

COMPARATION OF CLASSIFICATION ALGORITHM ON SENTIMENT ANALYSIS OF ONLINE LEARNING REVIEWS AND DISTANCE EDUCATION

Lila Dini Utami¹; Siti Masripah^{2*})

Sistem Informasi Akuntansi Kampus Kota Bogor
Universitas Bina Sarana Informatika
www.bsi.ac.id
lila.ldu@bsi.ac.id¹, siti.stm@bsi.ac.id^{2*})
(*) Corresponding Author

Abstract— As of January 27, 2021, confirmed cases of COVID-19 nationally stood at 1,024,298 people, this data is data that has been officially announced by the Indonesian Ministry of Health. Meanwhile, in Jakarta, there are 256,416 confirmed cases of COVID-19. In July 2021, there was a very significant increase, seeing the data caused the Central government to make a decision to continue the Large-Scale Social Restrictions (PSBB), followed by the Enforcement of Restrictions on Community Activities (PPKM), which affected all aspects, especially the education aspect. In the education aspect, the government applies distance and online learning. Of course, many people agree or disagree with this decision, because there must be sacrifices, both in terms of time and cost. Seeing these conditions makes the authors interested in discussing and processing public opinions on distance and online learning systems which certainly have positive and negative responses from learning implementers, to process the data the author uses Data Mining, namely using the Text Mining Classification method with several The classification algorithms are the Naïve Bayes Algorithm (NB), the k-Nearest Neighbor (k-NN) Algorithm and the Support Vector Machine (SVM) Algorithm to see which classification algorithm has the highest accuracy and diagnostic value in processing this opinion. After the calculations are done, the algorithm that is more suitable for analyzing reviews or opinions in this study is to use the Support Vector Machine (SVM) classification algorithm with the highest accuracy value of 87.67% and an AUC value of 0.939 with an Excellent Classification diagnostic level.

Keywords: Comparison, Online, PJJ, Classification, Sentiment Analysis

Abstrak— Per 27 Januari 2021 kasus terkonfirmasi COVID-19 secara Nasional berada diangka 1.024.298 jiwa, data tersebut adalah data yang telah diumumkan secara resmi oleh Kementerian Kesehatan RI. Sedangkan di Jakarta, kasus terkonfirmasi COVID-19 sebanyak 256.416 jiwa. Pada bulan juli 2021, terjadi peningkatan yang

sangat signifikan, melihat data tersebut menyebabkan pemerintah Pusat melakukan pengambilan keputusan untuk meneruskan Pembatasan Sosial Berskala Besar (PSBB), dilanjutkan dengan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) sehingga menyebabkan seluruh aspek terpengaruh terutama aspek pendidikan. Pada aspek pendidikan pemerintah memberlakukan pembelajaran Jarak Jauh dan Daring. Keputusan ini tentunya banyak masyarakat yang setuju maupun tidak setuju, dikarenakan pasti ada yang harus dikorbankan baik dari segi waktu maupun biaya. Melihat kondisi tersebut membuat penulis tertarik untuk membahas dan mengolah opini-opini masyarakat terhadap sistem pembelajaran jarak jauh dan daring yang pastinya memiliki tanggapan positif dan negatif dari para pelaksana pembelajaran, untuk mengolah data tersebut penulis menggunakan Data Mining, yaitu menggunakan metode Klasifikasi Text Mining dengan beberapa Algoritma klasifikasi yaitu Algoritma Naïve Bayes (NB), Algoritma k-Nearest Neighbor (k-NN) dan Algoritma Support Vector Machine (SVM) untuk melihat Algoritma klasifikasi mana yang nilai akurasi dan diagnosa paling tinggi dalam pengolahan opini ini. Setelah dilakukan perhitungan maka Algoritma yang lebih cocok untuk menganalisa review atau opini pada penelitian ini adalah menggunakan algoritma klasifikasi Support Vector Machine (SVM) dengan nilai akurasi tertinggi yaitu 87.67 % dan nilai AUC sebesar 0.939 dengan tingkat diagnosa Excellent Classification.

Kata Kunci: Komparasi, Daring, PJJ, Klasifikasi, Analisa Sentimen.

INTRODUCTION

Based on the official website corona.jakarta.go.id which was accessed on January 27, 2021, nationally confirmed cases of COVID-19 stated positive cases reaching 1,024,298 cases, while positive confirmed cases of COVID-19 for Jakarta were 256,416 people. (Pemprov DKI Jakarta,



2020b). And now in July 2021 there is an increase in confirmed cases of COVID-19 nationally reaching 3,372,374 cases (Pemprov DKI Jakarta, 2020a), there was a very high increase from January to July 2021.

Based on official information on the website <https://www.kemkes.go.id/> which was accessed on January 27, 2021, where information can be seen at (Kementerian Kesehatan Republik Indonesia, 2021), namely the COVID-19 situation (cumulative) with confirmed Positive Covid-19 as many as 1,024,298, recovered from Covid-19 positive as many as 831,330, while confirmed death from Covid-19 positive as many as 28,855 (Kementerian Kesehatan Republik Indonesia, 2021). Data on July 30, 2021, cumulative Covid-19 situation: Positive Covid-19 as many as 3,372,374, recovered (Positive of Covid-19) as many as 2,730,720 and died (Positive of Covid-19) as many as 92,311 (Kementerian Kesehatan Republik Indonesia, 2021). There is an increase in the number of confirmed positive Covid-19 from January to July 2021.

Seeing the positive confirmed cases of Covid-19 which continues to increase in number every day, the government must make several decisions for the good of the Indonesian nation, one of the decisions that is highlighted is regarding education in Indonesia. So the Indonesian Ministry of Education issued a Circular on the Implementation of Educational Policies in the Emergency Period for the Spread of Corona Virus Disease (COVID-19) number 4 of 2020, where the learning process is carried out from home using the Online and Remote method. (Pelaksanaan Kebijakan Pendidikan Dalam Masa Darurat Penyebaran Co Ro Naviru S D/Sease (Covid-19), 2020). Seeing the increasing number of Covid-19 cases, the government extended the period of application of the Community Activity Restrictions or PPKM until August 2, 2021 (Nurita, 2021). So that learning both at school and at the college level must be done online.

Online learning methods can be used as an alternative that can be done during a pandemic like this (Sadikin & Hamidah, 2020). However, education can currently be held with two methods, only certain learning that requires non-routine face-to-face meetings, such as final assignment guidance, or final exams, by paying attention to health protocols such as Circular Letter number 6 of 2020 concerning Implementation of Learning in Even Semesters of the Year. Academic 2020/2021 which said that education at higher education could be held in a mixture of two methods, namely face-to-face and online with the obligation to pay attention to health protocols (Ministry of Education and Culture No. 36603/A.A5/OT/2020, 2017).

There are many analytical methods used by researchers to analyze public opinion based on

information provided both from social media and from the results of questionnaires, one of which is the sentiment analysis method. (Laurensz & Eko Sedyono, 2021). Sentiment analysis sourced from text reviews is aimed at predicting, analyzing the public atmosphere, the mood of a problem (Que et al., 2020).

Sentiment analysis related to public opinion has also been widely studied, including opinions on products, opinions on an application, opinions on tourist attractions and so on. The following are some of the studies conducted related to sentiment analysis using several algorithms to analyze public opinion:

The Covid-19 virus has spread throughout the world including Indonesia, so a vaccine is needed to suppress the spread of the Covid-19 virus, responses related to this are quite diverse on social media so that sentiment analysis is carried out on the problem, sentiment analysis uses the Naïve Bayes classification algorithm and SVM, The results of this study conclude that public sentiment towards vaccination has a positive response (Laurensz & Eko Sedyono, 2021).

In other research on sentiment analysis, namely the application of the Naïve Bayes Algorithm for sentiment analysis on website-based TMII tourism, this is a study that classifies opinions obtained from Tripadvisor using the Naïve Bayes algorithm, by the researchers made a web-based application to facilitate the public in finding conclusions from a opinion (Sari & Hayuningtyas, 2019).

Muhammad Syarifuddin's research in the analysis of public opinion sentiment regarding Covid-19 on Twitter to provide views on the latest issues about Covid-19, where in his research using the nave Bayes algorithm and KNN, from the results of the two algorithms, the Nave Bayes accuracy level is obtained with a higher accuracy value. high that is 63.21% (Syarifuddin, 2020).

Research conducted by Fajar Ratnawari, which discusses the implementation of the Naïve Bayes Algorithm for sentiment analysis of film opinions on Twitter, said that the experimental results used the Nave Bayes algorithm which has an accuracy rate of 90%. (Ratnawati, 2018). The Naïve Bayes algorithm can be used to determine the quality of the type of airline, by producing the highest score based on the specified criteria.

Research conducted by Fattya and Andi which discusses the classification algorithm, namely Support Vector Machine, Decision Tree and Naive Bayes in analyzing the level of satisfaction of Telkomsel Prepaid customers, and produces the highest accuracy value on the C45 Algorithm (Ariani & Taufik, 2020)

In the research conducted by Maxsi and Dyah also discussed the classification of the K-Nearest Neighbor and Backward Elimination Algorithms, with the aim of measuring the level of accuracy to determine the next medical action in Mesothelioma disease. (Ary & Rismiati, 2019).

Research conducted by Masripah and friends, compared with two algorithms, namely the K-NN algorithm and Naive Bayes in sentiment analysis of a money transfer application, and resulted in the Naive Bayes algorithm having a higher accuracy value. (Masripah et al., 2020).

Based on this, the authors conducted research on sentiment analysis on online learning and distance learning systems which were formed due to the government's decision to carry out the teaching and learning process online and PJJ, because the Covid-19 virus was still increasing. Sentiment analysis data which contains whether or not they agree or disagree in carrying out online education and distance learning and what are the positive and negative opinions of the learning, by distributing questionnaires so that the opinions of respondents are not affected by the surrounding situation. Data processing uses a classification process with several algorithms. The purpose of this research is to find out the public's review of online learning and PJJ that is enforced due to the Covid-19 condition and determine what classification algorithm is accurate in processing opinion data using three classification algorithms, namely the Naive Bayes (NB) Algorithm, k-Algorithm. Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Algorithm, as well as to find out the high level of accuracy of the three algorithms by evaluating the classification model based on data calculations, with a confusion matrix and using ROC graphs with AUC values, as for the level of diagnostic accuracy is (Gorunescu, 2011):

Accuracy is 0.90-1.00= Excellent classification

Accuracy is 0.80-0.90 = Good classificaion

Accuracy is 0.70-0.80 = Fair classification

Accuracy is 0.60-0.70 = Poor classification

Accuracy is 0.50-0.60 = Failure

MATERIALS AND METHODS

In this study using a survey method, where the authors collect data through the use of structured data collection tools such as questionnaires or interviews (Odoh & Chinedum E, 2014).

The data collection used in this study was to distribute online questionnaires using Google Forms which were distributed to online learning and distance learning (PJJ) actors. The review data used is 300 data, consisting of 150 negative reviews and 150 positive reviews. The data is a collection of text stored in different folders. For 150 positive reviews, it is stored in the Positive Reviews folder and 150 negative reviews, stored in the Negative Reviews folder which will later be processed using Rapid Miner.

The following questions are available on the Questionnaire:

- P1 Setujukah anda pelaksanaan pembelajaran secara daring?
- P2 Menurut anda, apa Sisi Positif dari Pembelajaran Daring yang sudah berjalan saat ini?
- P3 Menurut anda, apa Sisi Negatif dari Pembelajaran Daring yang sudah berjalan saat ini?

In the questionnaire recapitulation, there are several attributes, including: Age, Gender, Class/Semester, Approval questions about online learning and PJJ, and at the end with the correspondent's opinion about the positive and negative sides of online learning and PJJ. Some examples of survey results conducted by distributing questionnaires can be seen in table 1.

Table 1. Questionnaire Recap

Usia	Jenis Kelamin	Kelas/Semester	Setujukah anda pelaksanaan pembelajaran secara daring?	Menurut anda, apa Sisi Positif dari Pembelajaran Daring yang sudah berjalan saat ini?	Menurut anda, apa Sisi Negatif dari Pembelajaran Daring yang sudah berjalan saat ini?
22	Laki-Laki	Smester 4	Setuju	Menjadi lebih aktif	Informasi dan sistem absen tidak jelas
35 tahun	Perempuan	1/2	Setuju	Anak belajar langsung keterampilan hidup	Kejenuhan, kebosanan yang melanda
52	Laki-Laki	-	Setuju	Menghindari Covid-19, belajar bisa tetap berjalan	Materi kurang jelas
22	Laki-Laki	11.6A.13	Ragu-Ragu	Meminimalisir resiko terkena covid-19	Pembelajaran kurang efektif



Usia	Jenis Kelamin	Kelas/Semester	Setujukah anda pelaksanaan pembelajaran secara daring?	Menurut anda, apa Sisi Positif dari Pembelajaran Daring yang sudah berjalan saat ini?	Menurut anda, apa Sisi Negatif dari Pembelajaran Daring yang sudah berjalan saat ini?
46 tahun	Laki-Laki	Starata 2	Setuju	Alternatif selama pandemi Covid-19, Kapanpun-dimanapun dapat terjawab.	Emosional inteartkif terbatas media.
18	Perempuan	Semester 3	Ragu-Ragu	Meminimalisir waktu	Semakin banyak tugas, sulit memahami
20 tahun	Perempuan	Semester 5	Ragu-Ragu	Dapat memunculkan kreativitas mahasiswa	Materi pembelajaran kurang dipahami
19	Laki-Laki	1	Tidak Setuju	Enak tiduran, garibet,bebas bisa dimana aja	Boros uang, bikin otak kurang fokus karena sulit memahami materi dari dosen, kurang sosialisasi dgn lingkungan kampus,tidak ada pengurangan biaya semester selama di daring apalagi yg masuk 2020, dan pengurangan jatah uang bulanan.
36	Perempuan	ibu Rumah tangga	Setuju	orang tua menjadi lebih mengenal pola belajar anak dan menjadi lebih fokus dalam belajar anak	terkadang dalam menjelaskan orang tua mengalami kesulitan, krn penjelasan dari guru terkadang kurang jelas
20	Perempuan	Semester 5	Tidak Setuju	Lebih menghemat energi uang juga waktu.. dan belajar juga lebih santai dan bebas	Karena terlalu santai lama2 belajar di rumah menjadi membosankan dan mungkin slide di setiap matkul pun jarang di baca jikapun di baca itu ketika ada tugas..dan selebihnya untuk bertanya pun terbatas tidak seperti di dalam kelas..

From the recapitulation of Microsoft Excel in table 1, they are separated into each folder. Negative reviews are put in the Negative folder, while positive reviews are put in the Positive folder, then processed using Rapid Miner to find out the results of Tokenize, Stopwords, and Stemming. The author compares the classification algorithm, the algorithm used is the Naïve Bayes (NB) Algorithm, the k-Nearest Neighbor (k-NN) Algorithm, and the Support Vector Machine (SVM) Algorithm using Rapid Miner tools then evaluates using the ROC graph of the AUC value for get the best accuracy value.

In Rapid Miner, related document files are entered into Process Document. After data obtained, the next step is to process the data with the pre-processing stage.

The pre-processing stage is carried out to facilitate the data analysis process stage (Laurenz & Eko Sedyono, 2021). The following pre-processing steps are carried out, including Tokenize, Filter Stopwords and Stemming (See Figure 4).

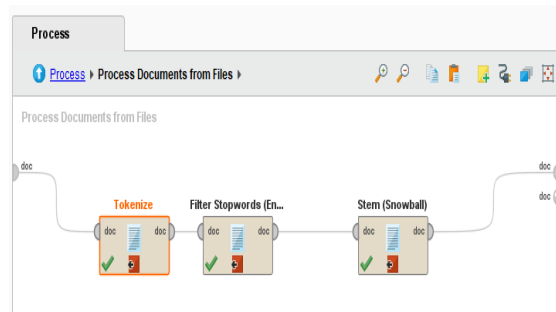


Figure 4. Stages of pre-processing using Rapid Miner

1. Tokenize

Tokenize is a process to separate words. The result of the separation is called a token (Somantri & Khambali, 2017).

Table 2. Example of the 16th Positive Document Tokenize

Proses Tokenize Dokumen Positif Ke - 16			
Dokumen	Tokenize		
"- Melihat kondisi sekarang yang sangat rentan dalam penyebaran virus,sangat cocok belajar secara daring,	Melihat kondisi sekarang yang sangat rentan dalam penyebaran virus sangat cocok belajar secara		



Proses <i>Tokenize</i> Dokumen Positif Ke - 16	
Dokumen	<i>Tokenize</i>
terjadinya kluster baru covid 19 - Kesehatan anak bisa dipantau langsung secara berkala, - keakraban keluarga menjadi semakin erat - materi pelajaran yang beragam, menambah semangat anak untuk mengulang-ulang materi yang berbentuk video"	meminimalisir terjadinya kluster baru covid Kesehatan anak bisa dipantau langsung secara berkala keakraban keluarga menjadi semakin erat materi pelajaran yang beragam menambah semangat anak untuk mengulang ulang materi yang berbentuk video

Table 3. Example of the 121st Negative Document Tokenize

Proses <i>Tokenize</i> Dokumen Negatif Ke - 121	
Dokumen	<i>Tokenize</i>
Memerlukan kuota yang banyak karena kuota pemerintah hanya dipakai untuk gogle classroom saja, susah dalam bertanya, dan sulit memahami beberapa pembelajaran jarak jauh, sulit untuk bekerja kelompok dengan keadaan pandemi yang tidak boleh keluar rumah, jaringan/sinyal yang kadang jelek mempersulit untuk kuliah lewat zoom atau gogle Meet.	Memerlukan kuota yang banyak karena kuota pemerintah hanya dipakai untuk gogle classroom saja, susah dalam bertanya, dan sulit memahami beberapa pembelajaran jarak jauh, sulit untuk bekerja kelompok dengan keadaan pandemi yang tidak boleh keluar rumah, jaringan/sinyal yang kadang jelek mempersulit untuk kuliah lewat zoom atau gogle Meet.

2. *Stopwords*

Stopwords can be interpreted as eliminating common words that do not have the required meaning or information, such as in, with, for and others. (Ernawati & Wati, 2018).

Table 4. Examples of 16 Positive Document Stopwords

Proses <i>Stopwords</i>	
Dokumen	<i>Stopwords</i>
"- Melihat kondisi sekarang yang sangat rentan dalam penyebaran virus, sangat cocok belajar secara daring, meminimalisir	Melihat kondisi sekarang yang sangat rentan dalam penyebaran virus sangat cocok belajar secara meminimalisir

Proses <i>Stopwords</i>	
Dokumen	<i>Stopwords</i>
terjadinya kluster baru covid 19 - Kesehatan anak bisa dipantau langsung secara berkala, - keakraban keluarga menjadi semakin erat - materi pelajaran yang beragam, menambah semangat anak untuk mengulang-ulang materi yang berbentuk video	terjadinya kluster baru covid Kesehatan anak bisa dipantau langsung secara berkala keakraban keluarga menjadi semakin erat materi pelajaran yang beragam menambah semangat anak untuk mengulang ulang materi yang berbentuk video

Table 5. Examples of 121 Negative Document Stopwords

Proses <i>Stopwords</i> Dokumen Negatif Ke - 121	
Dokumen	<i>Tokenize</i>
Memerlukan kuota yang banyak karena kuota pemerintah hanya dipakai untuk gogle classroom saja, susah dalam bertanya, dan sulit memahami beberapa pembelajaran jarak jauh, sulit untuk bekerja kelompok dengan keadaan pandemi yang tidak boleh keluar rumah, jaringan/sinyal yang kadang jelek mempersulit untuk kuliah lewat zoom atau gogle Meet.	Memerlukan kuota yang banyak karena kuota pemerintah hanya dipakai untuk gogle classroom saja susah dalam bertanya dan sulit memahami beberapa pembelajaran jarak jauh sulit untuk bekerja kelompok dengan keadaan pandemi yang tidak boleh keluar rumah jaringan sinyal yang kadang jelek mempersulit untuk kuliah lewat zoom atau gogle Meet

3. *Stemming*

Stemming is the process of changing a word into its root word by removing the affixes in the word in the document or changing the verb into a noun. Example: the word 'eliminated' after the affixes di- and -kan changes to "lost" (Saputra et al., 2015).

Table 6. Example of the 16th Positive Document Stemming

Proses <i>Stemming</i>	
Dokumen	<i>Stemming</i>
"- Melihat kondisi sekarang yang sangat rentan dalam penyebaran virus, sangat cocok belajar secara daring, meminimalisir	melihat kondisi sekarang yang sangat rentan dalam penyebaran virus sangat cocok belajar secara meminimalisir



Proses Stemming	
Dokumen	Stemming
terjadinya kluster baru covid 19 Kesehatan anak bisa dipantau langsung secara berkala, - keakraban keluarga menjadi semakin erat - materi pelajaran yang beragam, menambah semangat anak untuk mengulang-ulang materi yang berbentuk video"	terjadinya kluster baru covid kesehatan anak bisa dipantau langsung secara berkala keakraban menjadi semakin erat materi pelajaran yang beragam menambah semangat anak untuk mengulang ulang materi yang berbentuk video

Table 7. Stemming of the 121st Negative Documents

Proses Stemming Dokumen Negatif Ke - 121	
Dokumen	Tokenize
Memerlukan kuota yang banyak karena kuota pemerintah hanya dipakai untuk gogle classroom saja, susah dalam bertanya, dan sulit memahami beberapa pembelajaran jarak jauh, sulit untuk bekerja kelompok dengan keadaan pandemi yang tidak boleh keluar rumah, jaringan/sinyal yang kadang jelek mempersulit untuk kuliah lewat zoom atau gogle Meet.	Memerlukan kuota yang banyak karena kuota pemerintah hanya dipakai untuk gogle classroom saja susah dalam bertanya dan sulit memahami beberapa pembelajaran jarak jauh sulit untuk bekerja kelompok dengan keadaan pandemi yang tidak boleh keluar rumah jaringan sinyal yang kadang jelek mempersulit untuk kuliah lewat zoom atau gogle Meet

4. Evaluation of Confusion Matrix and ROC Curve
 To evaluate the model classification based on the calculation of which testing objects are predicted to be true and which are not true. This calculation is tabulated into a table called a confusion matrix (Gorunescu, 2011), can be seen in table 8.

Table 8. Confusion Matrix

Classification	Predicted Class	
	Class = YES	Class = NO
Observed Class	Class = Yes	a (true positive- TP) b (false negative- FN)
	Class = No	c (false positive- FP) d (true negative- TN)

Source: (Gorunescu, 2011)

After the data calculation process, an evaluation of the three algorithms is carried out by looking at the ROC curve to get the AUC value.

RESULTS AND DISCUSSION

This study uses the Rapid Miner 5.1 application to carry out the analysis process, with a dataset in the form of online learning reviews and PJJ as many as 150 positive reviews and 150 negative reviews, which were drawn from the results of a questionnaire using google form and recapitulated into excel. The next step is the Pre-processing process, namely Tokenize, Filter Stopwords and Stemming

b. Analysis using Naïve Bayes Algorithm (NB)

Naïve Bayes algorithm is an algorithm that can classify a certain variable using probability and statistical methods (Kurniawan, 2018).

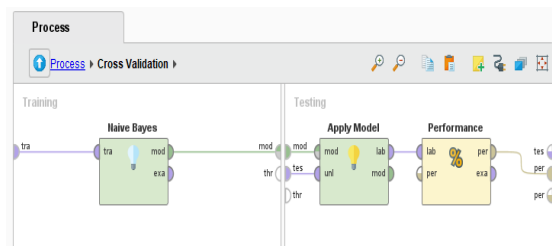


Figure 5. Nave Bayes Algorithm Training and Testing on Rapid Miner

The data processing process uses Nave Bayes on Rapid Miner, the results of the tests carried out can be seen in table 9. Based on the test results using the Nave Bayes algorithm, the accuracy value is 83.33%.

Table 9. Confusion Matrix Model for Naïve Bayes

Accuracy ;	83,33%	+/-	4,44%	(micro average:83,33%)
	True Positif	True Negatif	Class precision	
Pred.	121	21	85,21%	
Pred. Negatif	29	129	81,65%	
Class recall	80,67%	86,00%		

It can be seen in table 9, out of 150 data reviews, 121 data were predicted to be positive, and 29 data were predicted to be positive but turned out to be negative. In contrast, 129 data were predicted to be negative and 21 were predicted to be negative but



turned out to be positive. So that the calculation results get an accuracy value of 83.33%

From as many as 121 data, it is predicted that it is positive, from 150 data that is predicted to be positive. And as many as 129 data are predicted to be appropriate, namely negative from 150 data that are predicted to be negative. So that the calculation results for a positive precision of 80.67% and a negative precision of 86%.

And from 150 review data, 121 data are predicted to be in accordance with the prediction, namely positive. And 129 data are predicted accordingly, which is negative. So that the calculation results for positive recall are 85.21% and negative recall are 81.65%.

To evaluate using the ROC graph. Is a method for illustrating, determining classification according to ability or evaluating accuracy visually (Que et al., 2020). And for the results of the Graph Area Under Curve (AUC) Naïve Bayes Algorithm obtained AUC = 0.756 (See Fig. 6).

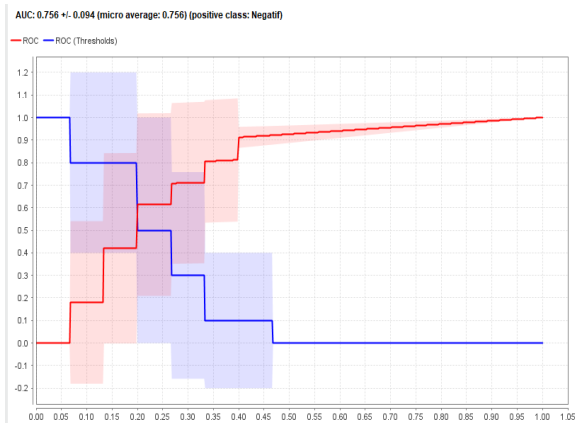


Figure 6. Graph of Area Under Curve (AUC) Naïve Bayes Algorithm

c. Analysis using the k-Nearest Neighbor Algorithm (*k-NN*)

According to Prasetyo in (Nasution & Hayaty, 2019), The k-Nearest Neighbor (*k-NN*) method is one of the top 10 most widely used data mining methods. This method performs classification based on the similarity of a data with other data. The simple principle of this method is “If an animal walks like a duck, quacks like a duck, and looks like a duck, then the animal is probably a duck”.

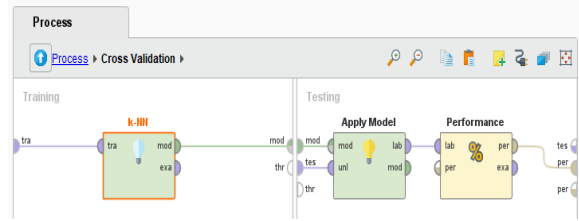


Figure 7. Training and Testing of the k-Nearest Neighbor (*k-NN*) Algorithm on Rapid Miner

In table 10, it can be seen from as many as 150 data on online learning reviews and PJJ, as many as 137 data are predicted to be appropriate, namely positive, and as many as 13 data are predicted to be positive but turn out to be negative. On the other hand, 122 data were predicted accordingly, namely negative and 28 were predicted to be negative but turned out to be positive

Table 10. Confusion Matrix Model for k-Nearest Neighbor(*k-NN*)

Accuracy ; 86,33% +/- 6,93% (micro average:86,33%)	True Positif	True Negatif	Class precision
Pred.Positif	137	28	83,03%
Pred. Negatif	13	122	90,37%
Class recall	91,33%	81,33%	

So that the calculation results get an accuracy value of 86.33%. So that the calculation results for a positive precision of 91.33% and a negative precision of 81.33%. And the calculation result for positive recall is 83.03% and negative recall is 90.37%.

As for the results of the AUC graph obtained by using the k-Nearest Neighbor (*k-NN*) algorithm, it is equal to = 0.911 (See Fig. 8)

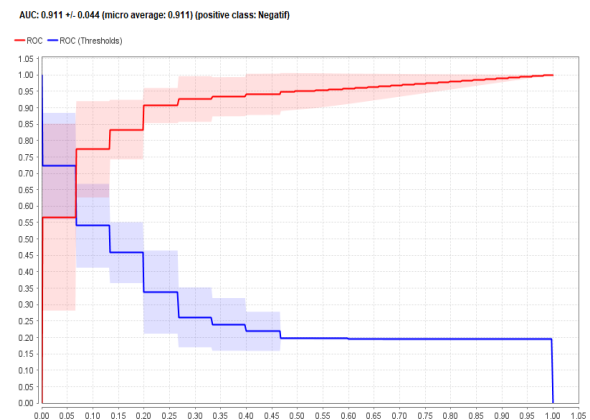


Figure 8. Graph of Area Under Curve (AUC) k-Nearest Neighbor (*k-NN*) Algorithm



d. Analysis using the Support Vector Machine Algorithm (SVM)

SVM is a machine learning algorithm that works on the principle of Structural Risk Minimization (SRM) with the aim of finding the best hyperplane that separates two classes in the input space (Rahutomo et al., 2018).

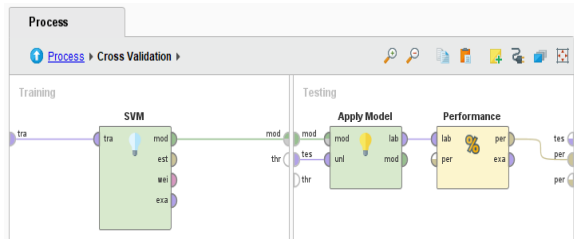


Figure 9. Training and Testing of Support Vector Machine Algorithm on Rapid Miner

In table 11, the Confusion Matrix Algorithm Support Vector Machine (SVM) shows as many as 150 data on online learning reviews and PJJ, as many as 143 data that are predicted to be positive, and as many as 7 data are predicted to be positive but turn out to be negative. On the other hand, 120 data were predicted accordingly, i.e. negative and 30 were predicted to be negative but turned out to be positive.

Table 11. Confusion Matrix Model for Support Vector Machine

Accuracy ; 87,67% +/- 6,49% (micro average:86,33%)			
	True Positif	True Negatif	Class precision
Pred.Positif	143	30	82,66%
Pred. Negatif	7	120	94,49%
Class recall	95,33%	80,00%	

So that the calculation results get an accuracy value of 87.67%. For the calculation results for a positive precision of 95.33% and a negative precision of 80.00%. For the calculation results for a positive recall of 82.66% and a negative recall of 94.49%.

And the results of the AUC graph obtained using the Support Vector Machine (SVM) algorithm with the AUC value = 0.939 (See Fig. 10).

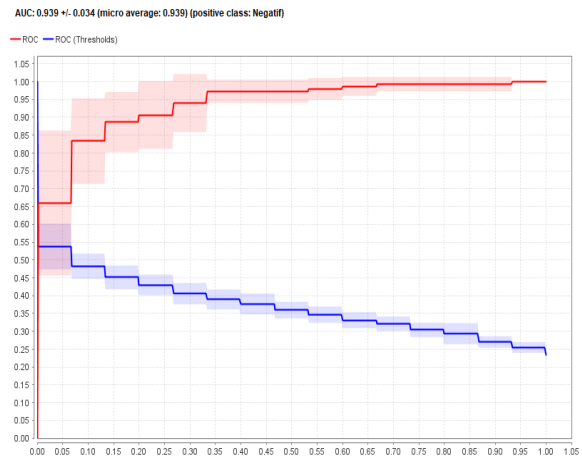


Figure 10. Graph of Area Under Curve (AUC) Support Vector Machine (SVM) Algorithm

CONCLUSION

Based on our research, by comparing three classification algorithms, namely the Naïve Bayes (NB) algorithm, the k-Nearest Neighbor (k-NN) Algorithm and the Support Vector Machine (SVM) algorithm, it can be seen that the results of the Naïve Bayes Algorithm (NB) get the Accuracy value: 83.33% +/- 4.44% (micro average: 83.33%), with an AUC value of 0.756. Meanwhile, the k-Nearest Neighbor (k-NN) algorithm gets an Accuracy value: 86.33% +/-6.93% (micro average: 86.33%) with an AUC value of 0.911. and for the Support Vector Machine (SVM) Algorithm, the Accuracy value: 87.67% +/-6.49% (micro average: 87.67%) with an AUC value of 0.939.

Comparison of each accuracy and AUC is shown in Figure 22 and Figure 23.

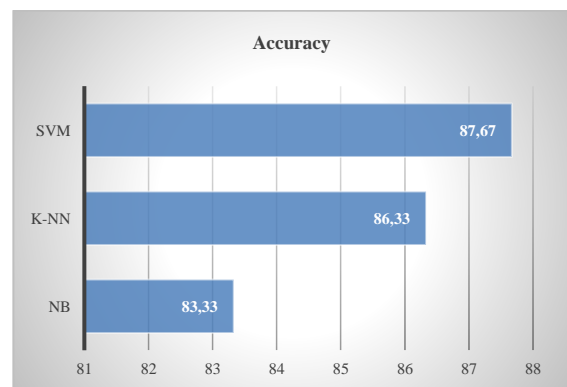


Figure 22. Comparison of the Accuracy of the Naïve Bayes Algorithm, k-NN, and SVM

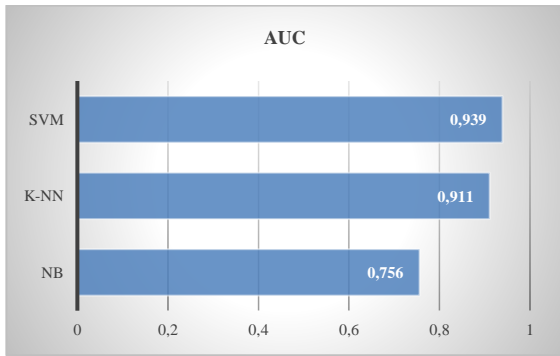


Figure 23. Comparison of AUC of Naïve Bayes Algorithm, k-NN, and SVM

So it can be concluded that online and distance learning opinion data processing uses three classification algorithms, namely the algorithm used, namely Naïve Bayes (NB), k-Nearest Neighbor (k-NN) Algorithm and Support Vector Machine (SVM) algorithm to obtain an algorithm that has an accuracy value. the highest level with the highest evaluation diagnosis level is using the Support Vector Machine (SVM) classification algorithm with an accuracy value of 87.67% and an AUC value of 0.939 with an Excellent Classification diagnostic level.

REFERENCE

- Ariani, F., & Taufik, A. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar. *SATIN-Sains Dan Teknologi Informasi*, 16(2), 46–55. <http://jurnal.stmik-amik-riau.ac.id/index.php/satin/article/view/666>
- Ary, M., & Rismiati, D. A. F. (2019). Ukuran Akurasi Klasifikasi Penyakit Mesothelioma Menggunakan Algoritma K-Nearest Neighbor dan Backward Elimination. *SATIN - Sains Dan Teknologi Informasi*, 5(1), 11–18. <https://doi.org/10.33372/stn.v5i1.444>
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, 6(1), 64–69. <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/>
- Gorunescu, F. (2011). Data Mining: Concepts, Models and Techniques. In *Data mining - Concepts, Models and Technique*. Springer. <https://doi.org/10.1007/978-3-642-19721-5>
- Pelaksanaan Kebijakan Pendidikan Dalam Masa Darurat Penyebaran Co Ro Naviru S D/Sease (Covid-1 9), 300 (2020).
- Kementerian Kesehatan Republik Indonesia. (2021). *Kementerian Kesehatan Republik Indonesia Untuk Indonesia yang Lebih Sehat*. Kementerian Kesehatan Republik Indonesia.
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*. <https://doi.org/10.25126/jtiik.201854803>
- Laurensz, B., & Eko Sedyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Masripah, S., Utami, L. D., Amalia, H., Nurlaela, D., Ryansayah, M., & Yusuf, L. (2020). Comparison of Text Mining Classification Algorithms in Interbank Money Transfer Application. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012088>
- Nasution, M. R. A. N., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226–235.
- Nurita, D. (2021). Ini 4 Pelonggaran Aktivitas PPKM Level 4 yang Diperpanjang Hingga 2 Agustus. *Nasional Tempo*.
- Odoh, D. M., & Chinedum E, D. I. (2014). Research Designs, Survey and Case Study. *IOSR Journal of VLSI and Signal Processing*, 4(6), 16–22. <https://doi.org/10.9790/4200-04611622>
- Pemprov DKI Jakarta. (2020a). *Data Pemantauan Covid-19 Jakarta*. Corona.Jakarta.Go.Id. <https://corona.jakarta.go.id/id/data-pemantauan>
- Pemprov DKI Jakarta. (2020b). *Wujudkan Jakarta Sehat, Aman, dan Produktif Jakarta Tanggap Covid-19*. Jakarta Smart City. <https://corona.jakarta.go.id/id>
- Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. <https://doi.org/10.22146/jnteti.v9i2.102>
- Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A. (2018). Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Polinema*, 4(2), 93–100. <https://doi.org/10.33795/jip.v4i2.152>
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng-Seri Informatika*, 3(1), 50–59. <http://ejournal.polbeng.ac.id/index.php/ISI/article/view/335>



- Sadikin, A., & Hamidah, A. (2020). Pembelajaran Daring di Tengah Wabah Covid-19. *Biodik*, 6(2), 109–119. <https://doi.org/10.22437/bio.v6i2.9759>
- Saputra, N., Adji, T. B., & Permanasari, A. E. (2015). Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM. *Jurnal Dinamika Informatika*, 5(11), 1–12.
- Sari, R., & Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal on Software Engineering (IJSE)*, 5(2), 51–60. <https://doi.org/10.31294/ijse.v5i2.6957>
- Somantri, O., & Khambali, M. (2017). Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naive Bayes dan Algoritme Genetika. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 6(3), 301–306. <https://doi.org/10.22146/jnteti.v6i3.332>
- Syarifuddin, M. (2020). *Laporan Akhir Penelitian Mandiri: Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn*.