International Journal of Advances in Data and Information Systems
Vol. 4, No. 2, October 2023, pp. 191~200
ISSN: 2721-3056, DOI: 10.25008/ijadis.v4i2.1306

Hoax Detection News Using Naïve Bayes and Support Vector Machine Algorithm

Nur Elyta Febriyanty[1], M. Amin Hariyadi[2*], Cahyo Crysdian[2]
[1]IKIP Budi Utomo, Indonesia
[2]Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia

Article Info

Article history:

Received Sep 18, 2023
Revised Oct 08, 2023
Accepted Oct 22, 2023

Keywords:

Detection
Hoax news
Support Vector Machine
Naive Bayes

ABSTRACT

Websites and blogs are well-known as media for broadcasting news in various fields such as broadcasting news. The validity of news articles can be valid or fake. Fake news is also known as hoax news. The purpose of making hoax news is to persuade, manipulate, and influence news readers to do things that contradict or prevent correct action. This study proposes to experiment with the Support Vector Machine and Naïve Bayes classifications to detect hoax news in Indonesian. This study uses a dataset from public data, namely news between valid news and hoaxes. The system can classify online news in Indonesian with the term frequency feature the machine vector Support algorithm and naïve Bayes classification. While the evaluation model used is the Confusion Matrix. The results of the comparison of the two models as a Support Vector Machine have an accuracy rate of 75,5%, and Naive Bayes has an accuracy rate of 88%. Therefore, for the classification of hoax news, we recommend the Naive Bayes model because it has a better level of accuracy than the Support Vector Machine.

Corresponding Author:

Mokhamad Amin Hariyadi,
Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia,
Jl. Gajayana No.50 Dinoyo Kec.Lowokwaru Kota Malang, Jawa Timur 65144
Email: adyt2002@uin-malang.ac.id

1. INTRODUCTION

Indonesia became the country with the sixth largest internet users in the world in 2014, and the country with the highest growth in internet users in 2016. At the end of 2018, internet users in Indonesia were categorized as heavy internet users, with data consumption of more than 5GB per month only for internet connection from mobile devices [1]. So with the increase in internet users, people can consume any information that spreads quickly. With this speed, of course, it produces positive and negative impacts, where the information circulating does not always provide the truth that is reality or can be calleda hoax.

A hoax is a piece of information or news that contains things that have not been identified or are not facts that happened[2]. To gain profit and achieve personal goals, hoaxes are often deliberately created and shared so that they can spread more quickly. Information obtained from hoaxes can certainly influence society because it creates doubts and confusion about the information received, and can damage the image of individuals and related groups. Headlines used by hoax information are often sensational and provocative. It is deliberately made to attract the interest and curiosity of readers. At least 30% to nearly 60% of Indonesian people are exposed to hoaxes when accessing and communicating through cyberspace. Meanwhile, only 21% to 36% can recognize or detect hoaxes. Most of the hoaxes found were related to political, health, and education issues.

Previous studies have discussed the detection of hoaxes in various information. Like Covid 19 [3] conducted experiments on the Naïve Bayes classification algorithm to classify hoax news Covid 19. Based on the results of research that has been carried out, the naïve Bayes model and cross validation can classify hoax news well, the resulting accuracy is 86.3% where 80- 90% fall into the criteria of good classification. There are also not too many data that are predicted to be wrong out of a total of 300 datasets, only 41 are declared wrong in labeling, not up to 2% of the total dataset. checking facts on social media platforms[4], Proposed a framework for detecting fake news based on feature extraction and feature selection algorithms and a set of sound classifiers detecting fake news based on feature extraction and algorithms. the extracted characteristics are reduced with the help of the chi-square algorithm and the analysis of variance algorithm (ANOVA). Using three data sets published online: Fake-or-Real-News, Media- Eval, and ISOT. This study uses five performance metrics to evaluate the proposed framework: accuracy, area under the curve, precision, recall, and f1 score. Our system achieves 94.6% accuracy for Fake-or-Real data sets[5]. Classification of Facebook and Twitter posts into hoaxes and non-hoaxes[6], [7]. The second compares different classification approaches to empirically find the most suitable text classifier for categorizing various sub-categories of news, and the three methods have been proposed and compared to detect duplicate news from corpus by involving various pre-processing techniques and widely used similarity measures, similarity cosine, and Jaccard coefficient. The results show that conventional text classifiers are still relevant and work well in text classification tasks because MNB has provided 89.5% accuracy [8].

In the Machine Learning approach, classifiers are built automatically by learning the category properties of a predefined training data set. It can process more complex and diverse information in dynamic situations. In this paper, we propose a naïve Bayes classifier that scales directly with the number of indicators and data points that can be used for binary and multiclass classification problems. We implement the presented schema using Machine Learning tools. The experimental results show a 90 percent increase in performance in the classification technique. In another study, the classification of hoax tweet topics. [9] The author created a data set consisting of 200 tweets about the selected topic, while conducting a truth assessment. Initially performing "text normalization" on tweets, exploring feature extraction techniques to classify news into categories, performing comprehensive linguistic analysis on tweets, extracting word sets to find visible patterns, and finally implementing the k-nearest neighbor algorithm to classify polarized news from credible. This method produces a method accuracy of more than 80%. Explores the application of natural language processing techniques to detect 'fake news', namely misleading news that comes from unreputable sources. Using datasets obtained from Signal Media and source lists from OpenSources.co, applying bi-gram term-inverse document frequency (TF-IDF) and probabilistic context-free grammar detection (PCFG) to a corpus of approximately 11,000 articles tested our dataset on several classification algorithms, Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees, and Random Forests. Found that the TF-IDF of bi-grams entered into the Stochastic Gradient Descent model identified non-credible sources with an accuracy of 77.2% [10]. Proposes a news text classification model based on the Latent Dirichlet Allocation (LDA) method and the Logistic Regression algorithm. Because the dimensions of the news text are too high, this model uses the topic model to reduce the dimensions of the text and get features. At the same time, this paper also conducts research on the Softmax regression algorithm to solve multi-class text problems in our life and make it a model classifier. This paper evaluates the proposed model on real news datasets and the experimental results show that the improved model has relatively good performance. This model can effectively reduce the dimensions of news text features and get 81% Recall, 82% Precision and 81% F1 Mesure[11].

The current solution is that the Ministry of Communication and Informatics of the Republic of Indonesia Kominfo has provided the kominfo.go.id site a content complaintservice from various social media sources suspected of containing hoax news. There is alsothe turnback hoax. id site managed by the Indonesian Anti Hoax Society which has provided definite information on hoax news that was previously spread.

Therefore, to detect information or news containing hoaxes, it is necessary to have tools that can detect the truth of the news whether it is a hoax or valid. A classification method is needed as a

way to reduce the spread of content containing hoaxes by utilizing text mining. The purpose of this study is to conduct experiments on text classification algorithms for detecting hoax news, classifying news in the Indonesian language network, namely valid and hoax using the term frequency (TF-IDF) as a feature in this experiment.

2. MATERIAL AND METHOD

2.1 Data

The data source of this research is qualitative data, which is secondary data, Data collection was carried out using crawlers manually from Indonesian-language online news portals such as Tribunnews, Detik, and CNN. The benchmark for the validity of news is based on articles published by Kominfo.The dataset was taken from January 2020 to July 2023. The data taken was 80% of the data in the hoax category and 20% of the data in the fact or valid category.Label manually by looking at the description and information provided at the news source. After obtaining the dataset, the data is divided into two, namely training data and testing data. The data in Table 1 is the amount of data, there is training 2.350 text data where there are 1394 valid news and 981 texts, and for data testing 456 valid news and 84 hoax news. The detailed descriptions of this research procedure are given in the following subsections.

Table 1 Total Dataset

Class	Training Data	Test Data
Valid	1400 Text	456 Text
Hoax	950 Text	84 Text
Amount total	2.350 Text	530 Text

Collect datasets on news portals manually via the internet, the author enters keywords that are entered on news portals. Data collection was carried out using crawlers manually from Indonesian language news portals such as Tribunnews, Tempo, Detik, Kompas, and CNN. This data must include news that has been verified as valid and some news that has been verified as hoax . Dataset is shown in Table 2.

Table 2 Dataset Class Category

News Titles	Class Category	Source Data	News Category
Warga Australia Disarankan Tetap Mengenakan Masker Hadapi Penyebaran Flu di Musim Dingin	Valid	Detik.com	Health News
Naik KA Tetap Wajib Booster dan Pakai Masker	Valid	Tempo.co.id	Social
Presiden Jokowi Sapa Masyarakat Indonesia Hingga Bertemu Kanselir Jerman	Valid	Kompas.com	Politik
Berdamai Dengan Corona yang Dimaksud Jokowi Adalah Pemerintah Telah Kalah dan Menyerah	Hoax	Hoax Dataset	Health News
Penerima Stiker Pada Pesan Whatsapp Dikenakan Sejumlah Biaya	Hoax	Hoax Dataset	Business
Mesin Mobil Sering Mati Diatas Rel Kereta Api	Hoax	Hoax Dataset	Travel

Moreover, data from Indonesian language news portals are also used to procure some data required for this study. In total, there are 2350 news titles from valid and hoax classes.

2.2 System Design

This study uses a design system the proposed algorithm is shown through a block diagram in the following Figure 3. In Figure 3 it can be explained as follows, Collection of valid news and hoax news. Text processing processes such as Case Folding (changing characters into lowercase), Stemming (getting basic words), Stop Word Removal (removing unnecessary words), and Tokenization (sentences are broken into tokens). From the text preprocessing process, get the basic words. Weighing using TF-IDF for weighting terms on datasets or documents using token sequences in the form of unigrams in this implementation gets the number of tokens from TF-IDF in just one word. The dataset in this study is divided into 4 sets, train-set, and test-set to be included in the classification algorithm with a ratio of 90% train-set and 10% test-set, 80% train-set and 20% test-set, train-set 70% and test-set 30%, train-set 60% and test-set 40%, train-set 50% and test-set 50%

of the entire dataset. Naïve Bayes Algorithm and Support Vector Machine then data training is carried out and produces learning which will later be used as a reference in the algorithm testing process.

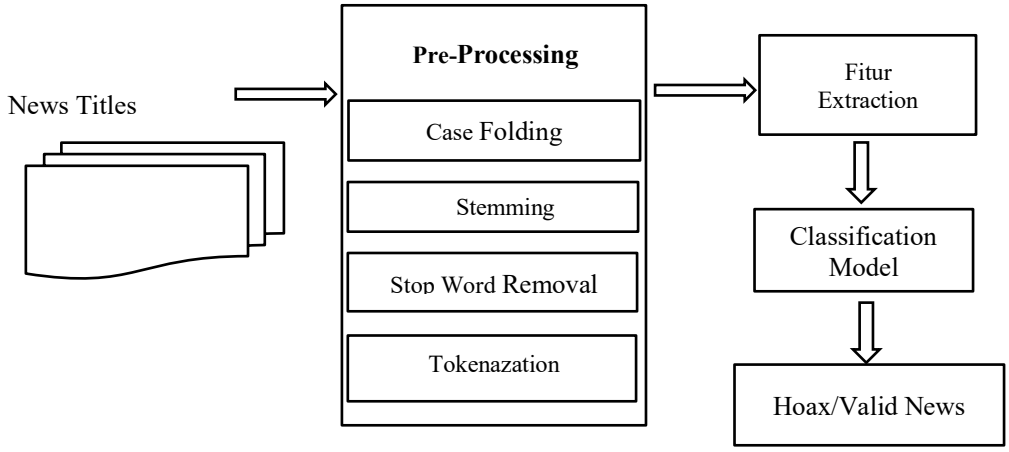

Figure 3. Blockstage Design System

1. Preprocessing Data. Preprocessing data is an important step in news classification because it affects the quality and performance of classification models. Several stages in preprocessing such as Figure 4.

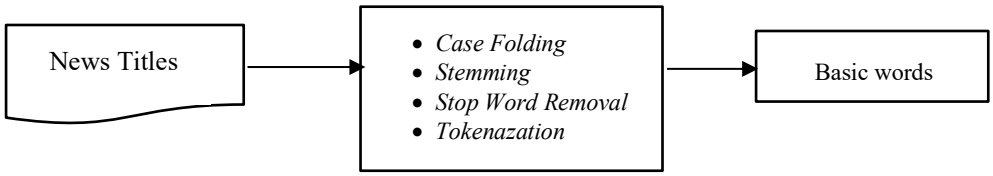

Figure 4 Preprocessing Data

   a. Case folding, preprocessing text data which involves converting all letter characters in the text to lowercase. The aim of this process is to convert the same text into a form that is uniform in terms of letter capitalization, making text analysis easier.
   b. Stemming, and preprocessing of text data which involves converting words into basic forms or root words by removing prefixes and suffixes from these words. The goal of the stemming process is to reduce variations in similar words as to identify more general patterns and meanings in the text.
   c. Stop word removal, The Stop Word Removal process involves removing common words that do not provide important information in the text. The aim of this process is to reduce text complexity and improve the quality of text analysis by focusing on more informative words.
   d. Tokenization. Preprocessing of text data which involves breaking text into smaller units, called tokens. Tokens can be words, phrases, or even characters, depending on the needs of the analysis. The purpose of this process is to facilitate further text processing and analysis.

Pre-processing is a data mining technique that transforms incomplete and inconsistent raw data into a machine-understandable format[12]. Each training and test news dataset goes through a pre-processing stage. This isdone to get the desired model. The preprocessing stages are case folding, stemming, stopword removal, and tokenizing. The following are the results preprocessing stage as shown in Table 3.

Table 3. Preprocessing Result

News Title	Damai Dengan Corona yang Maksud Jokowi Adalah Perintah Telah kalah dan Serah
Case Folding	damai dengan corona yang maksud jokowi adalah perintah telah kalah dan serah
Stemming	damai dengan corona yang maksud jokowi adalah perintah telah kalah dan serah
Stop Word Removal	damai corona maksud jokowi perintah kalah serah
Tokenization	"['damai', 'corona', 'maksud', 'jokowi', 'perintah', 'kalah', 'serah']"

Moreover, data from Indonesian language news portals are also used to procure some data required for this study. In total, there are 2350 news titles from valid and hoax class

3 Feature Extraction – TF-IDF

After the dataset has gone through the pre-processing stage, the dataset willbe calculated for the frequency of the word appearing in the document or what is known as the term frequency (TF). [13]The TF-IDF algorithm is usually used to extract features in tasks because of its simplicity and robustness. TheTF-IDF algorithm is divided into two TF terms, meaning how many words are in the current post, which is given by the equation. (1).

$$\text{TF (Word)} = \frac{\text{Number of repeat words appear in the document}}{\text{Total number of words in the document}} \quad (1)$$

where IDF refers to how important any term is in all headings. IDF scores the words, which is given by equation (2)

$$\text{IDF (Word)} = \frac{\text{Log (total of documents)}}{\text{Nmber of document where the appear}} \quad (2)$$

After the dataset has gone through the pre-processing stage, the dataset will be calculated for the frequency of the word appearing in the document or what is known as term frequency (TF). The results are obtained in Table 3.

Table 4. TF Result

Words	TF Hoax
damai	1
corona	1
maksud	1
jokowi	1
perintah	1
kalah	1
serah	1

4 Naïve Bayes Algorithm

The Model in Naïve Bayes is used to compare the result of each algorithm used using the Scikit Learn library to import Naïve Bayes in the Python programming language. This naïve Bayes method has the advantage of being able to work well on small datasets, being easy to create, and being fast in the calculation process. Meanwhile, the drawback of this method is that it does not apply if the probability value or possibility is 0 (zero). If it is zero then the probability of the prediction is also 0. As well as assuming independent variables. The flow in testing using the Naïve Bayes algorithm is proposed in this study, Figure 5 shows the flow of the Naïve Bayes algorithm performance.

Data preparation, Data preparation includes Data Preparation. Data preparation includes preprocessing of news text, and converting news text into a numerical representation using the TF-IDF method. Class Probability Calculation, Count the amount of training data with hoax labels (N_hoax) and non-hoax labels (N_Valid). Calculate the prior probability for the hoax class (P(hoax)) and the non-hoax class (P(Valid)) in equations (3) and (4):

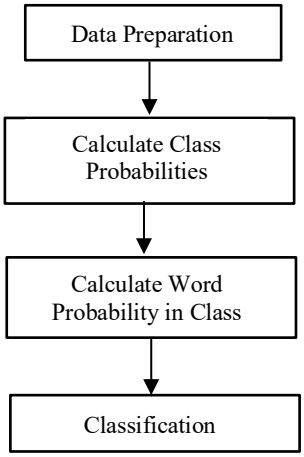

Figure 5. The Flow of The Naïve Bayes Algorithm Performance

$$P\ (hoax) = \frac{N\_Hoax}{N\_Hoax + N\_Valid} \quad (3)$$

$$P\ (Valid) = \frac{N\_Valid}{N\_Valid + N\_Hoax} \quad (4)$$

Calculation of the probability of the word appearing in the class, hoax (P(word|hoax)) and in the non-hoax class (P(word|valid)) in equations (5) and (6):

$$\text{P(word|}hoax\text{)} = \frac{(the\ number\ of\ word\ occurrences\ in\ the\ hoax\ class + 1)}{total\ words\ in\ class\ hoax + number\ of\ unique\ words} \quad (5)$$

$$\text{P(word|valid)} = \frac{(the\ number\ of\ word\ occurrences\ in\ the\ valid\ class + 1)}{total\ words\ in\ class\ valid + number\ of\ unique\ words} \quad (6)$$

Classification, Given the news text to be classified (text_test). Do preprocessing on the same text_test as on the training data. Convert text_test to the appropriate numeric representation. Calculate the posterior probabilities for test_text in the hoax class (P(hoax|test_text)) and in the non-hoax class (P(valid|test_text)) using the Bayes Theorem formula in equations (7) and (8):

$$\text{P(}hoax\text{|teks\_test)} = P(teks\_test|hoax) * \frac{P(hoax)}{P(teks\_test)} \quad (7)$$

$$\text{P(non\_}hoax\text{|teks\_test)} = P(teks\_test|valid) * \frac{P(valid)}{P(teks\_test)} \quad (8)$$

Comparing the values of P(hoax|text_test) and P(valid|text_test) to determine the hoax or valid class label.

5 Support Vector Machine Algorithm

The author uses the existing kernel in the Support Vector Machine to compare the results of each kernel used. SVM supports various types of kernel functions that allow the model to handle various types of data, including text data. Following is some of the steps in calculating the SVM algorithm stage. It is in Figure 6.

Data preparation, training data consisting of features and class labels (hoax or not hoax). then preprocessing the data, such as normalization or selection of relevant features. Kernel Selection Select the appropriate kernel type to map data into a higher feature space. Commonly used kernel types are linear kernels, polynomial kernels, and RBF (Radial Basis Function) kernels [14]. SVM training, Compute kernel matrix for training data. Classification, after obtaining the optimal weight and bias vectors, use the SVM prediction formula to classify new data. The prediction formula in equation (9).

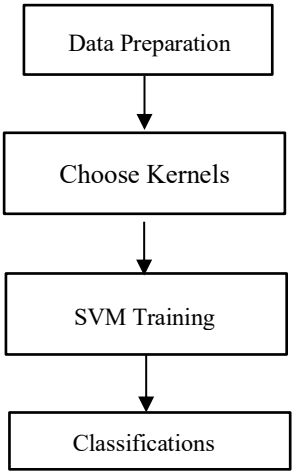

Figure 6. The Flow of The SVM Algorithm Performance

$$SVM = sign(w^T * x + b) \quad (9)$$

w = vector , x = fitur new data vector, b = bias

3. RESULTS AND DISCUSSION

In this section, explain the result of research conducted by researchers on the implementation of the Naïve Bayes algorithm and the Support Vector Machine for the classification of hoax news in Indonesian.

3.1. Testing Process

The comparison of the two data includes 90:10, 80:20,70:30, 60:40 dan 50:50. Total of all data there is training 2.350 text data where there are 1367 valid news and 981 texts, and for data testing 456 valid news and 84 hoax news. The comparison is used to experiment with Naïve Bayes and Support Vector Machine on hoax news detection. The more training data used, the greater the accuracy or directly proportional. Because the more data that is trained, the more precise the test result will better. From the table 5, we get a 50:50 data comparison with 1175:1175 data, 60:40 data comparison 1410:940, 70:30 data comparison 1645:705, 80:20 data comparison 1880:470, and 90 comparison: 10 total data 2115:235.

Table 5. Data Testing and Data Training

Dataset	News Comparation (%)		Number of News	
	Training	Testing	Training	Testing
2350	50	50	1175	1175
	60	40	1410	940
	70	30	1645	705
	80	20	1880	470
	90	10	2115	235

To measure performance accuracy, precision, recall, f1 score, and accuracy are how to approach relative to the right value. Accuracy = number of correct predictions/total number of predictions [15]. Existing approaches regard the problem of fake news as a classification problem that predicts whether an article is fake or not:

True Positive (TP): when a predicted fake news headline is actually classified as fake news.

True Negative (TN): when the predicted true news headline is actually classified as true news.

False Negative (FN): when the news headline prediction is actually classified as fake news.

False Positive (FP): when the predicted fake news headline is actual

The accuracy equation is given by the equation (10)

$$\text{Accuracy } (ACC) = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (10)$$

Precision is how close the measurements are, which is given by the equation (11).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (11)$$

Recall is how many correctly defined actual positives, given by Equation (12).

$$\text{Recall|Sensitivity|TPR} = \frac{\text{TP}}{\text{TP+FN}} \quad (12)$$

The definition of F1 is that if the costs of false positives and false negatives vary, we need precision and recall, as given by Equation (13).

$$F1 = \frac{\text{TP}}{(\text{TP} + \frac{1}{2}(\text{FN+FP})} \quad (13)$$

The comparison of these data will be used in this study and produce values of accuracy, precision, recall, and F1-score of naïve Bayes and SVM which are shown in Table 5.

Table 6. Accuration

Training : Testing	Naïve Bayes				Support Vector Machine			
	Accuration	Precission	Recall	F1-Score	Accuration	Precission	Recall	F1-Score
50:50	80%	100%	80%	100%	90%	100%	80%	90%
60:40	90%	100%	80%	90%	75%	100%	50%	66%
70:30	100%	100%	100%	100%	75%	100%	50%	66%
80:20	100%	100%	100%	100%	60%	100%	33%	50%
90:10	100%	100%	100%	55%	55%	100%	30%	60%

From the data displayed, get results from accuracy, precision, recall, and F1-Score from Naïve Bayes and SVM. Precision can reach 100% because the test results match the detection results using the algorithm. And presented in the graph in Figure 7.

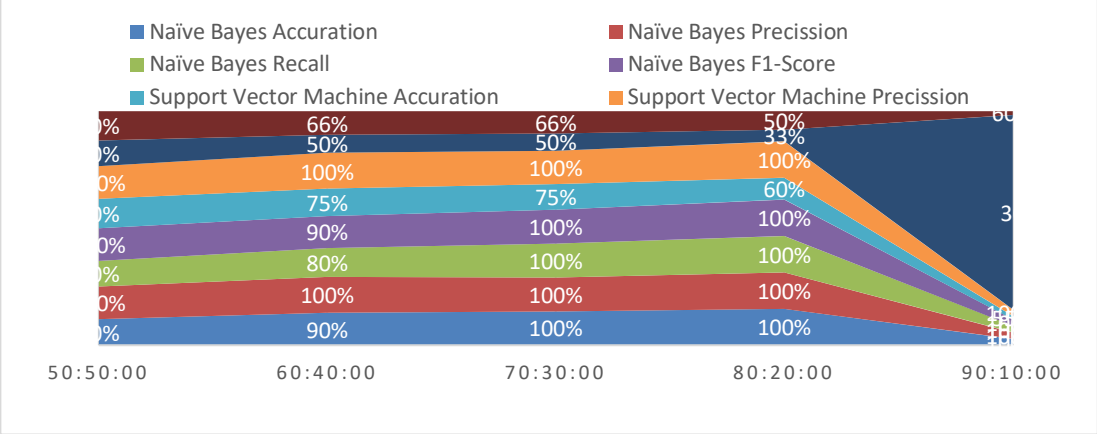

Figure 7. Accuracy Results

The result chart presents the confusion matrix calculation data in Naïve Bayes and SVM calculations which are displayed in the graph. From the draft above, some of the accuracy results do not reach 100% because if your data set has an unbalanced class distribution, the amount of valid data and hoax data comparisons is not balanced, so the model may be biased towards the majority class, thus affecting overall accuracy.

4. CONCLUSION

The results of this study indicate that after using the Support Vector Machine and Naive Bayes models to classify valid and hoax news, the results show that performance of Naive Bayes is superior to the Support Vector Machine. It is proven that from the test data used Naive Bayes has an accuracy value of 75,5%, and Support Vector Machine has an accuracy value of 88%. This research is expected to provide benefits for the development of public news service providers within KOMINFO and This research is expected to be one of the inputs for the government in developing matters related to public information disclosure policies. Based on the research results, it can be concluded that this model can be used as a reference when continuing with more complex prediction models, for example, in prediction models using web-based machine learning which is currently trending as a continuation of previously existing expert systems. The contributions in this research develop robust validation methods to measure model performance, including the use of more diverse test datasets and more comprehensive evaluation metrics. The next suggestion is to use a dataset that is even bigger and from various languages, and an even better document preprocessing model, In this model, the researcher only uses specific preprocessing in Indonesian, If you want to get even richer results you can continue with a deep learning algorithm model with more diverse testing methods.

ACKNOWLEDGEMENTS

Thank you to the Faculty of Science and Technology UIN Maulana Malik Ibrahim Malang. Thank you for the opportunity to be involved in writing the article. we are happy to be able to contribute and help achieve the desired writing goals. I hope the article can be useful and meet the expectations of other readers. If you have other needs in the future, don't hesitate to contact us. Thanks again for this opportunity.

REFERENCES

[1] M. Muttaqin, "Internet Usage Behavior of the ICT Young Workforce in the Border Region," *Pekommas*, vol. 4, no. 1, p. 11, Apr. 2019, doi: 10.30818/jpkm.2019.2040102.

[2] C. Juditha, "Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya Hoax Communication Interactivity in Social Media and Anticipation," vol. 3, no. 1.

[3] F. Prasetya and F. Ferdiansyah, "Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes," *json*, vol. 4, no. 1, p. 132, Sep. 2022, doi: 10.30865/json.v4i1.4852.

[4] T. W. S. Nagel, "Measuring fake news acumen using a news media literacy instrument," *JMLE*, vol. 14, no. 1, pp. 29–42, May 2022, doi: 10.23860/JMLE-2022-14-1-3.

[5] E. Elsaeed, O. Ouda, M. M. Elmogy, A. Atwan, and E. El-Daydamony, "Detecting Fake News in Social Media Using Voting Classifier," *IEEE Access*, vol. 9, pp. 161909–161925, 2021, doi: 10.1109/ACCESS.2021.3132022.

[6] "Book Bhala-Keskar 2020.pdf."

[7] "2020-Semi-Automatic_Classification_and_Duplicate_Detection_From_Human_Loss_News_Corpus.pdf."

[8] Venkatesh and K. V. Ranjitha, "Classification and Optimization Scheme for Text Data using Machine Learning Naïve Bayes Classifier," in *2018 IEEE World Symposium on Communication Engineering (WSCE)*, Singapore, Singapore: IEEE, Dec. 2018, pp. 33–36. doi: 10.1109/WSCE.2018.8690536.

[9] A. Dey, R. Z. Rafi, S. Hasan Parash, S. K. Arko, and A. Chakrabarty, "Fake News Pattern Recognition using Linguistic Analysis," in *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, Kitakyushu, Japan: IEEE, Jun. 2018, pp. 305–309. doi: 10.1109/ICIEV.2018.8641018.

[10] "gilda2017-Evaluating Machine Learning Algorithms for.pdf."

[11] Z. Li, W. Shang, and M. Yan, "News text classification model based on topic model," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, Okayama, Japan: IEEE, Jun. 2016, pp. 1–5. doi: 10.1109/ICIS.2016.7550929.

[12] "2019-A_Hybrid_Bidirectional_Recurrent_Convolutional_Neural_Network_Attention-Based_Model_for_Text_Classification.pdf."
[13] A. Dhar, N. Dash, and K. Roy, "Classification of text documents through distance measurement: An experiment with multi-domain Bangla text documents," in *2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall)*, Dehradun, India: IEEE, Sep. 2017, pp. 1–6. doi: 10.1109/ICACCAF.2017.8344721.
[14] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. A. Zami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia," *Jurnal Masyarakat Informatika*, vol. 13, no. 2, 2022.
[15] S. Raza and C. Ding, "Fake news detection based on news content and social contexts: a transformer-based approach," *Int J Data Sci Anal*, vol. 13, no. 4, pp. 335–362, May 2022, doi: 10.1007/s41060-021-00302-z.