SENTIMENT ANALYSIS ON TWITTER SHOPEECARE USING NAIVE BAYES, ADABOOST, AND SVM (EVOLUTION) ALGORITHM COMPARATIVE METHODS

Rizky Nugraha Pratama [1*]; Ghina Amanda Kamila [2]; Kresna Lazani T [3]; Ilham Fauzi [4]; Muhammad Reynaldo Oktaviano [5]; Dedi Dwi Saputra [6]

Program Studi Sistem Informasi[1,2,3,4,5,6]
Universitas Nusa Mandiri[1,2,3,4,5,6]
https://nusamandiri.ac.id/
11213113@nusamandiri.ac.id [1*]; 11212986@nusamandiri.ac.id [2]; 11213091@nusamandiri.ac.id [3];
11212960@nusamandiri.ac.id [4]; 11213088@nuamandiri.ac.id [5]; dedi.eis@nusamandiri.ac.id [6]

Abstract — *The growth of Indonesian e-commerce is increasing along with the growth of internet use in Indonesia. In 2015, there were 92 million internet users in Indonesia. One of the popular online shopping platforms in Indonesia is Shopee. The role of social media also does not escape the role of e-commerce players by utilizing one of them, namely social media twitter. the amount of customer enthusiasm in tweeting and Retweeting existing content, made us decide to research about Sentiment analysis on twitter social media accounts: Shopeecare uses the smote NB, ADboost, and SVM comparison methods. From the data, the comparison results from the test experiments used the Smote + Naive Bayes, Smote + Naive Bayes + Adaboost, and Smote + SVM models. It is known that the Accuracy, Precision, AUC values of the Smote + SVM algorithm are higher than other algorithms, namely Accuracy 76.24%, Precision 75.65%, AUC 0.822. From the results of the algorithm comparison, it shows that the algorithm in determining the sentiment of the complaint and not complaint analysis is better than other algorithms.*

***Keywords** : Shopee, Twitter, Sentiment Analysis, Naïve Bayes, Adaboost, SVM.*

Intisari— Pertumbuhan e-commerce Indonesia meningkat seiring dengan tumbuhnya penggunaan internet di Indonesia. Pada tahun 2015, terdapat 92 juta pengguna internet di Indonesia. Salah satu platform belanja online yang popular di Indonesia adalah Shopee. peran media sosial juga tidak luput dari para pelaku e-commerce dengan memanfaatkan salah satunya yaitu media sosial twitter. banyaknya antusiasme pelanggan dalam men-*tweet* dan Retweet konten yang ada, membuat kami memutuskan untuk meneliti tentang Analisis sentimen pada akun media sosial twitter: *Shopeecare* menggunakan metode komparasi SMOTE NB, ADboost, dan SVM. Dari data hasil perbandingan dari percobaan pengujian menggunakan model Smote + Naive Bayes, Smote + Naive Bayes + Adaboost, dan Smote + SVM. Diketahui bahwa nilai *Accuracy, Precision,* AUC dari algoritma Smote + SVM lebih tinggi dari algoritma lainnya, yaitu *Accuracy* 76.24%, *Precision* 75.65%, AUC 0.822. Dari hasil komparasi algoritma menunjukan bahwa algoritma tersebut dalam menentukan sentimen analisis complain dan not complaint lebih baik daripada algoritma lain.

Kata Kunci : Shopee, Twitter, Sentimen Analisis, Naïve Bayes, Adaboost, SVM.

INTRODUCTION

Referring to data from a study entitled The Opportunity of Indonesia initiated by TEMASEK and Google, Indonesia's e-commerce growth is increasing along with the growth of internet use in Indonesia. In 2015, there were 92 million internet users in Indonesia. In 2020, it is predicted that Indonesian internet users will increase to 215 million users(Harahap 2018). Of the total number of internet users, in 2015, there were 18 million online shoppers in Indonesia. By 2025, 119 million people are predicted to become online shoppers in Indonesia. So it's not surprising that this increase will increase the value of the Indonesian e-commerce market. TEMASEK and Google predict that the value of the Indonesian e-commerce market will reach $81 billion by 2025 (Zaenudin 2017).

One of the most popular online shopping platforms in Indonesia is Shopee. Reporting from Idn Times, according to Ipsos, a global market research company based in Indonesia released the latest research results related to competition in the e-commerce industry during the end of 2021. Based on the survey results, among the three main e-commerce players in Indonesia , namely Tokopedia, Shopee, and Lazada, it is known that Shopee ranks first(Perdana 2022).

Shopee is expanding its market in Indonesia by providing an easy shopping experience. The role of social media also does not escape the role of e-commerce players including shopee in marketing and informing their products and services. According to the daily compass (Syam 2022). Youtube, Whatsapp, Facebook, Instagram, TikTok and Twitter are the six most commonly used social media in Indonesia. Twitter is one of the most used social media because the features offered by twitter make it very easy for its users to spread information to other users. The like and retweet feature is one of twitter's mainstay features that is useful for spreading information to other users. Not only that, the trending topic feature is also the most widely used feature to see hot news that is being discussed, this feature works by summarizing keywords or hashtags that are being widely used by users in a region or throughout the world.

Based on previous research via twitter The results of the study for the Naïve Bayes Classification Algorithm for Sentiment Analysis of the Shopee Application resulted in an accuracy value of 71.50% and an AUC (Area Under Curve) Value of 0.500 (Masripah and Utami 2020). by looking at the amount of customer enthusiasm in tweeting and Retweeting existing content, we decided to research about sentiment analysis on twitter social media accounts: Shopeecare uses NB, Adaboost, and SVM comparison methods. This study is expected to find out the results of the percentage of accuracy between complaints or non-complaints contained in shopeecare user tweets.

MATERIALS AND METHODS

In this section, the system design in this study will be explained. The method was carried out through several stages. The data research that we used about 3 comparison methods, There's namely:
1. SMOTE + NB
2. SMOTE + NB + AdaBoost
3. SMOTE + SVM (Evolution)

The research flow can be explained through the diagram in Figure 1.

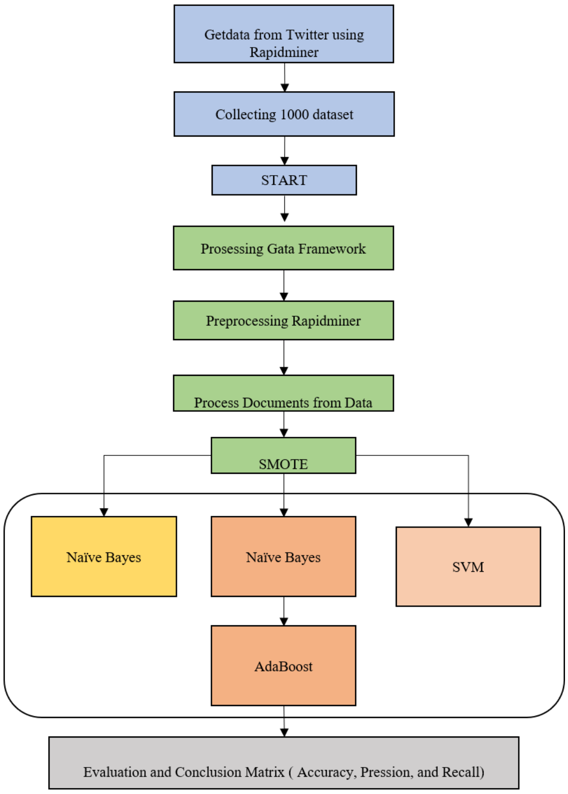

Figure 1. Research flow

As we can see about that flow we should take the data and collecting about 1000 dataset. After that the data should be analyze to complain and not complain. Next we can start the pre processing GataFramework and use the method of Smote combine of Naïve Bayes, AdaBoost, and SVM. Which is that three methods will be explain in this part.

1. Dataset

The data used in this study is data from the Twitter Shoppe Care Indonesia (https://twitter.com/ShopeeCare). The data used in this study are all the tweet data in the shopeecare account which contains reviews, criticisms, Services and comments on products and promos they were campaign in 2022. For the application we use is Rapidminer version 9.10.1. The Dataset Connection flow can be explained through the Figure 2.

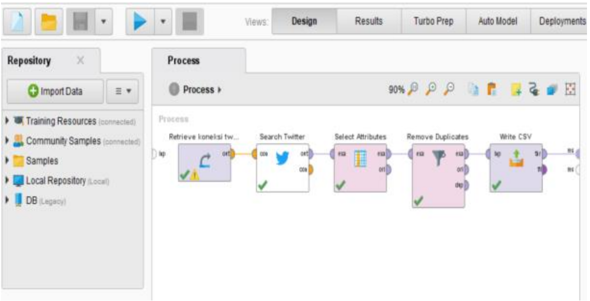

Figure 2. Dataset Connection

We took amounted about 1000 data which is tweets and commented are in Indonesian language and we follow that direction in the rapidminer application to gain all tweets and changes the following data obtained from twitter shopeecare. Firstly we should retrive connection with the link and username twitter that we want to analyzed. The Account twitter was @Shopeecare and after that we select the attribute data were setting in the currently date and filtered if the data contains duplicates tweets. Then we pick a location data where we want to saved. For the checking, we can click the Play button, it's for viewing the result if the following command works correctly. So finaly in the excel worksheet we should Analyze that to labeling sentences if contains complain or not complain. After that Process the Dataset is ready to the next Processing GataFrameworks part.

Table 1. Dataset Twitter Shopeecare

No	Text	Status
651	@ShopeeCare Min, akun dibatasi caranya biar kembali normal gimana?	not complaint
652	@ShopeeCare min, kalo akun dibatasi itu bisa kembali normal nggak? Kalo bisa caranya gimana?	not complaint
653	Selamat pagi @ShopeeID @ShopeePay_ID @ShopeeCare ,kenapa saya tdk bisa melakukan pembayaran via shopeepay ya? Pdhal saldo masih cukup, terima kasih https://t.co/0JOZIepIRZ	complaint
654	@ShopeeCare min cek dm	not complaint
655	@ShopeeCare Baca dm min	not complaint

2. Processing Gataframework

Gataframework is a framework for Indonesian text mining preprocessing that provides Indonesian stopword removal, Indonesian stemming, regular expressions(Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana 2019), transform: URL removal and annotation removal. Within this framework, preprocessing stages can be performed on individual text and uploaded files. To use them, you can use the following link www.gataframework.com. Below is the Shopeecare Twitter data table processed by Gataframework first step is parting the dataset into a 50 data because the server is limited and then we should upload that excel file. The next part we separate the sentences with unneeded words, using tools :

1. @Anotation removal to remove text that has @annotations, such as: @shopee_ID, @shopeecare @shoppe_pay or refer to another account.
2. *transformation:remove URL* to remove links contained in the tweet itself, such as : http://t.co//0JOZIepIRZ, or other inserted link sources.
3. *Regexp to remove punctuation marks, Such as : ?!,.*
4. *indonesian stemming* is the process of turning the word affixed into a base word. Such as: how to Be how to be, How to be a Way, limited to being a limit.
5. *Indonesian stopword removal* is the process of removing unimportant words, such as : this, that is, if, the way.

For more details please see table 2.

Table 2. Shopeecare twitter dataset after *Gataframework* process

No	Regexp	Indonesian Stemming	Indonesian Stop word removal
651	min akun dibatasi caranya biar kembali normal gimana	min akun batas cara biar kembali normal gimana	min akun batas biar normal gimana
652	min kalo akun dibatasi itu bisa kembali normal nggak kalo bisa caranya gimana	min kalo akun batas itu bisa kembali normal nggak kalo bisa cara gimana	min kalo akun batas normal nggak kalo gimana
653	selamat pagi kenapa saya tdk bisa melakukan pembayaran via shopeepay ya pdhal saldo masih cukup terima kasih	selamat pagi kenapa saya tdk bisa laku bayar via shopeepay ya pdhal saldo masih cukup terima kasih	selamat pagi tdk laku bayar via shopeepay pdhal saldo terima kasih
654	min cek dm	min cek dm	min cek dm
655	baca dm min	baca dm min	baca dm min

3. Preprocessing Rapidminer

Rapidminer is software for data processing. Using data mining principles and algorithms, Rapidminer extracts patterns from large data sets by combining statistical methods, artificial intelligence and databases(Afifah Cahayani Adha

Vol. 19, No.1 Maret 2022 | DOI https://doi.org/10.33480/techno.v19i1.3086

2019). Preprocessing is the process of converting raw data into a form that is easier to understand.

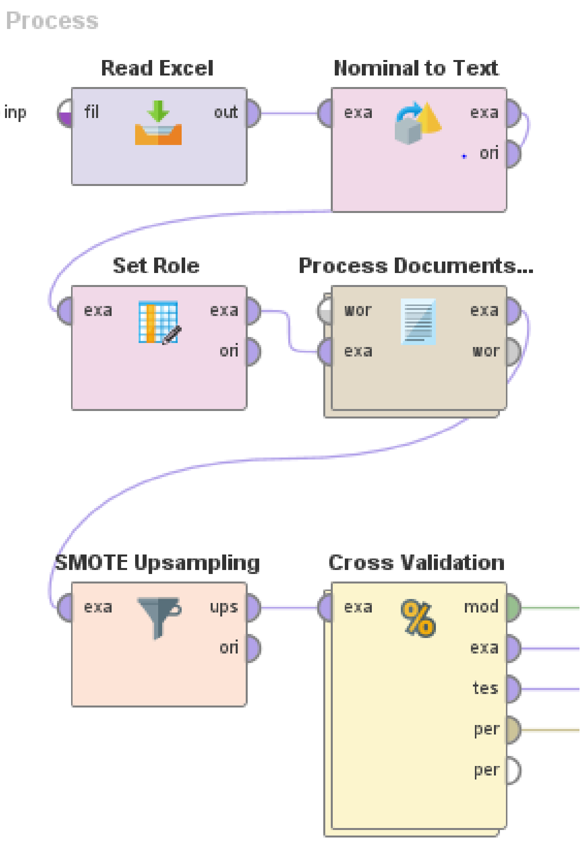

Figure 3. Preprocessing Rapidminer

In figure 3 shows the flow of the preprocessing process on the rapid miner, from this stage describes the steps starting from reading the dataset file, nominal to text is used to change the nominal attribute type to text, the role set is used to change the role of one or more attributes, this process document from data is used to extract information from documents with structured content, SMOTE Upsampling and coss validation.

4. Process Documents from Data

a. Transform Cases

In the transform cases stage, the goal of this stage is to change all letters in the dataset to all lowercase or all capital letters. In this study, the dataset will be converted to lowercase because the majority of the text is written, most of which are lowercase(Jaka 2015).

Table 3. Transform Cases

No	Document	Transform Cases
45	Beli baju di Solo Jateng, kirim ke Grobogan jawa tengah tapi kok sampai Kalimantan tengah di aplikasi nya	beli baju di solo jateng, kirim ke grobogan jawa tengah tapi kok sampai kalimantan tengah di aplikasi nya

b. Tokenize

The tokenize process serves to remove punctuation marks, symbols and non-letter characters in each dataset document(Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana 2019). All unnecessary characters will be discarded including redundant white space and all punctuation marks.

Table 4. Tokenize

No	Document	Tokenize
8	sistemnya emang sering ngawur udah seringkali kejadian begini dan pasti ngeles 'setelah kami cek tidak ada kendala di sistem kami ya...' log kalian ampas	sistemnya emang sering ngawur udah seringkali kejadian begini dan pasti ngeles setelah kami cek tidak ada kendala di sistem kami ya log kalian ampas

c. Filter Tokens (by Length)

Filter tokens is the process of taking important words from the tokens (Langgeni, Baizal, and W 2010)In this process, words that have a certain length will be deleted.

d. Stemming

In the process of grouping words into groups that have the same basic word and performing transformations for the weighting process by calculating the presence or absence of a word in the document(Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana 2019), with the aim that all words that have been selected as tokens in the previous stage will be converted into basic words.

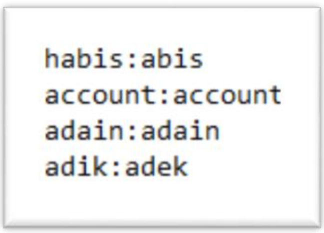
```
habis:abis
account:account
adain:adain
adik:adek
```

Figure 4. Stemming

e. Stopword Removal

Stopword removal is the removal of irrelevant words, such as conjunctions and others(Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana 2019), for example for, will, so, between, which are words that do not have their own meaning if they are removed and words related to adjectives related to sentiment analysis. At this stage will refine the previous dataset.

Table 5. Stopword Removal

No	Document	Stopword Removal
4	selamat pagi kenapa saya tdk bisa laku bayar via shopeepay ya pdhal saldo masih cukup terima kasih	selamat pagi tdk laku bayar via shopeepay pdhal saldo terima kasih

5. SMOTE

The Synthetic Minority Oversampling Technique (SMOTE) method is a popular method for dealing with imbalances. The SMOTE method is a development of the oversampling method, which is a technique that synthesizes new samples from minority classes to balance the dataset by resampling minority class samples(Kasanah, Muladi, and Pujianto 2019).

6. Cross Validation

Cross validation is a method used to get the best model. The explanation of the method used by this research is described as follows:

a. Naive Bayes

Naive Bayes is a method based on the Bayes theorem proposed by British scientist Thomas Bayes, this method belongs to a simple probabilistic classification algorithm that calculates a set of probabilities by adding up the frequencies and combinations of values from a given dataset. This algorithm assumes that object attributes are independent(Iswanto et al. 2021). The probabilities involved in generating the final estimate are calculated as the sum of the frequencies from the decision table. Here is the Bayes theorem equation(Iswanto et al. 2021):

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \text{..................(1)}$$

Description :

X :Data with unknown class

H :Hypothesis data with a certain class

P(H|X) :Probability of hypothesis H based on condition X

P(H) :Hypothesis probability H

P(X) :Probability X

P(X|H) :Probability of X based on the conditions on the hypothesis H

b. AdaBoost

Adaboost or Adaptive Boosting was first introduced by Yoav Freund and Robert Schapire. AdaBoost algorithm itself is an acronym for Adaptive Boosting, this algorithm is widely applied to prediction models in data mining. The whole point of the AdaBoost algorithm is to give more weight to improper observations (weak classification)(Zulhanif 2015). To build a model in the supervised learning algorithm using 2 variables, namely the independent variable and the target variable. Adaboost utilizes boosting to improve predictor accuracy. Adaboost and its variants have been successfully applied to several fields (domains) due to their strong theoretical basis, accurate predictions, and great simplicity. The steps in the Adaboost algorithm are as follows:

1. Input : A collection of research samples on the label $\{(x_i, y_i), \ldots, (x_n, y_n)\}$ a component learn algorithm, the number of cycles T.
2. Initialize : Weight of a training sample $W_i^1 = 1/N$, for all $i = 1, \ldots, N$
3. Do for : $t = 1, \ldots, T$
4. Use component learn algorithms to train a classification component h_t on the training weight sample.
5. Count his training errors on h_t: $\varepsilon_t = \sum_{i=1}^{n} W_i^t, y_i \neq h_t(x_i)$
6. Assign weights to classifier components $h_t = a_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$

7. Training sample weight update $w_i^{t+1} \frac{w_i^t \exp\{-a_t y_i h_t(x_i)\}}{C_t}, i = 1, \dots, N C_t$ is a normalization constant.

8. Output:

$$f(x) = sign\left(\sum_{t=1}^{T} a_t\, h_t(x)\right)$$

c. Support Vector Machine(Evolutionary)

SVM have been around since the 1960s, and this approach makes SVM a promising new method for classifying linear and nonlinear data. An SVM is an algorithm that transforms the original training data into higher dimensions in a new dimension using a non-linear mapping, and then searches for a linearly optimal hyperplane separator (i.e., the "" that separates tuples from one class to another decision boundary"). For sufficiently high dimensions, using a suitable non-linear mapping, the data from the two classes can always be separated by a hyperplane. SVM finds this hyperplane using support vectors ("basic" training tuples) and edges (defined by support vectors) (Muslim et al. 2018). The initial steps of an SVM algorithm are as follows(Ritonga and Purwaningsih 2018):

a) Training data is used for the training process. The training process uses SVM Multi-Class. Each classification model is trained by using the entire data, to find solutions to problems.

b) After training and producing a classification model, the next step is to test the classification model with data testing to determine the accuracy of the classification model.

c) After the model is generated from the classification, testing is carried out on data other than the dataset, testing using the resulting data.

d) Using the results of the implementation above, the researchers then analyzed and discussed the classification results generated by the Support Vector Machine (SVM) method.

7. Evaluasi

This study compares the accuracy level between the Smote + Naive Bayes, Smote + Naive Bayes + AdaBoost and Smote + SVM algorithm approaches which are evaluated using accuracy, precision, recall and AUC. The explanation of the evaluation is described as follows(Maragoudakis, Fakotakis, and Kokkinakis 2004)

a) Accuracy

Accuracy is an evaluation based on the number of correct prediction proportions. Accuracy can be measured by the following equation:

$$a = \frac{tp}{tp + fp} \text{..........................} (2)$$

b) Precision

Precision is the level of accuracy between the information requested by the user and the answer given by the system. The precision formula is as follows:

$$p = \frac{tp}{tp + fp} \text{..........................} (3)$$

c) Recall

Recall is the success rate of the system in retrieving information. The recall formula is as follows:

$$r = \frac{tp}{tp + fn} \text{..........................} (4)$$

d) AUC

AUC is used to measure discriminatory performance by estimating the probability of the output that has been obtained from a randomly selected sample from a positive or negative population, the greater the AUC value, the stronger the resulting classification. Since AUC is part of the unit area of a square, the resulting value will always be the same as it generates, between 0.0 and 1.0(Wisdayani, Indah Manfaati Nur 2019)

Berikut panduan untuk mengklasifikasikan keakuratan diagnosa menggunakan AUC(Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana 2019):

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 clasification.

RESULTS AND DISCUSSION

1. Smote and Naive Bayes

The test results of the Smote and Naive Bayes algorithms carried out in this study were to measure the performance of Accuracy, precision, recall and AUC from the results of training and testing datasets that had gone through the data pre-processing process. The following are the results of testing the Smote and Naive Bayes algorithms. As in the table below.

Table 6. Accuracy Smote and Naive Bayes

	true Complaint	true Not_Complaint	class precision
pred. Complaint	261	41	86.42%
pred. Not_Complaint	242	462	65.62%
class recall	51.89%	91.85%	

Table 6 shows that the Smote and Naive Bayes algorithms produce an Accuracy value of 71.87%.

Table 7. Precision Smote and Naive Bayes

	true Complaint	true Not_Complaint	class precision
pred. Complaint	261	41	86.42%
pred. Not_Complaint	242	462	65.62%
class recall	51.89%	91.85%	

Table 7 shows that the Smote and Naive Bayes algorithms produce a Precision value of 65.73%.

Table 8. Recall Smote and Naive Bayes

	true Complaint	true Not_Complaint	class precision
pred. Complaint	261	41	86.42%
pred. Not_Complaint	242	462	65.62%
class recall	51.89%	91.85%	

Table 8 shows that the Smote and Naive Bayes algorithms produce a recall value of 91.84%.

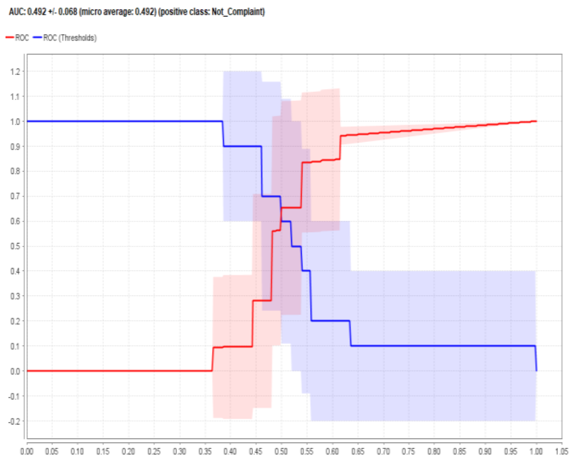

Figure 5. AUC Smote and Naive Bayes

Figure 5 shows that the calculation results on the ROC curve, depicting the ROC curve for the Smote and Naive Bayes algorithms. The ROC curves of the Smote and Naive Bayes algorithms have an AUC of 0. 492. The curve illustrates that the prediction of not complaint 242 is considered a review complaint, the prediction of complaint 261 is included in the review complaint, while the prediction of not complaint 462 is included in the review not complaint and the prediction of complaint 41 is included in the review not complaint . These results show that the Smote Naive Bayes algorithm method is included in the category of failure clasification.

2. Smote, Naive Bayes and Adaboost

The test results of the Smote, Naive Bayes and AdaBoost algorithms carried out in this study are to measure the performance of Accuracy, precision, recall and AUC from the results of training and testing datasets that have gone through the data pre-processing process. The following are the results of testing the Smote and Naive Bayes algorithms. As in the table below.

Table 9. Accuracy Smote, Naive Bayes and AdaBoost

	true Complaint	true Not_Complaint	class precision
pred. Complaint	261	41	86.42%
pred. Not_Complaint	242	462	65.62%
class recall	51.89%	91.85%	

Table 9 shows that the Smote, Naive Bayes and AdaBoost algorithms produce an Accuracy value of 71.87%

Table 10. Precision Smote, Naive Bayes and AdaBoost

	true Complaint	true Not_Complaint	class precision
pred. Complaint	261	41	86.42%
pred. Not_Complaint	242	462	65.62%
class recall	51.89%	91.85%	

Table 10 shows that the Smote, Naive Bayes and AdaBoost algorithms produce a Precision value of 65.73%.

Table 11. Recall Smote, Naive Bayes and AdaBoost

	true Complaint	true Not_Complaint	class precision
pred. Complaint	261	41	86.42%
pred. Not_Complaint	242	462	65.62%
class recall	51.89%	91.85%	

Table 11 shows that the Smote, Naive Bayes and AdaBoost algorithms produce a recall value of 91.84%.

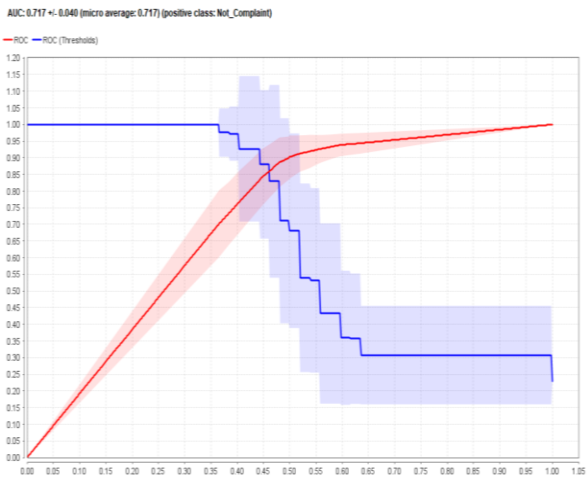

Figure 6. AUC Smote, Naive Bayes and AdaBoost

Figure 6 shows that the calculation results on the ROC curve, depicting the ROC curve for the Smote, Naive Bayes and AdaBoost algorithms. The ROC curves of the Smote, Naive Bayes and AdaBoost algorithms have an AUC of 0.717. The curve illustrates that the prediction of not complaint 242 is considered a review complaint, the prediction of complaint 261 is included in the review complaint, while the prediction of not complaint 462 is included in the review not complaint and the prediction of complaint 41 is included in the review not complaint . These results show that the Smote, Naive Bayes and AdaBoost algorithm method is included in the category of fair clasification.

3. Smote and Support Vector Machine (Evolutionary)

The test results of the Smote and SVM algorithms carried out in this study are to measure the performance of Accuracy, precision, recall and AUC from the results of training and testing datasets that have gone through the data pre-processing process. The following are the results of testing the Smote and Naive Bayes algorithms. As in the table below.

Table 12. Accuracy Smote and SVM

	true Complaint	true Not_Complaint	class precision
pred. Complaint	376	112	77.05%
pred. Not_Complaint	127	391	75.48%
class recall	74.75%	77.73%	

Table 12 shows that the Smote and SVM algorithms produce an Accuracy value of 76.24%.

Table 13. Precision Smote and SVM

	true Complaint	true Not_Complaint	class precision
pred. Complaint	376	112	77.05%
pred. Not_Complaint	127	391	75.48%
class recall	74.75%	77.73%	

Table 13 shows that the Smote and SVM algorithms produce a Precision value of 75.65%.

Table 14. Recall Smote and SVM

	true Complaint	true Not_Complaint	class precision
pred. Complaint	376	112	77.05%
pred. Not_Complaint	127	391	75.48%
class recall	74.75%	77.73%	

Table 14 shows that the Smote and SVM algorithms produce a recall value of 77.72%.

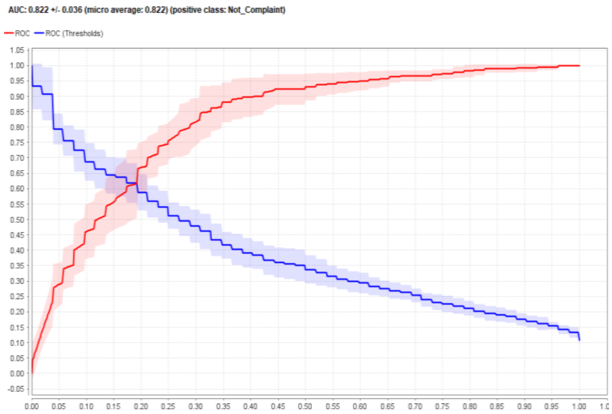

Figure 7. AUC Smote and SVM

Figure 7 shows that the calculation results on the ROC curve, depicting the ROC curve for the Smote and SVM algorithms. The ROC curves of the Smote and SVM algorithms have an AUC of 0.822. The curve illustrates that the prediction of not complaint 127 is considered a review complaint, the prediction of complaint 376 is included in the review complaint, while the prediction of not complaint 391 is included in the review not complaint and the prediction of complaint 112 is included in the review not complaint. These results show that the Smote and SVM algorithm method is included in the category of good clasification.

4. Results

The following is a comparison of the experimental results using the Smote + Naive Bayes, Smote + Naive Bayes + Adaboost, and Smote + SVM models.

Table 15. Comparison of Algorithms

Algoritma	Accuracy	Precision	Recall	AUC
Smote + Naive Bayes	72.87%	65.73 %	91.84%	0.492
Smote + Naive Bayes + AdaBoost	71.87%	65.73 %	91.84%	0.717
Smote+SVM (Evolution)	76.24%	75.65 %	77.72%	0.822

Based on Table 15 above, it can be seen that the Accuracy, Precision, AUC values of the Smote + SVM algorithm are higher than other algorithms, namely Accuracy 76.24%, Precision 75.65%, AUC 0.822. The results of the comparison of the algorithms show that the algorithm in determining the sentiment analysis of complaints and not complaints is better than other algorithms. The comparison of the Smote + Naive Bayes, Smote + Naive Bayes + Adaboost, and Smote + SVM model algorithms is depicted graphically in Figure 8.

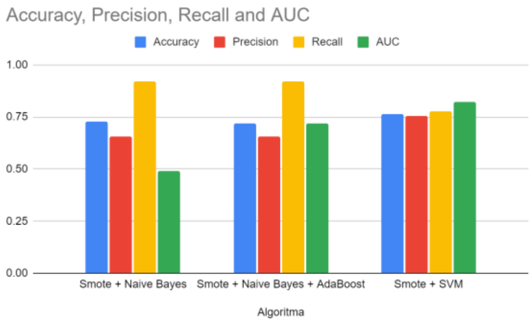

Figure 8. Comparison of Algorithms

CONCLUSION

From the comparison data from testing experiments using the Smote + Naive Bayes, Smote + Naive Bayes + Adaboost, and Smote + SVM models. It is known that the Accuracy, Precision, AUC values of the Smote + SVM algorithm are higher than other algorithms, namely Accuracy 76.24%, Precision 75.65%, AUC 0.822. The results of the comparison of the algorithms show that the algorithm in determining the sentiment analysis of complaints and not complaints is better than other algorithms.

REFERENCE

Afifah Cahayani Adha. 2019. "APLIKASI PENCARIAN HUBUNGAN ANTAR POKOK BAHASAN PADA AYAT AL-QURAN MENGGUNAKAN ALGORITMA APRIORI." Universitas Islam Negeri Sultan Syarif Kasim Riau.

Harahap, Dedy Ansari. 2018. "Perilaku Belanja Online Di Indonesia: Studi Kasus." *JRMSI - Jurnal Riset Manajemen Sains Indonesia* 9(2):193–213. doi: 10.21009/jrmsi.009.2.02.

Iswanto, Hery, Erni Seniwati, Yuli Astuti, and Dina Maulina. 2021. "Comparison of Algorithms on Machine Learning For Spam Email Classification." *IJISTECH (International Journal of Information System and Technology)* 5(4):446. doi: 10.30645/ijistech.v5i4.164.

Jaka, Aris Tri. 2015. "Preprocessing Text Untuk Meminimalisir Kata Yang Tidak Berarti Dalam Proses Text Mining." *Informatika UPGRIS* 1:1–9.

Kasanah, Anis Nikmatul, Muladi Muladi, and Utomo Pujianto. 2019. "Penerapan Teknik SMOTE Untuk Mengatasi Imbalance Class Dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN." *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 3(2):196–201. doi: 10.29207/resti.v3i2.945.

Langgeni, Diah Pudi, Z. K. Abdurahman Baizal, and

Yanuar Firdaus A. W. 2010. "Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection." *Seminar Nasional Informatika 2010* 2010(semnasIF):1–10.

Maragoudakis, Manolis, Nikos Fakotakis, and George Kokkinakis. 2004. "A Bayesian Model for Shallow Syntactic Parsing of Natural Language Texts." *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004* 847–50.

Masripah, Siti, and Lila Dini Utami. 2020. "Algoritma Klasifikasi Naïve Bayes Untuk Analisa Sentimen Aplikasi Shopee." *Swabumi* 8(2):114–17. doi: 10.31294/swabumi.v8i2.8444.

Muslim, M. A., S. H. Rukmana, E. Sugiharti, B. Prasetiyo, and S. Alimah. 2018. "Optimization of C4.5 Algorithm-Based Particle Swarm Optimization for Breast Cancer Diagnosis." *Journal of Physics: Conference Series* 983(1):22–27. doi: 10.1088/1742-6596/983/1/012063.

Perdana, Hana Adi. 2022. "Survei: Shopee Rajai Pasar E-Commerce Indonesia Di 2021." *Idn Times* 1–8.

Riska Aryanti, Atang Saepudin, Eka Fitriani, Rifky Permana, Dede Firmansyah Saefudin. 2019. "Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway." *Jurnal Teknik Komputer* V No. 2Agu. doi: 10.31294/jtk.v4i2.

Ritonga, Alven Safik, and Endah Supeni Purwaningsih. 2018. "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding)." *Ilmiah Edutic* 5(1):17–25.

Syam, Ali Musri. 2022. "Pemanfaatan Twitter Sebagai Sarana Informasi _ KKN Tematik UPI - Kompasiana."

Wisdayani, Indah Manfaati Nur, Rochdi Wasono. 2019. "Perbandingan Algoritma K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Di Kabupaten Pati Jawa Tengah." *Universitas Muhammadiyah Semarang*.

Zaenudin, Ahmad. 2017. "Profil Konsumen Belanja Online Di Indonesia." *Tirtoid*.

Zulhanif. 2015. "Algoritma AdaBoost Dalam Pengklasifikasian." Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS.