA TANGERANG SELATAN. *NEO POLITEA*, 3(2), 1–9.
- Sharma, A., & Dey, S. (2012). A comparative study of feature selection and machine learning techniques for sentiment analysis. *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, 1–7.
- Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of sentimental reviews using machine learning techniques. *Procedia Computer Science*, 57, 821–829.
- Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6), 282–292.