

SENTIMENT ANALYSIS ARTICLE NEWS COORDINATOR MINISTER OF MARITIME AFFAIRS USING ALGORITHM NAIVE BAYES AND SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION

NIA KUSUMA WARDHANI¹, REZKIANI², SIGIT KURNIAWAN³, HENDRA SETIAWAN⁴,
GRACE GATA⁵, SISWANTO TOHARI⁶, WINDU GATA⁷, MOCHAMAD WAHYUDI⁸

¹Fasilkom, Mercubuana University, Jl. Meruya Selatan, Kembangan, Jakarta Barat . Indonesia.

^{2,3,6,7} Graduate School Master's Degree Computer Science Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri), Jl. Salemba Raya No. 5 Jakarta Pusat. Indonesia

⁴Information System, Sekolah Tinggi Manajemen Informatika dan Komputer Bani Saleh (STMIK Bani Saleh). Jl. M. Hasibuan Bekasi Timur. Indonesia.

^{5,8}Information Technology, Budi Luhur University, Jl. Ciledug Raya No. RT.10, RT.10/RW.2, Petukangan Utara, Pesanggrahan, Kota Jakarta Selatan. Indonesia.

E-mail: ¹Nia.kusuma@mercubuana.ac.id, ²qimunkqi@gmail.com, ³kurniawan.sgt@gmail.com,
⁴Hendrasetiawan007@gmail.com, ⁵gatasmara@gmail.com, ⁶siswantobl@gmail.com,
⁷windu@nusamandiri.ac.id, ⁸wahyudi@nusamandiri.ac.id

ABSTRACT

News has become a basic human need along with the development of technology and the internet. The development of technology and the internet is causing the change of publication pages from a print media to the internet. The use of online media today is not only for reading news articles, but also can be used to see the issues that occur can even be used to see the performance of a political figure. The classification of the contents of news articles into a new knowledge that is a negative or positive conclusions about the content of the news contained in a news site. It is possible by using sentiment analysis that is by document classification with text mining. The algorithm used in this research is Naive Bayes and Support Vector Machine with Particle Swarm Optimization. NB has an accuracy value of 89.50% with AUC of 0.500 while the NB PSO obtains an accuracy of 92.00% with AUC of 0.550. SVM has an accuracy value of 87.50% with AUC of 0.979, while SVM PSO has an accuracy value of 90.50% and AUC of 0.975. The best application of optimization in this model is NB PSO can provide solution to the classification problem in this case of sentiment analysis. NB PSO algorithm provides solutions to the analysis of sentiments from the content of various online media news optimally.

Keywords: *Sentiment Analysis, Text Mining, Classification, Naive Bayes, Support Vector Machine, Particle Swarm Optimization, AUC*

1. INTRODUCTION

News has become a basic human need along with the development of technology and the internet. The development of technology and the internet is causing the change of publication pages

from a print media to the internet. The news presented is usually in text form on digital media grouped based on the content of the discussion of each news category [20]. The use of online news sites today is not only limited as a medium for reading news articles, but also can be used to look at

issues that are going on. News is not only used to read articles about entertainment, sports, economy and so on, but news articles can be used to see the performance of a person. A political figure must have their respective responsibilities in accordance with the positions they have and can not be separated from the pros and cons of the performance that has been done. There are some of people who question the performance of political figures who do many things, even can be said to take over the responsibility of others [14].

Based on the explanation it is necessary a way to be able to classify a content from a news article into a new knowledge that is a negative or positive conclusions about the content of news that is in a news site. It is possible using sentiment analysis. The magnitude of the effects and benefits of sentiment analysis led to research and application-based sentiment analysis growing rapidly [3]. The application of machine learning method is used to classify the polarity of a news from a very large data source. To do so, can use one of the functions of text mining, in this case is the document classification [13].

In the previous research, on the analysis of sentiments, there is research on the analysis of sentiments against President Jokowi. In the research, using “search techniques” in data collection “Boolean searching” with operator “AND”. The data obtained are labeled positive, neutral and negative then corrected by linguists. After it is done preprocessing either it turns the word does not default to default or commonly called normalization using dictionary and search for root word that is stemming with the help of Sastrawi Master application. Furthermore, there is also the N-Gram tokenize, Unigram, Bigram, and Trigram to the sentence, then eliminating the commonly used words and do not have valuable information in a context or commonly called stopword removal, and retain emoticons because emoticons are symbols that show expression someone into the writing. Algorithm classification used is Naive Bayes Classifier and Support Vector Machine [17].

In this research, will be discussed the steps taken to conduct the process of sentiment analysis of news articles about political figures namely the Coordinator Minister of Maritime Affairs Period 2016-2019. It starts from the preprocessing phase to the phase of sentiment analysis with Naive Bayes Classifier and Support Vector Machine with Particle Swarm Optimization and how to measure the quality of the analysis results from each of the classification algorithms. Particle Swarm Optimization (PSO) is an optimization technique used to apply and modify some parameters and increase attribute weights.

2. RELATED WORKS

Related research used as reference is a research that also uses Text Mining by classification method. [13] did a classification of sentiments against Indonesian presidential candidates and vice presidents of 2014 revealed through Twitter's social network. There was several phases to perform sentiment analysis, including data collection, preprocessing data, POS Tagging, opinion extraction using rule based and opinion classification using Naïve Bayes Classifier method.

[3] tried to analyze Jakarta Governor's 2017 Gubernatorial election to be discussed in real world and cyberspace, especially in social media Twitter. The method used in this research, for preprocessing data using tokenisasi, cleansing and filtering, to determine the class sentiment with Lexicon Based method. For the process of classification using Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) methods

[1] did research about the analysis of sentiment against former President @SBYudhoyono timeline. [1] used three classes of positive, negative and neutral sentiments. The model used to determine sentiments comes from the previous research timeline.

[15] described the national exam (UN) as a government policy to evaluate the level of education on a national scale to measure the competence of students who graduate with other schools at the

same educational level. The process of analysis and data processing for document classification used two classification methods: quintuple method and one of the methods for learning machine, Naive Bayes Classifier (NBC) method. The data used to classify the document in the form of a text document news about the implementation of the UN, which was taken from the online news media (detik.com).

[4] conducted a sentiment analysis research on product reviews. used the Naive Bayes classification combined with Neural Network on the standard dataset will improve the accuracy and performance of sentiment classification.

[21] conducted a research that discusses unlabeled data through the use of hashtag Twitter with Support Vector Machine. In this research also used 1- grams and tag part of speech (POS Tagging) which is applied as a feature.

[2]This research discussed binary classifications that grouped into two classes. The algorithm used is SVM with validation process using 10-Fold cross-validation and confusion matrix. Use of Partical Swarm Optimization (PSO) is used to improve the selection of parameters.

3. LITERATURE SURVEY

3.1. Text Mining

Text mining is a new and exciting field of research that tries to solve excess information problems using data mining techniques, machine learning, Natural Language Processing (NLP), Information Retrieval (IR), and knowledge management. Text mining involves preprocessing phases of document collections such as text categorization, information extraction, term extraction [5].

An important goal of text mining is to get high quality information from the text. According to Feldman and Sanger that the stages of text mining in general are text preprocessing and feature selection. The preprocessing text stage is the initial stage of text mining. This

phase includes all routines, and processes for preparing the data to be used on the operation of discovery knowledge of text mining systems. The feature selection phase aims to reduce the dimension of a collection of text, or in other words remove the words that are considered unimportant or do not describe the contents of the document so that the process of classification is more effective and accurate [10].

3.2. Sentiment Analysis

Sentiment analysis is a process for determining the sentiments or opinions of a person embodied in text form and can be categorized as positive or negative sentiments [6]. The analysis of sentiment refers to the broad field of natural language processing, linguistic computing and text mining aimed at analyzing the opinions, sentiments, evaluations, attitudes, judgments and emotions of a person whether the speaker or author pertains to a topic, product, service, organization, individual, or certain activities [10]. Sentiment analysis can also be interpreted as textual information is generally divided into fact (objective) and opinion (subjective) information [16]. The sentiment analysis is also a computational research of opinions, sentiments and emotions expressed textually [9].

3.3. Naïve Bayes

Naive Bayes classification is a statistical classification that can be used to predict the probability of membership of a class. According to Wu and Kumar that Naive Bayes is a popular classification method and is included in the top ten algorithms in data mining. Naive Bayes uses a branch of mathematics known as probability theory to look for the greatest opportunity of classification possibilities, by looking at the frequency of each classification in training data [11]. NBC method takes two phase in the process of classification of text, the training phase and the classification stage. In the training phase, the process of analyzing the sample of the document in the form of vocabulary selection and then doing the probability determination for each category based on the sample document. At the classification stage, the category

value of a document is determined based on the term that appears in the document classified [7].

3.4. Support Vector Machine

SVM is a learning machine method that works on the principle of Structural Risk Minimization (SRM) in order to find the best hyperplane that separates the two classes in input space [12]. The following are the strengths of the Support Vector Machine (SVM) among others [18]. Has a high generalization capability:

1. Able to produce good classification models although trained with relatively few data sets with simple parameter settings. SVM has a clear concept and formulation with few parameters to be set.
2. Relatively easy to implement because SVM determination can be formulated in the QP (Quadratic Programming) problem.
- 3.

3.1 Particle Swarm Optimization

According to Basari that Particle Swarm Optimization (PSO) is a very simple optimization technique to apply and modify some parameters. In Particle Swarm Optimization (PSO) there are several techniques to perform optimization such as increasing attribute weight (attribute weight) to all attributes or variables used, select attributes (attribute selection) and feature selection [8]. Particle Swarm Optimization (PSO) begins with a population consisting of a number of randomly generated particles (stating potential candidates). Further updating the position and speed of each particle fly iteratively to produce a new, better solution. Particle Swarm Optimization (PSO) will stop when the optimum solution has been found or certain conditions have been reached.

4. RESEARCH METHODOLOGY

4.1. Data Collection Method

The data to be used is data from news articles concerning the Coordinator Minister of Maritime Affairs Period of 2016-2019. The data is obtained from online news site www.news.detik.com, www.antaraneews.com, www.kompas.com, www.tribunnews.com, [\[putan6.com\]\(http://putan6.com\) and \[www.tempo.com\]\(http://www.tempo.com\). News data used in this research as much as 200 data. The data was collected from the beginning Minister of Coordination Kemaritiman that is dated July 28, 2016 until November 23, 2017. News data is categorized into positive news data based on articles related only to maritime. While news articles that are not related to maritime discussion will be categorized as negative news data.](http://www.li</p></div><div data-bbox=)

4.2. Data Processing

First step in data processing, text mining must go through several phases called preprocessing. Preprocessing steps in this research are Tokenize, Filter Tokens (By Length), Stopwords Removal, Transform Cases [19].

4.3. Proposed Method

The proposed method or model is using two Naive Bayes algorithms, Support Vector Machine with each using the feature selection of Particle Swarm Optimization (PSO). In addition, the proposed method uses the CRISP-DM method that aims to obtain a pattern or knowledge of the data used to solve a problem and used for appropriate decision making.

4.3. Experiments and Test Results

In testing the software model used to retrieve data and tools to calculate accuracy is RapidMiner version 7.3. The dataset used for model testing is obtained from news articles about the figure of the Minister of Coordination of the Ministry of Marine and Fisheries Period 2016-2019 as much as 200 data, which then classified positive news and negative news. Test results produce accuracy value and for testing method will be made application planning using PHP and HTML programming language.

4.4. Evaluation and Validation

Standard validation for this research is 10 fold cross-validation where this process divides data randomly into 10 parts. The testing process begins with the formation of models with data in the first

section. The model formed will be tested on the remaining 9 sections of the data. The process performed after testing is measuring the performance of the text mining classification

algorithm used. In this research the performance is measured using Accuracy and AUC and will be shown in the form of ROC curve.

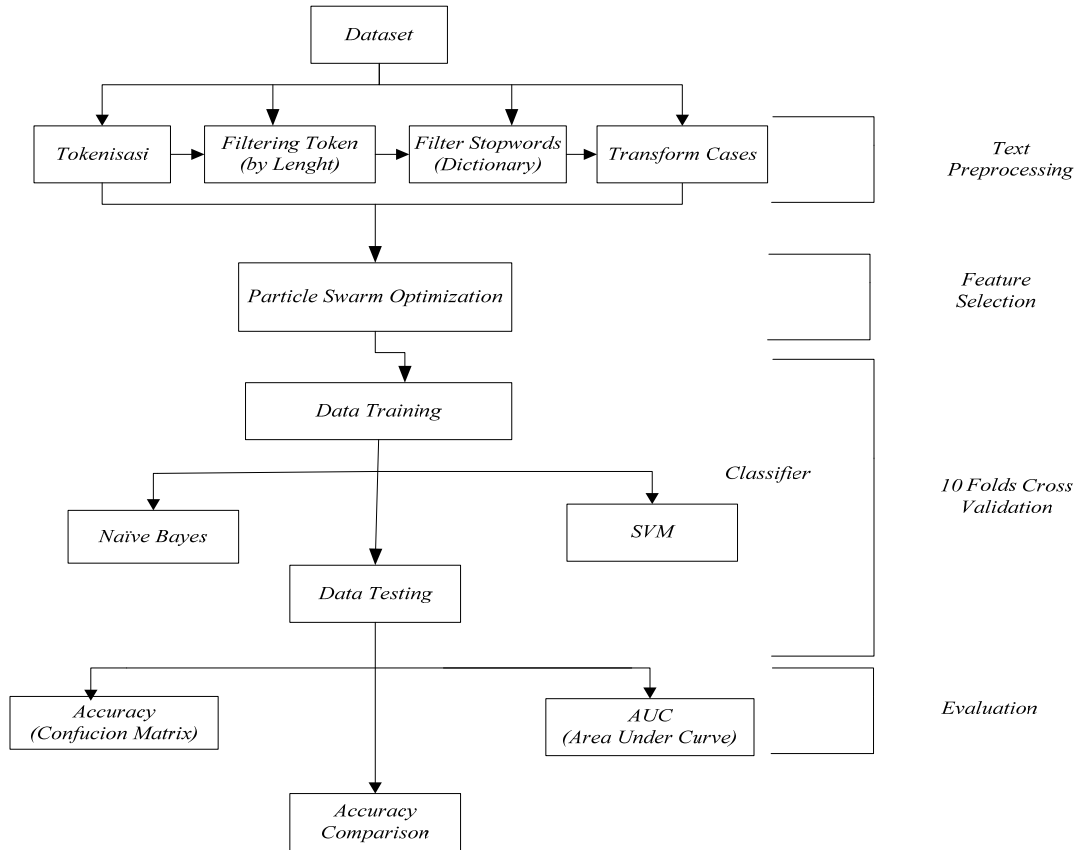


Figure 1: Research Framework

5. RESULTS

5.1. Business Understanding Phase

In this phase, need to understanding of the object of research conducted by digging information through several online news sites against the Minister of Coordination of Ministry for the Period 2016-2019. Motivation in this phase of the news presented usually in the form of text on digital media grouped by content of the discussion of each news category. Online media is now not only limited as a medium for reading news articles, but also can be used to see the issues that occur can even be used to see the performance of a political figure. This sentiment analysis is conducting to look for classification methods that can assist in

determining positive and negative news articles. At this phase, it is also an understanding to find the best classification method in order to assist during data processing and to improve the performance of the classification method can be done by using feature selection.

5.2. Data Understanding Phase

Data obtained from online news sites are www.news.detik.com, www.antaranews.com, www.kompas.com, www.tribunnews.com, www.liputan6.com and www.tempo.com. The data was collected from the beginning Minister of Coordination Kemaritiman that is on July 28, 2016 - November 23, 2017. Primary data obtained by 273 data news articles (Detiknews

with 136 data, Antaraneews with 40 data, Kompas with 17 data, Tribunnews with 50 data and Liputan6 with 22 data and Tempo with 8 data). By using the data source obtained, created dataset with attribute that is news data containing all news data about Minister of Maritime Coordination Period 2016-2019 period and sentiment which is label 100 positive data and 100 negative data from each news. All news data is grouped into one either positive news or negative news and stored in the form of .xlsx extension. These are examples of news snippets taken from Indonesian news, as follows:

Menko Maritim berencana mengembangkan salah satu pelabuhan di bagian timur Sumatera. Pelabuhan yang terletak di dekat Pelalawan, Riau bakal dikembangkan untuk mempermudah logistik. "Daerahnya dekat pulau apa saya lupa, dekat Pelalawan lah. Di pantai timur Sumatera," jelas Kemenko Maritim, Jakarta Pusat, Jumat (2/12/2016).

5.3. Data Preparation Phase

A data preparation process aimed at obtaining clean and ready data for use in research. In the text mining the initial stage to be performed is the preprocessing phase.

1. Tokenize

In this tokenize process, all the words in each document will be collected and removed punctuation, and also removed the symbols, special characters or not letters. The results of the tokenize phase are as follows:

Menko Maritim berencana mengembangkan salah satu Pelabuhan di bagian timur Sumatera pelabuhan yang terletak di dekat Pelalawan Riau bakal dikembangkan untuk mempermudah logistik Daerahnya dekat pulau apa saya lupa dekat pelalawan lah Di pantai timur Sumatera jelas di Kemenko Maritim Jakarta Pusat Jumat

2. Filter token (by length)

In this process, words that have a length of less than 4 or more than 25 will be removed, such as

those which are not, *say, jd, ga, ane, gan* which are words that have no meaning apart if separated in other words and not related to adjectives associated with sentiment. The results of the Filter Token by Length phase are as follows:

Menko Maritim berencana mengembangkan salah satu Pelabuhan bagian timur Sumatera pelabuhan yang terletak dekat Pelalawan Riau bakal dikembangkan untuk mempermudah logistik Daerahnya dekat pulau saya lupa dekat pelalawan pantai timur Sumatera jelas Kemenko Maritim Jakarta Pusat Jumat

3. Stopword Removal

At this stage, the operator used is a stopword filter (dictionary) because of the dataset used in Indonesian language. In this process first made a list of words that include stopwords and then the file will be uploaded into the operator filter stopword (dictionary). In this stage, irrelevant words will be deleted like a word but, for, with, which are words that have no special meaning.

Menko Maritim berencana mengembangkan salah Pelabuhan Timur Sumatera pelabuhan terletak pelalawan riau dikembangkan mempermudah logistik daerahnya pulau lupa Pelalawan pantai timur Sumatera Kemenko Maritim Jakarta Pusat Jumat

4. Transform Cases

In this process, irrelevant words will be changed, such as words containing uppercase letters converted into lowercase letters so that they can relate to sentiments.

menko maritim berencana mengembangkan salah pelabuhan timur sumatera pelabuhan terletak pelalawan riau dikembangkan mempermudah logistik daerahnya pulau lupa pelalawan pantai timur sumatera kemenko maritim jakarta pusat jumat

5.4. Modelling Phase

Rapidminer 7.3 model designs are used below:

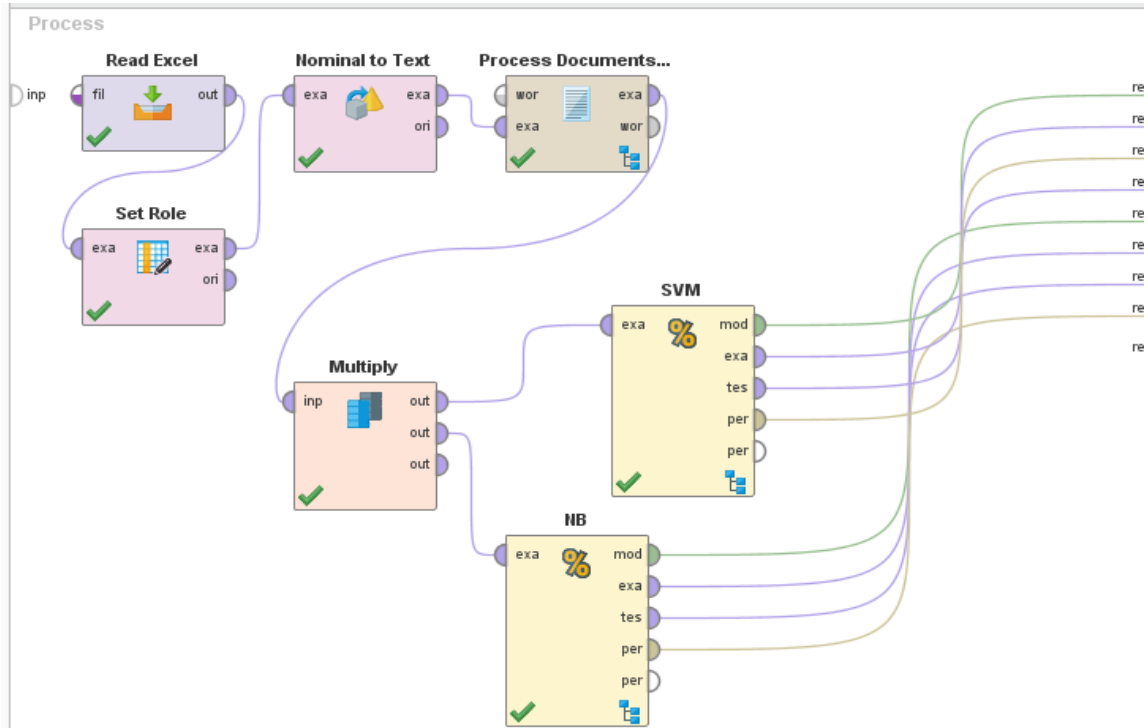


Figure 2: Design Model Comparison Of Algorithms (Svm And Nb)

In Figure 2 it can be explained that the data that has been collected is entered into a Microsoft Excel application that will be read by the Read Excel operator, then on the Set Role operator with Role Status. Before being processed in the Process Document operator, the operator Nominal to Text is given. Process Document consists of Tokenize operator, Tokens

(by length) Filter, Stopword Filter, and Transform Cases described in Figure 3. The results of the previous process were duplicated using the Multiply operator for 10-Fold validation using the Cross Validation operator. The validation is in it to test SVM algorithm (Figure 4) and Naïve Bayes (Figure 5).

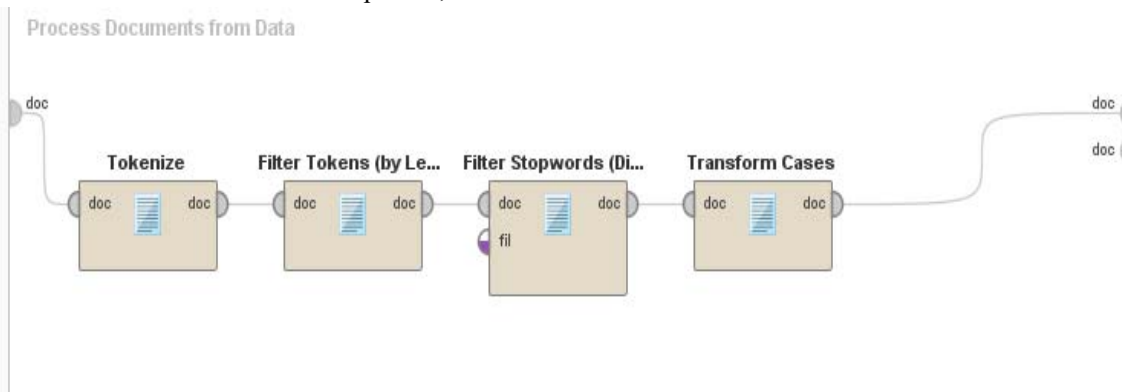


Figure 3. Process Document

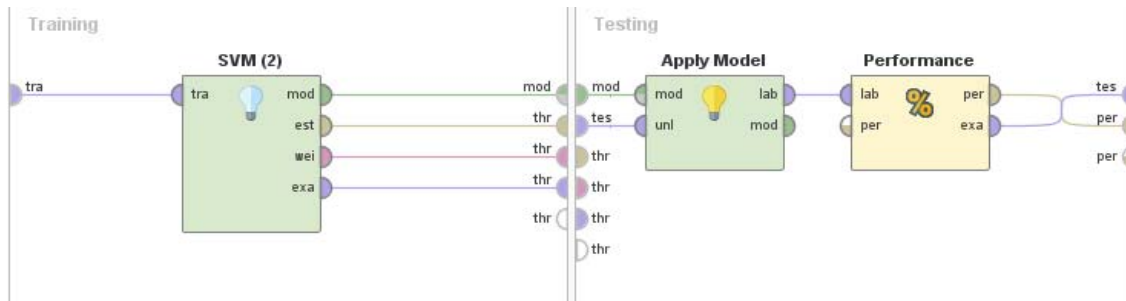


Figure 4. Process Inside SVM Cross Validation

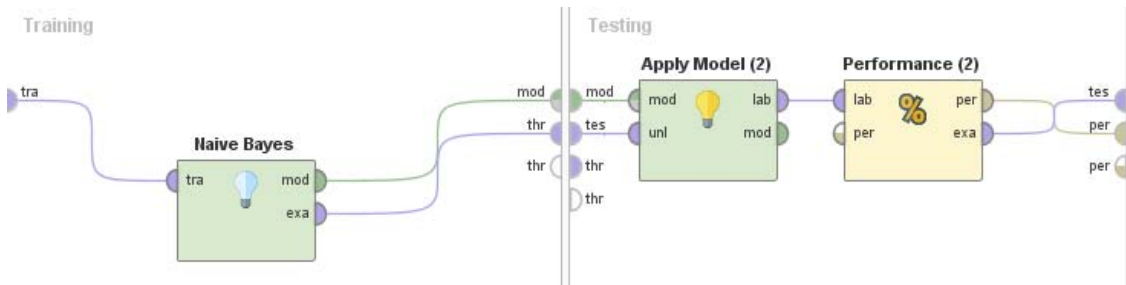


Figure 5. Process Inside Naïve Bayes Cross Validation

The design for PSO model for each algorithm is like below:

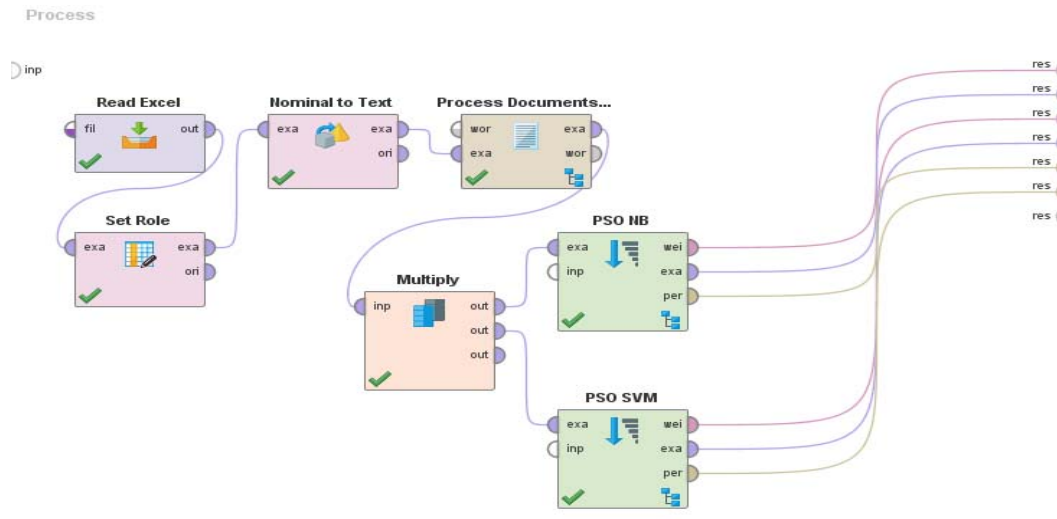


Figure 6. Design Model Comparison Of Algorithms (SVM PSO And NB PSO)

In Figure 6. The process of reading from the Read Excel operator up to Multiply operator has the same process as before. While the PSO operator is placed after the Multiply operator is needed to improve the performance of each algorithm that has been tested as in figure 7 and figure 8.

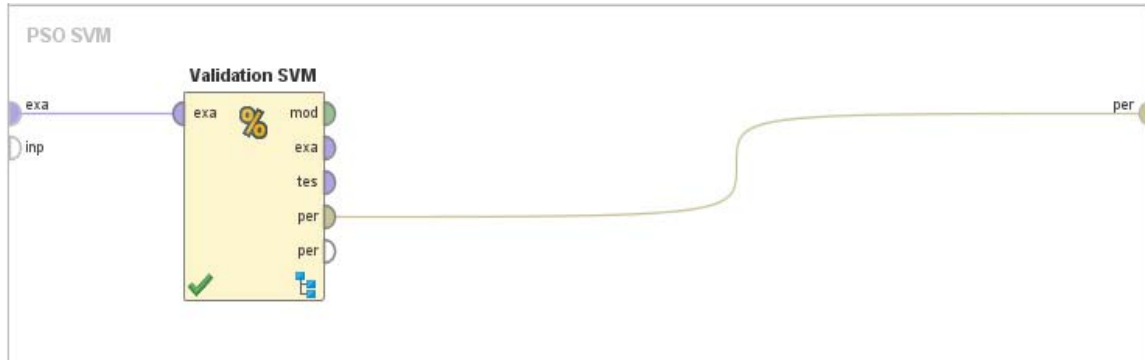


Figure 7: Cross Validation Inside Pso For Svm

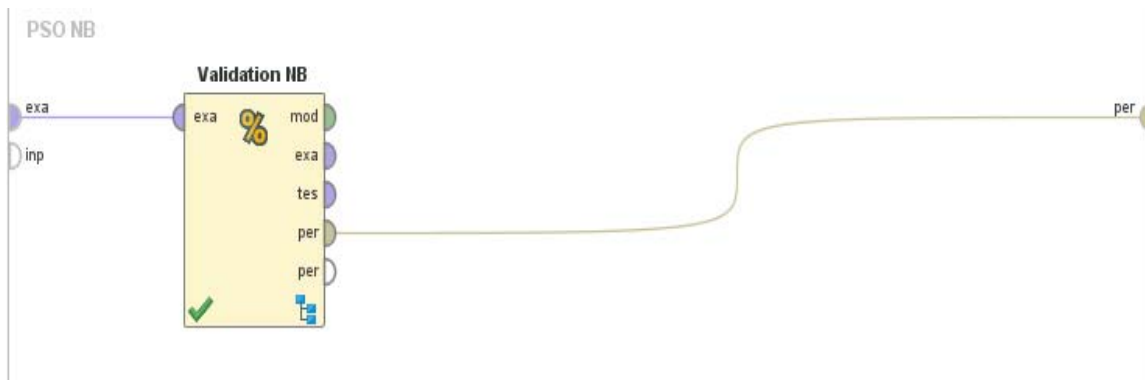


Figure 8: Cross Validation Inside PSO for NB

While the contents of figure 7 there is an algorithm as in figure 4 and the contents of figure 8 as in figure 5.

5.5. Evaluation Phase

The evaluation stage aims to determine the value of the usefulness of the model that has been successfully created in the previous step. For evaluation use 10-fold cross validation. In this test, the data used is the data that has been through preprocessing. The data is retrieved from the Read Excel operator, this is done because the dataset is stored in Excel form. Process documents from files to convert files into documents. The validation process consists of training data and data testing. At this stage also use the Set Role that serves to determine the field in the class and then use Particle Swarm Optimization (PSO) so that the resulting

accuracy is higher. From the previous modeling results The following will explain the ROC Curve and Confusion Matrix of each algorithm. AUC Classification Category has value: 1). 0.90 - 1.00 = *excellent classification*. 2). 0.80 - 0.90 = *good classification*. 3). 0.70 - 0.80 = *fair classification*. 4). 0.60 - 0.70 = *poor classification*. 5). 0.50 - 0.60 = *failure*.

The ROC and Confusion Matrix results from each algorithm are as follows:

1. ROC Curve

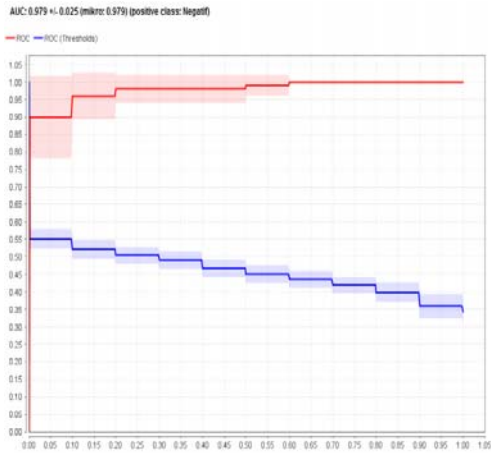


Figure 9: ROC Curve (SVM)

Figure 9 explain about ROC curve SVM obtains the AUC (Area Under Curve) value is 0.979 where the diagnosis is Excellent Classification.

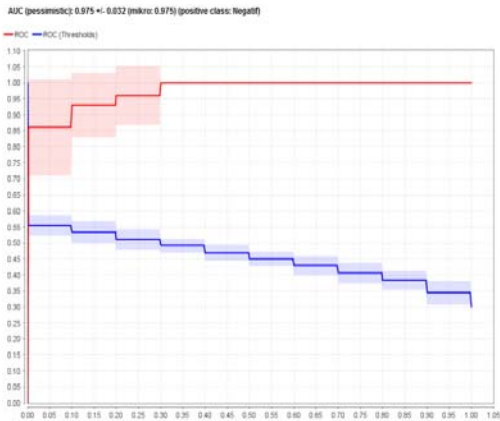


Figure 10. ROC Curve (SVM PSO)

Figure 10 explain about ROC curve SVM PSO obtains the AUC (Area Under Curve) value is 0.975 where the diagnosis is Excellent Classification.

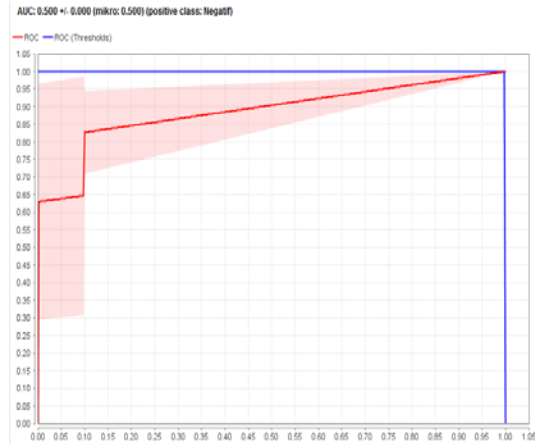


Figure 11: ROC Curve (NB)

Figure 11 explain about ROC curve NB obtains the AUC (Area Under Curve) value is 0.500 where the diagnosis is Failure.

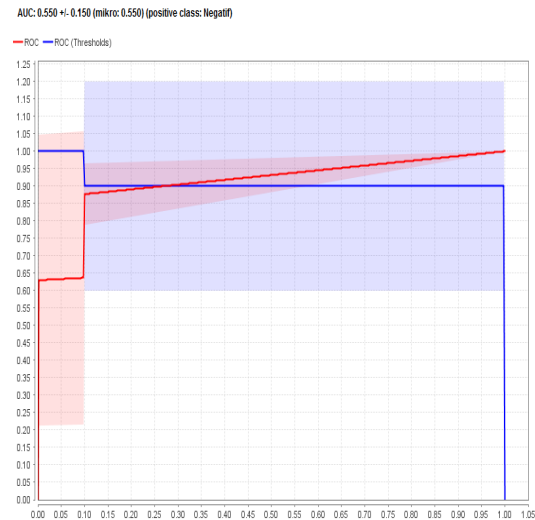


Figure12: ROC Curve (NB PSO)

Figure 12 explain about ROC curve NB obtains the AUC (Area Under Curve) value is 0.550 where the diagnosis is Failure.

2. Confusion Matrix

Confusion Matrix provides performance classification assessment based on true and false objects. The Confirmation Matrix contains actual and predicted information on the classification system.

Table 1. Confusion Matrix Svm

accuracy: 87.50% +/- 6.42% (mikro: 87.50%)			
	true Positive	true Negative	class precision
pred. Positive	77	2	97.47%
pred. Negative	23	98	80.99%
class recall	77.00%	98.00%	

Table 1. Explain about confusion matrix that SVM get true positive is 77 data, false negative is 2 data, false positive is 23 data and true negative is 98 data.

Table 2. Confusion Matrix Svm Pso

accuracy: 90.50% +/- 6.87% (mikro: 90.50%)			
	true Positive	true Negative	class precision
pred. Positive	82	1	98.80%
pred. Negative	18	99	84.62%
class recall	82.00%	99.00%	

Table 2. Explain about confusion matrix that SVMPSO get true positive is 82 data, false negative is 1 data, false positive is 18 data and true negative is 99 data.

Table 3. Confusion Matrix Nb

accuracy: 89.50% +/- 6.10% (mikro: 89.50%)			
	true Positive	true Negative	class precision
pred. Positive	98	19	83.76%
pred. Negative	2	81	97.59%
class recall	98.00%	81.00%	

Table 3. Explain about confusion matrix that NB get true positive is 98 data, false negative is 19 data, false positive is 2 data and true negative is 81 data.

Table 4. Confusion Matrix Nbpso

accuracy: 92.00% +/- 6.00% (mikro: 92.00%)			
	true Positive	true Negative	class precision
pred. Positive	97	13	88.18%
pred. Negative	3	87	96.67%
class recall	97.00%	87.00%	

Table 5. Explain about confusion matrix that NB PSO get true positive is 97 data, false negative is 13 data, false positive is 3 data and true negative is 87 data.

The comparison of accuracy results and AUC Algorithm has been used as follows:

Table 5. Comparison Of Accuracy Results And Auc

Algoritma	Accuracy	AUC
SVM	87.50%	0.979
SVM + PSO	90.50%	0.975
NB	89.50%	0.500
NB +PSO	92.00%	0.550

In this research, the results of the calculation of the SVM method get an accuracy value of 87.50% while the results of SVM calculation (PSO) get an accuracy value of 90.50%. The NB method calculation results produce 89.50% accuracy value, while the NB (PSO) calculation results get an accuracy value of 92.00%. SVM and NB accuracy has a difference of about 2%, while SVM (PSO) and NB (PSO) have a difference in value of about 1.5%. Based on Table 4.8, it can be concluded that PSO-based NB accuracy has higher accuracy compared to SVM, SVM (PSO), NB and PSO can increase accuracy values for SVM and NB methods. The text classification model used can make it easier to find positive news articles and negative news articles. Based on news data that is processed using the Rapidminer tool, news article data will be separated into words that have weight on each word. These words will be used to see words that relate to sentiments that often appear and have the highest weight and can be used to find out positive news articles and negative news articles.

5.2. Deployment Phase

In the deployment phase, application design using the news dataset of the figure of the Coordinator Minister of Maritime Affairs Period 2016-2019. At this phase, the results obtained from all previous phases are used in real terms

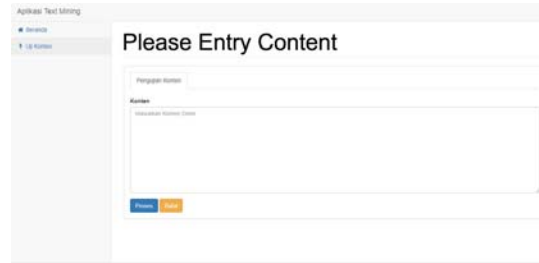


Figure 13: Graphical User Infarce – Form Input

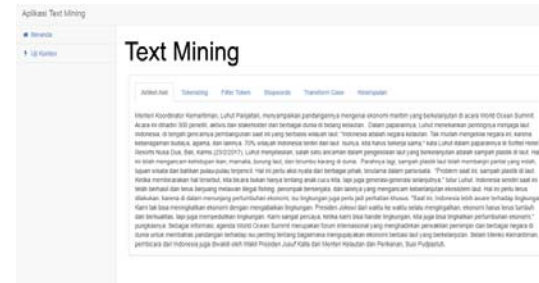


Figure 14: Graphical User Interaface - Result

Figures 13 and 14 are a view of a GUI that has been created. Figure 8 is describe the content test menu page. On this page the process of testing of the text. Figure 9 is results content for tokenize, token filter, stopwords and transform cases.



Figure 15. The Result of Text Mining

6. CONCLUSION

Based on experiments that have been done then the results obtained from testing data NB, NB (PSO), SVM and SVM (PSO) are compared. Accuracy of NB (PSO) gets higher accuracy compared to SVM, SVM (PSO) and NB. Data sources obtained as collections data are more than one online media source when compared to the literature taken as a reference. So it can be

concluded that the best application of optimization in this model is Naive Bayes based Particle Swarm Optimization (PSO) can provide solution to the problem of classification in the case of sentimental analysis of the Coordinator Minister of Maritime Affairs Period 2016-2019. NB PSO algorithm provides solutions to the analysis of sentiments from the content of various online media news optimally.

REFERENCES

- [1] Alindu, Paulina. 2013. Twitter Used by Indonesian President: An Sentiment Analysis of Timeline. Kupang : Information Systems International Conference (ISICO) 2-4 Desember 2013.
- [2] Basaria , Abd. Samad Hasan, Burairah Hussina, dan I. Gede Pramudya Anantaa,Junta Zeniarja. 2012. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Malaysia : Procedia Engineering 53 (2013) 453 – 462 Malaysian Technical Universities Conference on Engineering & Technology 2012, MUCET 2012 Part 4 Information And Communication Technology.
- [3] Buntoro, Asrofi Ghulam. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. Jakarta : Integer Journal Vol 1 No 1 Maret 2016:32-41.
- [4] Dhande, Lina L, dan Dr. Girish Patnaik. 2014. Review of Sentiment Analysis using Naive Bayes and Neural Network Classifier, ISSN :2319-8885. India : International Journal of Scientific Engineering and Technology Research Vol. 03 Issue.07 Mei 2014 Page 1110-1113.
- [5] Feldman, R., & Sanger, J. (2007). The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge: Cambridge University Press.
- [6] Hadna, Nurrin Muchammad Shiddieqy, Paulus Insap Santosa dan Wing Wahyu Winarno. 2016. Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter, ISSN :2089-9815. Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi (Sentika 2016).
- [7] Hamzah, Amir. 2012. Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Text Berita dan Abstrak Akademis, ISSN:1979-911X. Yogyakarta : Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III.
- [8] Indrayuni, Elly. 2016. Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization, ISSN: 2338 – 8161.Pontianak : Jurnal Evolusi Volume 4 Nomor 2 – 2016.
- [9] Ipmawati, Joang,Kusrini dan Emha Taufiq Luthfi. 2016. Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen, ISSN :2461-0690.Jakarta : IJSE – Indonesian Journal on Software Engineering Vol. 2 No 2.
- [10] Manalu, Boy Utomo. 2014. Skripsi Analisis Sentimen Pada Twitter Menggunakan Text Mining. Medan : Universitas Sumatera Utara.
- [11] Mukminin, Amirul, Dwiza Riana. 2017. Komparasi Algoritma C4.5, Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah, ISSN: 2355-6579 E-ISSN: 2528-2247. Bandung : JURNAL INFORMATIKA, Vol.4 No.1 April 2017, pp. 21-31.
- [12] Nugroho, Anto Satriyo, Arief Budi Witarto, dan Dwi Handoko.2003. Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika. <http://asnugroho.net/papers/ikcsvm.pdf>.
- [13] Nurhuda, Faishol, Sari Widya Sihwi dan Afrizal Doewes. 2013. Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier, ISSN :2301-7201. Jakarta : Jurnal ITSMART Vol. 2 No. 2 Desember 2013.
- [14] Putera,Andri Donnal. 2017. Disebut Sering Ambil Alih Tanggung Jawab Kementerian Lain, Apa Kata ?. <http://ekonomi.kompas.com/read/2017/11/01/20000526/disebut-sering-ambil-alih-tanggung-jawab-kementerian-lain-apa-kata->.
- [15] Rupilele, Frits Gerit John, Danny Manongga, dan Wiranto Herry Utomo.2013. Sentiment Analysis Of National Exam Public Policy

- With Naive Bayes Classifier Method (NBC),
ISSN : 1992-8645 E-ISSN : 1817-3195. __.
Journal of Theoretical and Applied
Information Technology 10th December 2013.
Vol. 58 No.1.
- [16] Safari, Cecep Barkah. 2012. Skripsi Analisis Sentimen Menggunakan Metode Klasifikasi SVM.Jakarta : Universitas Pendidikan Indonesia.
- [17] Saputra, Nurirwan, Teguh Bharata Adji, dan Adhistya Erna Permanasari.2015. Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan SVM.Yogyakarta : Jurnal Dinamika Informatika Volume 5, Nomor 1, November 2015.
- [18] Suyanto. 2017. Data mining Untuk Klasifikasi Dan Klasterisasi Data.Bandung:Informatika Bandung.
- [19] Utami, Lilyani Asri. 2017. Analisis Sentimen Opini Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization.Jakarta : Jurnal Pilar Nusa Mandiri Volume 13 No.1, Maret 2017.
- [20] Widodo, Pangestu, Januar Adi Putra, Suwanto Afiadi, Agus Zainal Arifin, dan Darlis Herumurti. 2016. Klasifikasi Kategori Dokumen Berita Berbahasa Indonesia Dengan Metode Kategorisasi Multi-Label Berbasis Domain Specific Ontology, ISSN : 2407 – 391.Surabaya : Jurnal Ilmiah Teknologi Informasi Terapan Volume II, No 2, 15 April 2016.
- [21] Zgheib, Wassim A dan Aziz M. Barbar.2017. A Research using Support Vector Machines to Classify the Sentiments of Tweets.Lebanon : International Journal of Computer Applications (0975 – 8887) Volume 170 – No.2, July 2017.