BinusMaya Sentiment Analysis with Naive Bayes and KNN: Exploring User Satisfaction of BinusMaya as a Learning Platform

Moh. Khoirul Umam Al - Amin
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
moh.amin@binus.ac.id

Angel Marcelina Br Sipahutar
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
angel.sipahutar@binus.ac.id

Dimas ramadhan
Computer Science Department,
Bina Nusantara University
Jakarta, Indonesia
dimas.ramadhan@binus.ac.id

David David
Computer Science Department,
Bina Nusantara University
Jakarta, Indonesia
david01@binus.edu

***Abstract*— Rapidly developing technology has brought significant changes to the education sector, especially with the emergence of Learning Management Systems (LMS) such as BinusMaya, a learning platform developed by Bina Nusantara University. BinusMaya integrates various learning methods, including online, offline, and independent, to support teaching and learning activities. However, despite offering various benefits, BinusMaya faces challenges such as material accessibility issues, real-time synchronization inconsistencies, and notification failures. This research aims to evaluate user satisfaction with BinusMaya through sentiment analysis using two machine learning algorithms: Naïve Bayes and K-Nearest Neighbors (KNN). Data was collected from 621 user reviews via a questionnaire and then processed to identify positive or negative sentiment. After the data preprocessing, the two algorithms are applied to build a sentiment classification model. The results show that Naïve Bayes has an accuracy of 85% and an F1 score of 87%, better than KNN, which has an accuracy of 80% and an F1 score of 83%. This analysis revealed that BinusMaya had almost a balance of positive and negative reviews, with Naïve Bayes identifying 66 positives and 54 negative reviews and KNN identifying 62 positives and 58 negative reviews. Although BinusMaya is generally well-received as a learning platform, there is an urgent need for continuous improvement to address issues experienced by users. In the future, more sophisticated algorithms such as Support Vector Machines (SVM) and artificial neural networks are recommended for more in-depth analysis. This research also highlights the importance of collecting more prominent and representative data to provide more comprehensive and reliable results. The findings from this research provide important insights for BinusMaya developers in their efforts to improve the application and contribute significantly to the literature regarding sentiment analysis in the education sector.**

Keywords*— Learning Management Systems (LMS), BinusMaya, Sentiment Analysis, Naïve Bayes, K-Nearest Neighbors (KNN), User Satisfaction

I. INTRODUCTION

The educational sector is one of the sectors that experience the impact of rapid technological advancement. Through this technological development, innovations in education emerge to assist educational institutions and students in teaching and learning activities[1]. One of the innovations that has emerged is the development of digital learning platforms, also known as Learning Management Systems (LMS)[2]. Almost all educational institutions in Indonesia utilize Learning Management Systems (LMS) developed based on websites or mobile applications. BinusMaya, is one of the Learning Management Systems (LMS) based on website and mobile application platforms.

Binus Maya is an online learning platform in the form of a Learning Management System (LMS) developed by Binus University. This platform is designed based on Multi-Channel Learning (MCL), which combines several learning methods, such as online learning, offline learning, and self-study. In addition to that, BinusMaya is also a multi-platform application. Its development aims to assist learning activities at Bina Nusantara University. BinusMaya is not only a learning platform but also a communication tool among students, faculty members, academic services, and lecturer services. However, despite its numerous benefits, some issues need to be addressed. For instance, some learning materials cannot be accessed through the BinusMaya application. Additionally, the sync time of the BinusMaya application in real-time could be more consistent, which can be frustrating for users. There are also occasional notification failures, which can affect the convenience of Binus Maya application users.

Ensuring user satisfaction is paramount for any application to be deemed suitable and effective. Positive feedback from users is a clear indicator of an application's success[3]. Therefore, to address any issues related to user satisfaction in the BinusMaya application, it is essential to employ an analysis method to determine users sentiments towards the application, whether positive, neutral or negative[4]. This will enable the developers to make necessary improvements and enhance the user experience[5].

To conduct sentiment analysis, textual data is required, which will then be processed and classified to determine whether the text expresses positive or negative sentiment[5]. Various algorithms can be used in classification, such as Logistic regression, Support Vector Machine (SVM),K- Mean, Naive Bayes, Decision Tree, KNN, and many more. Research on sentiment analysis to analyze user satisfaction in an application has been widely conducted[6]. One example is sentiment analysis of Shopee which is an e-commerce application using the Naive Bayes algorithm conducted by Dany Pratmanto, utilizing data obtained through crawling the Google Play Store[7].

Most research on sentiment analysis utilizes data obtained through social media such as Instagram, Twitter, or TikTok[3]. However, only some researchers conduct sentiment analysis using questionnaires with specific target populations. Therefore, this study focuses on user preferences and user satisfaction with the BinusMaya Application, with its target population being students of Bina Nusantara University. The benefits of this research are that it assists BinusMaya developers in understanding positive or negative opinions about the application and improves its performance. Besides, this research can be used to develop further theories.

II. LITERATURE REVIEW

This research used several methods and techniques, starting from collecting data using questionnaires distributed to all users of the LMS BinusMaya application. This research involves a critical process called text mining, aiming to obtain quality information from the text data obtained for subsequent sentiment analysis.

Sentiment analysis is a method used to identify or categorize text as positive or negative based on expressed sentiment[8]. Sentiment analysis usually involves the use of machine learning techniques, and classification algorithms to identify keywords, phrases, or contexts that indicate a particular sentiment. In general, there are five stages in conducting sentiment analysis, namely data collection through scraping, pre-processing, feature selection, classification, and evaluation.

One of the classification algorithms that is often used in case sentiment analysis is K-Nearest Neighbors (KNN), and Naïve Bayes[7]. Although both can be used to perform classification, they have differences. The Naïve Bayes algorithm is classified as an unsupervised model while the K-Nearest Neighbor is classified as supervised[3], [9]. These two algorithms can be described as follows:

A. Naïve Bayes Method

Naïve Bayes is one of the most effective and efficient inductive learning algorithms for machine learning and data mining, especially in case of sentiment analysis. Aside from that. Naïve Bayes is a text mining technique often used in sentiment analysis. It operates on a probabilistic model where documents are expected to be distributed differentially among several groups based on the phrases they include [3]. The Naïve Bayes method applies Bayes' theorem. The equation for Bayes' theorem is as follow:

$$p(H|X) = \frac{p\left(\frac{X}{H}\right)p(H)}{p(X)}$$

Wheres :
X = evidence
H = the hypothesis
P(H | X) = the probability that hypothesis H is true given the evidence X
P(H) = for the prior probability of hypothesis H
P(X) = the prior probability of the evidence.

From the equation, the formula for the Bayesian theorem can be rewritten as :

$$P(A|x_1, x_2, \ldots . x_n) = \frac{P(A|x_1, x_2, \ldots .. x_n|A)P(A)}{P(x_1, x_2, \ldots .. x_n)}$$

In the case of sentiment analysis, the basic principles of Naïve Bayes are used as follows.

$$P(w_t|c_k) = \frac{\sum_{i=1}^{|D|} N_{i,t}\, P(c_k|d_i) + 1}{\sum_{i=1}^{|D|}\sum_{i=1}^{|D|} N_{i,t}\, P(c_k|d_i) + |V|}$$

The number of occurrences of the word wt in the dataset di, while |v| is vocabulary size and |D| is the dataset size. Therefore, the Naïve Bayes rule for classification is formulated as,

$$c = \arg max_{c_k}\,(P(c_k)\prod_{i\epsilon|V|}(P(w_t|c_k))^{N_{ij}}$$

The goal is to determine the likelihood of a phrase appearing in a particular feedback or review. The categorization is based on the highest probability of a phrase belonging to a specific class. Therefore, Naive Bayes and sentiment analysis illustrate that using the Naive Bayes method to perform sentiment analysis involves breaking down each word and assigning it a sentiment. [3], [4].

B. K-Nearest Neighbors Method

K-Nearest Neighbors (KNN) is the simplest and most widely used data mining classification method. K-Nearest

Neighbor (KNN) is a non-parametric supervised learning algorithm for classifying objects or datasets based on the proximity or similarity of characteristics properties of one data to another data. KNN works by taking several K nearest data as a reference in determining new data classes[10]. The K-Nearest Neighbor (KNN) algorithm is used because it is easy for a process because the process is carried out based on a weighting approach that is simple and easy to implement, adjust, and circumvent the process and has a high accuracy value [11].

III. RESEARCH MODEL

This research utilizes several sampling techniques in collecting data from BinusMaya users, such as collecting data by distributing questionnaires using Google Forms, which are then distributed to BinusMaya users. The model in this research describes the process and relationship between text mining, especially in carrying out sentiment analysis. The algorithms used, validation and testing methods and user satisfaction are based on research that has been conducted. In general, the data mining process in this research follows a workflow similar to the workflow of creating a machine learning model.

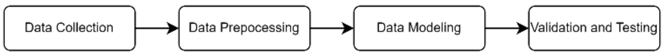

Fig.1. Data Mining Workflow

A. Data Collection

The data collection process in this research uses a questionnaire, Questionnaires were distributed from April 17, 2024, to June 10, 2024.

B. Data preparation

After the data collection process, the data that has been obtained is then carried out data preparation so that the data is ready to be used for further sentiment analysis. The data preprocessing process goes through several stages, as below,

1. Feature Selection
   This process removes several unnecessary features from previously obtained data. In this research, the feature selection process was carried out by removing the username, NIM, and class features so that only two features were used, namely reviews and user ratings of BinusMaya.

2. Data Cleaning
   At this stage, the previously obtained data is removed for duplicates or missing values. Numbers, symbols, and emoticons are also removed.

3. Data Transformation
   After the feature selection is carried out, the data transformation process is then carried out. Namely, all the words in the feature review are changed to lowercase, and there are no longer capital letters in the words, so the words become uniform. A filtering process is also carried out, namely by removing stop words in sentences. In its application, stopwords are removed by utilizing a list of words, including stopwords from the Sastrawi and Natural Language Toolkit (NLTK) libraries. This filtering process also normalizes data with typos or slang that cannot be identified.

4. Data Reduction
   This process is carried out to reduce the dimensions of the data to become data that is easier to process and analyze. In this research, the data reduction process was carried out using the stemming and tokenization methods. Stemming itself is a process of combining variant word forms (affixes) into a general form.

C. Modeling

They use two frequently used classification algorithms: Naïve Bayes and KNN. The use of Naïve Bayes is based on the accuracy of Naïve Bayes, which is quite accurate in the case of sentiment analysis, and KNN was chosen because the KNN algorithm is distance-based and very easy to apply. The model creation in this research involves applying k-fold cross-validation, where the model will be trained and tested k times. [3]

D. Testing and Evaluation

The final stage is evaluating the model that has been created. Each model will have its evaluation results. The results of the sentiment analysis process for each model will produce a confusion matrix. The confusion matrix evaluates overall model performance. Many classification metrics can be derived from the confusion matrix, starting from accuracy, precision, recall, etc.

IV. RESULTS AND DISCUSSION

A. Data Collection

The data collection process in this research uses a questionnaire, which is then distributed to BinusMaya users. The questionnaire was created using a Google form and consisted of five questions: name of the questionnaire taker, Student Identification Number (NIM), rating of BinusMaya, and review or review of BinusMaya. Questionnaires were distributed from April 17, 2024, to June 10, 2024. This research used 621 ratings and review data from BinusMaya users. The data is then saved as a Comma-Separated Values (CSV) file.

B. Data Preparation

The data collected and obtained through questionnaires certainly still has a lot of noise, so it is necessary to go through data preprocessing or data preparation processes. Data preparation includes feature selection, a data cleaning process, data transformation, and data reduction, as shown in Fig. 3 and table I.

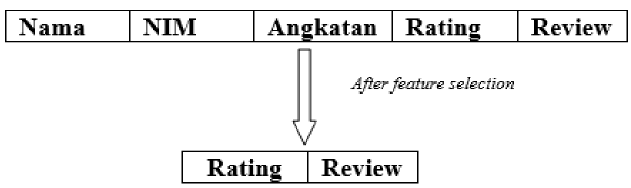

Fig. 3. Feature Selection Procedure

C. Modeling

After the data preparation, the data will be used for the modeling process. Modeling was carried out using the Naïve Bayes and KNN algorithms. In this research, the KNN algorithm itself uses a k value of 5. To apply these two algorithms, there is no need to do modeling from scratch because we can directly call and use the two algorithms via the skit-learn library.

D. Testing and Evaluation

1. Training and testing process

After preparing and modeling the data, the training data is validated using the cross-validation method.

TABLE I. Data Preparation Procedure

Procedure	Data
Real Data Respondents	Aplikasi kadang2 error ketika gw pakai, dan lemot :)
After Data Cleaning	Aplikasi kadang error ketika gw pakai dan lemot
After Data Filtering	aplikasi kadang error ketika saya pakai lemot
After Stemming Data	aplikasi kadang error ketika saya pakai lemot
After Data Tokenization	[aplikasi], [kadang], [error], [ketika], [saya], [pakai], [lemot]

This research uses the most straightforward cross-validation technique, namely k-fold cross-validation, in evaluating the model. The k-fold method works by randomly collecting K-unique data with the same number of samples created, and models built sequentially on each (K-1) dataset or folds, validating the remaining data sets, and tracking the evaluation matrix. Previously, In this study, a k value of 20 was used. After dropping missing values and duplicate data, from a total of 623 data, 600 data remained. The 600 data were then divided into two parts, namely training and test data. It can be seen in Table II that the number of training data is 480 data, and the test data is 120 data. The data will then be processed using two different classification algorithms.

TABLE II. Numbers of Training and Testing Data

	20	
Data	Train	Test
	480	120

2. Alghoritms Results and Analysis Data

After carrying out the training and testing process using two classification algorithms, it was found that using the naïve Bayes algorithm with k-folds cross-validation of 20 folds, the accuracy was 85%, and the f1-score was 87%. Meanwhile, when using KNN with a value of k= 5 and 20 folds, the accuracy was 80%, and the f1 score was 83%. Carrying out a classification to determine user satisfaction with BinusMaya as a learning platform obtained the following results.

TABLE III. Sentiment Analysis Results

K	Naive Bayes		KNN	
20	Positive	Negative	Positive	Negative
	66	54	62	58

The results of sentiment analysis show that using the Naïve Bayes algorithm, there were 66 positive and 54 negative reviews. Besides that, using the KNN algorithm, the results were 62 positive reviews and 58 negative reviews.

V. Conclussion

This research has focused on conducting sentiment analysis of the BinusMaya Learning Management System (LMS) to evaluate user satisfaction. By analyzing user feedback, the study sought to classify sentiments as positive or negative using two machine learning algorithms: Naïve Bayes and K-Nearest Neighbors (KNN). The research employed a rigorous methodology, including data collection through questionnaires, data preprocessing, and subsequent analysis using these algorithms. The findings reveal insightful trends regarding user perceptions of the BinusMaya application, offering both practical implications for developers and theoretical contributions to sentiment analysis research. The results from the sentiment analysis indicate that the Naïve Bayes algorithm outperformed the K-Nearest

Neighbors (KNN) algorithm. Naïve Bayes achieved an accuracy of 85% and an F1-score of 87% using a 20-fold cross-validation, while KNN, with k=5, achieved an accuracy of 80% and an F1-score of 83% under the same cross-validation conditions. These metrics suggest that Naïve Bayes is more effective in handling the text data involved in this study, likely due to its probabilistic approach, which can better manage the varied and nuanced nature of user reviews compared to KNN's reliance on proximity-based classifications.

The sentiment analysis outcomes are pivotal in understanding the overall user satisfaction with BinusMaya. According to the Naïve Bayes algorithm, there were 66 positive and 54 negative reviews, while the KNN algorithm identified 62 positive and 58 negative reviews. This close ratio between positive and negative sentiments suggests that while the platform is generally well-received, a significant portion of users experience challenges or issues with the application.

Given the nearly equal distribution of positive and negative feedback, it is clear that BinusMaya is a valuable tool for educational purposes but requires continuous improvements. Enhancements could focus on the issues users have pointed out, such as accessibility of learning materials, real-time synchronization, and notification reliability.

To further refine the sentiment analysis of BinusMaya and improve its user satisfaction, future research could explore the use of more advanced algorithms. Support Vector Machines (SVM) and neural networks, particularly models like Long Short-Term Memory (LSTM) networks or transformers, could offer more sophisticated analysis capabilities. These methods are well-suited to capturing the complexities of natural language, potentially providing more nuanced insights into user feedback. Additionally, employing ensemble methods, which combine multiple algorithms to improve prediction accuracy, could also be beneficial. These methods often outperform individual models by balancing their respective strengths and weaknesses.

One notable limitation of this study is the relatively small dataset used for analysis. With only 621 reviews collected, the sample size may not be fully representative of BinusMaya entire user base. This limitation underscores the need for collecting more extensive and diverse data to ensure a more comprehensive understanding of user sentiments. Larger datasets would enable more robust training and testing of machine learning models, leading to more reliable and generalizable results.

VI. Reference

[1] B. H. Situmorang, A. Chairunnas, and A. Talib Bon, "Sentiment Analysis of User Preferences on Learning Management System (LMS) Platform Data."

[2] D. Araújo Lima, D. A. Lima, F. B. Alves, J. F. Ribeiro, J. F. Ribeiro, and V. P. Magalhães, "Experience report: students sentiment analysis in using Google Classroom as a virtual learning environment Experience report: students' sentiment analysis in using Google Classroom as a virtual learning environment", doi: 10.6084/m9.figshare.19853920.

[3] Universitas Indonesia. Fakultas Ilmu Komputer, STMIK Primakara, Institute of Electrical and Electronics Engineers. Indonesia Section., and Institute of Electrical and Electronics Engineers, *2019 International Workshop on Big Data and Information Security (IWBIS) : Nusa Dua, Bali, Indonesia, October 11, 2019*.

[4] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012102.

[5] V. Christanti Mawardi and E. Darmaja, "Logistic Regression Method for Sentiment Analysis Application on Google Playstore," *International Journal of Application on Sciences, Technology and Engineering*, vol. 1, no. 1, pp. 241–247, Feb. 2023, doi: 10.24912/ijaste.v1.i1.241-247.

[6] O. F. Prihono and P. K. Sari, "Comparison Analysis Of Social Influence Marketing For Mobile Payment Using Support Vector Machine," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 367–374, Oct. 2019, doi: 10.22219/kinetik.v4i4.921.

[7] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012043.

[8] Z. Rais, F. T. T. Hakiki, and R. Aprianti, "Sentiment Analysis of Peduli Lindungi Application Using the Naive Bayes Method," *SAINSMAT: Journal of Applied Sciences, Mathematics, and Its Education*, vol. 11, no. 1, pp. 23–29, Mar. 2022, doi: 10.35877/sainsmat794.

[9] A. Ridok, "SENTIMENT ANALYSIS FOR REVIEW MOBILE APPLICATIONS USING NEIGHBOR METHOD WEIGHTED K-NEAREST NEIGHBOR (NWKNN)," *Journal of Environmental Engineering & Sustainable Technology JEEST*, vol. 03, no. 01, pp. 23–32, 2016, [Online]. Available: http://jeest.ub.ac.id

[10] A. Salma and W. Silfianti, "Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors," *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, pp. 158–162, 2021.

[11] S. Masturoh, R. L. Pratiwi, M. R. R. Saelan, and U. Radiyah, "APPLICATION OF THE K-NEAREST NEIGHBOR (KNN) ALGORITHM IN SENTIMENT ANALYSIS OF THE OVO E-WALLET APPLICATION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 8, no. 2, pp. 78–83, Jan. 2023, doi: 10.33480/jitk.v8i2.3997.