

Original Research Paper

Hoax Classification Corona Virus (COVID-19) News in Indonesian using the Support Vector Machine (SVM) Method

Wiranto Herry Utomo and Karno Juni Prayoga

Department of Information Technology, President University, Bekasi, Indonesia

Article history

Received: 29-04-2021

Revised: 01-06-2021

Accepted: 24-06-2021

Corresponding Author:

Wiranto Herry Utomo

Department of Inf. Technology,

President University, Bekasi,

Indonesia

Email: wiranto.herry@president.ac.id

Abstract: The development of information technology more widely, rapidly and quickly provide convenience to the public in access information. Internet is a container or online media makes the information hasn't been verified or proved to be true which is rapidly spreading in the community. The purpose of this study is to facilitate in determining the hoax news or facts about corona virus in Indonesia and appoint the performance text mining classification with SVM algorithm. Stages The hoax classification process is carried out with the preprocessing then weighting is carried out using TF-IDF method and classified using the support vector machine algorithm then tested by cross validation and k-folds testing. The data used in this study consisted of 535 text message containing information about facts and news text 425 contains information on hoaxes. The results obtained on testing the highest accuracy to 7 on the k-fold 9 with accuracy of 83.82%. Thus, the SVM algorithm can be used in the classification of Corona Virus Hoax news or COVID-19.

Keywords: Hoax News, Corona Virus, SVM Algorithm, K-folds, Cross Validation

Introduction

The development of information technology is increasingly widespread, quick and fast enough to give the public convenience in accessing information. According to one website, the number of internet users in Indonesia is increasing. Indonesia is one of the countries with the highest internet access, number 3 in the world. About 64% or 174 million Indonesians actively access the internet with an average duration of 7 h 59 min and almost 8 h a day (Wearesocial.com, 2020).

Internet is a container or online media makes the information has not been verified or proved to be true which is rapidly spreading in the community. The public as consumers of information is said to have not been able to distinguish between true or false information (hoax). This hoax news is widely spread on the internet through several platforms. There are forms or manifestations of hoax news, not only in writing, but in the form of photos, sounds or videos. Some hoax news spread in the community, one of which is news about the corona virus or COVID-19, which is increasing in number along with the ongoing pandemic. According to the Ministry of Communication and Informatics (Kemkominfo) found 1,125 hoax news on social media related to the corona virus or COVID-19 (Kumparan.com, 2020).

Hoax are pervert and harmful information because they pervert perceptions humans by communicate false information as truth. Hoax can aim to influence readers with false information so that readers take action according to the contents of the hoax (Afriza and Adisantoso, 2018). At this time, we often have difficulty how to distinguish between fake news (hoax) or true news. To find out the news that is spread is included in hoax news or true news, therefore a text classification algorithm is needed. Document classification is grouping a set of models that describe and differentiate data classes according to the categories contained in the document. The purpose of classification is to predict the class of objects whose class and data type characteristics are not known yet (Afriza and Adisantoso, 2018). Before a text data or document is classified, analysis is carried out at the preprocessing stage.

The stages in preprocessing include tokenizing, case folding, stop words and stemming. One of the classification techniques that are often used is the Support Vector Machines (SVM) method. Of the several classification techniques that are most often used is the Support Vector Machines (SVM) method. Previous research related to hoaxes was carried out by Rahmat and Areni (2019) who did the detection of hoax news using the Support Vector Machines (SVM) method with 100 training data and 20 test

data resulting in an accuracy of 85%. Likewise, research conducted by Honakan *et al.* (2018) who analyzed and implemented the Support Vector Machine method with the Kernel String in Classifying Indonesian News by grouping news into 3 parts or classes, namely government, economy and sports resulted in an accuracy of 47.43%. Research conducted by Jaya and Aulia (2018) classifies scientific paper documents using the Support Vector Machine algorithm using 150 training data and 50 test data resulting in 90% accuracy. Support Vector Machine is a technique for classifying. The advantage of a Support Vector Machine (SVM) is that it is relatively easy to implement because the determination of Support Vector can be formulated in the QP problem. The way the Support Vector Machine (SVM) method works is by looking for the hyperplane with the largest margin. The data closest to the dividing line in each class is called a Support Vector (Jaya and Aulia, 2018; Susilo *et al.*, 2020; Asiyah, 2016; Feldman and Sanger, 2007; Gupta and Bhathal, 2018; Maulina and Sagara, 2018; Prayoga *et al.*, 2019; Williams and Simoff, 2006; Zy, 2017).

Literature Review

This study discusses the classification of hoax news in Indonesian using the Support Vector Machine (SVM) algorithm. The dataset used in this research is Indonesian language news taken from the archives of the Anti-Defamation Forum for Hasut and Hoax (FAFHH) on the turnbackhoax.id site. This study used 100 training data and 20 test data. Algorithm that applies to Support Vector Machine (SVM). The results obtained from this study are accurate at 85% (Rahmat and Areni, 2019). This study contains the implementation of the Support Vector Machine (SVM) algorithm for classification of Indonesian news. The word weighting used is tf-idf and tf-chi square. The data is divided into 3 parts or classes, namely government, economy and family. Text mining is used to make it easier to manage information and classify existing data. The results obtained in this study amounted to 47.43% (Honakan *et al.*, 2018). This research discusses the classification of scientific paper documents. Prior to the conference, scientific papers must be grouped according to their categories. Classification is used to classify documents according to their categories. This study uses 150 training data and 50 test data and is divided into 5 categories of computer science, namely computer systems, data mining, graphics and design, computer interaction and information security. The results obtained from this study are an accuracy of 90% (Jaya and Aulia, 2018).

Methodology

Object of Research

In preparing this research the author conducted research on the classification of hoax news on the corona virus or COVID-19 using SVM method. The object of

research is in the form of news about the corona virus or COVID-19 taken from the website turn back hoax and cekfakta. tempo. The choice of objects was based on the fact that starting from January 2020, there was a lot of news about the corona virus or COVID-19 which was still confusing. So that it took a lot of attention from the world community including the Indonesian people.

Framework of Research

In this study, the data set used was the dataset taken from the website turnbackhoax.id. The data taken is hoax news. Initial data processing is preprocessing. by doing the cleansing process to remove noises, after that the process of normalizing words to eliminate non-standard words, then at the preprocessing stage there is also a document process. The document process uses the Rapid Miner which includes transform cases, tokenize, filter stop words, stemming, using the TF-IDF indicator. Meanwhile, the proposed method (Proposed Method) is a classification process using SVM algorithm. The objectives of this study are to determine the results classification based on the proposed model, namely the cross validation technique with indicators using the k-folds value, which will then be measured/measurement evaluation using confusion matrix.

Research Stages

The research stages that will be carried out are as follows (Fig. 1 - 27).

Data Collection

Data collection is done by crawling the websites turnbackhoax.id and cekfakta.tempo.co.id. The data collected is text data in Indonesian which is taken from the website, which is about the COVID-19 hoax.

B. Preprocessing

Preprocessing is a process to get clean data so that the next process can be carried out, this process is done manually at this stage which is done as follows.

Data Selection

At this stage it aims to select news, namely to retrieve news data used, news data which consists of data that is not needed, in this study it will only take news about the COVID-19 or corona virus.

Cleansing

The cleansing stage in this study is to remove words, characters and symbols that are not needed in the research process, namely as follows:

- a. HTML Character (<, >, dll)
- b. Url (<http://website.com>)
- c. E-mail (nama@website.com)
- d. Punctuation (“,”,”,”,”?””, dll)

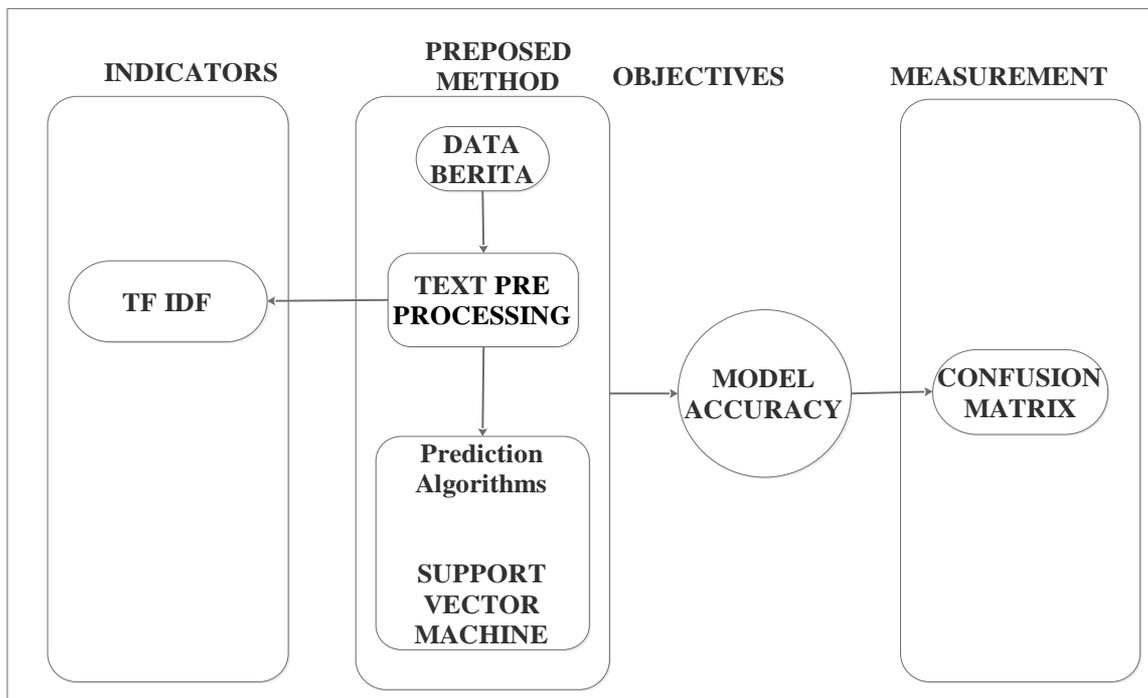


Fig. 1: Framework

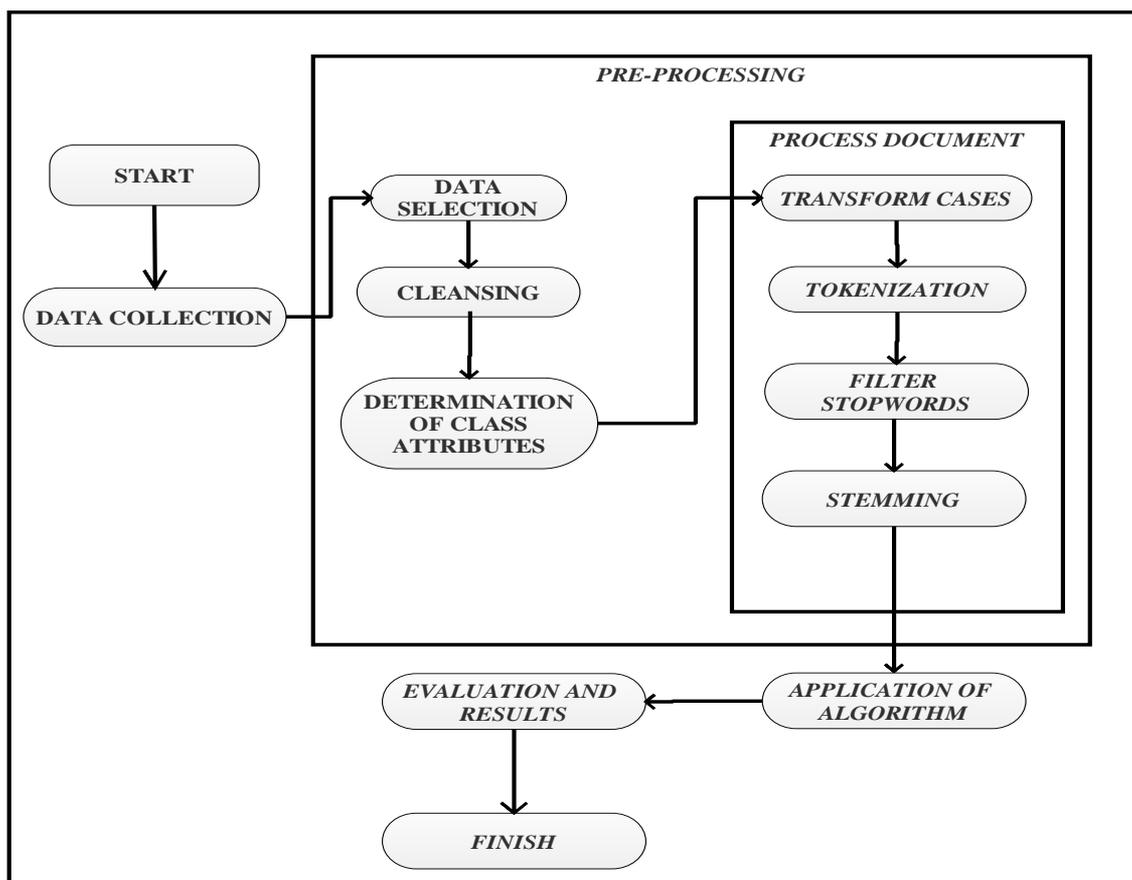


Fig. 2: Research stages

Determination of Attribute Class

The news data that has undergone processing at the preprocessing stage will then determine the class of its attributes, the determination of the attribute class is given according to the subjectivity of the researcher. The division of attribute classes in the study is divided into facts and hoaxes.

Data Sharing

Data sharing aims to obtain training data and test data which will later be used in the testing process using the Rapid Miner Tools. The data used amounted to 958 Indonesian language news texts obtained from the website account turn back hoax and cekfakta.tempo.co. The data consists of 535 fact information news texts and 425 hoax information news texts. Later, each of these data will be divisible into 80% as training data and 20% and test data or 760 respectively as training data and 198 as test data.

C. Process Document

Pre-processing at this stage is carried out using the Rapid Miner tool, where the process uses several series of sub-process operators in the process of document from data in the series. Pre-processing using Rapid Miner is applied to training data and testing data. The components used are Transform Cases, Tokenize, stop words Filters, Stemming. Here is the explanation.

Transform Cases

The stages of transform cases in this study are to change or uniform all forms or character letters to lower all letters (lower cases) in documents or text, because they have non-uniform letterforms.

Tokenization

In this process, separating all the words in each document into word pieces (terms).

Filter Stop words

At this stage, the words that are not relevant will be deleted, words that do not have a special meaning if they are separated by another word and is not bound by the adjective associated with sentiment.

Stemming

Stemming is a process of searching for basic words by removing affixes. In this process the words will be grouped into several groups which have the same root word. In this research, the stemming used is from the stemmer literary library which is built based on the Nazief and Andriani algorithms.

D. Application of Algorithms

Support Vector Machine (SVM) is a learning algorithm (machine learning). The basic concept of this algorithm is actually a harmonious combination of computational theories that have existed decades before, one example is the hyperplane margin. The working principle of the Support Vector Machine (SVM) is to determine the dividing line or hyperplane with the largest margin value. Hyperplane is a boundary line or decision boundary for data between classes, while margin is a distance between hyperplane and the closest data in each class. The data closest to the hyperplane in each class can later be referred to as a support vector (Nugroho *et al.*, 2008). Algorithms that will be used in this research that Support Vector Machine (SVM) is relatively easy to implement method for classification of documents.

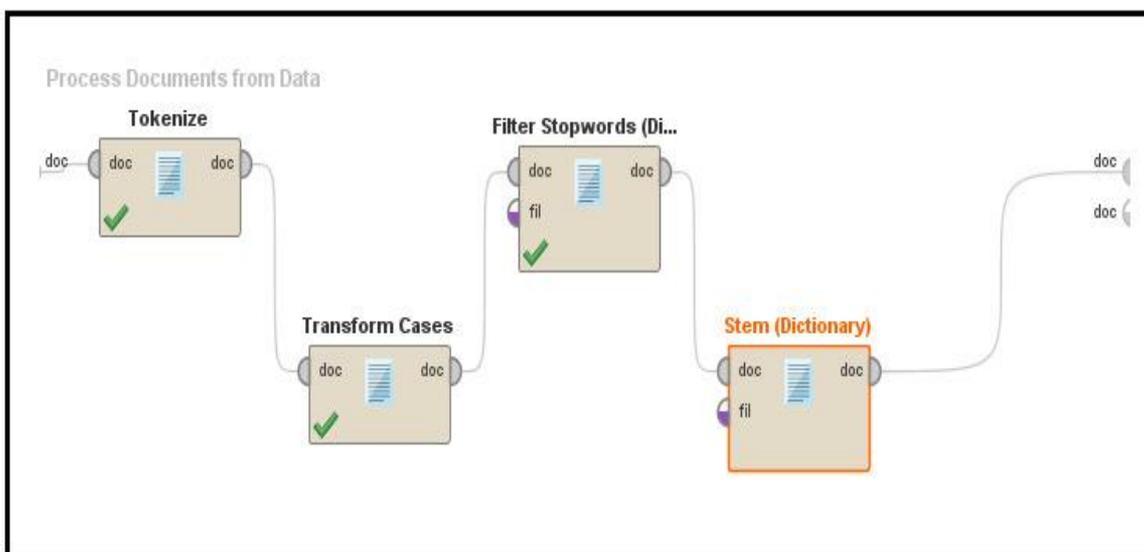


Fig. 3: Process document

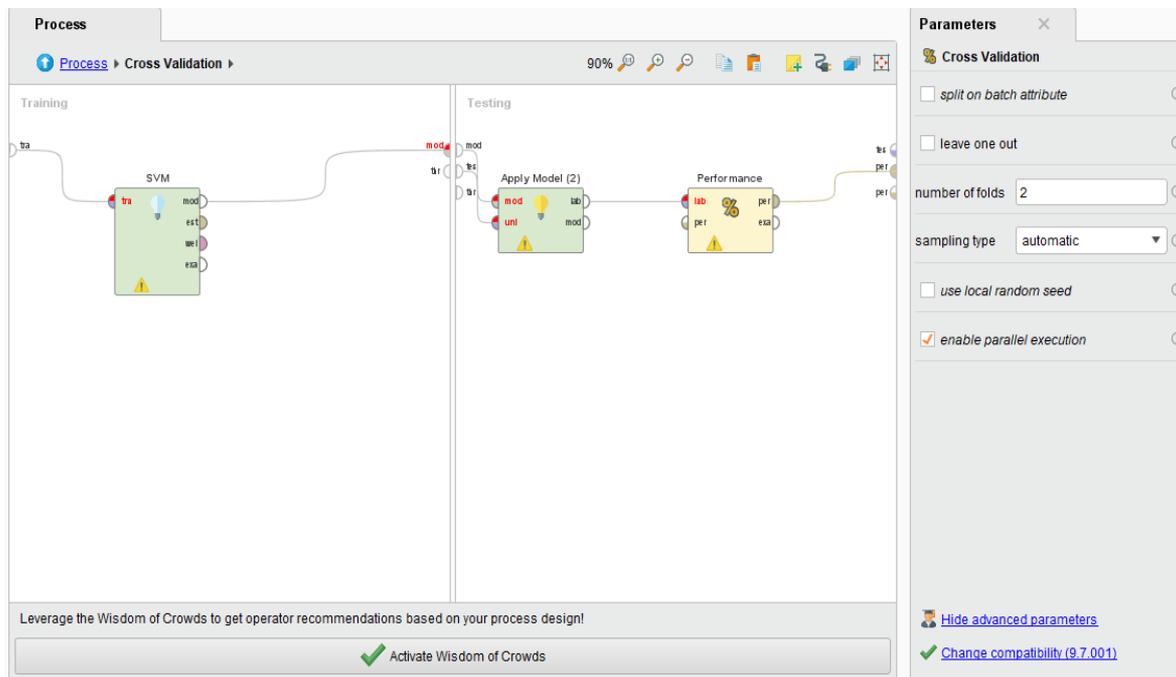


Fig. 4: SVM process

E. Evaluation Dan Validation of Result

In this study, an evaluation was carried out to determine the accuracy and performance of SVM algorithm for the classification of hoax news. Validation aims to compare the results of the accuracy of the method or model used with existing results. The validation technique used in this study is Cross Validation and K-folds. This test aims to determine the accuracy and performance of each testing technique

Results and Discussion

Testing Process

The test is carried out following a predetermined flow, namely the Support Vector Machine (SVM) classification method. This testing process uses Rapid Miner by combining several appropriate process operators, namely read excel, set roles, nominal to text, process document form data, cross validation and apply model as well. The main image of some of the experimental processes can be seen in the image below.

The testing process is also applied the k-fold method that aims to determine the value of the accuracy of each test dataset in order to get the best accuracy. This test is carried out in each of each dataset which is divided into 10 test datasets which are divisible into training data and testing data. K-fold the value used in this study is from 2 to 10. To be more explicit about the testing process following is a description of the testing string Table 1 - 11.

Testing

A. Testing Dataset 1

In this test testing first dataset predetermined. This test uses a string of models that have been described previously. Here is one example of a string of first testing process.

Based on the above test results obtained an accuracy of test results dataset unity. The result can be seen in the confusion matrix below.

The image above is the test results of one of the first dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as shown below.

B. Testing Dataset 2

In this test testing second dataset predetermined. This test uses a string of models that have been described previously. Here is one example of the series of the second test process.

Based on the above test results obtained an accuracy of test results the second dataset. The result can be seen in the confusion matrix below.

The image above is the test results of one of the second dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as shown below.

C. Testing of Dataset 3

In this test testing third dataset predetermined. This test uses a string of models that have been described previously. Here is one example of the series of the third test process.

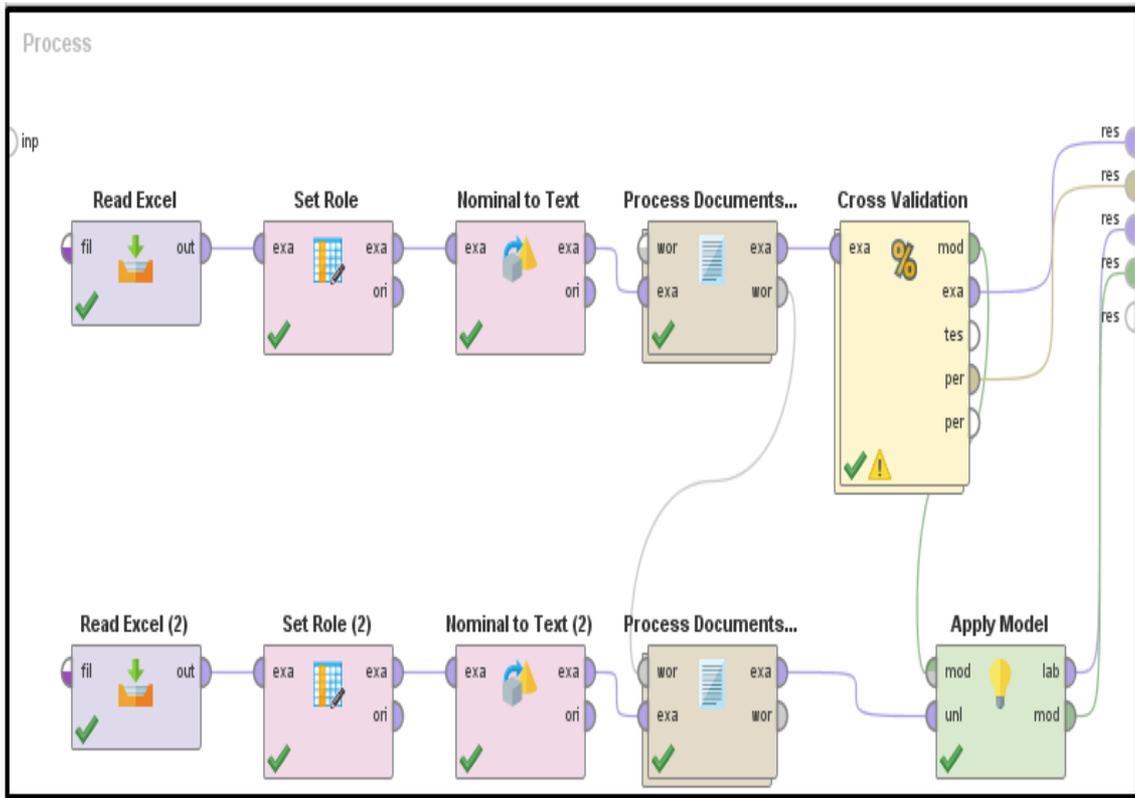


Fig. 5: Main testing process

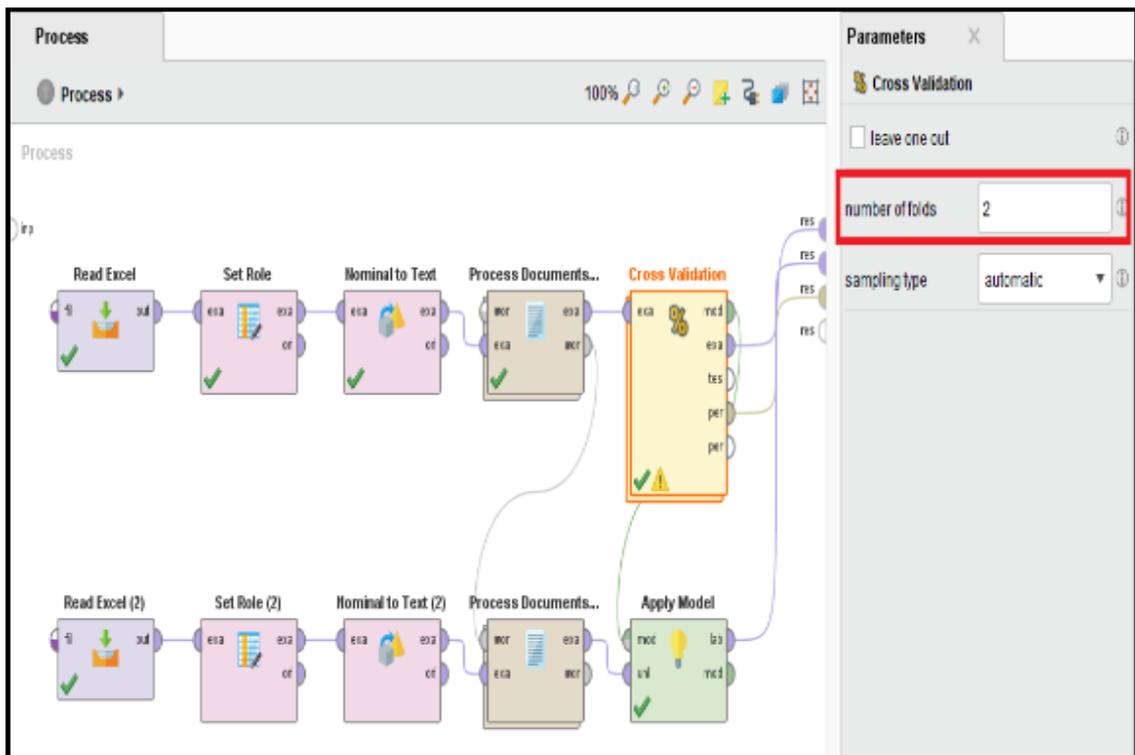


Fig. 6: Series of k-folds test

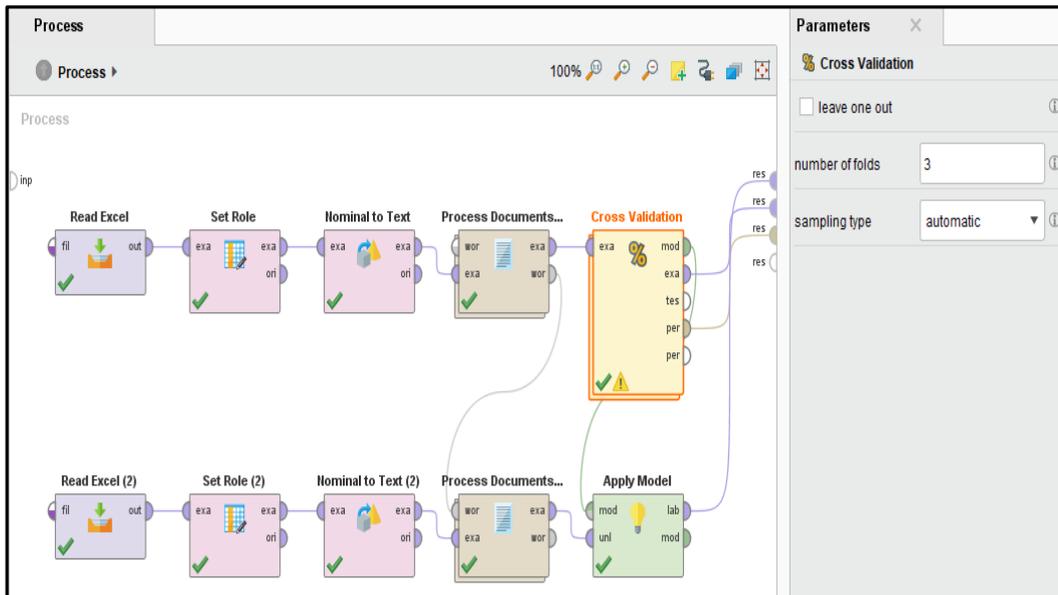


Fig. 7: Main testing process

accuracy: 80.79% +/- 1.34% (micro average: 80.79%)

	true hoax	true fakta	class precision
pred. hoax	315	118	72.75%
pred. fakta	28	299	91.44%
class recall	91.84%	71.70%	

Fig. 8: Accuracy testing results

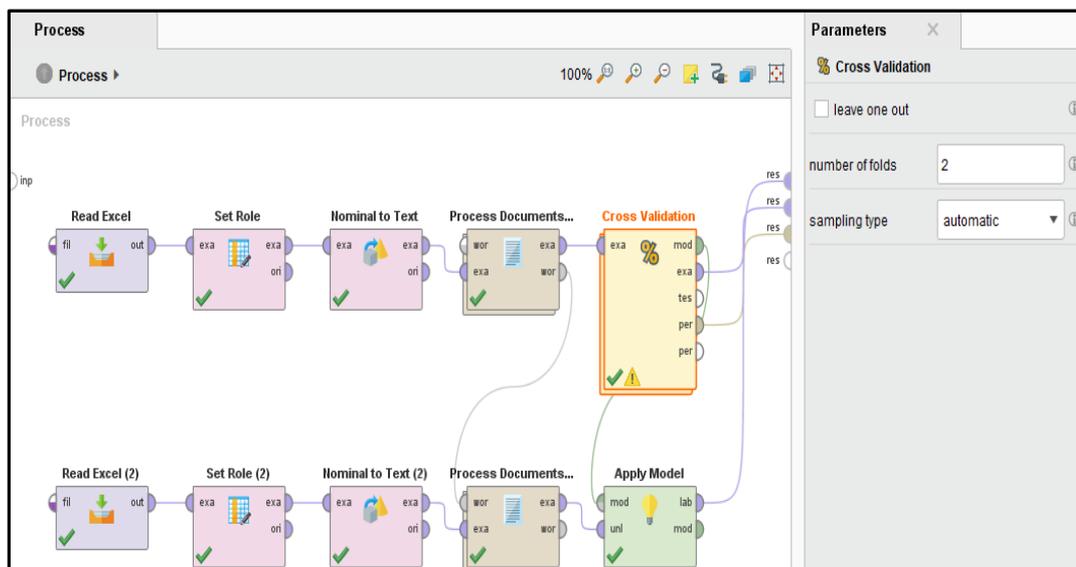


Fig. 9: Testing process dataset 2

accuracy: 77.50% +/- 0.92% (micro average: 77.50%)

	true hoax	true fakta	class precision
pred. hoax	266	107	71.31%
pred. fakta	64	323	83.46%
class recall	80.61%	75.12%	

Fig. 10: Accuracy testing results

Table 1: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	78.3
2	3	80.8
3	4	81.97
4	5	81.05
5	6	80.79
6	7	81.18
7	8	83.19
8	9	81.58
9	10	81.58

Based on the above test result is obtained with the highest accuracy on the value of k-fold 7 with a value of 83.19% accuracy

Table 2: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	77.50
2	3	81.32
3	4	81.71
4	5	80.53
5	6	82.37
6	7	81.45
7	8	82.24
8	9	83.16
9	10	82.11

Based on the above test result is obtained with the highest accuracy on the value of k-fold 9 with a value of 83.16% accuracy

Based on the above test results obtained an accuracy of test results the third dataset. The result can be seen in the confusion matrix below.

The image above is the test results of one of the second dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as shown below.

D. Testing of Dataset 4

In this test testing dataset fourth predetermined. This test uses a string of models that have been described previously. Here is one example of a string of four testing process.

The image above is the test results of one of the fourth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

E. Testing of Dataset 5

In this test testing dataset fifth predetermined. This test uses a string of models that have been described previously. Here is one example of a string of five testing process.

Based on the above test the accuracy of the results obtained a dataset fifth test results. The result can be seen in the confusion matrix below.

Figure above is the result of testing of one fifth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

F. Testing of Dataset 6

In this test testing dataset sixth predetermined. This test uses a string of models that have been described previously. Here is one example of a string of six testing process.

Based on the above test the accuracy of the results obtained a dataset fifth test results. The result can be seen in the confusion matrix below.

Figure above is the result of testing of one sixth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

G. Testing of Dataset 7

In this test testing dataset seventh predetermined. This test uses a string of models that have been described previously. Here is one example of a string of seven testing process.

Based on the above test the accuracy of the results obtained a dataset fifth test results. The result can be seen in the confusion matrix below.

Figure above is the result of testing of one sixth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

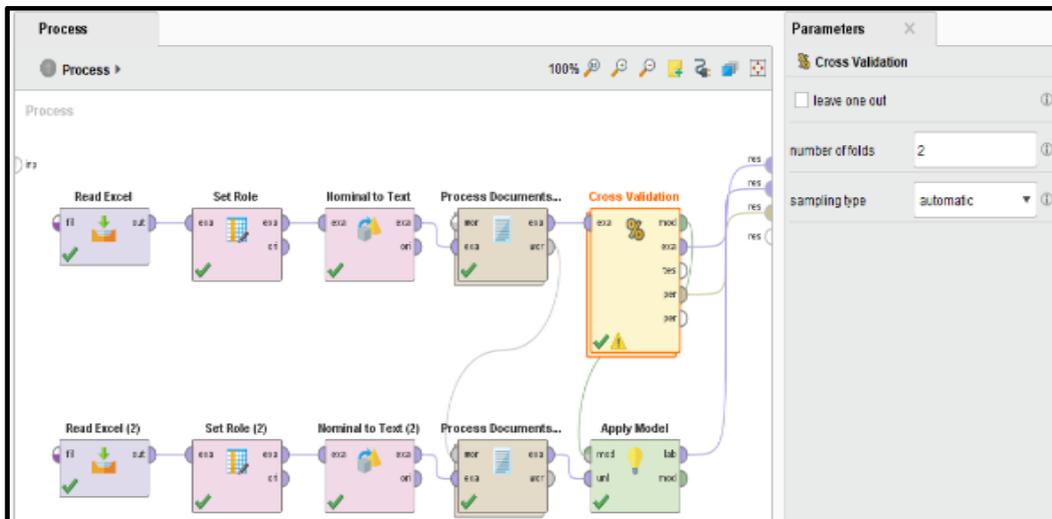


Fig. 11: Testing process dataset 3

accuracy: 76.71% +/- 2.24% (micro average: 76.71%)

	true hoax	true fakta	class precision
pred. hoax	286	121	70.27%
pred. fakta	56	297	84.14%
class recall	83.63%	71.05%	

Fig. 12: Accuracy testing results dataset 3

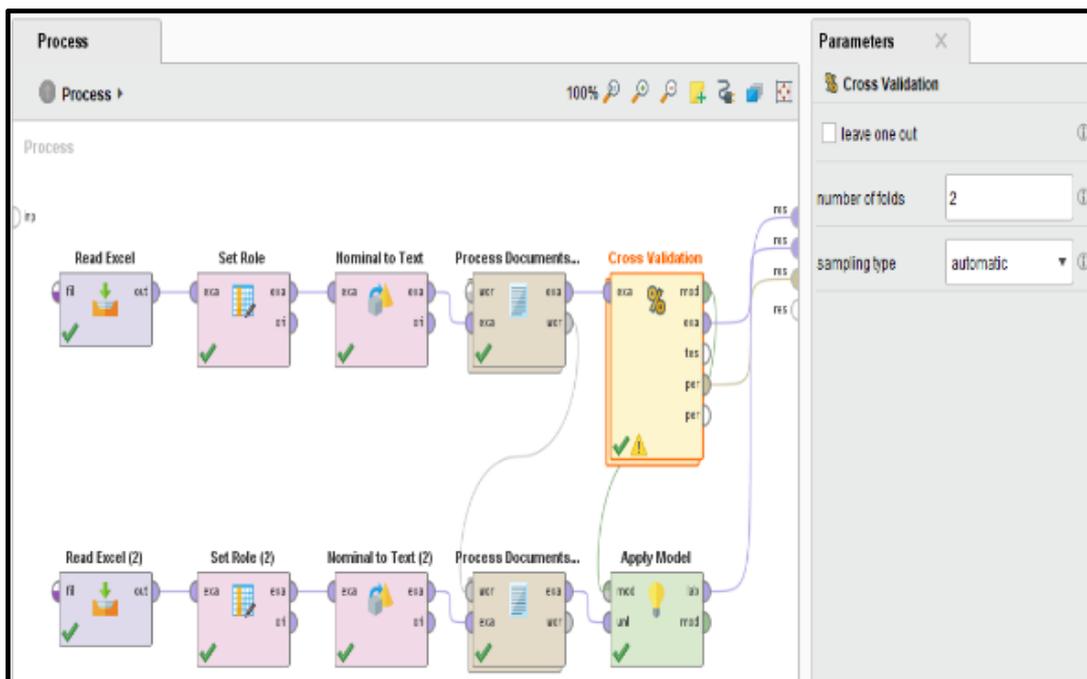


Fig. 13: Testing process dataset 4

accuracy: 75.92% +/- 1.71% (micro average: 75.92%)

	true fakta	true hoax	class precision
pred. fakta	297	57	83.90%
pred. hoax	126	280	68.97%
class recall	70.21%	83.09%	

Fig. 14: Accuracy testing results dataset 4

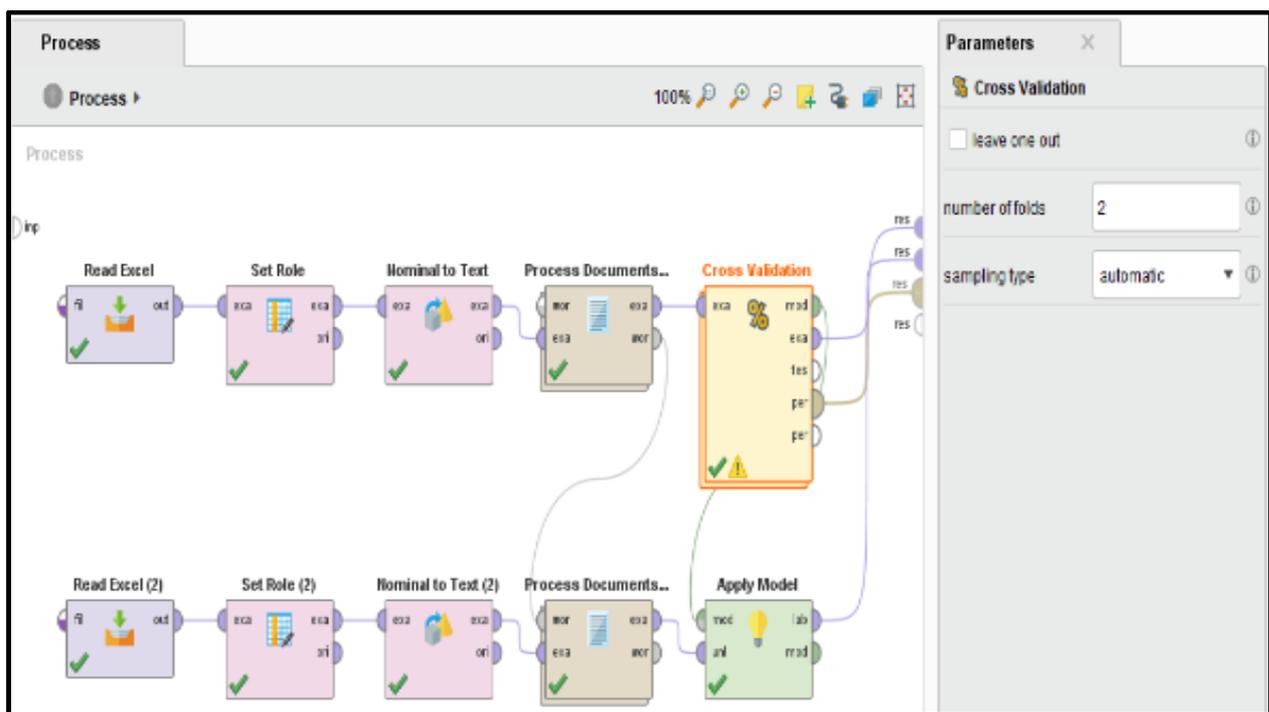


Fig. 15: Testing process dataset 5

accuracy: 79.74% +/- 4.74% (micro average: 79.74%)

	true fakta	true hoax	class precision
pred. fakta	295	21	93.35%
pred. hoax	133	311	70.05%
class recall	68.93%	93.67%	

Fig. 16: Accuracy testing results dataset 5

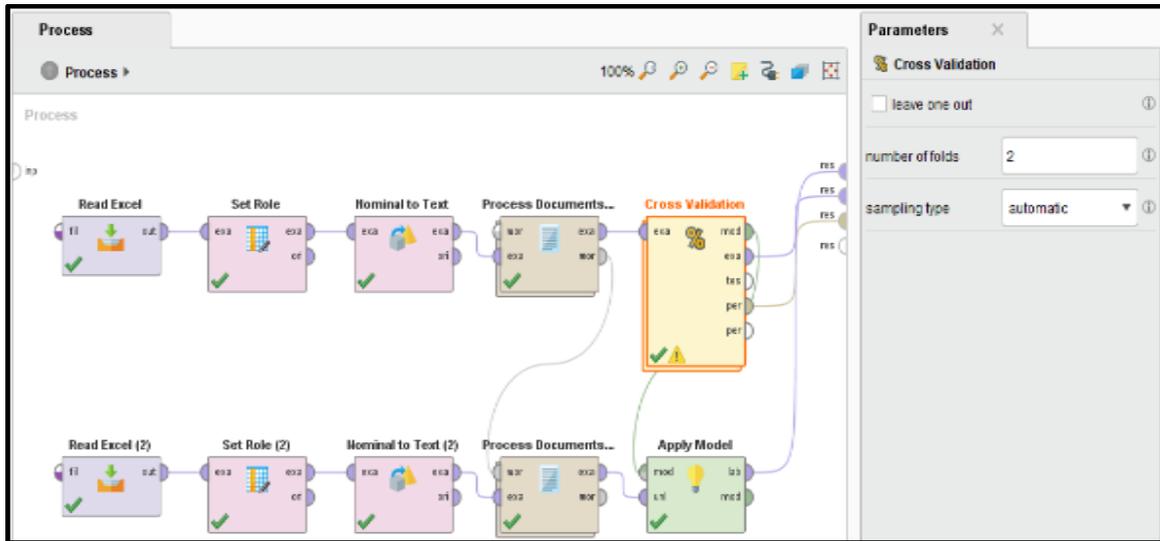


Fig. 17: Testing process dataset 6

accuracy: 81.97% +/- 1.18% (micro average: 81.97%)

	true fakta	true hoax	class precision
pred. fakta	321	40	88.92%
pred. hoax	97	302	75.69%
class recall	76.79%	88.30%	

Fig. 18: Accuracy testing results dataset 6

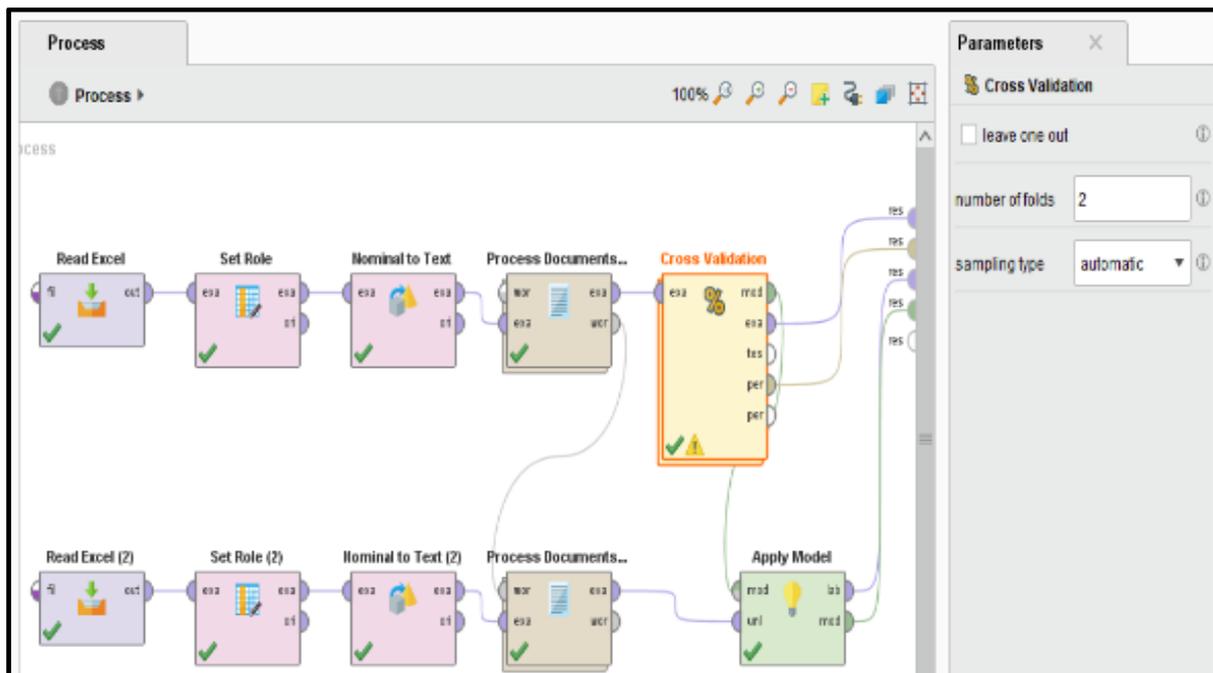


Fig. 19: Testing process dataset 7

accuracy: 80.39% +/- 1.97% (micro average: 80.39%)

	true fakta	true hoax	class precision
pred. fakta	342	59	85.29%
pred. hoax	90	269	74.93%
class recall	79.17%	82.01%	

Fig. 20: Accuracy testing result

Table 3: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	76.71
2	3	79.61
3	4	79.87
4	5	78.42
5	6	80.26
6	7	79.61
7	8	78.55
8	9	80.26
9	10	80.13

Based on the above test result is obtained with the highest accuracy on the value of k-fold 6 and 9 with a value of 80.26% accuracy

Table 4: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	75.92
2	3	78.82
3	4	77.89
4	5	80.92
5	6	79.08
6	7	80.66
7	8	79.08
8	9	80.79
9	10	80.53

Based on the above test result is obtained with the highest accuracy on the value of k-fold 5 with a value of 80.92% accuracy

Table 5: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	79.74
2	3	78.03
3	4	78.29
4	5	79.87
5	6	78.55
6	7	77.76
7	8	79.87
8	9	79.61
9	10	80.92

Based on the above test result is obtained with the highest accuracy on the value of k-fold 10 with a value of 80.92% accuracy

Table 6: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	81.97
2	3	82.63
3	4	81.97
4	5	81.05
5	6	80.79
6	7	81.18
7	8	83.29
8	9	81.58
9	10	81.58

Based on the above test result is obtained with the highest accuracy on the value of k-fold 8 with a value of 83.29% accuracy

Table 7: Accuracy testing result dataset 7

No	k-fold value	Testing accuracy (%)
1	2	80.39
2	3	79.87
3	4	81.05
4	5	82.63
5	6	82.76
6	7	82.37
7	8	80.79
8	9	83.82
9	10	83.55

Based on the above test result is obtained with the highest accuracy on the value of k-fold 9 with a value of 83.82% accuracy

Table 8: Accuracy testing result

No	k-fold value	Testing accuracy (%)
1	2	77.11
2	3	80.00
3	4	81.32
4	5	80.00
5	6	80.26
6	7	80.39
7	8	82.24
8	9	79.87
9	10	81.84

Based on the above test result is obtained with the highest accuracy on the value of k-fold 8 with a value of 82.24% accuracy

Table 9: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	80.92
2	3	79.21
3	4	81.45
4	5	80.39
5	6	80.13
6	7	81.32
7	8	81.32
8	9	80.79
9	10	81.58

Based on the above test result is obtained with the highest accuracy on the value of k-fold 10 with a value of 81.58% accuracy

Table 10: Accuracy testing results

No	k-fold value	Testing accuracy (%)
1	2	74.21
2	3	79.08
3	4	78.29
4	5	79.21
5	6	79.21
6	7	79.34
7	8	78.82
8	9	78.16
9	10	79.34

Based on the above test result is obtained with the highest accuracy on the value of k-fold 10 with a value of 79.34% accuracy

Table 11: Comparative testing results

No	Testing	Accuracy (%)
1	Test 1	83.19
2	Test 2	83.16
3	Test 3	80.26
4	Test 4	80.92
5	Test 5	80.92
6	Test 6	83.29
7	Test 7	83.82
8	Test 8	82.24
9	Test 9	81.58
10	Test 10	79.34

Based on the above test results obtained the highest accuracy in testing to 7 on the k-fold to 9 is equal to 83.82%. The following table charts the comparison accuracy on the test results

H. Testing of Dataset 8

In this test testing dataset eighth predetermined. This test uses a string of models that have been described previously. Here is one example of a string of eight testing process.

Based on the above test the accuracy of the results obtained a dataset fifth test results. The result can be seen in the confusion matrix below.

Figure above is the result of testing of one sixth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

I. Testing of Dataset 9

In this test testing dataset ninth predetermined. This test uses a string of models that have been described previously. Here is one example of a string of nine testing process.

Based on the above test the accuracy of the results obtained a dataset fifth test results. The result can be seen in the confusion matrix below.

Figure above is the result of testing of one sixth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

Figure above is the result of testing of one sixth dataset with k-fold value. Tests performed on the k-fold to 10 the obtained results as follows.

Accuracy Analysis

After several series of tests following is a table of the comparative value of each individual test carried out using cross validation techniques and k-fold.

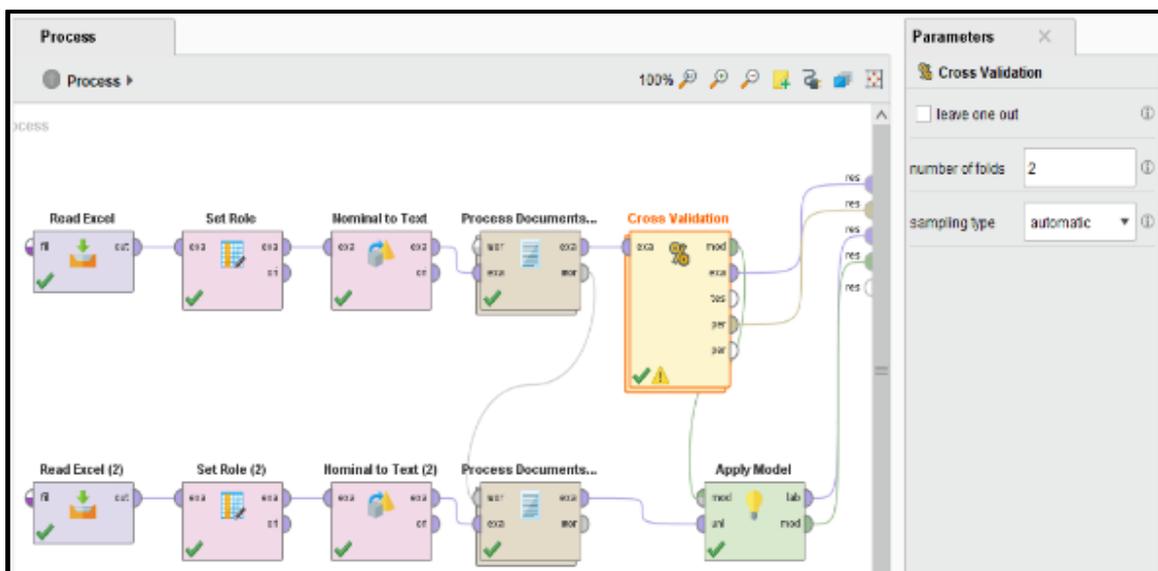


Fig. 21: Testing process dataset 8

accuracy: 77.11% +/- 2.89% (micro average: 77.11%)

	true hoax	true fakta	class precision
pred. hoax	282	116	70.85%
pred. fakta	58	304	83.98%
class recall	82.94%	72.38%	

Fig. 22: Accuracy testing result

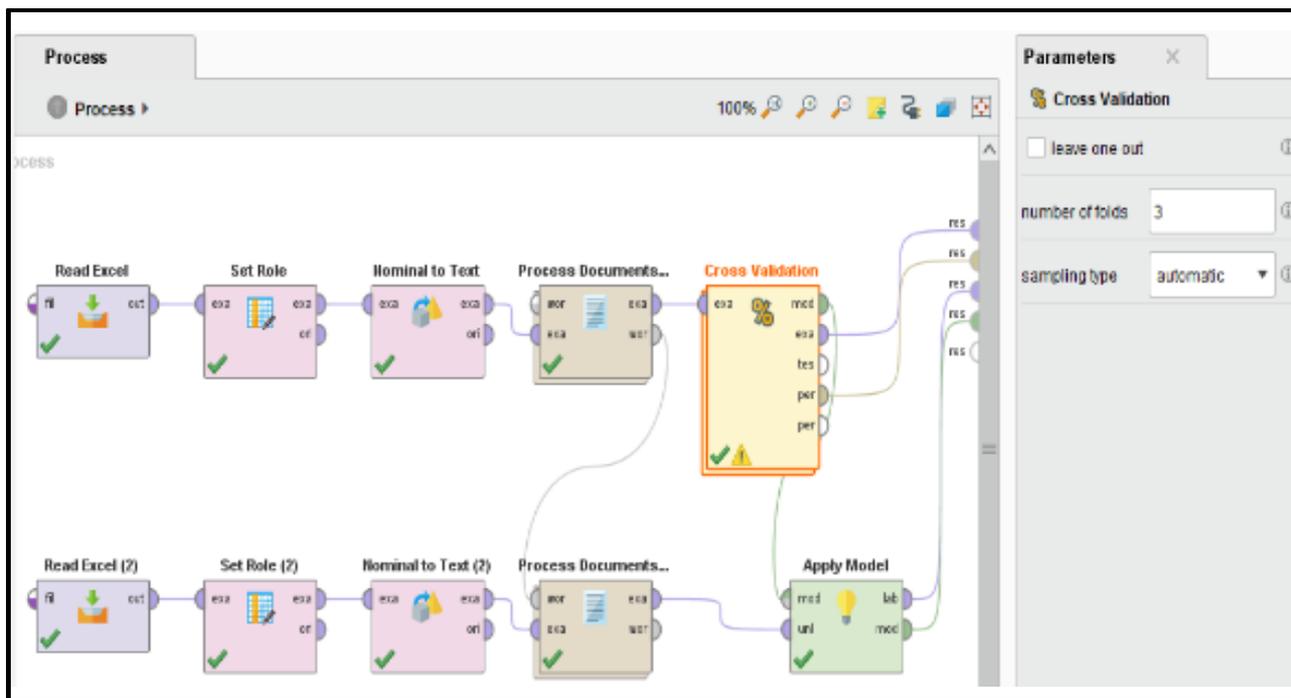


Fig. 23: Testing process dataset 9

accuracy: 79.20% +/- 4.81% (micro average: 79.21%)

	true fakta	true hoax	class precision
pred. fakta	306	36	89.47%
pred. hoax	122	296	70.81%
class recall	71.50%	89.16%	

Fig. 24: Accuracy testing result

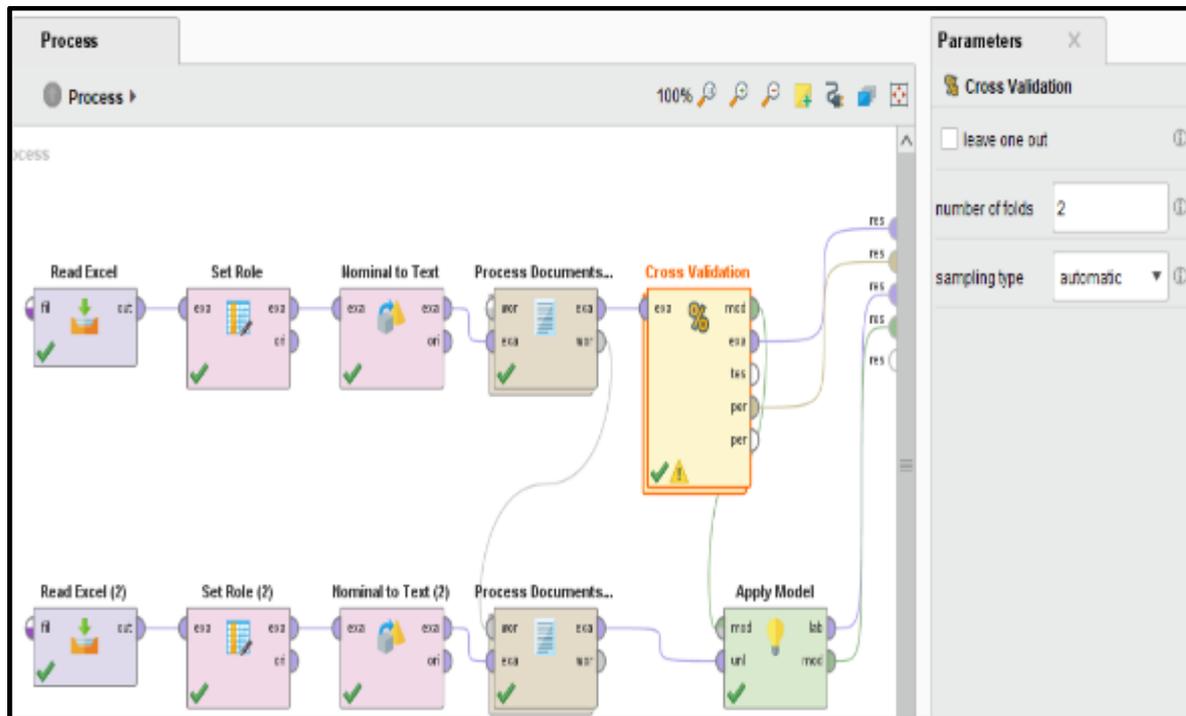


Fig. 25: Testing process dataset 10

accuracy: 74.21% +/- 2.11% (micro average: 74.21%)

	true hoax	true fakta	class precision
pred. hoax	320	175	64.65%
pred. fakta	21	244	92.08%
class recall	93.84%	58.23%	

Fig. 26: Accuracy testing result

J. Testing of Dataset 10

In this test, the tenth dataset that has been determined previously was tested. This test uses a string of models that have been described previously. Here is one example of a string of ten testing process.

Based on the above test the accuracy of the results obtained a dataset fifth test results. The result can be seen in the confusion matrix below.

Analysis of Classification Results

After following several test series obtained a classification results generated by testing with Rapid Miner studio.

Based on the amount of data testing in the test amounted to 198 datasets obtained prediction results with 29 different classifications with the classification that has been set. Thus the prediction accuracy rate obtained by:

$$\begin{aligned}
 \text{Akurasi Klasifikasi (\%)} &= \frac{\text{hasil benar} - \text{hasil analisis}}{\text{Hasil benar}} \times 100\% \\
 &= \frac{198 - 29}{198} \times 100\% \\
 &= 85,35\%
 \end{aligned}$$

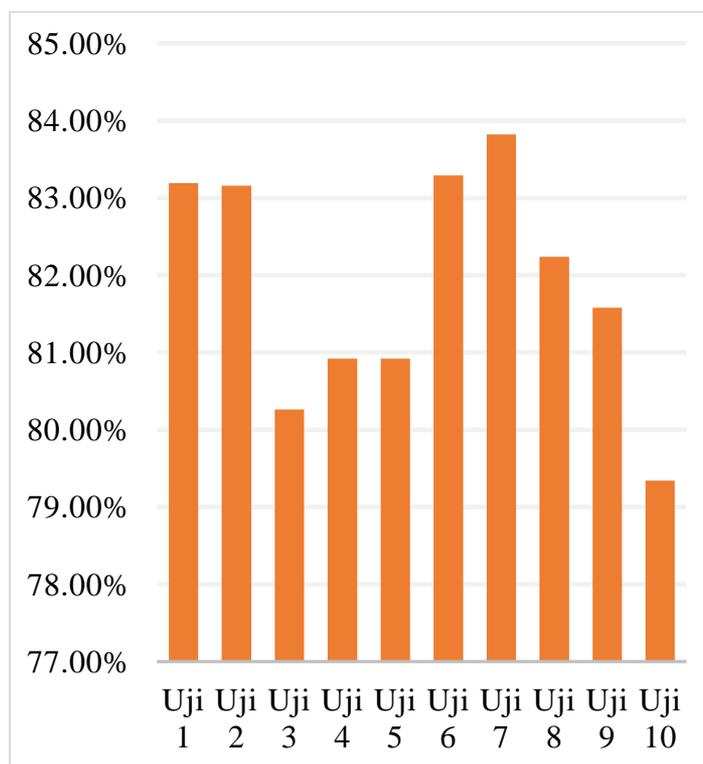


Fig. 27: Research stages

Conclusion

Based on the results of testing that has been done in this study that the method of Support Vector Machine can be used for classification of virus hoaxes news COVID-19. Tests conducted by the preprocessing stage and pembobota using the TF-IDF said. The dataset is divided into 535 text message containing information on 425 facts and news text containing information on hoaxes. In this study using cross validation and testing techniques using k-fold value. Testing is done by dividing the 10 datasets. The first test resulted in accuracy of 83.19%, the second test resulted in accuracy of 83.16%, the third test result in accuracy of 80.26%, the fourth test result in accuracy of 80.92%, the fifth test produces accuracy of 80.92%, the sixth test yield of 83.29% accuracy, testing the accuracy of the seventh yield 83.82%, generating an eighth test accuracy of 82.24%, the ninth test produces an accuracy of 81.58%, the tenth test produces an accuracy of 79.34%. Thus the highest accuracy results obtained on testing to 7 with accuracy of 83.82%.

Suggestion

In this study still has many shortcomings, therefore the author gives suggestions for future research are:

1. Using other classification algorithms

2. The use of feature selection to improve the accuracy of the results
3. Using a broader hoax news
4. It is expected to form a system that can help people in knowing the information is fact or hoax

Author's Contributions

Wiranto Herry Utm: Conceived and designed the experiments think of the ideas presented in the research topic encourage author 1 to investigate (certain aspects) and monitor the findings of the study. Develop theory and conduct the experimental designs. Discuss the results and contribute to the final manuscript.

Analyzed and interpreted the data perform data analysis, existing data is made in a simple form so that it is easy for readers to understand and interpret.

Contributed materials analysis tools or data perform data analysis is done using appropriate and accurate research met hoods check originality of research results perform data analysis using data collection techniques using the correct methodological prince.

Wrote the paper give guidance author 2 to write the manuscript of journal explain the main ideas related to the research topic and problem.

Check out the sources cited, the accidental plagiarism, clear and specific language, proper grammar, content, spelling and punctuation checks. revise the article.

Karno Juni Prayoga: Conceived and designed the experiments develop theory and conduct the experimental designs. Discuss the results and contribute to the final manuscript.

Performed and experiment did an experiment, analyze and making simulation and develop theory create manuscript with support of Author 1.

Analyzed and interpreted the data perform data analysis and data interpretation after collecting and correlating data obtained after conducting research perform data analysis, existing data is made in a simple form so that it is easy for readers to understand and interpret conduct data analysis and then interpret the data to look for meaning and wider implications regarding the findings in research.

Contributed materials analysis tools or data perform data analysis is done using appropriate and accurate research methods perform data analysis using data collection techniques using the correct methodological principles check out the sources cited, the accidental plagiarism, clear and specific language, proper grammar, content, spelling and punctuation checks.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved

References

- Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, H., ... & Yuniastuti, E. (2020). Coronavirus disease 2019: Tinjauan literatur terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1), 45-67. <http://www.jurnalpenyakitdalam.ui.ac.id/index.php/jpd/article/view/415>
- Afriza, A., & Adisantoso, J. (2018). Metode Klasifikasi Rocchio untuk Analisis Hoax. *Jurnal Ilmu Komputer dan Agri-Informatika*, 5(1), 1-10. <http://journal.ipb.ac.id/index.php/jika/article/view/21737>
- Asiyah, S. N. (2016). Klasifikasi berita online menggunakan metode support vector machine dan k-nearest neighbor (Doctoral dissertation, Institut Teknologi Sepuluh Nopember). <https://repository.its.ac.id/62883/>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press. ISBN-10: 0521836573
- Gupta, G., & Bhathal, G. S. (2018). Sentiment analysis of english tweets using data mining: Data mining, sentiment analysis. BookRix. <https://www.amazon.com/sentiment-analysis-english-tweets-mining-ebook/dp/b07ccmjntm>
- Honakan, H., Adiwijaya, A., & Al Faraby, S. (2018). Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia. *e-Proceedings of Engineering*, 5(1). <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/6088>
- Jaya, I., & Aulia, I. (2018). Klasifikasi Dokumen Karya Ilmiah Menggunakan Algoritma Support Vector Machine. <http://repositori.usu.ac.id/handle/123456789/6617>
- Kumparan.com. (2020). Kominfo Temukan 1.125 Hoaks di Medsos Terkait Corona, 2020. https://kominfo.go.id/content/detail/25831/kominfo-temukan-1125-hoaks-di-medsos-terkait-corona/0/sorotan_media
- Maulina, D., & Sagara, R. (2018). Klasifikasi artikel hoax menggunakan support vector machine linear dengan pembobotan term frequency–Inverse document frequency. *Jurnal Mantik Penusa*, 2(1). <http://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/322>
- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2008). Support Vector Machine (Teori dan Aplikasinya dalam Bioinformatika), *IlmuKomputer.com*, pp. 842-847. https://www.academia.edu/24381027/Support_Vector_Machine_Teori_dan_Aplikasinya_dalam_Bioinformatika_1
- Prayoga, K. J., Nugroho, A., & Wiyatno, T. N. (2019). Komparasi feature selection particle swarm optimization (ps) dengan genetic algorithm (ga) terhadap algoritma naïve bayes pada analisis sentiment twitter. *e-Prosiding SNasTekS*, 1(1), 299-314. <https://journal.unusida.ac.id/index.php/snts/article/view/99>
- Rahmat, M. A., & Areni, I. S. (2019, July). Hoax Web Detection For News in Bahasa Using Support Vector Machine. In *2019 International Conference on Information and Communications Technology (ICOIACT)* (pp. 332-336). IEEE. <https://ieeexplore.ieee.org/abstract/document/8938425/>
- Wearesocial.com. (2020). Indonesian Digital Report 2020, 2020. <https://wearesocial.com/>
- Williams, G. J., & Simoff, S. J. (Eds.). (2006). *Data mining: Theory, methodology, techniques, and applications* (Vol. 3755). Springer. ISBN-10: 3540325484
- Zy, A. T. (2017). Comparison Algorithm Classification Naive Bayes, Decision Tree, and Neural Network for Analysis Sentiment. *Pelita Teknologi*, 12(1), 5-12. <https://www.jurnal.pelitabangsa.ac.id/index.php/pelitakno/article/view/173>