Journal of Information Systems and Informatics
Vol. 5, No. 1, March 2023 e-ISSN: **2656-4882** p-ISSN: **2656-5935**
DOI: 10.51519/journalisi.v5i1.438 Published By DRPM-UBD

Hoax News Analysis for the Indonesian National Capital Relocation Public Policy with the Support Vector Machine and Random Forest Algorithms

Aang Kisnu Darmawan[1], Mohammad Waail Al Wajieh[2], Mohammad Bhanu Setyawan[3], Tri Yandi[4], Hoiriyah[5]

[1,4,5]Department of Information System, Universitas Islam Madura, Pamekasan, Indonesia
[2]Department of Information Technology, Institut Sains dan Teknologi Annuqayah, Sumenep, Indonesia
[3]Department of Informatics, Universitas Muhammadiyah Ponorogo, Ponorogo, Indonesia
Email: [1]ak.darmawan@gmail.com, [2]wail.aw@istannuqayah.ac.id, [3]m.banu@umpo.ac.id, [4]triyandi2604001922@gmail.com, [5]hoiriyah.file.uim@gmail.com

Abstract

The decision of the Indonesian government to relocate the nation's capital outside Java to the North Penajam Paser Regency has sparked controversy and misinformation on social media platforms. While sentiment analysis studies have been conducted on this topic, no research has yet analyzed the issue of hoaxes related to the relocation of the national capital. This study aims to fill this gap by analyzing hoaxes related to the relocation of the Indonesian national capital on Twitter. The study utilizes data crawling, filtering with Hoax Booster Tools (HBT) ASE, data labeling, preprocessing, and TF-IDF weighting. The data is then classified using Support Vector Machine (SVM) and Random Forest (RF) algorithms, and the results of both algorithms are compared. The study found that 85% of tweets had a positive sentiment and 15% had a negative sentiment. Furthermore, the SVM algorithm outperformed the RF algorithm with an accuracy of 95.24% compared to 86.90%. This study contributes to the understanding of the hoax issues related to the relocation of the Indonesian state capital and provides recommendations for government policies to address community concerns.

Keywords: hoax news analysis, national capital relocation, support vector machine, random forest

1. INTRODUCTION

In 2019, the Indonesian Government began discussing the transfer of the Indonesian State Capital (IKN) from Java Island to a new location outside of Java. Subsequently, the President of Indonesia announced the transfer of the IKN to the North Penajam Paser Regency in the National Medium Term Development Plan for the 2020-2024 fiscal years [1]. The newly established IKN will be named

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

"Nusantara," as announced by the Head of Bappenas during a special committee meeting for the Draft State Capital Law (RUU IKN) on January 17, 2022 [2].

The policy of relocating the country's capital has sparked controversies and misinformation on social media platforms. Despite several studies that have conducted sentiment analyses to capture public opinion on this issue, there has been no research on hoaxes or hoax analysis related to relocating the national capital. As opinions for and against the policy become more pronounced on various social media platforms, there is a need to investigate hoaxes related to this issue to provide a more complete understanding of public perception.

The theory of sentiment analysis is a fascinating area of study that seeks to classify the polarity of text in sentences, features/aspects, or documents, and determine whether the opinions expressed are positive or negative [3]. This study employs classification methods, including the Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbor (KNN) algorithms. Each algorithm has its strengths and weaknesses in classifying text data. Previous research by Darwis et al. (2020) [4] demonstrated that the SVM algorithm outperforms other methods. Other studies conducted by Hidayat & Syafrullah (2017) [5] and Darwis et al. (2021) [6] revealed that Naïve Bayes can be effectively used for sentiment analysis, yielding high accuracy results. Additionally, Putra & Juanita's research (2021) [7] achieved an accuracy of above 85% for user sentiment towards two different applications.

Various studies have explored public opinion on the policy of relocating the Indonesian state capital. For instance, Lestari et al. (2022) [8] used Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms with 10-fold cross-validation and achieved 1,141 positive and 591 negative sentiments. Their results indicated that SVM (accuracy = 85.71%) outperformed Naïve Bayes (NB) (76.7%) and KNN (52.74%). Arsi and Waluyo (2021) [9] analyzed 1,236 tweets using SVM and reported very good accuracy (96.68%). Similarly, Lestari (2022) [10] used SVM with Chi-square feature selection and achieved an accuracy and precision of 90%. Other studies have also explored sentiment analysis with different algorithms, including NB, Logistic Regression, and Neighbor Weighted K-Nearest Neighbor (NWKNN).

Despite the many studies on public opinion about the capital relocation policy, no research has been conducted on analyzing hoaxes related to the topic. This study seeks to address this gap by comparing the performance of SVM and Random Forest (RF) algorithms in hoax analysis on Twitter. SVM and RF were chosen due to their excellent performance in text mining [8],[9],[11],[18]. The study's contribution is twofold: first, it scientifically evaluates the performance of SVM and RF for hoax analysis, and second, it offers practical recommendations on government policy regarding the capital relocation.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2. METHODS

This study uses text mining techniques to classify public sentiment analysis on tweet posts that are positive or negative.. The data obtained from the results of Crawling tweet data on the Twitter application is then processed and input into the database, then classified using two algorithms, namely the Support vector machine (SVM) and Random Forest (RF) algorithms. The data collection method uses data obtained from crawling results on Twitter based on the keywords entered by utilizing the search API (Application Programming Interface) provided by Twitter. In this study, the methods used to obtain information data and to solve problems are as follows:

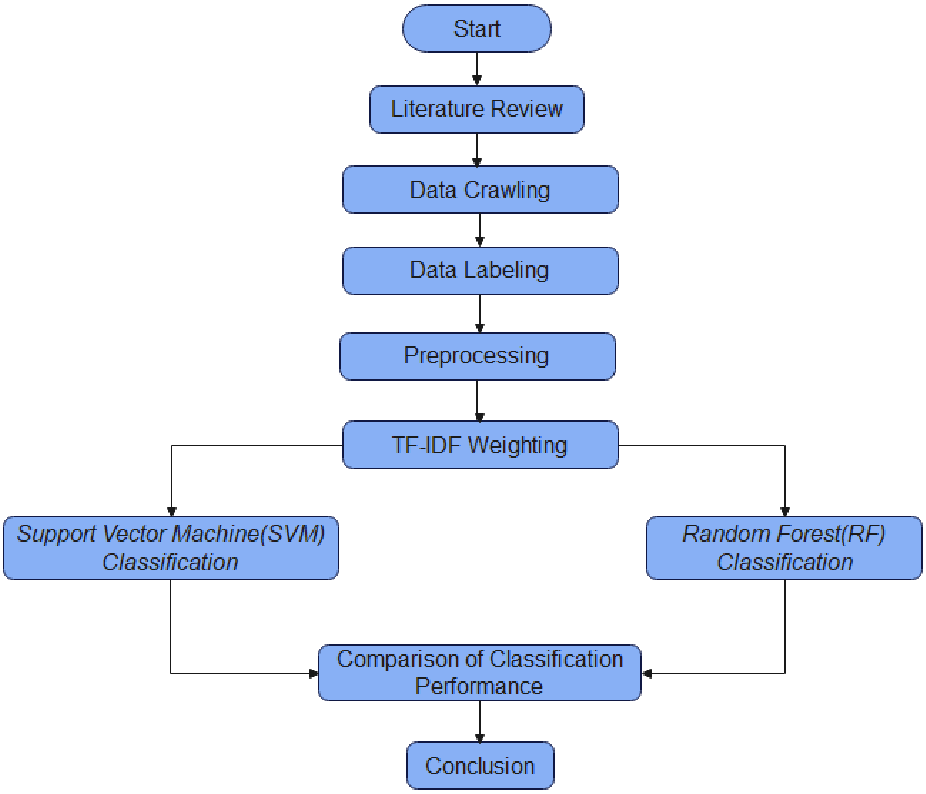

Figure 1. Research flowchart

2.1 How to Search Literature Review

The literature study was carried out by reading papers in accredited national journals and reputable international journals regarding research reports on the topic of hoaxes on the issue of moving the national capital. Journal paper searches were carried out by exploring the "garuda portal" website for national journals

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

with the keywords "sentiment ibukota negara indonesia" and "hoax ibukota negara indonesia", while the Publish or Perish (PoP) v.8.6 software and the mendeley.com portal were used to search. Reputable international journal paper with the keyword "sentiment of the Indonesian state capital moving".

2.2 Data Collection

2.2.1 Data Crawling

Tweet data collection from Twitter uses the Rapid Miner application with the queries #Ibukotanegara and #IKNNusantara. Collection of hoax and non-hoax datasets is carried out by collecting data using the hoax buster application, by utilizing several operators and tools available on the Hoax buster tolls (HBT) application. The following are the tools used:

1. Search tool operators on ASE (c) tolls to display reliable information.

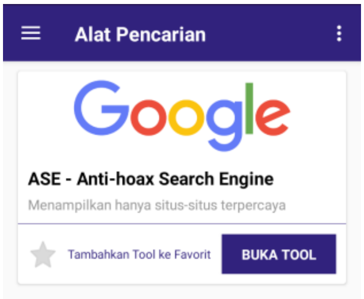

Figure 2. Tools for anti-hoax search engines

2. Search tool operators on hoax search tools are used to search hoax sites and display the results.

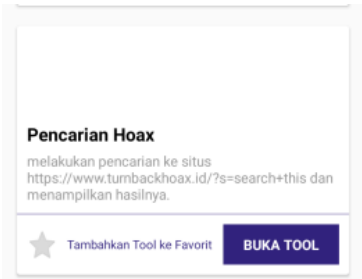

Figure 3. Hoax search tools

3. Social media operators at to foller. me to analyze the Twitter public account of hoax news spreaders.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

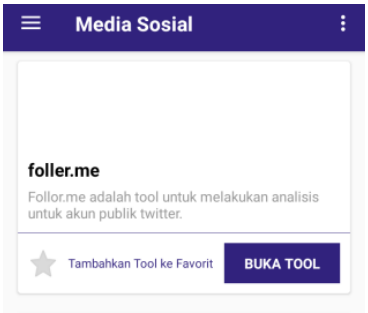

Figure 4. Hoax spreader analysis tools

2.3 Data Analysis

2.3.1 Data Labeling

Data labeling is done by collecting 2 (two) datasets in the form of tables, namely, first, manually labeling tweet data with hoax and non-hoax labels based on the results of analysis of tweet data information using hoax buster tolls (HBT). Then, label the tweet sentiment with positive and negative labels randomly (random).

2.3.2 Preprocessing

Processing is done to change tweets or unstructured text data so that it becomes structured data for the needs of sentiment analysis on the topic of moving the capital city. Preprocessing in this study, implementing a series of stages sequentially, namely:

1. Case folding is the stage for changing the elements of capital letters in documents to become standard, namely lowercase.
2. Cleaning is used to clean characters like https: @, #, and URL links.
3. Tokenizing is done to break sentences in the document into words.
4. Stopword Removal is done to remove words with low information content.
5. Stemming is done to remove affixes, both prefixes, and suffixes so that they return to the basic word forms.

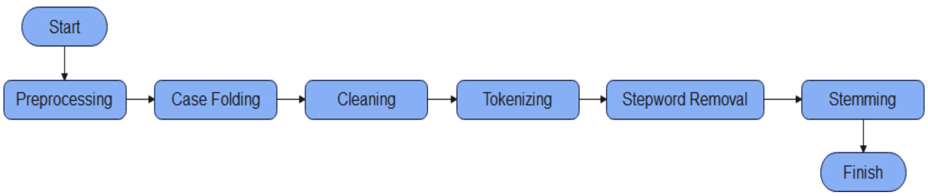

Figure 5. Preprocessing data flow

2.3.3 TF-IDF Weighting

In this stage, the weighting of words in the word tweet is composed using Term Frequency (TF) with Inverse Document Frequency (IDF). The word weight

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

barometer is done by looking for the results of words in the tweet where each word appears, then it is done by word cloud visualization.

$$W_{i,j} = tf_{i,j} \times \log(N/df_1) \quad (1)$$

Description:
$tf_{i,j}$ = The number of words –i in the jth document
N = Total documents
Df_1=Number of documents containing the word

2.3.4 Support Vector Machine (SVM) Classification

The Support Vector Machine (SVM) procedure is a method that can be used to analyze sentiment in this study. The results to be determined in this way are the classification of the positive class and the negative class obtained based on the weight of each feature of the text document.

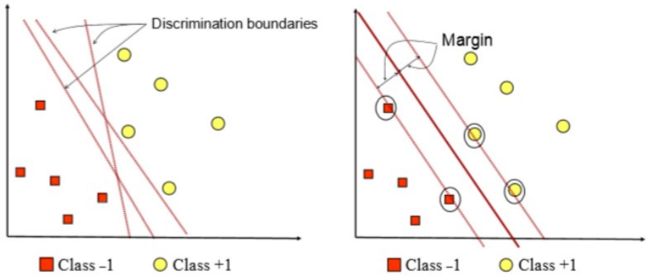

Figure 6. Hyperplane SVM

2.3.5 Random Forest (RF) Classification

This method is used to build a decision tree consisting of root nodes, internal nodes, and leaf nodes by obtaining attributes and data randomly by applicable rules, attributes, and acquisition value information.

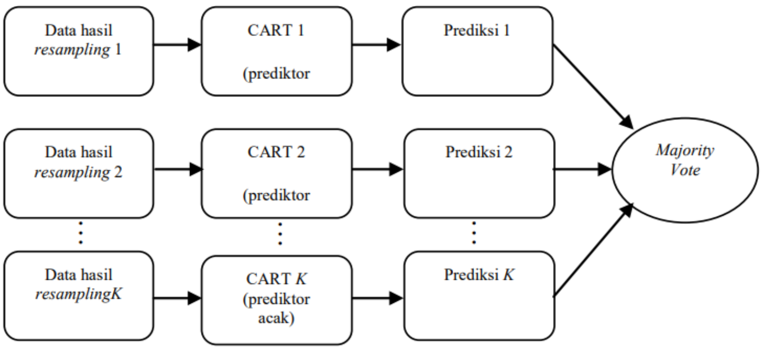

Figure 7. Workflow of Random Forest

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2.3.6 Comparison of Classification Performance

This classification comparison process is carried out by comparing the performance of the two algorithms, namely the Support Vector Machine (SVM) and Random Forest (RF). This comparison is carried out in several stages, namely:

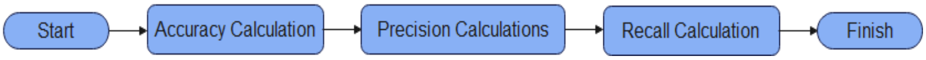

Figure 8. Comparison process flow

1. Accuracy
   Accuracy is defined as the closeness of the value of the correct prediction results from the actual value. In calculating the accuracy using the confusion matrix to calculate the percent accuracy of the classification model in the following formula.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

2. Precision
   Precision is a comparison to measure the accuracy of all dataset results and find out the actual category classification results. Precision is the result of the correct calculation (TP) divided by the amount of data identified by the system. The precision formula is as follows.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

3. Recall
   The success rate of a system in detecting a group becomes a recall parameter. The recall is a calculation of the division between the correct amount of data and the amount that should be. The recall formula is as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

Formula description:
TP (True Positive)
TN (True Negative)
FP (False Positive)
FN (False Negative)

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2.4 Make a Result Conclusion

Conclusions are drawn by focusing on the suitability of the problem formulation, the research questions and, the results obtained. The results obtained in this study are a comparison between the performance of the Support Vector Machine (SVM) and Random Forest (RF) algorithms.

3. RESULTS AND DISCUSSION

3.1 Results

3.1.1 Data Crawling and Data Aggregation

The data mining process was carried out from January 1 to February 31 2022. Before this research data was generated, the initial process was to synchronize to connect the Twitter application with the Rapid miner application so that clawer data was generated automatically using the #ibukotanegara and #IKNNusantara queries which then added operators select attribute to select just tweet sentiment text data then save it.

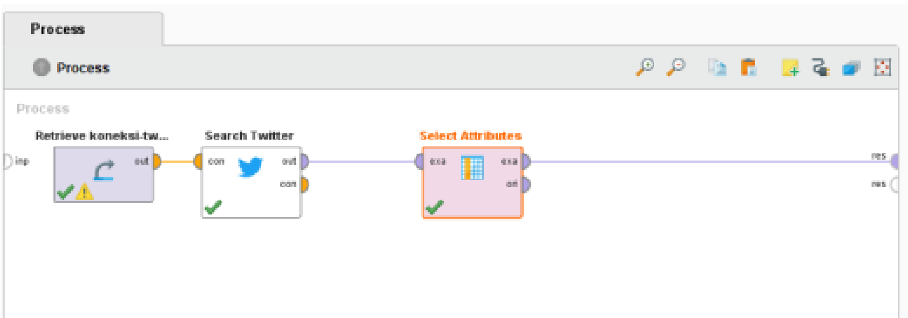

Figure 9. Crawling tweet data

The data generated in this study uses as much as 1,500 (one thousand five hundred) data by combining the results of tweet crawling data with different queries and then saving them in Comma Separated Values (CSV) format. The operators used in this process are:

1. Search Twitter, This operator is the main operator in data mining (data crawler). This operator can be integrated into the Twitter application by synchronizing through making connections in the view repository, some tools twoections are used to make connections to the Twitter application.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

In crawling this data, use the input query #state capital and #IKNnuasantara in the view parameters.

2. Select attributes, used to select the required data. In this case the data needed is text data sentiment.

The data aggregation stage is carried out using the execute process operator. In this operator, there is data from crawling results in the form of #ibukotabaru and #IKNnusantara queries. In merging the crawled data, it requires a support operator, namely the append operator, which in the main function of this operator is a data aggregation operator. The final stage in the data merging process requires the CSV write operator which is used to store the results of the data merging in CSV form. The following figure is a simulation of the process of merging data with different queries.

.

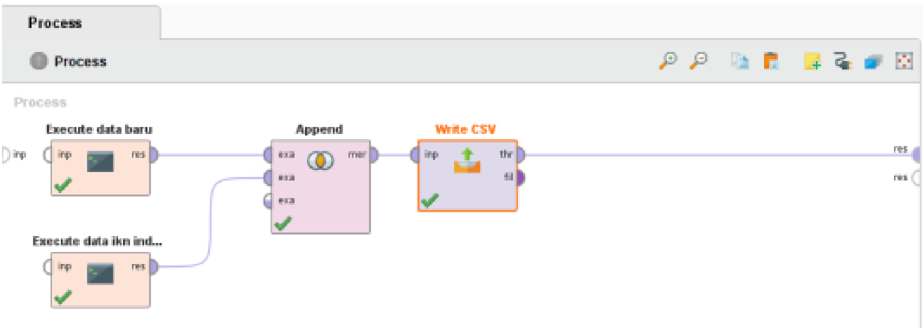

Figure 10. Data aggregation

3.1.2 Data Labelling

Data labeling is done manually with the provision that the label is a hoax with the reason that it contains words or news content that does not match the facts originating from the TurnBackHoax.id site, while non-hoax news is labeled originating from the online news sites kompas.com and detik.com which are News with news content contains actual hat occur in society. In this case using the ASE (Anti hoax search engine) tolls, Foller.Me tolls and hoax search tolls on the Hoax Buster Tools (Htools) application which is integrated with MAFINDO (Indonesian anti-slander community) or the TurnBackHoax.id website and to detect hoaxes so that the results of this analysis formed a data set of information or news including hoaxes or not. Then in the tweet sentiment data labeling is done manually with the provision that positive sentiment is in the form of an element of support and does not contain elements of harsh words, and vice versa in labeling negative sentiment.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Table 1. Labeling of hoaxes and non-hoaxes

NO	TWEET	LABEL
1	Peneliti inggris sebut calon ibu kota baru RI rawan Tsunami	Non Hoax
2	Dana haji untuk pembangunan IKN	Hoax
3	Rocky gerung mengatakan indonesia akan bangkrut jika memaksakan bangun IKN	Non-hoax
4	Pemindahan IKN ke Kaltim Perbanyak Kesempatan Kerja Bagi Masyarakat	Hoax
5	766 persen masyarakat Indonesia puas dengan kinerja Presiden Jokowi IKN Untuk Negeri	Non Hoax
6	Hasil Survei APSSI soal IKN, Sebanyak 48,2 Persen Masyarakat Minta Ditunda, Dominan Dampak Negatif	Non-Hoax
7	Ide Pemindahan Ibu Kota Sudah Ada Sejak Era Soekarno	Non Hoax
8	IKN Nusantara akan gunakan kendaraan tanpa awak sebagai transportasi publik	Non-Hoax
9	Peringati hari Kebangkitan Nasional, 8 organisasi pemuda lintas agama ikrar kebangsaan di Titik Nol IKN	Non-Hoax
10	Terdapat enam kluster ekonomi sebagai penggerak utama untuk mewujudkan visi Superhub Ekonomi IKN	Non-Hoax

Table 2. Sentiment labeling of tweets

NO	TWEET	LABEL
1	Melalui pembangunan kelestarian lingkungan tetap akan dijaga dan pemerataan ekonomi akan terwujud di Kalimantan.	Positive
2	Lanjutkan untuk Iindonesia Maju ibu kota Baru.	Positive
3	Dukung IKN Nusantara Penggerak Ekonomi Indonesia di Masa Depan.	Negative
4	Pemindahan IKN ke Kaltim Perbanyak Kesempatan Kerja Bagi Masyarakat.	Negative
5	Ratusan rakyat tak butuh ibu kota negara yg baru. Rakyat hanya butuh Rezim baru.	Negative
6	Pantas Elon Musk Kepincut Investasi di Indonesia Pemerintah Tawarkan Kawasan dekat IKN.	Negative
7	Pembangunan IKN Nusantara di Kalimantan Timur merupakan wujud betapa pentingnya posisi Kaltim secara geografis ekonomis hingga identitas sosial budaya bagi Indonesia.	Positive
8	Minyak naik itu karna maksa pingin bangun IKN investor kabur karna indonesia indeks korupsinya naik terus demokrasi menurun.	Negative
9	Iya njir masih ada aja yang nolak IKN Untuk Negeri padahal udah jelas keren banget dan manfaat untuk indonesia.	Positive
10	Alhamdulillah ditangan Jokowi Indonesia menjadi lebih baik IKN Untuk Negeri.	Positive

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.1.3 Data Preprocessing

The data processing is carried out in several stages, namely the process of sterilizing tweet data using a series of operators in the Rapid miner application. In the process of sterilizing this data, it produces clean data from several elements of the word RT (Retweet), Hastags, mentions, website links, and word attributes that are considered not important using the Replace/Cleaning operator.

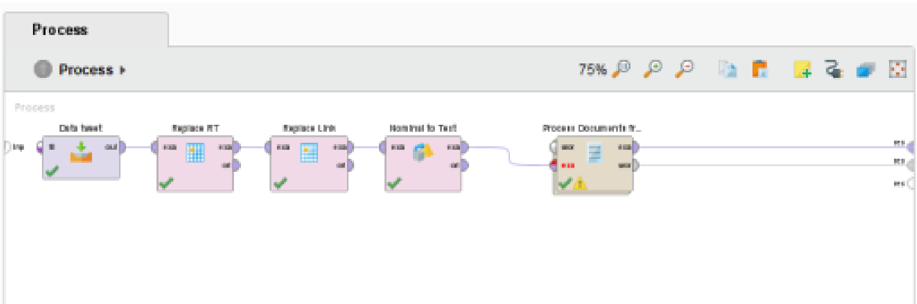

Figure 11. Preprocessing data

Next, the data sterilization stage is carried out using several operators which are packaged in sub-process operators to streamline the appearance of the view process in the Rapid miner application. The following is an explanation of some of the operators packaged in the sub-process operators used:

1. Tokenize for word separation in sentences. The process of tokenization of this text is a fragment of a tweet into a mapping of characters or word units that suit your needs. In the tokenization process, data filtering is carried out to retrieve words that have a minimum of 3 characters and a maximum of 25 characters, this step is carried out to maximize the analysis process. Words that are less than 4 characters or more than 25 characters will be deleted automatically.
2. Transform cases/case folding to change words to lowercase. This process will change the text that still contains uppercase or capital letters to lowercase letters or all lowercase letters. In this case, the aim is to carry out a classification model process that has letter uniformity so that errors do not occur in the next process.
3. Stopword filter/Stopword removal using an Indonesian dictionary to remove low-weight words. This stage takes important words from the tokenization results by using a stoplist algorithm (removing less important words). Stopwords are non-descriptive words that are discarded in the Indonesian language approach. The following examples of language stopwords are “yang”,” dan”,”di”,”dari”.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

4. Filter the token (light) to remove words that are too long and too short. In this process, the deletion of words is too short and is not considered important. For example, "ya","aq".

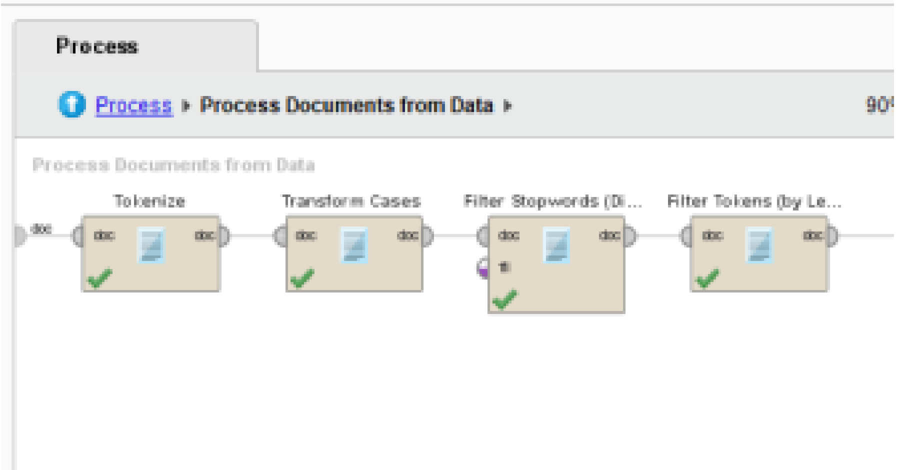

Figure 12. Sub-process operator process from data

3.1.4 TF-IDF Weighting

In this weighting is done to find out the number of occurrences of the word in the tweet data. The following is the result of the TF-IDF weighting process.

Table 3. TF-IDF weighting results

WORD	ATTRIBUTE NAME	TOTAL OCCURRENCES	DOCUMENT OCCURRENCES
Indonesia	Indonesia	532	468
Ibu kota negara	Ibukotanegara	422	422
Pimpin	Pimpin	229	225
Pembangunan	Pembangunan	156	149
Nusantara	Nusantara	145	140
Negeri	Negeri	113	112
Kota	Kota	56	48
Dukung	Dukung	47	47
Maju	Maju	47	47
Semangat nusantara	Semangat nusantara	45	45
Pemindahan	Pemindahan	44	44
Kalimantan	Kalimantan	38	32
Masyarakat	Masyarakat	38	37
Pemerataan	Pemerataan	38	38
Dana	Dana	35	35

Journal of Information Systems and Informatics
Vol. 5, No. 1, March 2023
p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Based on the results of the TF-IDF weighting, the results of the word Indonesia were found with a total of 532 occurrences (occurrence of data) out of a total of 468 occurrences of documents (total documents), then in the second most position, namely the word capital city of the country with a total of 422 occurrences (occurrence of data). of the total document occurrences (total documents) as many as 422.

3.1.5 Word Cloud visualization

Visualization was carried out to determine the appearance of the word text in the data resulting from the TF-IDF weighting process. This wordcloud visualization is the result of data processing on the previous TF-IDF weighting. Based on the results of this workload, it is set using neutral parameters, so that the size of the text that appears is equivalent. The results of the word cloud visualization are as follows:

Figure 13. Visualization results with the word cloud

3.1.6 Classification with Support Vector Machine (SVM)

The implementation of the SVM algorithm was carried out to classify tweet sentiment data and tweet data in the form of hoax and non-hoax news elements in this study followed by adding the apply model operator so that it could read the data. In the classification process, the prediction of data is determined by learning the algorithm so that the predictive data is valid. The following is a road map of the learning process of the algorithm in determining the predictions on the required labels.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

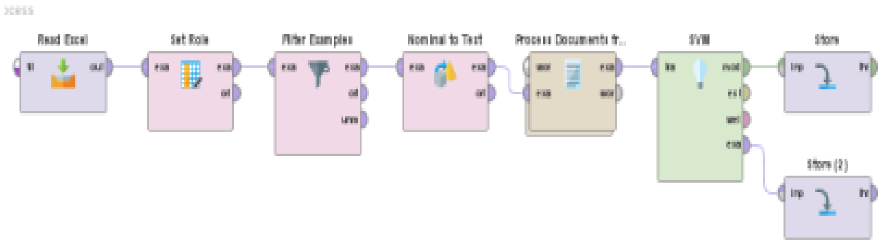

Figure 14. Algorithm learning process.

In the learning process of the SVM algorithm, there is a system operator composition as shown above. This process requires a store operator that is used to store the learning results of the algorithm. The learning outcomes of this algorithm are stored in the process repository and data repository. In inputting training data from manual hoax labeling results through detection results in the hoax buster application, the data in manual labeling is as much as 40% of the data. In this process learning the SVM algorithm so that the best prediction results can be found in data labeling and the process of measuring performance in the SVM algorithm.

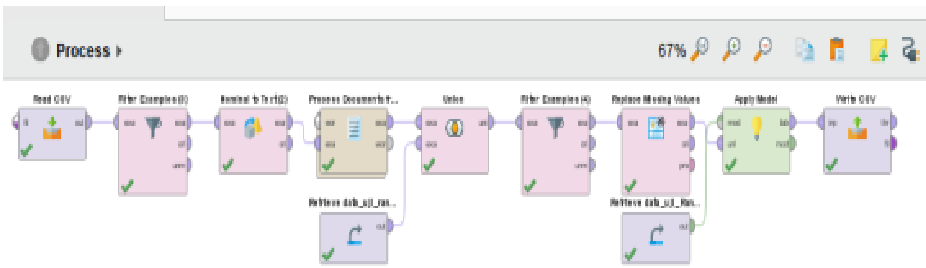

Figure 15. Data prediction classification process

This image is a road map for the process of testing data training and data testing to find the right labeling prediction results on sentiment data and hoax data. In this process, the predicted results of the SVM classification on the data used are then stored in CSV format so that they can be continued in the algorithm performance measurement process.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

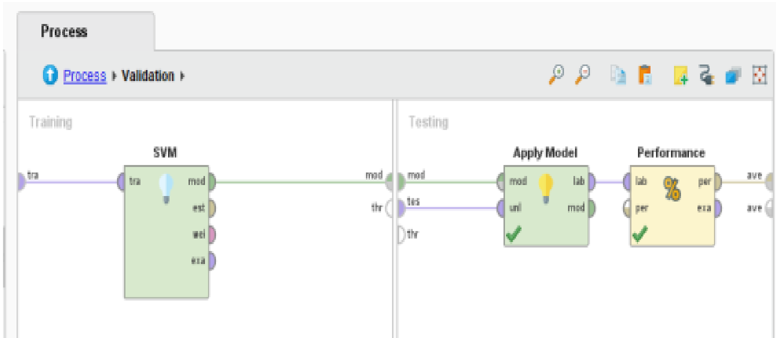

Figure 16. SVM classification road map

The results of the classification of hoax data from tweet sentiments from a total of 169 data on the Support vector machine algorithm using 30% training data and 70% testing data so that you get Performance Vector results with an accuracy value weight: 95.24%, a precision value weight: 95.18 % (positive class: non-hoax), weighted recall value: 100.00% (positive class: nonhoax), weighted AUC value (optimistic): 0.610 (positive class: nnon-hoax AUC: 0.610 (positive class: non-hoax), weighted AUC value (pessimistic): 0.610 (positive class: non-hoax). In this classification process, 10 hoax news and 159 non-hoax news were found.

The results of the classification of tweet sentiment data from a total of 366 data in the Support vector machine algorithm using 60% training data and 40% testing data so that you get Performance Vector results with an accuracy value weight: of 84.07%, a precision value weight: 66.67% (positive class: negative), recall score: 6.67 % (positive class: negative), AUC score (optimistic): 0.819 (positive class: negative) AUC: 0.817 (positive class: negative), AUC score (pessimistic): 0.816 (positive class: Negative) with the Confusion Matrix arrangement in the following table:

Table 4. Confusion matrix

	Accuracy	**True**	**Hoax**	**Non-hoax**
		Hoax	1	0
		Non-hoaxes	4	79
	Precision	**True**	**Hoax**	**Non-hoaxes**
Hoax classification		Positive	1	0
		Non-hoaxes	4	79
	Recall	**True**	**Hoax**	**Non-hoaxes**
		Hoax	1	0
		Non-hoaxes	4	79
	Accuracy	**True**	**Positive**	**Negative**
		Hoax	151	28

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

		Non-hoaxes	1	2
Classification of tweet sentiments	**Precision**	**True**	**Positive**	**Negative**
		Positive	151	28
		Negative	1	2
	Recall	**True**	**Positif**	**Negative**
		Positive	151	28
		Negative	1	2

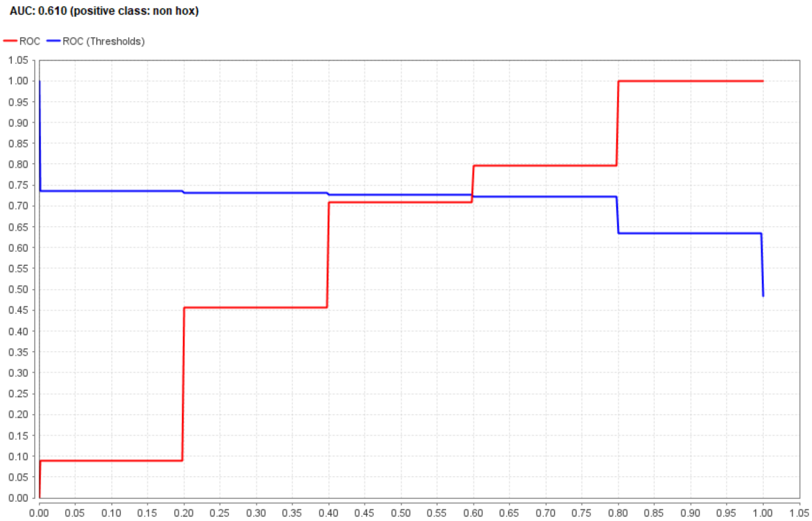

Figure 17. ROC Results of the classification of hoax news predictions

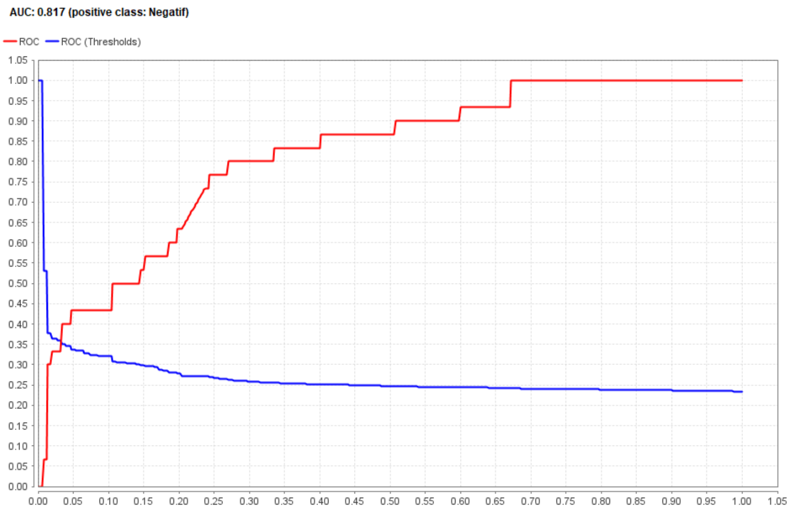

Figure 18. ROC results of svm sentiment classification

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3.1.7 Classification with Random Forest (RF)

Implementation of the RF algorithm is carried out to classify research data followed by adding the apply model operator so that it can read the data.

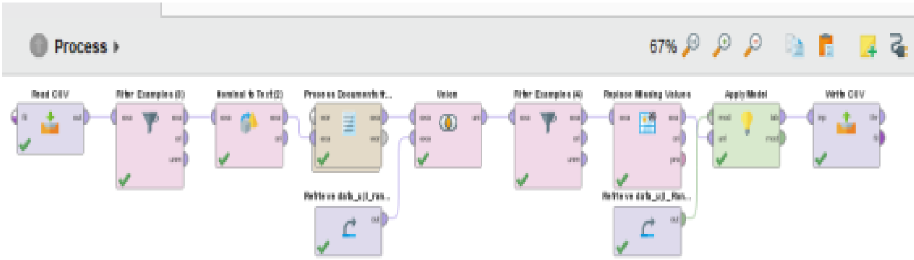

Figure 19. Data prediction classification process

This image is a road map for the process of testing data training and data testing to find the right labeling prediction results on sentiment data and hoax data. In this process, the results of the prediction of the Random forest q classification on the data used are then stored in CSV format so that they can be continued in the algorithm performance measurement process.

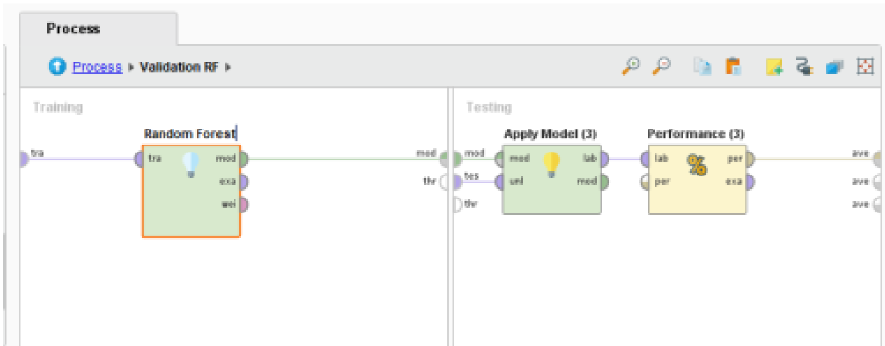

Figure 20. Random Forest classification road map

The results of the classification of hoax data from tweet sentiments from a total of 169 data on the Random Forest algorithm using 30% training data and 70% testing data so that you get Performance Vector results with an accuracy value weight: of 86.90%, a precision value weight: 87.95 % (positive class: non-hoax), weighted recall value: 98.65% (positive class: non-hoax), AUC value-weighted (optimistic): 0.746 (positive class: non-hoax) AUC: 0.744 (positive class: nonhoax), weighted AUC value (pessimistic): 0.742 (positive class: non-hoax). In this classification process, 21 hoax news and 148 non-hoax news were found.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The results of the classification of the Random forest algorithm using 60% of training data and 40% of testing data results in a Performance Vector with an accuracy weight of 83.52%, a precision value of unknown (positive class: negative), and AUC (optimistic) weight. : 0.799 (positive class: Negative) AUC: 0.798 (positive class: Negative), AUC value weight (pessimistic): 0.796 (positive class: Negative) with the Confusion Matrix arrangement in the following table:

Table 5. Confusion matrix

Hoax classification	**Accuracy**	**True**	**Hoax**	**Non-hoax**
		Hoax	0	1
		Non-hoaxes	10	73
	Precision	**True**	**Hoax**	**Nonhoax**
		Positive	0	1
		Non-hoaxes	10	73
	Recall	**True**	**Hoax**	**Nonhoax**
		Hoax	0	1
		Non-hoaxes	10	73
Classification of tweet sentiments	**Accuracy**	**True**	**Positive**	**Negative**
		Hoax	152	30
		Non-hoaxes	0	0
	Precision	**True**	**Positive**	**Negative**
		Positive	152	30
		Negative	0	0
	Recall	**True**	**Positive**	**Negative**
		Positive	152	30
		Negative	0	0

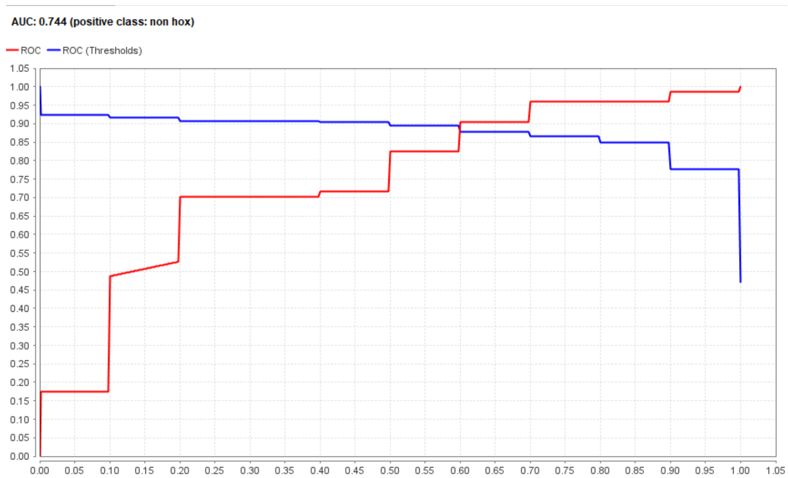

Figure 21. ROC classification results for predictions of hoax news

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

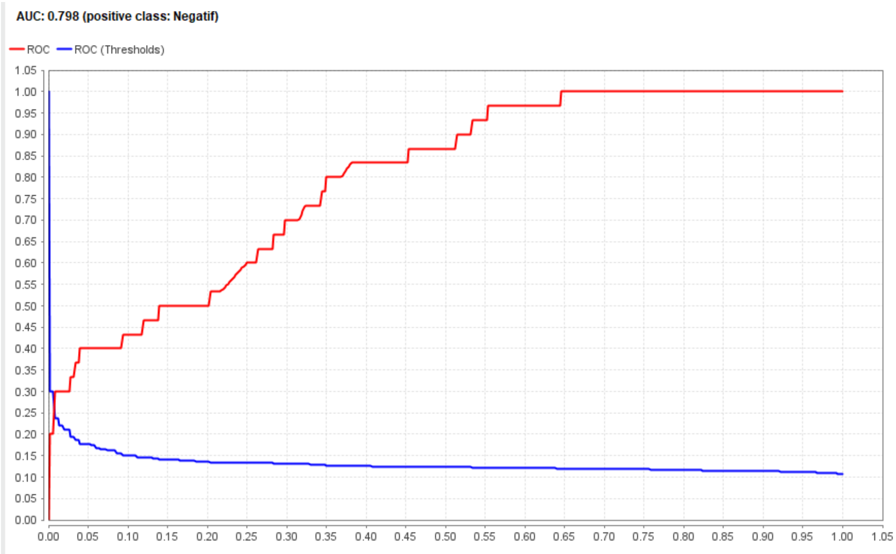

Figure 22. ROC Random Forest sentiment classification

3.1.8 Algorithmic Performance Comparison

In the process of comparing the SVM and RF algorithms, this is done by following several process operators, in this case, the main operator used is the multiply operator. This multiply is an operator that can combine classification algorithms so that the resulting output is the needs of the comparison.

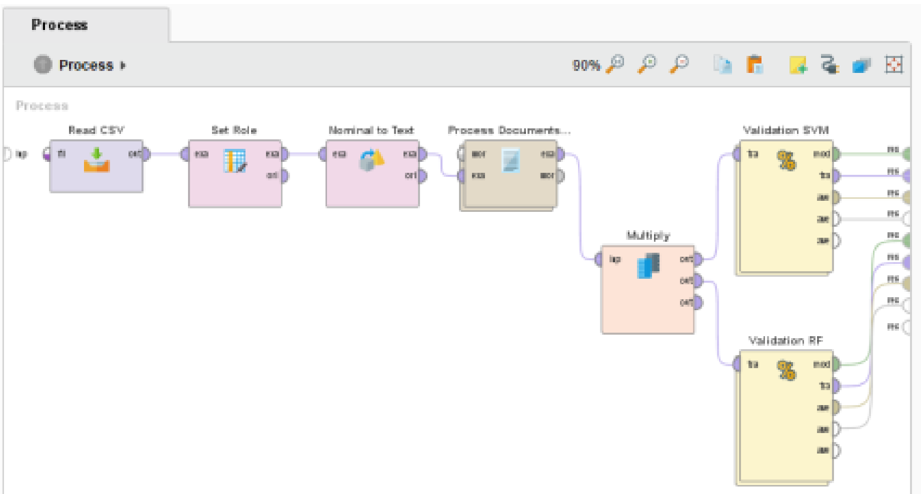

Figure 23. Algorithmic comparison road map

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

At this comparison stage, the best results were found from a comparison of performance in hoax news classification and tweet sentiment classification performance in the data used. Accuracy, precision, recall, and also A, UC (Area under the curve) the performance. In the comparison of the hoax classification performance of these two algorithms, the Support vector machine algorithm gets the best accuracy value with an accuracy value weight: 95.24% while the Random Forest algorithm gets an accuracy value weight of 86.90%. the weight of the precision value of the SVM algorithm is also better than the RF algorithm with a weight value of 95.18% as well as the weight of the recall and AUC values in the SVM algorithm is greater than the RF algorithm. In the tweet sentiment classification performance of the data used, the performance of the SVM algorithm dominates the weight of the resulting value, namely with a weight value of 84.07% accuracy, 66.67% precision, AUC 0.817. what is the RF algorithm getting an accuracy value of 83.52%, and the weight value of precision: unknown AUC: 0.798. The following is a comparison diagram of the performance of the algorithms in the hoax classification process and tweet sentiment classification in the data used. The following is a comparison diagram of the classification performance of the SVM and RF algorithms.

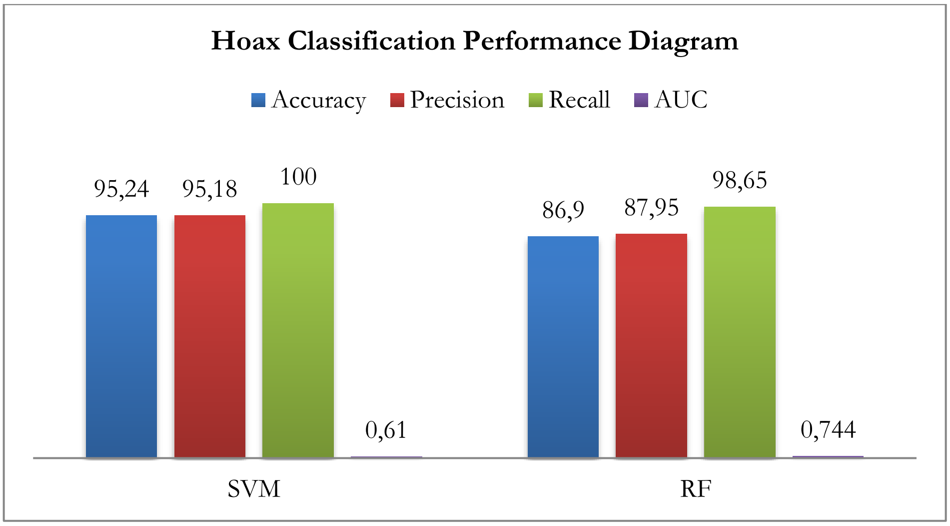

Figure 24. Comparison of hoax classification performance diagram for SVM and RF algorithm

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

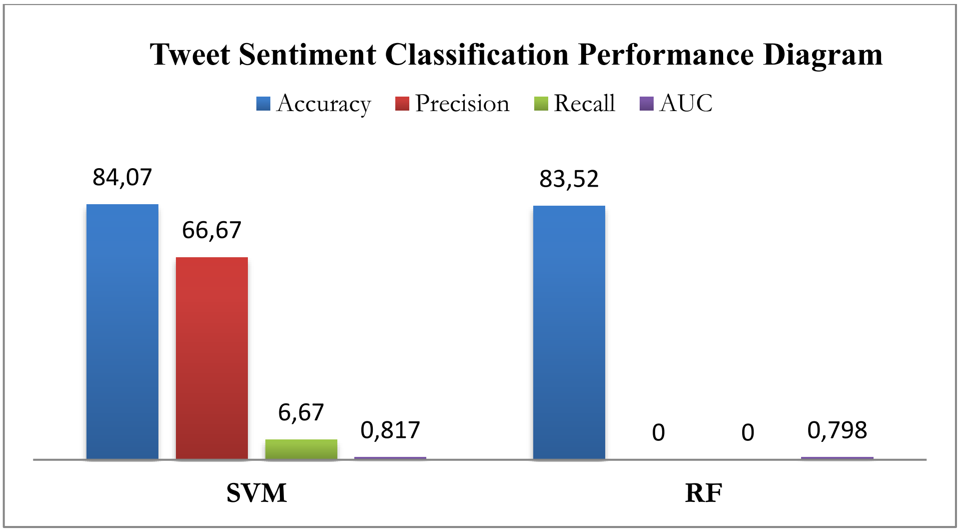

Figure 25. Comparison of tweet sentiment classification performance diagram with SVM and RF algorithm

3.2 Discussions

This study aims to compare the performance of the Support Vector Machine (SVM) and Random Forest (RF) algorithms in the Hoax Analysis of the policy of moving the Indonesian state capital from Jakarta to North Penajamthe Paser Regency. The results showed that the SVM (95.24%) algorithm is better than RF (86.90%). These results corroborate the research of Lestari et al (2022), that the accuracy of SVM (85.71%) is better than NB (76.70%) and KNN (52.74%)[8]. Also research by Lestari, et al (2022) on SVM with Chi-Square Feature selection shows that SVM is a very good algorithm with an accuracy value of 90%, an averaecision value of 90%, 86% recall, and an f1-score of 88%[10]. The research results of Arsi and Waluyo (2021)[9], also confirm that SVM is a good algorithm with accuracy = 96.68%, precision = 95.82%, recall = 94.04% and AUC = 0.,979. Meanwhile, SVM with PSO resulted in an increase in accuracy of 2.09% from the previous accuracy of 79.06% to 81.15% in the "Good Classification" category[11].

In a study conducted by Baita et al. (2021) explained that the SVM algorithm has better performance than the KNN algorithm, with English data sourced from Twitter[17]. However, Pertiwi's research (2019)[19] explains that KNN has higher accuracy than SVM, NN, Naïve Bayes fand, or English text classification on Twitter. Furthermore, Nasution & Hayaty (2019)[20] explained that SVM accuracy is superior by 88.70% without using K-Fold Cross Validation. As for the processing time calculation of 0.016s without K-Fold Cross Validation. The combination of SVM with unigram tokenization, Indonesian stopword lists, and emoticons has a good accuracy value[21].

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

In summary, our study demonstrates that SVM outperforms RF in the context of Hoax Analysis concerning the policy of moving the Indonesian state capital from Jakarta to North Penajamthe Paser Regency. These findings are consistent with earlier studies that have demonstrated the superior performance of SVM over other machine learning algorithms in various contexts.

4. CONCLUSION

In conclusion, this study utilized the Support Vector Machine (SVM) and Random Forest (RF) algorithms to analyze the hoax issue of relocating the Indonesian state capital on Twitter. The data was filtered by Hoax Booster Tools (HBT) ASE, labeled, preprocessed, and classified using the TF-IDF weighting method. The results indicated that 85% of the tweets had a positive sentiment while 15% were negative. Moreover, the SVM algorithm outperformed the RF algorithm, achieving a classification accuracy of 95.24% compared to RF's 86.90%.

Further research should explore a variety of methods in Traditional Machine Learning (TML) and Deep Learning (DL) and compare their performance to gain insight into various algorithms. Additionally, future studies should not limit themselves to Twitter and instead examine a more diverse range of social media platforms such as Facebook, Instagram, and YouTube comments. The focus of the research should be on Social Networks Analytics, Hoax and Community Detection, Fake and Social Media Forensics, Trend Analysis, Actor and Narration Analysis, Topics & Geolocation Analysis, Emotion & Demographic Analysis, and Bot and Diffusion Analysis concerning Policy Issues on the Relocation of the National Capital.

ACKNOWLEDGEMENT

Thank you to the Information Systems Study Program and the Institute for Research and Community Service (LPPM) at Universitas Islam Madura for their support in 2022.

REFERENCES

[1] F. Hadi and R. Ristawati, "Pemindahan Ibu Kota Indonesia dan Kekuasaan Presiden dalam Perspektif Konstitusi," JK, vol. 17, no. 3, pp. 530–557, Nov. 2020, doi: 10.31078/jk1734.

[2] Tim Kompas, "Kepala Bappenas Umumkan Nama Ibu Kota Baru: Nusantara. Kompas.Com." 2022. [Online]. Available: https://nasional.kompas.com/read/2022/01/17/12302621/kepala-bappenasumumkan-nama-ibu-kota-baru-nusantara

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

[3] D. A. Ramadhan, "Analisis Sentimen Program Acara Di Sctv Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine," e-Proceeding of Engineering, vol. 6, no. 2, p. 9736, 2019, doi: https://doi.org/10.34818/eoe.v6i2.10708.

[4] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," Edutic, vol. 7, no. 1, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.

[5] A. M. Hidayat and M. Syafrullah, "Algoritma Naïve Bayes Dalam Analisis Sentimen Untuk Klasifikasi Pada Layanan Internet PT.XYZ," Jurnal TELEMATIKA MKOM, vol. 9, no. 2, 2017.

[6] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," JTK, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.

[7] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," JATISI, vol. 8, no. 2, pp. 636–646, Jun. 2021, doi: 10.35957/jatisi.v8i2.962.

[8] S. Lestari, M. Mupaat, and A. Erfina, "Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter," JUSIFO: J. Sistem Inf., vol. 8, no. 1, pp. 13–22, Jun. 2022, doi: 10.19109/jusifo.v8i1.12116.

[9] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," JTIIK, vol. 8, no. 1, p. 147, Feb. 2021, doi: 10.25126/jtiik.0813944.

[10] S. Lestari, "ANALISIS SENTIMEN IBU KOTA NEGARA BARU INDONESIA PADA TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN SELEKSI FITUR CHI SQUARE," Nusa Putra University, SUKABUMI, 2022.

[11] P. Arsi, R. Wahyudi, and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia," RESTI, vol. 5, no. 2, pp. 231–237, Apr. 2021, doi: 10.29207/resti.v5i2.2698.

[12] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial," J.Gauss, vol. 9, no. 3, pp. 237–246, Aug. 2020, doi: 10.14710/j.gauss.v9i3.27963.

[13] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," Bulletin EEI, vol. 9, no. 4, pp. 1620–1630, Aug. 2020, doi: 10.11591/eei.v9i4.2352.

[14] M. I. D. Sakariana, "Analisis Sentimen Pemindahan Ibu Kota Indonesia Dengan Pembobotan Term BM25 Dan Klasifikasi Neighbor Weighted K-Nearest Neighbor," JTIIK, vol. 4, no. 3, pp. 748–755, Mar. 2020.

[15] A. Sa'rony, P. P. Adikara, and R. C. Wihandika, "Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan

Algoritme Term-Based Random Sampling dan Metode Klasifikasi Naïve Bayes," JTIIK, vol. 3, no. 10, pp. 10086–10094, Oktober 2019.

[16] A. H. Dyo fatra, N. H. Hayatin, and C. S. K. Aditya, "Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia," JR, vol. 2, no. 11, p. 1562, Dec. 2020, doi: 10.22219/repositor.v2i11.933.

[17] A. Baita, Y. Pristyanto, and N. Cahyono, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," Information System Journal (INFOS), vol. 4, no. 2, 2021.

[18] A. Muzakir, H. Syaputra, and F. Panjaitan, "A Comparative Analysis of Classification Algorithms for Cyberbullying Crime Detection: An Experimental Study of Twitter Social Media in Indonesia," Sci. J. Informatics; Vol 9, No 2 Novemb. 2022DO - 10.15294/sji.v9i2.35149 , Oct. 2022, [Online].

[19] M. W. Pertiwi, "Analisis Sentimen Opini Publik Mengenai Sarana Dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, KNN dan SVM," vol. 14, no. 1, 2019.

[20] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," JI. Jurnal. Informatika, vol. 6, no. 2, pp. 226–235, Sep. 2019, doi: 10.31311/ji.v6i2.5129.

[21] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine," Jurnal Dinamika Informatika, vol. 5, no. 2, 2016.