Twitter Comment Predictions on Dues Changes BPJS Health In 2020

Riza Fahlapi[1)], Yan Rianto[2)]
[1)2)] STMIK Nusa Mandiri, Indonesia
[1)]riza.fahlapi@gmail.co.id , [2)]y_rianto@yahoo.co.jp

Abstract: The Social Security Administering Body (BPJS) is a facility established by the government in providing services to citizens in The field of health welfare. The Spirit of cooperation in the utilization of health services which is very much currently a constraint in the budget is still insufficient in covering health services as a whole. For this reason, government policy is following with PERPRES No. 75 in 2019, the Government officially raised the BPJS Health contributions for 2020. The increase in BPJS Health contributions certainly caused a lot of comments. Namely Twitter, one of the social media that is used by the public to express disapproval or support for this government policy. This study, testing was carried out related to the prediction of comments from social media on community responses to the increase in BPJS Health contributions taken by the government. In the test carried out 3 (three) input algorithms. For every single algorithm including getting results through the K-NN method with an accuracy of 71.83% and AUC value of 0812, for the Naïve Bayes method produces an accuracy of 81.63% and AUC value of 0586. As for the C 4.5 method, the accuracy is 65.37% and the AUC value is 0628. While testing conducted through the Ensembles Vote method which combines the 3 algorithms above gives the best results with an accuracy of 80.10% and AUC value is 0871 for Twitter comment predictions.

Keywords: BPJS Health, dues, predictions, comments, Twitter, Ensembles Vote

INTRODUCTION

Twitter is widely used for its ease of uploading text in micro social media blogs. Twitter in limited comments can be categorized as the core information of a post Twit uploaded on Twitter. A lot of comments on Twitter by a particular account, famous or viral because the follower ALLOWS the RT as well as more and more information is the trend, update and or become viral news. For that in some research related detection conducted by some researchers to use the widely used Twitter dataset. A specific technique used algorithm-based, to detect the impact of a thing or information with a method that can be researched in writing.

Twitter comments on BPJS Health for the ministry factor, but in this case, the author discusses the increase in the BPJS health premium in the year 2020. In the study, research on community opinion on social media Twitter (Yanis, 2018). While the method used in the author's research is currently in the process to preprocessing the data including using tokenization, cleansing, filtering, stemming, and Stopword. For the process of Ensemble Vote algorithms are also used include K-NN, C4.5 and Naïve Bayes. The Data used is user behaviour and Twitter comments.

Evaluating Twitter's response to comments, frequencies, re-tweets, forward tweets, themes, and accounts that gave the post in Twitter, the authors of the analysis in assessing BPJS health dues increase in 2020 using several algorithms such as K-Nearest Neighbor (K-NN), Naïve Bayes, C 4.5, Ensemble Vote to ensure prediction as well as a winning method of the best Social Security Administering Agency (BPJS) health is one of the types of facilities in the field of health welfare. Indonesian Citizen (WNI) and foreign citizen working in Indonesia for more than 6 (six) months, must participate in contributing to BPJS Kesehatan. BPJS Kesehatan participants who become contributors to the service by paying the premium and there are also recipients of dues assistance. With the spirit of cooperation, the utilization of health services very much nowadays become an obstacle in the budget still not enough in covering the health service thoroughly.

LITERATURE REVIEW

Scope of this research is to test the Ensemble Votes algorithm which contains the algorithm of K-NN, C 4.5, Naïve Bayes. Used through sourcing Twitter comments over 2000 Twitter posts. The increase of dues BPJS health certainly resulted in many comments regarding this policy. One of them is using social media as a medium to

* Corresponding Author

devote the dismiss destination and support to this government step. This research, it is expected based on information sourced from Twitter by classifying to project the general public's response to the increase of the BPJS health dues taken by the government. On the other hand.

BPJS Health Dues increase has been officially signed by the president through Presidential Regulation (PERPRES) No 75 the year 2019 on amendments to Perpers No. 82 the year 2018 about health insurance on 24 October. In article 34 the regulation mentioned the increase in BPJS health dues up to 100%. Class I from IDR 80,000 to IDR 160,000, class II from IDR 51,000 to IDR 110,000, class III from Rp 25,500 to Rp 42,000. And for premium, the maximum limit for the company becomes IDR 600,000/month. BPJS Health dues increase is certainly inflicting polemics in the middle of society. Community response with a hike that will begin on January 1, 2020. The public is pairing the quality of service that is assessed still many problems with the plan of tuition increase up to 100 percent.

METHOD

Social Security is the right of every person whose fulfilment is guaranteed by the Constitution and the prevailing laws and regulations. Its implementation is mainly associated with the government's efforts in alleviating poverty which is done gradually according to the ability of the state, private, community in its financing. Metode Financing involves the participants themselves, through the payment obligations of dues called social insurance.

In conducting text mining, the text of documents used must be prepared in advance. Based on the inconsistency of the text data structure, the text mining process requires some initial stages. One of the implementations of text mining is the Text Preprocessing (INFORMATIKALOGI, 2016) stage, including:
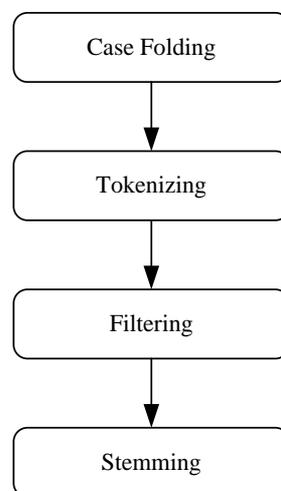
Fig 1. Text Mining process stages (text preprocessing) Gata Framework
Source: (INFORMATIKALOGI, 2016)

Gata Framework is an online text processing tool created by Dr Windu Gata, M. Kom to convert unstructured text data into structured text data so that the data is ready for use in predetermined modelling techniques. These Tools can be accessed with the address http://www.gataframework.com. Predictions are a systematic process of estimating something that is most likely to happen in the future based on past and present information so that mistakes (the difference between what is happening with the approximate outcome) can be minimized. Predictions do not necessarily give a definite answer to the incident that will occur, but rather seek to find the answer as close as possible that will happen (Herdianto, 2013:8). The definition of predictions equals forecasts or forecasts. According to the Big English Dictionary, predictions are the result of predicting or forecasting activities or estimating future value by using past data. Predictions show what will happen to a particular circumstance and is an input for the planning and decision-making process. Predictions can be based on scientific or subjective methods. Take for example, weather predictions are always based on data and the latest information based observations included by satellites. Likewise, predictions of earthquakes, mountains erupt or disaster in general. However, predictions like football matches, sports, etc. are generally based on subjective views with their viewpoints that predict them.

K-Nearest Neighbor (K-NN)

This algorithm is to classify the new object based on the attributes and sample training. You are matched and based only on memory Drive a query point, it will be found several K objects or (training points) closest to the

query point. The classification of K objects K-Nearest Neighbor (K-NN) algorithm using classification as a prediction value of the new instance query. The K-Nearest Neighbor (K-NN) method algorithm works based on the shortest distance from the instance query to the training, sample to determine its K-NN. The sample training is projected into a dimension space, where each dimension represents the features of the data. This space is divided into sections based on sample training classification. A point in this space is characterized by the class C if Class C is the most widely encountered classification on the closest neighboring fruit from that point. Close or far neighbors are usually calculated based on Euclidean Distance. Euclidean distances are most commonly used calculating distances. The Euclidean distance serves to test the size that can be used as an interpretation of proximity distance between two objects. Faith Matrix D (A,B) is a scalar distance of both vector A and B of the matrix with dimension size. The greater the value D will the further the level of similarity between the two individuals and vice versa if the value of D is increasingly smaller it will be closer to the level of the likeness between the individuals.

The best k value for this algorithm depends on a high k-value reduces the effect of the noise in the classification but makes the boundary between each classification more blurred. A good k value can be chosen with the optimization of a parameter, for example by using cross-validation. A special case where the classification is predicted based on the closest training data (in other words, K = 1) is called the Nearest Neighbor algorithm. The accuracy of the K-NN algorithm is strongly influenced by the presence or absence of irrelevant features or if the weight of the feature is not equal to its relevance to the classification.

C4.5 Algorithm

C4.5 is a flowchart-like structure where any internal nodes (nodes that are not a leaf or not the outer node) are tests on the attribute variable, each branch is the result of the test, while the outer nodes are the label (Han et al, 2012). Algorithm C4.5 and the decision tree are two models are inseparable, because to build a decision tree required algorithm C4.5. The C 4.5 and ID3 algorithms were created by a researcher in the artificial intelligence field named J. Rose Quinlan in the late 1970s. The C 4.5 algorithm makes the decision tree from top to bottom, where the top attribute is the root, and the bottom one is called leaf (leaf). In general, the algorithm process flows C 4.5 to build the decision tree in data mining are:
- Select the attribute as the root node.
- Create a branch for each value.
- Divide the case into branches.
- Repeat the process for each branch until all cases in the branches have the same class.

Naïve Bayes

This method uses the Bayes theorem, which was discovered by the 18th century Thomas Bayes (Suyanto,2017). The Naive Bayes (NB) classification is statistical classifications which can be used to predict the probability of a class membership. The NB method takes two stages in the text classification process, namely the training stage and the classification stage. At the training stage, the analysis process of the sample document in the form of vocabulary selection, which is a word that may appear in the collection of sample documents which can be a representation of the document. Next is the probability determination for each category based on sample documents. At the classification stage determined the value of the category of a document based on the term that appears in the document classified (Hamzah,2012).

Ensembles Vote

An Ensemble is a supervised learning algorithm, as it can be trained and then used to make predictions. The hypotheses contained not need to have space from the models built and can prove to have more flexibility in functions that can represent algorithms. This flexibility, some techniques of ensemble models tend to reduce problems relating to the installation of training data.

Experimentally, the ensemble seems to produce better results when there is significant diversity between models. Therefore, many ensemble methods attempt to promote diversity among the models they incorporate. While it may be non-intuitive, a more random algorithm (like a random decision tree) can be used to produce a stronger ensemble than a very intentional algorithm (such as a decision tree that reduces entropy). It uses a variety of powerful learning algorithms but has proven to be more effective than using techniques that attempt to knock down models to promote the diversity of algorithms used (Alhamad, Azis, Santoso, & Taliki, 2019).

E. Model CRISP-DM

In CRISP-DM, a data mining project has a life cycle divided into six phases (Fig. 2). The entire sequential phase depends on the output of the previous phase. The important relationship between phases is depicted with arrows. For example, if the process is in the modelling phase. Based on the behaviour and characteristics of the model, the process may have to return to the data preparation phase for further improvement of data or to move forward to the evaluation phase (Luthfi, 2009).
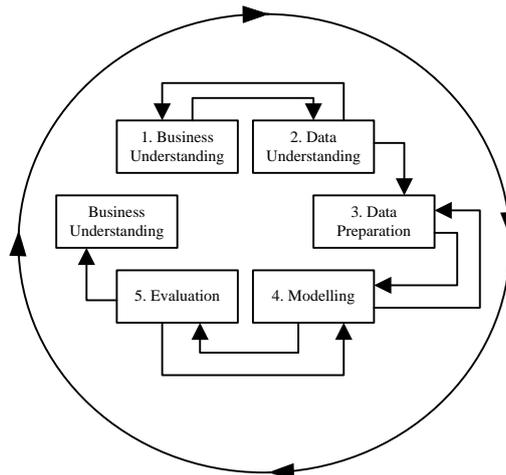
Fig 2. CRISP-DM Model Concept problems.
Source: (Luthfi, 2009)

In sensitivity and specificity can be used as a statistical measure of the performance of binary classification, sensitivity and specificity used to measure the best models and to select the most efficient models. The sensitivity of measuring true positive proportions is correctly identified, the specificity of measuring the true negative proportion identified correctly. False positives are known as type 1 errors, occurring when cases that should be classified as negatively classified as positive. False is a negative known as type 2 error, occurring when the case should be classified as positive classified as negative (Bramer, 2013).

Table 1
Confusion Matrik Table

Correct	Classified as	
	+	-
+	*True Positive*	*False negative*
-	*False positive*	*True Negative*

Cross-validation is a statistical method that can be used to evaluate the performance of a model or algorithm where data is separated into two subsets i.e. learning process data and validation/evaluation data. Models or algorithms are trained by a subset of learning and validated by a subset of validations.

The Area under Curve (AUC) as one indicator to evaluate the performance of the classifier. The AUC is the area under the ROC curve. The AUC has significant potential to increase empirically convergency of the experiment, as it separates the predictive performance from the operation of the condition, and is a predictive general measure. Furthermore, the AUC has clear statistics of its interpretation.

Table 2
Value AUC Table

AUC	Meaning	Symbol
0.90 – 1.00	*excellent classification*	
0.80 – 0.90	*excellent classification*	
0.70 – 0.80	*fair classification*	
0.60 – 0.70	*Poor classification*	
< 0.60	*Poor classification*	

Sumber: Gorunescu (Wahono, 2015)

A ROC (Receiver Operation Characteristic) curve according to Provost and Fawcett explains another measurement using the ROC curve depicting a trade-off of true positive against false positives.
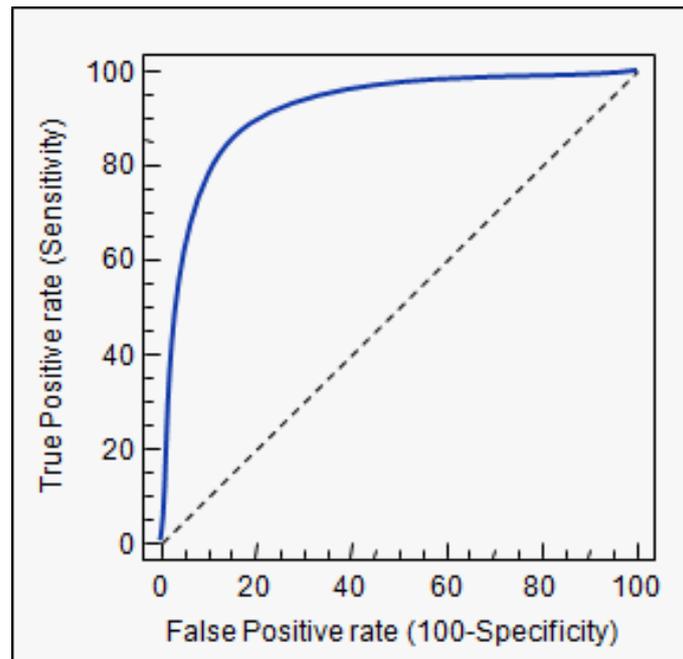
Fig 3. ROC Curve

RapidMiner is open-source software. RapidMiner is a solution for analysis of data mining, text mining and prediction analysis. RapidMiner uses a variety of desktop techniques and predictions in providing insight to the user so that it can make the most informed decisions. RapidMiner has approximately 500 data mining operators, including operators for input, output, data preprocessing and visualization. RapidMiner is a standalone software for data analysis and as a data mining machine that can be integrated into its products. RapidMiner is written using Java language so it can work in all operating systems (Aprilla Dennis, 2013)

Twitter is one of the most well-known microblogging services and allows its users to write or create statuses and comment Twitter or tweets. Social Media Twitter is used to express any opinion or opinion of a product, service or another thing. Twitter was created by Jack Dorsey in 2006 and first rolled out in the virtual world of July 2006 with a http://www.Twitter.com address still in use today.

There are several stages for this study :
- The first stage of this study is the process of collecting comments data related to BPJS. by more than 2000 public comment data consisting positive and negative comment.
- The initial data processing beginning with the text preprocessing stage. In the weighted case will be done using the Term Frequency Iners Document Frequency (TF-IDF).
- The Ensembles Vote Algorithm in there proceed K-NN, C4.5, NB Algorithm for classifications. The result of classifying the algorithm method will be compared and the result will be re-optimized to obtain the best algorithms method.
- Accuracy algorithm will be measured by Confusion Matrix and ROC curve and accuracy for the result.

DISCUSSIONS

Business Understanding

At the stage of business Understanding, conducted at the business understanding stage, the understanding of the research object. , is the research is to predict the response of Twitter users to the increase in the BPJS health dues on Twitter comments to get the best results and benefits to the development of BPJS Kesehatan in the future for the advancement of services BPJS Healthcare according Government policy (www.hukumonline.com/pusatdata,2019) PERPRES No. 75 the year 2019, the government is raising the dues BPJS Kesehatan for the year 2020.

 the following conclusions Understanding of the research object is done by digging the information through Twitter against the research object. The motivation in this phase is the news presented usually in the form of text on digital media grouped by the contents of the discussion of the respective categories of news. Twitter is now not only limited as a medium to read the essence of comments only, but can also be used to see the issues that occur can even be used to see the influence of a policy that is taken by the government. This comment prediction was done to find a classification method that can help in determining the comments of positive and negative news articles.
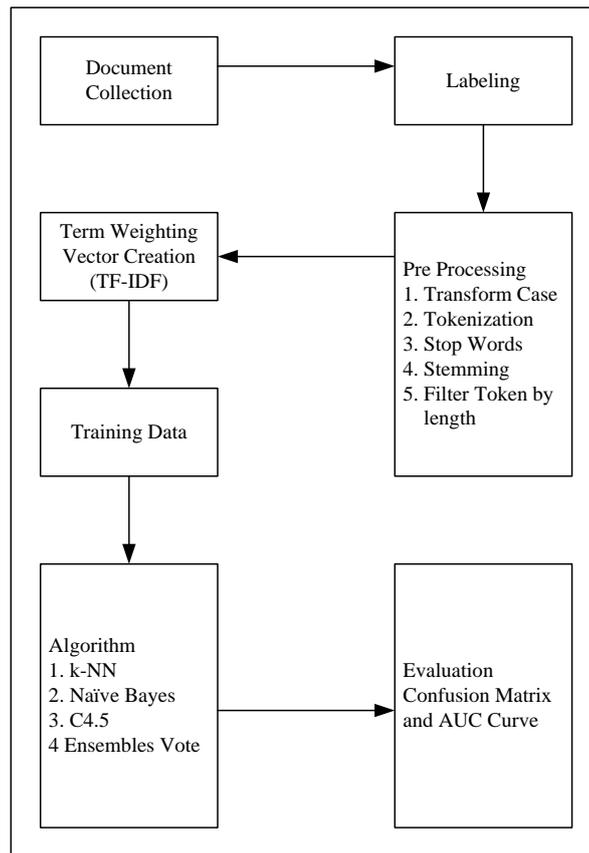
Fig 4.Framework

Data Understanding

At the data understanding stage, a raw data retrieval process is done according to the required attributes. Data obtained from the online news media site www.twitter.com.
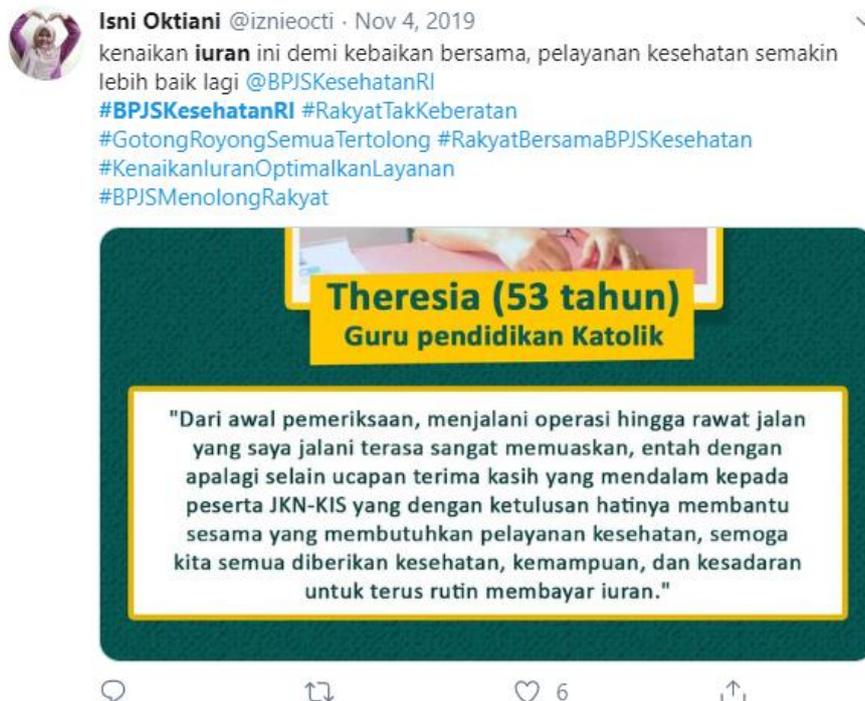
Fig 5. Twitter comments Example

Data that has been labelled by others is then taken as much as 2091 comment data based on each Twitter user to do the cleansing process to remove the duplicate data and eliminate other data that is not related to BPJS health. By using the data source obtained, created a dataset with an attribute that is a comment data containing all the comments data about BPJS Health which is a label consisting of 1046 positive comment data and 1045 negative comment data of each data. The comment Data is prompted and stored in the form of an. xls extension.

Table 3
Twitter Comment Data

No.	Twitter Comment	Sentiment
1	nonton berita "mahasiswa demo iuran bpjs diturunkan" kalo pelayanan kesehatan turun ntar demo lagi aku bingung dengan warga +62	negative
2	Seruan Aksi: 16 Desember 2019 BPJS GAGAL BPJS GAGAL BPJS GAGAL Sistem kesehatan yang sangat liberal, menyalahi kons…	negative
3	Mau kekantor bpjs kesehatan tapi jauh?? Tenang, ada mobile customer service & bisa melalui mobile JKN yg bisa di download dari appstore/ playstore	positive

Data Preparation

The data preparation phase is the stage with the data setup process aimed at obtaining clean data and ready for use in research. In the text mining the initial stage that will be done is the text preprocessing stage, at this stage researchers use the Tools Gata Framework. Here are the steps are done in text preprocessing :

Text Preprocessing with Tools Gata Framework :
• Case Folding
Not all text documents are consistent in the use of capital letters. Case Folding roles are needed to convert the entire text in a document into a standard form (lowercase).
• Tokenizing
Tokenizing will cutting phase of the input string based on each word that follows it as (???, "",-,).
• Cleansing
Remove the mention of the mentioned username (@username), remove the hastag (#hashtag), remove the punctuation, delete the number because only the text data is used, remove the link (https://), remove special characters such as symbols or emoticons, eliminate foreign words because only take words that speak Indonesian only. For the word in front of it, there is a word "no" will be normalized by using an underscore "_" to have a clear meaning, such as "not good" to "tidak_baik" because the word "tidak" means negative while the word "good" means positive.
• Filtering
The filtering stage is the stage of taking important words from the token results. Can use a stoplist algorithm (discarding less important words) or wordlist (save important words). At this stage slang words (in Indonesian) must be normalized to their standard form, for example "utk" to "untuk".
• Stemming
The stemming process is the process of finding the root of the word that results from the filtering process. Root search for a word or commonly called a root word can reduce the index results without having to eliminate the meaning. For example "naik kan" to "naik".

Modelling Stages

Is the stage of selecting mining techniques by determining the algorithm to be used. The tool used is RapidMiner version 9.1. The results of testing the model conducted are to correctly classify email complaints and not email complaints using the Naïve Bayes algorithm and Support Vector Machine to get the best accuracy value. Following is the design of the Rapidminer model used, i.e :
• Model Testing with Algorithms
It is the selection phase of mining techniques by determining the algorithm to be used. The tools used are RapidMiner version 9.5. The results of the model testing conducted are classifying positive news articles and negative news articles using the k-Nearest Neigbour algorithm, C4.5, Naïve Bayes and Ensemble Vote to get the best accuracy value.
Following is the design of the RapidMiner model that is used i.e :
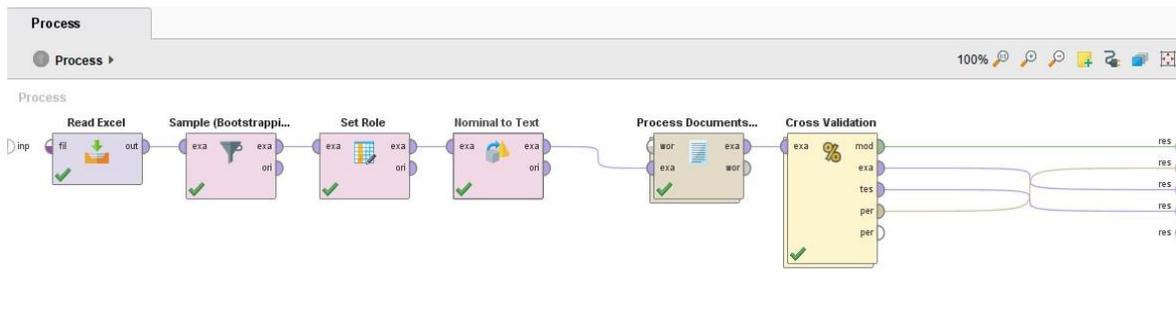
Fig 8. Design Algorithm Models Evaluation Model

The evaluation stage aims to determine the usability value of the model that was successfully created in the previous step.

The following is the design process of testing the k-Nearest Neighbor algorithm model used, i.e :
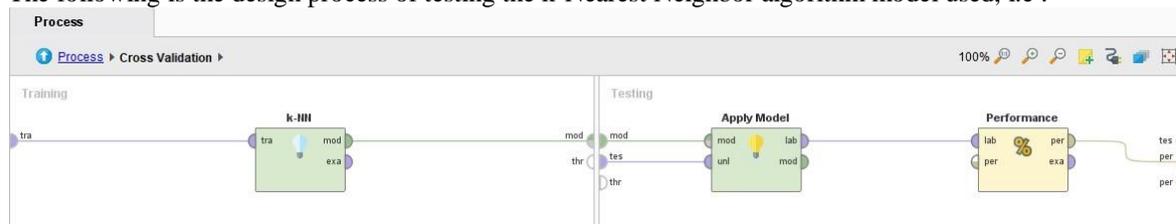
Fig 9. Design Process for k-NN

Following this is the design process of testing the Naïve Bayes method model used, i.e :
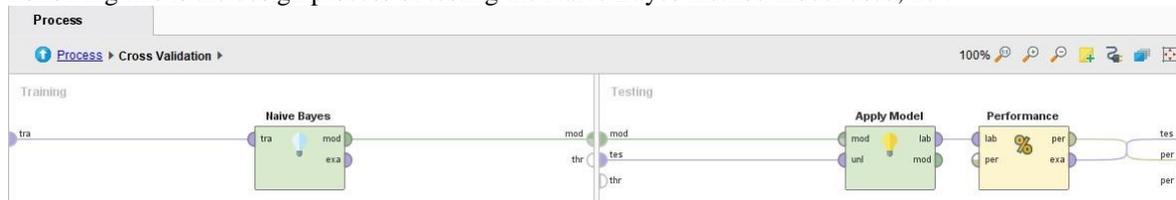
Fig 10. Design Process for Naïve Bayes

The following is the design process of testing the C4.5 method model used :
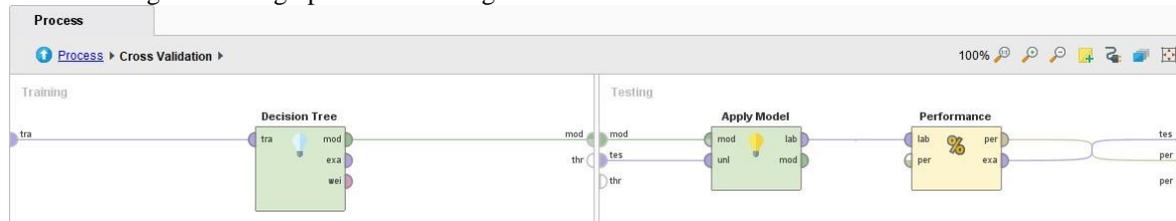
Fig 11. Design Process for C4.5

The following is the design process of testing the Ensembles Vote method model used :
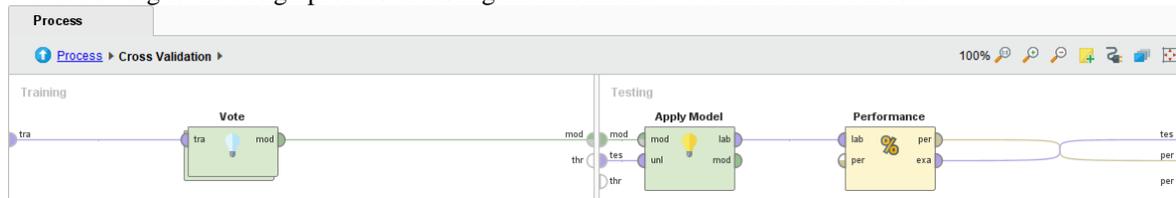
Fig 12. Design Process for Ensembles Vote

The Ensembles Vote algorithm is used to conduct training on the input dataset then the results are used for testing using the same dataset. The following is the design process of testing the Ensembles Vote method model used :
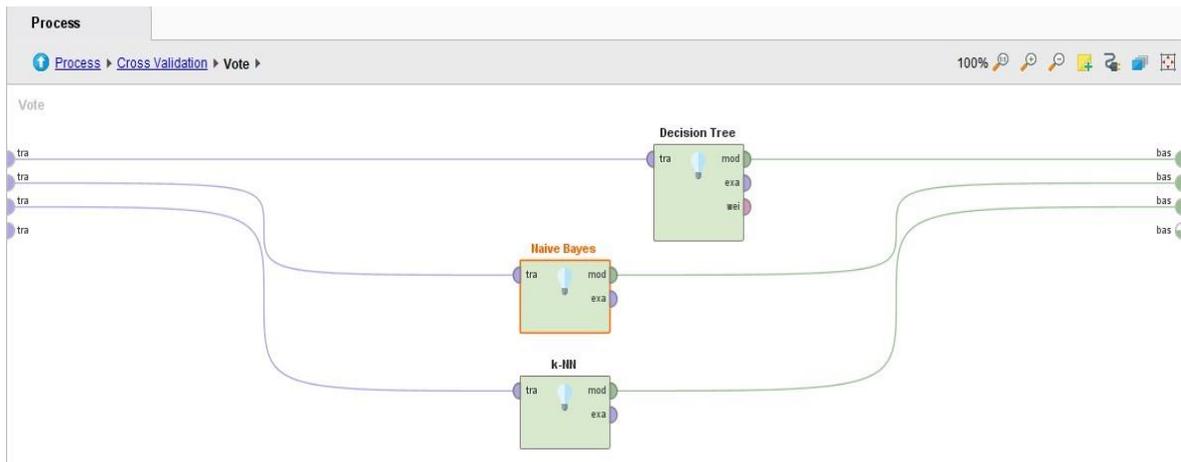
Fig 13. Model Process Design for Ensembles Vote

In this test, the data used is clean data that has been through preprocessing. The data is taken from the Read Excel operator, Process documents from files to convert files to documents. From the results of modelling that has been done before. The following will explain the ROC curve and the Confusion Matrix of each algorithm.

The ROC curve and the Confusion Matrix of the k-Nearest Neighbor (k-NN) algorithm
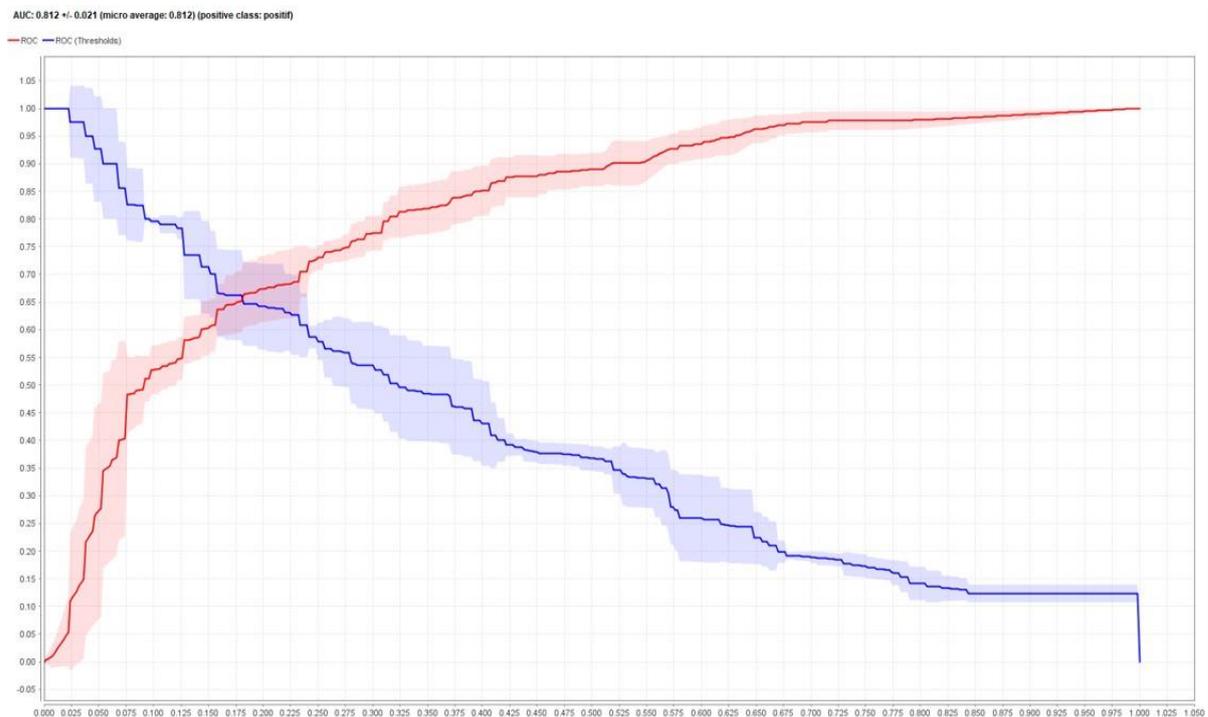
Fig 14. The k-Nearest Neighbor ROC curve

Table 4
Confusion Matrix k-Nearest Neighbor

accuracy: 81.20% +/- 0.02% (mikro: 81.20%)			
	true Positif Negatif	true Negatif	class precision
pred. Positif	730	256	74.04%
pred. Negatif	333	772	69.86%
class recall	68.67%	75.10%	

ROC and Confusion Matrix curves from the Naïve Bayes algorithm (NB)
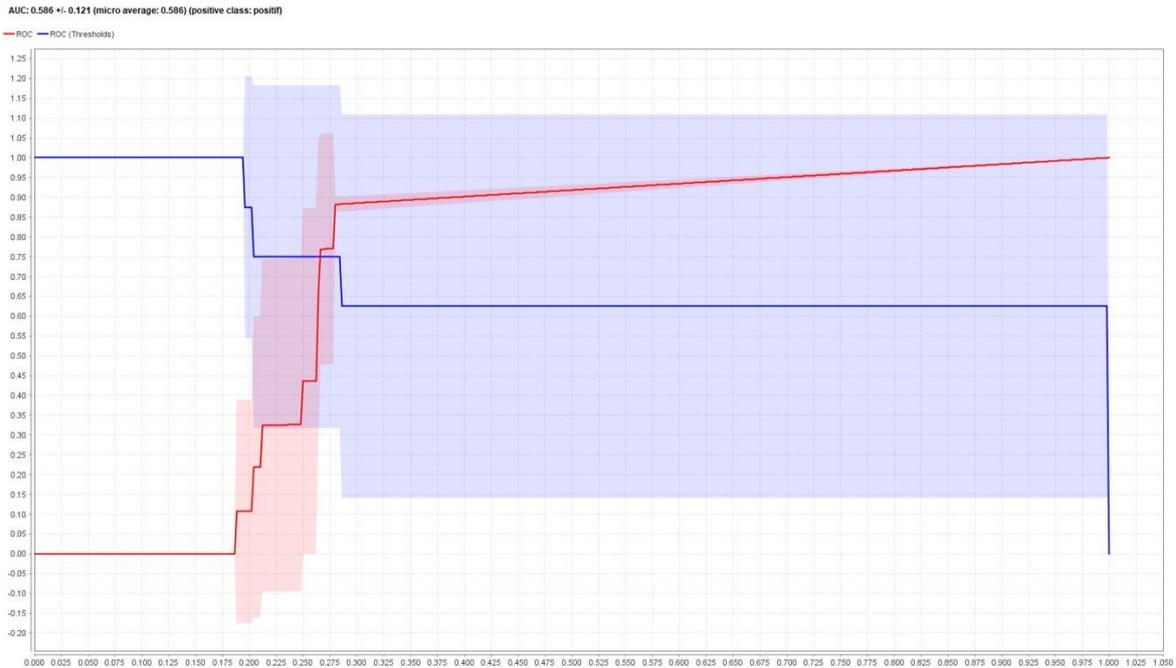
Fig 15. ROC Curve Naïve Bayes (NB)

Table 5
Confusion Matrix Naïve Bayes

accuracy: 81.63% +/- 1.50% (mikro: 81.64%)			
	true Positif	**true Negatif**	**class precision**
pred. Positif	808	129	86.23%
pred. Negatif	255	899	77.90%
class recall	76.01%	87.45%	

ROC and Confusion Matrix curves from the C4.5 algorithm

Fig 16. ROC curve C4.5

Table 6
Confusion Matrix C4.5

accuracy: 65.37% +/- 1.96% (mikro: 65.38%)			
	true Positif	**true Negatif**	**class precision**
pred. Positif	1042	703	59.71%
pred. Negatif	21	325	93.93%
class recall	99.02%	31.61%	

The ROC curve and the Confusion Matrix of the Ensembles Vote algorithm
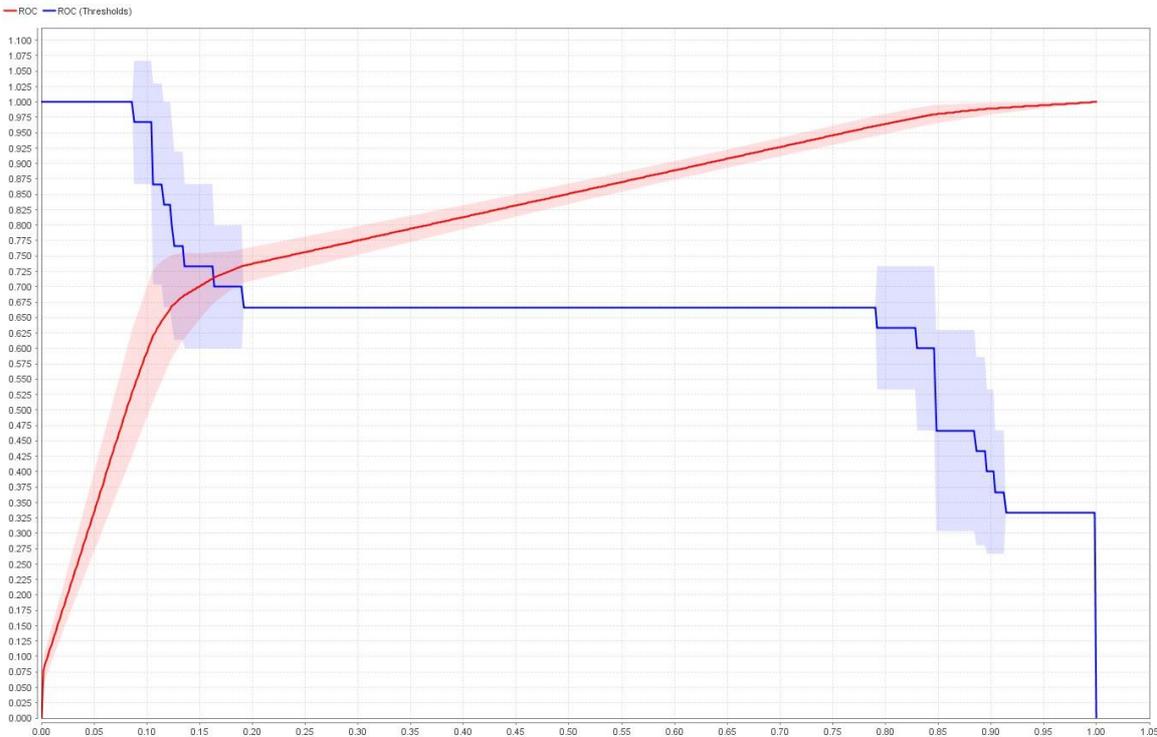
Fig 17. ROC curve Ensembles Vote

Table 7
Confusion Matrix Ensemble Vote

accuracy: 80.10% +/- 0.19% (mikro: 80.11%)			
	true Positif	**true Negatif**	**class precision**
pred. Positif	931	284	76.63%
pred. Negatif	132	744	84.93%
class recall	87.58%	72.37%	

The comparison of the results of the comparison of accuracy and AUC Algorithms that have been used as follows :

Table 8
Comparion Of Accuracy And AUC

Algorithm	Accuracy	AUC
k-Nearest Neighbor	71.8	0.812
Naïve Bayes	81.6	0.586
C4.5	65.3	0.628
Ensemble Vote	80.1	0.811

RESULT

In this research, the k-NN method calculation results get an accuracy value of 71.83% while the Naïve Bayes calculation results get an accuracy value of 81.63%. The calculation results of the C4.5 method produce an accuracy value of 65.37%, while the results of the Ensembles Vote calculation get an accuracy value of 97.93%. C4.5 accuracy has a difference of around +/- 16%, compared to using the Naïve Bayes algorithm. While the k-NN and NB methods have a difference of 9.80% addressing this difference based on the Vote carried out through the Ensembles Vote process based on Table 8, it can be concluded that the accuracy of C4.5 gets the smallest accuracy compared to k-NN, NB. And the best results for Accuracy and AUC comment prediction models are Ensemble Vote with Accuracy 80.10 and AUC 0.811 which are used to make it easier to find positive comments and negative comments.

Based on the comment data that is processed using the RapidMiner Tools, the comment data will separate into words that have weight in each of the words. These words will be used to see words related to sentiments that often appear and have the highest weight and can be used to find out positive comments and negative comments. trials.

CONCLUSION

The Conclusions that can be summarized regarding the prediction of comments on BPJS Health is to make predictions for comments on the increase in BPJS Health contributions. The use of k-Nearest Neighbor (k-NN) algorithm is 71.83% accuracy with AUC 0.812 compared to Naïve Bayes algorithm with 81.63% accuracy with AUC 0.586, and for C4.5 algorithm the accuracy is 65.37% with AUC 0.628, after making predictions through each algorithm above, the authors also make predictions through the Ensemble Vote algorithm method to ensure that the prediction results are with an accuracy value of 80.10% with AUC 0.811. The best comment on this model is Ensembles Votes which gives an accuracy value of 80.10% with AUC 0.811 in the prediction of BPJS Health comments. The accuracy values mentioned with some of the models above have been identified tend to be correct because they have a high prostase. As for the validation, which is more important and determines predictions in the above conditions, it has a good classification interpretation.

ACKNOWLEDGMENT

REFERENCES

Agrawal, C. C. (2018). Machine learning for text. In *Machine Learning for Text*. https://doi.org/10.1007/978-3-319-73531-3

Amriana, A., Joefrie, Y. Y., & Meidji, F. N. (2019). Penerapan Data Mining Untuk Pengelompokan Hasil Diagnosa Penyakit Pasien Pengguna BPJS Kesehatan (Studi Kasus Pada Rsud Undata Palu). *ScientiCO : Computer Science and Informatics Journal*, *1*(1), 51. https://doi.org/10.22487/j26204118.2018.v1.i1.11901

Amrullah, A. A., Tantoni, A., Hamdani, N., Bau, R. T. R. L., Ahsan, M. R., & Utami, E. (2016). Reviewatas Analisis Sentimen Pada Twitter Sebagai Representasi Opini Publik Terhadap Bakal Calon Pemimpin. *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call for Papers Unisbank (Sendi_U)*, *2*(1), 978–979.

Amriana, A., Joefrie, Y. Y., & Meidji, F. N. (2019). Penerapan Data Mining Untuk Pengelompokan Hasil Diagnosa Penyakit Pasien Pengguna BPJS Kesehatan (Studi Kasus Pada Rsud Undata Palu). *ScientiCO : Computer Science and Informatics Journal*, *1*(1), 51. https://doi.org/10.22487/j26204118.2018.v1.i1.11901

Aristianto, D., & Finandhita, A. (2015). *Penerapan Data Mining Untuk Mengklasifikasikan Penyakit Berdasarkan Wilayah Dengan Metode Klasifikasi dan Algoritma C4.5.*

Aprilla Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana Editor, D. C., & Sanjaya, R. (n.d.). *Identitas Belajar Data Mining dengan RapidMiner Hak Cipta © pada Penulis Hak Guna mengikuti Open Content model Desain sampul: Dennis Aprilla C.*

Bina Nusantara, U. (2018). 10 Fold-Cross Validation.

Bruno, L. (2019). 済無No Title No Title. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *3*(1), 1. https://doi.org/10.14421/jiska.2018.31-01

Dzikrulloh, N. N., & Setiawan, B. D. (2017). Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri). *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(5), 378–385.

Ertek, G., Tapucu, D., & Arin, I. (2013). Text mining with rapidminer. In *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. https://doi.org/10.1201/b16023-21

Ferdiana, R., Jatmiko, F., Purwanti, D. D., Sekar, A., Ayu, T., Dicka, W. F., … Indonesia, B. (2019). *Dataset Indonesia untuk Analisis Sentimen*. *8*(4), 334–339.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. https://doi.org/10.1016/C2009-0-61819-5

Heristian, S., Program, M., Ilmu, S., Tinggi, S., Informatika, M., Komputer, D. A. N., & Mandiri, N. (2019). *ANALISIS SENTIMEN BERITA ONLINE TERHADAP EKONOMI INDONESIA DENGAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE*.

Hermanto, B., & SN, A. (2017). Klasifikasi Nilai Kelayakan Calon Debitur Baru Menggunakan Decision Tree C4.5. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *11*(1), 43. https://doi.org/10.22146/ijccs.15946

Ignatow, G., & Mihalcea, R. (2018). Sentiment Analysis. https://doi.org/10.4135/9781483399782.n14

Informatikalogi. (2017). Algoritma Naïve Bayes | INFORMATIKALOGI. Retrieved from Informatikalogi website: https://informatikalogi.com/algoritma-naive-bayes/

Kabari, L. G., & Onwuka, U. C. (2019). Comparison of Bagging and Voting Ensemble Machine Learning Algorithm as a Classifier. *International Journals of Advanced Research in Computer Science and Software Engineering*, *9*(3), 19–23. Retrieved from www.ijarcsse.com,

Kesehatan, K., & Indonesia, R. (2011). *Kementerian Kesehatan Republik Indonesia Tahun 2011*.

Kemenkes no. 82. (2018). *Peraturan Presiden Republik Indonesia No. 82 Tahun 2018 Tentang Jaminan Kesehatan* (p. 74). p. 74.

Luthfi, K. A. E. T. (2009). *Algoritma Data Mining Yogyakarta*. Yogyakarta, Indonesia: Andi Offset.

Lestari, M. (2014). Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Mendeteksi Penyakit Jantung. *Faktor Exacta*, *7*(September 2010), 366–371.

Maimon, O, Rokach, L. (2011). Data mining and knowledge discovery handbook. In *Choice Reviews Online* (Vol. 48). https://doi.org/10.5860/choice.48-5729

Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *2*(8), 2739–2743.

Nawawi, H. M., Purnama, J. J., Hikmah, A. B., Komputer, S. I., Informasi, S. S., Bina, U., & Informatika, S. (2019). Komparasi Algoritma Neural Network dan Naive Bayes untuk Memprediksi Penyakit Jantung (Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri; Vol. 15). https://doi.org/10.33480/pilar.v15i2.669

Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Teknologi & Informasi ITSmart*, *2*(2), 35. https://doi.org/10.20961/its.v2i2.630

peraturan badan penyelenggara jaminan sosial. (2018). *Peraturan Badan Penyelenggara Jaminan Sosial Kesehatan No 1 Thn 2018*. 1–16.

Princy Christy, A., & Rama, N. (2019). Improving prediction capability of ensemble of classifiers through weighted average probabilities. *International Journal of Engineering and Advanced Technology*, *8*(5), 2540–2543.

Rana, S., & Singh, A. (2017). Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. *Proceedings on 2016 2nd International Conference on Next Generation Computing Technologies, NGCT 2016*, (October), 106–111. https://doi.org/10.1109/NGCT.2016.7877399

Rosdiansyah, D. (2014). Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor dan Pendekatan Lexicone. *Tugas Akhir Jurusan Teknik Informatika*, 1–15.

Rosyidi, R. (2019). *PERBANDINGAN ALGORITMA K-NN DAN CART PADA DATA*. *4*(2), 169–177.

Salini, A., Jeyapriya, U., College, S. M., & College, S. M. (2018). A Majority Vote Based Ensemble Classifier for Predicting Students Academic Performance. *International Journal of Pure and Applied Mathematics*, *118*(24), 1–11.

Sentimen, A., Menilai, U., Terhadap, M., Kementrian, K., Pembangunan, K., Dan, M., … Mandiri, N. (2018). *KOORDINATOR PEMBANGUNAN MANUSIA DAN KEBUDAYAAN MENGGUNAKAN SUPPORT VECTOR MACHINE DAN NAIVE BAYES*.

Studi, P., Santoso, I., Komputer, I., Mining, D., & Bayes, N. (n.d.). *Program Studi Ilmu Komputer (S2) STMIK Nusa Mandiri i*.

Vanich Sajee. (2015). *Data Mining: Data Mining (เหมืองข้อมูล)* (Vol. 2). Retrieved from http://sajeegm301.blogspot.com/2015/11/data-mining.html

Vinodhini, G. R. C. (2012). Sentiment Analysis and Opinion Mining : A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(6), 282–292.

Vulandari Tri Retno. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Retrieved from www.infogavamedia@yahoo.com

Wahono, R. S. (2015). Proses Data Mining. In *Data Mining*. Retrieved from http://romisatriawahono.net/dm

Wahyudi, M., & Kristiyanti, D. A. (2016). Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization. *Journal of Theoretical and Applied Information Technology*, *91*(1), 189–201.

Yanis, R. Y. (2018). Sentiment Analysis of Bpjs Kesehatan Services To Smk Eklesia and Bina Insani Jailolo Teachers. *Jurnal Terapan Teknologi Informasi*, *2*(2), 25–34. https://doi.org/10.21460/jutei.2018.22.105

Zhang, Y., Zhang, H., Cai, J., & Yang, B. (2014). A weighted voting classifier based on differential evolution. *Abstract and Applied Analysis*, *2014*. https://doi.org/10.1155/2014/376950