Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Spam Message Classification Using the Naïve Bayes Algorithm Based on RapidMiner

Muhamad Yusup[1]*, Mochamad Isham Fadillah[2], Rifky Adinanta Fauzanie[3], Risca Lusiana Pratiwi[4], Rani Irma Handayani[5], Euis Widanengsih[6]

[1,2,3,4,5,6] *Program studi informatika, Fakultas teknik dan Informatika, Universitas Bina Sarana Informatika*
15230269@bsi.ac.id [1*], 15230095@bsi.ac.id[2] , 15230769@bsi.ac.id [3], risca.ral@nusamandiri.ac.id [4], rani.rih@nusamandiri.ac.id [5], euis.ewh@bsi.ac.id [6]

Abstract

This study implements the Naïve Bayes algorithm for classifying spam and non-spam (ham) messages using the RapidMiner Studio platform. The dataset used was obtained from the SMS Spam Collection Dataset on the Kaggle platform, which consists of 5,759 messages with a distribution of 4,075 ham messages and 1,291 spam messages. The research stages included text pre-processing, model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The experimental results showed that the Naïve Bayes model achieved an accuracy of 89.64% with a precision of 56.93%, a recall of 100%, and an F1-score of 72.56%. The research findings indicate that the Naïve Bayes algorithm is effective in detecting spam messages with adequate accuracy, and prove that RapidMiner is an efficient tool for implementing machine learning methods in text classification.

Keywords: *Naïve Bayes, RapidMiner, Spam Classification, Text Mining, Machine Learning, Natural Language Processing*

1. Introduction

Spam messages are a significant problem in digital communication because they can lead to data misuse, the spread of false information, and disruption for users [1]. Spam messages are often sent in bulk via SMS, email, or instant messaging applications for promotional or fraudulent purposes. Therefore, spam message detection is important to improve the security and comfort of users when communicating electronically [2].

Various approaches have been used to detect spam messages, including rule-based methods, machine learning, and deep learning [1], [3]. The machine learning-based approach has proven to be more efficient in recognizing word patterns and spam message characteristics than manual methods [4]. One of the most widely used algorithms in text classification is Naïve Bayes because it has a simple yet effective probabilistic structure for large data sets [5].

The Naïve Bayes algorithm works by calculating the probability of a message being classified as spam or non-spam based on the distribution of words that appear in the dataset. A number of studies have proven the effectiveness of this algorithm in detecting spam messages with a high degree of accuracy. [6], [2]. However, most previous studies still focus on programming-based implementation without utilizing visual platforms such as RapidMiner, which can make it easier for users to build models and analyze classification results interactively.

Based on this, this study aims to implement a RapidMiner-based Naïve Bayes algorithm to classify spam and non-spam messages. This study also analyzes model performance based on evaluation metrics such as accuracy, precision, recall, and F1-score using the SMS Spam Collection dataset from Kaggle.

2. Literature

Research on spam message detection has been extensively conducted using machine learning and natural language processing (NLP) approaches. Methods such as Support Vector Machine (SVM), Decision Tree, and Random Forest have been widely used, but the Naïve Bayes algorithm remains a popular choice due to its model simplicity, computational efficiency, and stable performance on large-scale text data. [1].

Research in [5] applying the Naïve Bayes algorithm for SMS message classification using a public dataset and demonstrating a high level of accuracy with text pre-processing steps such as tokenization, stopword removal, and stemming. Other studies [3] Optimizing the

performance of Naïve Bayes by combining Term Frequency–Inverse Document Frequency (TF-IDF) and parameter adjustment, which successfully increased spam detection accuracy to 95% in the email dataset.

Comparative analysis in [6] shows that the Naïve Bayes algorithm performs better than Logistic Regression in classifying SMS messages, with an accuracy rate of 97%. Meanwhile, a comprehensive survey on [7] Reviewing various spam detection approaches between 2015 and 2024, including rule-based methods, machine learning, and deep learning. The survey results confirm that Naïve Bayes remains an efficient method for spam detection with low computational requirements.

Research in Indonesia also supports these findings. A study on [2] applying Naïve Bayes for SMS spam classification with high accuracy on local datasets, while the results on [4] shows that Naïve Bayes is still competitive compared to other algorithms such as Random Forest and Bagging in classifying Indonesian-language SMS messages.

In general, previous studies have shown that Naïve Bayes has competitive capabilities in spam classification for both SMS and email messages. However, most of these studies have not highlighted implementation using visual platforms such as RapidMiner, which can facilitate experimentation and analysis of results. Therefore, this study focuses on the application of the RapidMiner-based Naïve Bayes algorithm using the SMS Spam Collection Dataset to evaluate the effectiveness of this method in detecting spam and non-spam (ham) messages.

3. Research Method

3.1 Dataset

The dataset used in this study is the SMS Spam Collection Dataset accessed through the Kaggle platform. The dataset consists of 5,759 SMS messages that have been manually labeled into two categories: ham (4,075 messages) and spam (1,291 messages). The data was divided into training data and test data with a ratio of 80:20 using the Split Data operator in RapidMiner.

3.2 Naïve Bayes Algorithm

The Naïve Bayes algorithm is a probabilistic classification method based on Bayes' theorem with the assumption of feature independence. The mathematical formulation of this algorithm is:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \quad (1)$$

P(C|X) : probability of class C based on data X,
P(X|C) : probability of feature X appearing in class C,
P(C) : initial probability of class C,
P(X) : probability of feature X appearing in the dataset.

In the context of spam classification, this algorithm calculates the probability of a message belonging to the spam or ham category based on the occurrence of certain words.

3.3 Process in Rapidminer

The experiment was conducted using the following series of operators in RapidMiner:

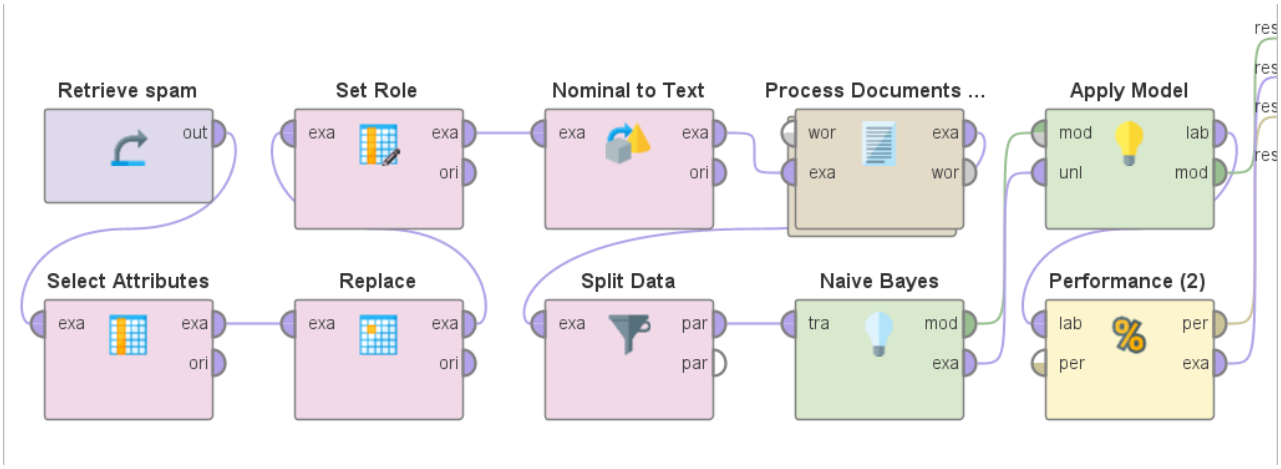

Fig. 1: Shows the series of operators used in the classification process in RapidMiner)

1. Retrieve: Load the dataset from the CSV file.
2. Set Role: Set column v1 as the label/target attribute.
3. Nominal to Text: Convert the message attribute to text type.
4. Process Documents from Data: Perform text preprocessing, which includes:
   a. Transform Cases: Convert to lowercase
   b. Tokenize: Break text into tokens (words).
   c. Stopwords filter: Removes common words.
   d. stemming (Porter): Reducing words to their basic form.
5. Split Data: Membagi data menjadi data latih (80%) dan data uji (20%).
6. Naïve Bayes: Train the model with training data.
7. Apply Model & Performance: Apply the model to test data and measure its performance.

4. Results and Discussion

Based on the results of testing the Naïve Bayes model on the SMS Spam Collection dataset using RapidMiner, an accuracy of 89.64%, precision of 56.93%, recall of 100%, and F1-score of 72.56% were obtained. These values indicate that the model performs quite well in classifying spam and non-spam (ham) messages.

Table 1 : Results of the Naïve Bayes Model Evaluation

Matrix	Velue %
Accuracy	89,64
Precision	56,93
Recall	100,00
F1-score	72,56

4.1 Evaluasi Model Berdasarkan Confusion Matrix

Table 2: Model Evaluation Based on Confusion Matrix

	Ham	Spam
Ham Prediction	4075	0
Spam Prediction	556	735

From the table, it can be seen that the model successfully classified all spam messages correctly (100% recall), with no spam messages incorrectly detected as ham. However, there were 556 ham messages that were misclassified as spam, resulting in a precision value for the spam class of only 56.93%.

This performance shows that the model tends to be sensitive to spam messages (high recall), but less selective in determining actual spam (low precision). In other words, the model prefers to consider a message as spam rather than missing actual spam (false negative = 0), even though this results in an increase in the number of false positives.

4.2 Performance Analysis

An accuracy of 89.64% indicates that, overall, most messages were correctly classified by the model. However, the significant imbalance in the dataset, with 86.3% of messages being ham and 13.7% being spam, is thought to be a factor affecting the model's precision performance. The Naïve Bayes model tends to follow the distribution of the majority of data, thus predicting the ham class more often.

The F1-score value of 72.56% represents the balance between precision and recall. This value is quite good for text classification cases with unbalanced datasets, indicating that the model has sufficient stability in recognizing word patterns that distinguish spam and ham messages.

In addition, feature analysis results show that several words (tokens) contribute significantly to the classification process. Words that often appear in spam messages, such as promotional words, gifts, or numbers indicating offers, have significantly different probabilities compared to common words in ham messages. This supports the effectiveness of Naïve Bayes in utilizing the probabilistic distribution of words to determine message classes.

Although the results obtained show fairly good performance, this study has several limitations. The model was only tested using one algorithm, namely Naïve Bayes, without comparison with other methods such as Support Vector Machine (SVM) or Random Forest. In addition, this study has not performed parameter optimization or data balancing, which has the potential to increase precision values. Therefore, further research can focus on exploring other algorithms and applying advanced preprocessing techniques to obtain more optimal results.

Overall, the Naïve Bayes model proved to be efficient and easy to implement on the RapidMiner platform. Text preprocessing steps such as tokenization, stopword removal, and stemming play an important role in improving accuracy by reducing noise in the text data.

5. Conslusion

Based on the test results, the Naïve Bayes algorithm implemented using RapidMiner was able to classify spam and non-spam messages with an accuracy rate of 89.64% and an F1-score of 72.56%. These values indicate that the model performs well in detecting spam messages despite the data imbalance between the spam and ham classes.

The accuracy value obtained in this study is comparable to the results of previous studies [5], which reached 88.7% using a similar algorithm. This shows that the application of Naïve Bayes in RapidMiner is capable of producing competitive performance compared to similar studies, even with lower implementation complexity.

However, this study has limitations because it only uses one algorithm without comparison with other methods. Therefore, further research can focus on the application of parameter optimization and data balancing techniques to improve the precision and stability of the model.

References

[1] H. A. Al-Kaabi, A. Darroudi, A. K. Jasim, H. Alaa, and A.-K. Hussain, "Survey of SMS Spam Detection Techniques: A Taxonomy," *Alkadhim Journal for Computer Science*, vol. 4, no. 2, 2024, doi: 10.53523/ijoirVolxIxIDxx.

[2] A. Sauddin, T. Azisah Nurman, N. Aeni, and S. Rahayu Sudarta, "Klasifikasi Spam SMS Menggunakan Naïve Bayes Classifier dan K-Nearest Neighbors."

[3] S. Charan Lanka, K. Pujita, K. Akhila, S. Mondal, P. Vidya Sagar, and S. Bulla, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Optimization of Naïve Bayes Classifier for Spam E-Mail Detection." [Online]. Available: www.ijisae.org

[4] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 10, no. 3, pp. 432–437, Dec. 2021, doi: 10.32736/sisfokom.v10i3.1302.

[5] D. A. Anggraini, M. Ikhsan, and S. Suhardi, "Implementation of the Naïve Bayes Algorithm in the SMS Spam Filtering System," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 2, pp. 838–849, May 2024, doi: 10.47709/cnahpc.v6i2.3875.

[6] P. A. Raharja, M. F. Sidiq, and D. C. Fransisca, "Comparative Analysis of Multinomial Naïve Bayes and Logistic Regression Models for Prediction of SMS Spam," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1290, Jul. 2022, doi: 10.30865/mib.v6i3.4019.

[7] E. Triana, A. Irma Purnamasari, A. Bahtiar, and E. Tohidi, "Journal of Artificial Intelligence and Engineering Applications Improved Spam Email Detection Performance Based on Naïve Bayes Approach TF-IDF Vectorizer with Multi-Metric Optimization," 2025. [Online]. Available: https://ioinformatic.org/