

Text Mining for Customer Sentiment Using Naive Bayes and SMOTE Methods on TokopediaCare Twitter

Rico Budiyanto¹, Indah Purnamasari², Dedi Dwi³ Saputra³ ^{1,2,3} Faculty of Information Technology, Universitas Nusa Mandiri

¹11211942@nusamandiri.ac.id, ²indah.ihi@nusamandiri.ac.id, ³dedi.eis@nusamandiri.ac.id

Abstract

At this time, buying and selling online has become part of the lives of the Indonesian people and the world, especially during the pandemic, marketplace users are increasing and slowly replacing traditional markets. Tokopedia as one of the largest marketplaces in Indonesia has the largest users in the 3rd quarter of 2019. Customer complaints to Tokopedia services can be submitted through Social Media such as Twitter and also other media. Complaints submitted via Twitter to the TokopediaCare are still manually identified by Tokopedia customer services one it takes a long time to respond to customer complaints, because customer services need so many time to classified where is complaint or not complaint tweet. Text mining is used to process customer complaint data through text or sentences submitted by tweets using the Naïve Bayes method and the Syntethic Minority Oversampling Technique Method (SMOTE) feature for the implementation of machine learning can help identify the classification of complaints submitted via Twitter automatically. The use of the Naïve Bayes method is added with the Syntethic Minority Oversampling Method feature which is considered better for generating predictions on tweets submitted by customers.

Keywords: Tokopedia, Text Mining, Naive Bayes, SMOTE, Sentiment Analysis

1. Introduction

To provide customers satisfaction, Tokopedia provides access to customer service through social media, one of which is Twitter. Access to customer service through Twitter makes it easy for customers to submit complaints without having to call customer service, besides that customers who are active on Twitter social media do not need to open an application to make a complaint. They done this to achieve customer satisfaction, in order to create customer loyalty to Tokopedia services. Customer satisfaction is something that is expected by the company when the goods or services have been marketed [1]. Sentiment Analysis or opinion mining according to Liu is based on the broad field of Natural Language processing, text mining and linguistic computation which aims to analyze opinions, sentiments, attitudes, emotions, evaluate someone's judgment, namely the author or speaker regarding a discussion, product, individual, organizational services or certain activities that describe a big problem. There are several names used for Sentiment Analysis, sentiment mining, affect analysis, review mining, emotional analysis, and so on [2].

Data mining is a process to get interesting patterns and knowledge from some data. Data origins can include databases, data warehouses, the Web, repositories and other information, or data that has been dynamically streamed into the system [3]. Currently, complaints submitted via Twitter to Tokopedia services via Twitter TokopediaCare are still being analyzed manually, because there is no tool that can predict a sentence in a tweet, so we need a tool that can help in dividing the classification between complaints and non-complaints. this is due to the limited technology and literature that discusses text



International Journal of Information System & Technology Akreditasi No. 158/E/KPT/2021, Vol. 6, No. 1, (2022), pp. 134-144

mining in detecting patterns of humanities characteristics, especially sentiment analysis in Bahasa. Naive Bayes as one of the methods in machine learning that applies probability [4] and compares by adding the SMOTE (Synthetic Minority Oversampling Technique Methode) feature to resolve imbalanced data [5] to find out how effective these methods and techniques are in predicting tweets from users, In this case we takes a case study on the TokopediaCare Twitter account.

2. Research Methodology

Following are the stages of research on Text Mining for Sentiment Analysis of Customers Using the Naive Bayes and SMOTE:

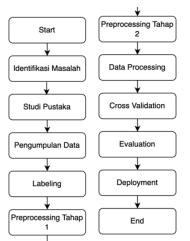


Figure 1. Stage Of Research

2.1. Problem Identification

At this stage, problem identification is carried out based on direct observations or observations on the TokopediaCare Twitter account to find out the problems that exist in the Tokopedia services.

2.2. Literature Review

After identifying the problem, a literature review is also carried out on pre-existing research related to machine learning, data mining, marketing and customer service and others related to the research to be carried out.

3. Results and Discussion

3.1. Collecting Data

That is collecting data through the RapidMiner application using a connection to a Twitter account and operator Retrieve Twitter by connecting the Twitter API through the developer's Twitter account. Twitter data can be accessed through the Twitter REST (Representational State Transfer) API which has been provided by Twitter by first submitting a request to Twitter to obtain data access from Twitter by registering as a developer account.

a) Population and Data Sample

Data population in this study is tweet data from customers Tokopedia marketplace users that mention to the TokopediaCare twitter account from November 2021 to April 2022 with a total sample of 2584 sentences (tweet)

b) Data Type

Data Type used in this study is secondary data (public data) originating from the interaction of Tokopedia users with the TokopediaCare account on Twitter social media.

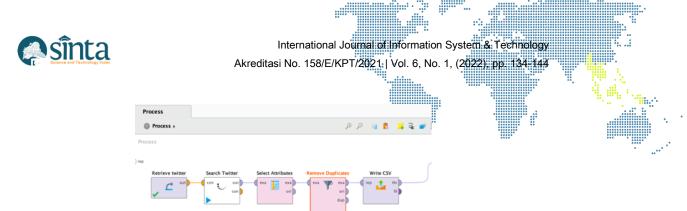


Figure 2. Collecting Data Process

3.2. Labeling

At this stage, the labeling of each sentence in the tweet that has been collected has been carried out so that there are results of classification of complaints and non-complaint categories. Labeling of 2584 datasets by Cityzen or Tokopedia Users and the data result with the classification of "complaint" is 601 data, while for the classification of "not complaint" as many as 1983 data.

	utoSave 📭 🏠 🛱 🛱 🎾 v 🔿 … 🔹 💁 Tugas Tokped 🗸						
Inse	rt Draw PageLayout Formulas Data Review View 🖓 Tell me						
X	Arial (Body) v 10 v A^ A' = = = 🗞 v 🔅 Wrap Text v General v 📰 v						
0 ×	B 7 II v H v A v E E E E E E E E E Karas & Caster v F v 94 9 62 92 Conditional Format	Cell Insert Delete Fo					
*	× √ fx	yles					
₩ A	A ∀ JA 8	c					
<u> </u>							
	wab dulu salah saya dimana? saya minggunakan softwarn, softwarmya aga? saya menggunakan 2 akun, akun yang mana? balah kalo kerja itu pake otak, biar gak tolol bunhet. https://t.co/RG2P2:031V	Complaint					
e	TokopediaCare @rikiarya757 @TokopediaCare coba dong kalo krrja itu pake otak, orang nanya apa dijawab apa						
	wab dulu salah saya dimana? saya menggunakan software, softwarenya apa? saya menggunakan 2 akun, akun yang mana? Abdah lada buda buda buda buda ab buda buda bud	Complaint					
	ibalah kalo kerjaitu pake otak, biar gak tolol banhet. https://t.co/BG2P25033V TokopediaCare @Ammalindra @TokopediaCare coba dong kalo kerjaitu pake otak, orang nanya apa dijawab apa	Complaint					
ľ	nonderner ein Businisten ein onderner ein der onder eine kulter in hand der der der der eine alle ein bestellte						
	wab dulu salah saya dimana? saya menggunakan software, softwarenya apa? saya menggunakan 2 akun, akun yang mana? Ibalah kalo kerjaitu pake otak, biar gak tolol banhet. https://t.co/8G2P2303/V	Complaint					
0	TokopediaCare @Josietlim @TokopediaCare coba dong kalo krrja itu pake otak, orang nanya apa dijawab apa						
	wab dulu salah saya dimana? saya menggunakan softwaren ya ina ana ana ana ana ana ana ana ana an	Complaint					
	TokopediaCare @henriechen @TokopediaCare.coba dong kalo krrja itu pake otak, orang nanya apa dijawab apa	congrant					
	jawab dulu salah saya dimana? saya menggunakan softwaren softwaren ya apa? saya menggunakan 2 akun, akun yang mana? 55 cobalah kalo kerja itu pake otak, biar gak tolol banhet. https://t.co/BG2P2:033V						
	TokopediaCare @jokowahdoyo @TokopediaCare coba dong kalo krrja itu pake otak, orang nanya apa dijawab apa						
	wab dulu salah saya di mana? saya menggunakan software, softwarenya apa? saya menggunakan 2 akun, akun yang mana? ibalah kalo kerja itu pake otak, biar gak tolol banhet. https://t.co/BG2P2i03/V	Complaint					
	TokopediaCare @HardyMynhart @TokopediaCare coba dong kalo krrja itu pake otak, orang nariya apa dijawab apa						
	wab dulu salah saya di mana? saya menggunakan softwaren, softwarenya apa? saya menggunakan 2 akun, akun yang mana? ibalah kalo kerja itu pake otak, biar gak tolol banhet. https://t.co/BG2P2i03iV	Complaint					
	TokopediaCare coba dong kalo krrja itu pake otak, orang nanya apa dijawab apa						
	wab dulu salah saya dimana? saya menggunakan software, softwarenya apa? saya menggunakan 2 akun, akun yang mana?						
	wao uso sain saya umanar saya menggunaan suriwan, suriwan, suriwan, ang saya menggunaan ziawan, akan yang manar Ibalah kalo kerja leng iatu pake taki, biang ak tolo banket, https://t.os/80279031V	Complaint					
	sanan sy belum di respon penjual. Padahal dichat penjual menyanggupi bisa dikirim pada tgi pembelian. Kio belum dikirim berarti penjual bohong Aumirandah ato bisa dikategorikan pembehongan						
	blik. Harusnya batal otomatisi tu 1x24 jam @TokopediaCare https://t.co/VUSXMdIADM TokopediaCare Tokopedia. Ini kurirnya gimana sih ? Saya belum terima pesanan saya tapi dikatakan barang sudah diterima ? https://t.co/OUtwNv7Wyi	Complaint Complaint					
	rootopota, are rootopeta, ini kuinnya gimana sin ranya otuun terima pesana saya tupi otxatakan baarag tubana rutun //r.co/uoukwv/vyo Brokopedia Brokopedia Zima dari kenaria sampai hari ini tubi panu pembagaran premi, aren mulu kenaga ya? https://t.co/MXKHp?niy	Complaint					
	TokopediaCare mint transaksi digital udah 1 jam kok masih di proses terus ya	Complaint					
	TolopediaCare Sudah saya kirimkan kode invoceny di DM, mohon bantuannya min. Ini penggunanya beda orangloh. Masa dibatalkan begitu saja oleh sistem kma terindikasi mlakukan kecurangan. io para admin @tolopediacare @8NI mohon bantuannya, saya togua Tapcash via @tokopedia namun sakio tapcashnya belum bertambah. https://t.co/wWASDOWyS	Complaint Complaint					
20/4 h	and a state up a state of the	A A A A A A A A A A A A A A A A A A A					
Sh	+ +						
y 60	of 2584 records found						

Figure 3. Labeling Complaint Sentences

	AutoSave 🕘 🚥 🏠 🛱 🦻 🏷 v 🗇 … 🔹 💁 Tugas Tokped 🗸			
ə İn	ert Draw Page Layout Formulas Data Review View 🖓 Tell me			
X	Arial (Body) v 10 v A^ A = = = 🕸 v 🔅 wrap Text v General v 📰 v 🔤 v			
Ū.		/ · ·		
3	B I U v ⊞ v A v ≣ Ξ Ξ Ξ Ξ Ξ Ξ Μerge & Center v 🔤 v % 9 1% -% Conditional Format C Formatting as Table Sty			
÷	$\times \checkmark f_x$			
Α	В	с		
	Hatiku kau tembusôf**			
2542	Disaat yang lain sudah wafat di Y	iot Complaint		
2543	@TokopediaCare Disaat ada yang promo hargaðŸr-ðYlsEðYlsE canda mimin tokpedðŸr3ðYls	iot Complaint		
		iot Complaint		
2545	@TokopediaCare Skakmat 1	iot Complaint		
2546	@TokopediaCare cek dm min urgent	iot Complaint		
	Mimin masih tunggu jawaban kamu sampai 14 Maret 2022.			
2547	Yang belum ikutan, gass langsung kirim jawaban, tag dan ajak temen kamu follow @TokopediaCare 89's	iot Complaint		
		Not Complaint		
		Not Complaint		
		Not Complaint		
		iot Complaint		
2571	@TokopediaCare Yah hanya yang merasakan yang tau Yang punya pasar emang perdali mw seller jelas banget gak bener https://t.co/QOvzfoteCw P	iot Complaint		
		iot Complaint		
		iot Complaint		
		Not Complaint		
2577	@TokopediaCare yang mau beli pelatihan prakerja lewat tokped bisa dm aku yaal bakal dapat double bonus loh I kapan lagi ya kan?? 1	Not Complaint		
2578	@TokopediaCare Cek DM ya Kak 7	iot Complaint		
	Max narya dong, kalo misalkan penjual kirim balik barang bermasalah ke pembeli dengan metode DFOD (ongkir dibayar pembeli) itu nanti ongkirnya diganti asuransi kah atau kita sebagai pembeli tetep Jayar engkir tanga diganti oleh asuransi ???			
		Not Complaint		
		iot Complaint		
2581	@TokopediaCare Mohon ki solusi itu yang make sense. Masuk akai gak? https://t.co/2ENPBQnaen //	iot Complaint		
	Di Nakopatian Senga berutna (PT) Magosa Cardeda du Januari Berlin Manor na Janga hash Mine Jago ya Makomini Al Manor Na Janga Hang Manari Manari Manor Manari br>Manari Manari			
2582		iot Complaint		
		iot Complaint		
	@DauglasHack @TokopediaCare Deal setuju banget dengan kata2 ini Ini buktinya Seller bermasalah. Transaksi dari kapan. Nunggu dana balik sampe tgl 14. Duit di tahaaaaan euy sm @tokopedia			
	https://t.co/8YQ0ju7JWT	iot Complaint		
2584				

Figure 4. Labeling Not Complaint Sentences



Figure 5. Preprocessing Phase 1 Process

a) Annotation Removal

On the Gata Framework website, it is done to remove the mention or @ sign in the tweet sentence so that the resulting sentence does not have the @ sign.

- b) Transformation Remove URL In this process, the URL is removed in the sentence contained in the tweet on the Gata Framework web.
- c) Tokenizasion Regular Expression (Regexp) At this stage, the sentences on a dataset are broken into words on the Gata Framework website.
- d) Indonesian Stemming In this process, the sentence that contains the affix is removed so that the word that contains the affix becomes the base word on the Gata Framework website.
- e) Indonesian Stop Word Removal

The last process in preprocessing phase 1 is to do Indonesian Stopword removal on the Gata Framework website.

No	Text	Status	@Anotation Removal	Transformati on: Remove URL	Regexp	Indonesian Stemming
1.	@TokopediaCare Siap ka, tolong dibantu ya terima kasih	Not Complai nt	siap ka, tolong dibantu ya terima kasih	siap ka, tolong dibantu ya terima kasih	siap ka tolong dibantu ya terima kasih	siap ka tolong bantu ya terima kasih
2.	@tokopedia Pearl pink, warnanya manis bangett kaya minto @TokopediaCare	Not Complai nt	pearl pink, warnanya manis bangett kaya minto care	pearl pink, warnanya manis bangett kaya minto care	pearl pink warnanya manis bangett kaya minto care	pearl pink warna manis bangett kaya minto care
••••			•••••			
2584	@DauglasHack @TokopediaCare Deal setuju banget dengan kata2 ini Ini buktinya Seller bermasalah. Transaksi dari kapan. Nunggu dana balik sampe tgl 14. Duit di tahaaaan euy sm @tokopedia https://t.co/BYQ0j u7JWT	Not Complai nt	deal setuju banget dengan kata2 ini ini buktinya seller bermasalah. transaksi dari kapan. nunggu dana balik sampe tgl 14. duit di tahaaaaan euy sm https://t.co/byq0ju7jw t	deal setuju banget dengan kata2 ini ini buktinya seller bermasalah. transaksi dari kapan. nunggu dana balik sampe tgl 14. duit di tahaaaaan euy sm	deal setuju banget dengan kata ini ini buktinya seller bermasala h transaksi dari kapan nunggu dana balik sampe tgl duit di tahaaaaan euy sm	deal tuju banget dengan kata ini ini bukti seller masalah transaksi dari kapan nunggu dana balik sampe tgl duit di tahaaaaan euy sm

Tabel 1. Preprocessing Result using Gata Framework Website



International Journal of Information System & Technology Akreditasi No. 158/E/KPT/2021 | Vol. 6, No. 1, (2022), pp. 134-144

3.4. Preprocessing Phase 2

At this stage we use Rapid Miner Software, several functions, methods and operators are carried out in data processing, namely read excel, which is reading data that has been carried out in preprocessing phase 1, then process Nominal to Text, namely converting nominal attributes into string attributes, Process Document from data in which there are several operators to process the existing data set, SMOTE Upsampling to overcome the imbalance data and without SMOTE upsampling as a comparison of the level of accuracy, Precission, Recall, and AUC of the data set owned, then the process is carried out using the Cross Validation operator.

a) Read Excel

This stage is carried out by importing data that has been preprocessed in phase 1 into the Rapid Miner application using the read excel operator.

			//Local Reposi	tory/tugas/skrip	si2* - RapidMine	er Studio Educat	ional 9.10.000 @	Ricos-MacBool	k-Pro.local				
2 📁 🖬 - 🜔		0			Import Data -	Select the cells	to import.				erators	, etc 🔑 All Studi	• •
Repository ×					Select the	e cells to in	nport.						
O Import Da	1												
Training Resources (connected)	Sheet:	result (2) 🔻	Cell range: A	1	Select All	Define	header row:	1.0			Iguratio	on Wizard	0
Samples												eprocessing 1.xis	
Sommunity Samples (connected)		A		c	D	E	F	G	н	1		eprocessing Like	1.
Local Repository (Local)	1	No	Text	Status	@Anotation	Transformat	Regexp	Indonesian _	Indonesian _		2	sheet number	- 0
B Expany	2	1	@Tokopedia	Not Complaint	siap ka, tolo	siap ka, tolo	siap ka tolon	siap ka tolon	ka tolong ba		-		
	3	2	@tokopedia	Not Complaint	pearl pink, w	pearl pink,	pearl pink w	pearl pink w	pearl pink w			1	đ
	4	3	@tokopedia	Not Complaint	care pearl pl	care pearl pi	care pearl pi	care pearl pl	care pearl pi				
	5	4	@jaejunct@t	Complaint	care ga bisa	care ga bisa	care ga bisa	care ga bisa	care kali jam			A1	9
	6	5	RT @tokope	Not Complaint	rt menangka	rt menangka	rt menangka	rt menang tx	rt menang tx			SYSTEM	
	7	6	@Tokopedia	Complaint	tolong di foll	tolong di foll	tolong di foil	tolong di foll	tolong follow			STSTEM	
	8	7	RT @Tokope	Not Complaint	rt hai kak, te	rt hai kak, te	rt hai kak ter	rt hai kak ka	rt kait kenda				0
	9	8	Meski mayori	Not Complaint	meski mayor	meski mayor	meski mayor,	meski mayor	mayoritas m		100		
Operators ×	10	9	@oktochristia	Not Complaint	mohon maaf	mohon maaf	mohon maaf	mohon maaf	mohon maaf			Enter value 💌 📓	1 0
nominal to te	11	10	Hei @Tokop	Complaint	hei saya mau	hei saya ma	hei saya mau	hei saya ma	hei cek resi s				
* - I YURS (LL)	12	11	@Tokopedia	Not Complaint	sudah dm	sudah dm	sudah dm	sudah dm	dm			SYSTEM	• 0
Set Positive Value	13	12	@Tokopedia	Complaint	min jam segi	min jam segi	min jam segi	min jam gin	min jam gin			English (United Stat	- 6
Numerical to Binomina	14	13	@bukansiap	Complaint	care pernah	care pernah	care pernah	care pernah	care claim ja			English (United Stat	
Numerical to Polynomi	15	14	@jaejunct @t	Complaint	care sama ni	care sama ni	care sama ni	care sama ni	care aneh ka		lanke		.0
Nominal to Binominal	16	15	@tokopedia	Complaint	care kenapa	care kenapa	care kenapa	care kenapa	care jam clai				
Nominal to Text	17	16	RT @Tokope	Not Complaint	rt hai kak, b	rt hai kak, b	rt hai kak bo	rt hai kak bo	rt info email			🕞 Edit List (0)	d
Nominal to Numerical	18	17	@Tokopedia	Not Complaint	tolong cek d	tolong cek d	tolong cek d	tolong cek d	tolong cek d				
Nominal to Date	19	18	Наі @Токор	Complaint	hai kalau sa	hai kalau sa	hai kalau sa	hai kalau Sa	sabtu mingg		as miss	lings	a
Text to Nominal	20	19	@Tokopedia	Not Complaint	cek dm min	cek dm min	cek dm min	cek dm min	cek dm min				
Date to Nominal	21	20	RT @Tokope	Not Complaint	rt asalkan di	rt asalkan di	rt asalkan di	rt asal dia b	rt masuk gol		× 11		

Figure 6. Import data from excel process using read excel Operator

b) Nominal To Text

This operator functions to convert the selected nominal attribute type to text and maps all attribute values to the appropriate string

É RapidMiner Studio File Edit Process Vi	ew Connections	Settings Extensi	ions Help		8	😑 🔘 🕲	□ ⊙ ∩ ∗	#0 奈 @ Q ₽	Tue 17 May 09.18
• • •	//Local Repository	/tugas/skripsi2* -	- RapidMiner	Studio Education	al 9.10.000 @ I	Ricos-MacBook	-Pro.local		
🔰 😑 Η · 🕨 · 🔳	Views:	Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operatorsetc	🔎 All Studio 👻
Repository ×	Process						Parameters	×	
🗘 Import Data 🛛 🗉 👻	Process			p p	- An 🖪 🗖	a 🖝 🗵	Nominal to Tex	t	
Training Resources (connected)	Process						attribute filter type	single	v 0
Samples	Read Ex		al to Text			res (attribute	Indonesian Stop word re	moval v 0
Local Repository (Local)	Dinp 🛄 🎽		exa ori			ie (invert selection		d
DB (Legacy)							include special a	ttributes	d
Operators ×									
process docum									
Extensions (17)	^								
🐨 🔁 Operator Toolbox (4)									
Text Processing (4)									
Apply Model (Documents)									
Dictionary-Based Sentiment (Documents)									
Extract Topics from Documents (LDA)									
Split Document into Collection									
Text Processing (11)									
Create Document									
Read Document		dom of Crowds to g	get operator re	commendations ba	ised on your pro	cess design!			
Write Dorument	<u>v</u>		Activate Wis	dom of Crowds			Change compatil	bility (9.10.000)	
No results were found.									

Figure 7. Nominal To Text operator

c) Process Document From Data

Several processes are used to clean the data so that it becomes a vector that can be used as an algorithm calculation, including:



- 1. Transform Case on the parameter tab we change to transform to lower case is to change the character to lowercase.
- 2. Tokenize to break sentence into word
- 3. Filter Tokens (by Length) to select words to be processed into words with a minimum of 4 character and 25 characters maximum
- 4. Filter Stopwords (Dictionary) to remove words that have low information from a text.
- 5. Steming (dictionary) to remove affixes in the form of confixes, prefixes, or suffixes in a word so that it returns to the basic word.

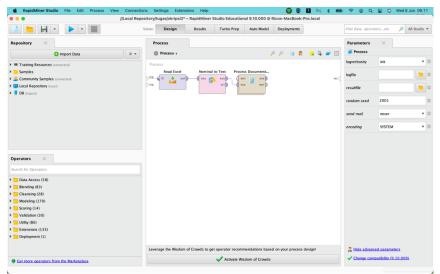


Figure 8. Process Document From Data operator

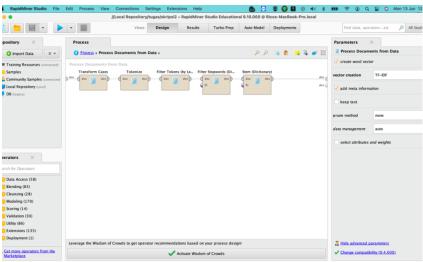


Figure 9. Inside Process Document From Data operator

3.5. Evaluation

At this stage an evaluation of the performance vector of machine learning is carried out with reference to the Confusion Matrix in the form of accuracy, precision, recall, AUC (Optimistic), AUC, AUC (pessimistic) and stemming words that have no value or reduce the level of performance in performance. vectors. The following is an explanation of:

a) Accuracy

The ratio of predictions to the true class ("complaint" and not "complaint") to the entire data. Accuracy can answer questions about the percentage of



International Journal of Information System & Technology Akreditasi No. 158/E/KPT/2021. Vol. 6, No. 1, (2022), pp. 134-144

complaints and not complaints submitted via twitter to the TokopediaCare account.

Accuracy = (TP+TN) / (TP + FP + FN + TN)

b) Precission

The ratio of true positive prediction to the overall positive predicted outcome. Precision will answer the question of what percentage of "complaint" are submitted from all customers who are predicted to "complaint". Precisson = (TP) / (TP + FP)

c) Recall (Sensitivity)

the ratio of true positive predictions compared to the overall true positive data. Recall can answer the question "what percentage of customers have been predicted as "complaint" compared to all customers who actually "complaint" [7].

Recall = (TP) / (TP + FN)

- d) AUC (Area Under The Curve)
 - AUC (Area Under The Curve) serves to make it easier to compare one model with another, AUC is the area under the ROC (Receiver Operating Characteristics) curve or as an integral of the ROC [8].

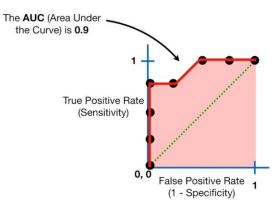


Figure 10. AUC Sample

Process Process >	PP 👂 🗽 🛢 🚅 🜌
	nts Cross Validation res res res res res res res res

Figure 11. Cross Validation Operator

We use this operator is to perform training and also testing on data sets in which there is a Naive Bayes algorithm operator for training data for evaluation process. This use for testing the data to determine the level of accuracy, performance, recall and AUC (Area Under Curve) in the dataset. Inside cross validation operator there is training and testing section we use operator Naive Bayes to training the data and apply model to test Naive Bayes Algorithm and we got the performance

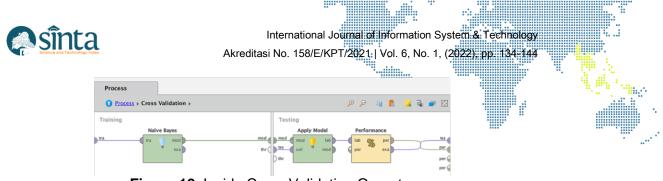


Figure 12. Inside Cross Validation Operator

The use of the Naive Bayes Algorithm in this study resulted in the following data:

- 1. Accuracy : 83.13%
- 2. Precission : 65%
- 3. Recall : 60.56%
- 4. AUC : 0.801%



Figure 13. Added SMOTE Upsampling Operator

And we use SMOTE for comparation of Naive Bayes Algorithm and result is:

- 1. Accuracy : 81.59%
- 2. Precission : 82.73%
- 3. Recall : 84.17%
- 4. AUC : 0.852%

3.6. Deployment

At the deployment stage, database input is carried out from the dataset that has been preprocessed in stages 1 and stage 2, deployment is also the application of machine learning into the php programingfrom the Gata Framework so that it can directly predict tweets addressed to the TokopediaCare Twitter account. Depolyment based on the results of the evaluation of the process of testing the model between the Naive Bayes algorithm and the Naive Bayes algorithm model added with the Syntethic Minority Over Sampling Technique Method (SMOTE) feature. Because this weight will be used in predicting a sentence or tweet in the application

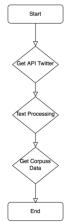


Figure 14. insert preprocessing result into database



International Journal of Information System & Technology Akreditasi No. 158/E/KPT/2021. | Vol. 6, No. 1, (2022), pp. 134-144

-2

.

ľ

					····	
ction New Query Table	View Function Others Us	er Query Backup	Automation M	fodel Charts		View
tokopediacare	Objects bobotcom	plaint@sentim				
information_schema		7 🚛 🔍 🖓	B 4			
performance_schema	id attribute	parameter	not_complaint	complaint		
phpmyadmin	4678 abai	mean	0	2880540478		
sentimencomplaint	4679 albai	standard deviation	0.001	0.012827302		
 Tables 	4680 abisin	mean	460541047623.3	0		
	4681 abisin	standard deviation		0.001		
bobotcomplaint	4682 absen	mean	504286434694.5	0		
dashboard_user	4683 absen	standard deviation				
graph_model	4684 aceh	mean	504286434694.9	0		
graph_query	4685 aceh	standard deviation				
initial_company	4686 adain	mean	863910337910.3			
master_department	4687 adain	standard deviation				
master_group	4688 aktifasi	mean		404802670.1		
master_group_detail	4689 aktifasi	standard deviation		0.017053508		
	4690 aktifin	mean		9024887979		
master_module	4691 aktifin	standard deviation		0.02292633+		
master_profil	4692 aktipkan	mean		8769065458		
imaster_unit	4693 aktipkan	standard deviation mean		0.019288924 8591198762:		
master_user	4694 alesannya					
master_user_detail	4695 alesannya 4696 alhamduillah	standard deviation mean	0.004767604561	0.024981252		
replace_character	4697 alhamduillah	mean standard deviation		0.001		
report_query	4698 alhasil	mean	0.001040264006			
table 18	4699 alhasil	standard deviation				
	4700 aman	mean	0.005962554781			
tbl_results	4700 aman	standard deviation				
💳 testajah	4702 ampun	mean	458852706218.8	0		
> 🤜 Views	4703 ampun	standard deviation		0.001		
> 🗲 Functions	4704 asia	mean	393989013741.3			
> 🕢 Events	4705 asia	standard deviation		0.001		
>	4706 awas	mean	0	3804488682		
> Backups	4707 awas	standard deviation		0.01668303!		
test	4708 baca	mean	0.009480198923			
	4709 baca	standard deviation	0.094611552223	0.011956947		
	4710 balik	mean	0	0.004878261		
	4711 balik	standard deviation	0.001	0.057471234		
	4712 balikpapan	mean	3225631487551.	0		
	4713 balikpapan	standard deviation	0.014354023500	0.001		

Figure 15. Insert Text Mining Result into Database

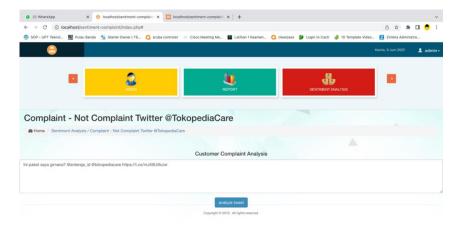


Figure 16. Get API Twitter

Deployment results to retrieve data from the @Tokopedia Care Twitter account using an API (Application Programming Interface). Then the next step, when you press the Analyze Tweet button, it will go to the text processing stage.

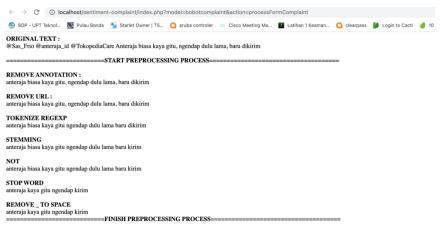
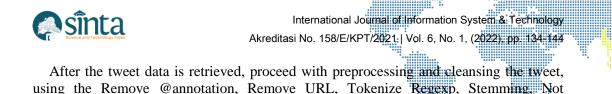


Figure 17. Preprocessing Result on application



FINISH PREPROCESSING PROCESS	
PROBABILITY OF WORD	
0 anteraja	
1 kaya kaya Not Complaint : 0.00823279995568006, Complaint : 0.011780314690692314 kaya Not Complaint : 0, Complaint : 0.00673997608950266	
2 gitu	
3 ngendap	
4 kirim kirim Not Complaint : 0.021534167761419094, Complaint : 0.09014604229356253 kirim Not Complaint : 0.0022669019149307, Complaint : 0 =====PROBABILITY OF WORD====================================	
SUMMARY WEIGHT OF WORD POSITIVE or NEGATIVE	
Not Complaint : 0.03203386963203 Complaint : 0.10866633307376	
KESIMBUI AN: Complaint	

Transformation Negative, Stopword, Remove _ to Space techniques.

Figure 18. Word Weight Calculation

In Figure 15 stages of word weight assessment in tweets addressed to the @TokopediaCare account, the results of the calculation produce a Complaint conclusion because the complaint weight of the words in the tweet is greater than Not Complaint.

4. Conclusion

The results of the research that has been carried out regarding customer sentiment towards Tokopedia services carried out via Twitter on the @TokopediaCare account, there are the following conclusions, the use of the Naive Bayes Algorithm in the training and testing process of the processed dataset obtains an accuracy value of 83.13%, which is higher than the addition of the Syntethic Minority Over Sampling Technique Method (SMOTE) feature. In the implementation of Machine Learning to predict tweet sentences used at the time of deployment using the Naive Bayes algorithm and the addition of the Syntethic Over-Sampling Technice Method (SMOTE) feature, it is more effective in overcoming data imbalances because the resulting weight can reach above 80%, namely accuracy 81.59 %, Precision 82.73%, Recall 84.17%, and AUC of 0.852% while the use of the Naive Bayes algorithm that does not add SMOTE features is 83.13% accuracy, Precision 65.00%, Recall 60.56%, and AUC is 0.801%.

References

- [1] N. Ruhyana and D. Rosiyadi, "Klasifikasi Komentar Instagram untuk Identifikasi Keluhan Pelanggan Jasa Pengiriman Barang dengan Metode SVM dan Naïve Bayes Berbasis Teknik Smote," *Fakt. Exacta*, vol. 12, no. 4, p. 280, 2020, doi: 10.30998/faktorexacta.v12i4.4981.
- [2] B. Liu, "Sentiment Analysis and Opinion Mining," no. May, 2012.
- [3] S. Agarwal, Data mining: Data mining concepts and techniques. 2014.
- [4] Mochammad Haldi Widianto, "Algoritma Naive Bayes," *Https://Binus.Ac.Id*, 2019. https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/ (accessed Apr. 09, 2022).
- [5] Arwan, V. Ardiana, L. Reza Ariana, F. Samuel, D. Ramdani, and Aditya, "Synthetic Minority Over-sampling Technique (SMOTE) Algorithm For Handling Imbalanced Data," *binus.ac.id*, 2018.
- [6] D. D. Saputra et al., "Optimization Sentiments of Analysis from Tweets in



International Journal of Information System & Technology Akreditasi No. 158/E/KPT/2021 | Vol. 6, No. 1, (2022), pp. 134-144

myXLCare using Naïve Bayes Algorithm and Synthetic Minority over Sampling Technique Method," J. Phys. Conf. Ser., vol. 1471, no. 1, 2020, doi: 10.1088/1742-6596/1471/1/012014.

- [7] R. Arthana, "Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning," *Medium.com*, 2019. https://rey1024.medium.com/mengenal-accuracy-precission-recall-dan-specificityserta-yang-diprioritaskan-b79ff4d77de8.
- [8] R. Arifin, "Memahami ROC dan AUC," *Medium.com*, 2019.