Comparison Method Of Classification Naives Bayes And Classification Nearest Neighbor On Analysis Twitter Posts Jams Traffic In Bandar Lampung

R Wijaya¹, U Rizal¹

^{1.} Information Systems, Computer Science Faculty, Bandar Lampung University, Indonesia

Abstract. Twitter became one of the popular social media in cyberspace. Twitter can be used as a means of disseminating information in their status, this is an opportunity for the parties to disseminate information. This study aims to classify the status of which contain information congestion on Twitter. This study focused on the information content congestion tweet on Twitter. Classification algorithm used is Naive Bayes classifier and K-Nearest Neighbor classification model that produces information-based decision trees and rules. This research will prove Classification is done with Twitter data method, Naive Bayes classifier and K-Nearest Neighbor to classify tweets containing information on traffic congestion in Bandar Lampung. Keywords: *Naive Bayes classification, classification Nearest Neighbor, Twitter, Tweets.*

1. Introduction

1.1 Background

Every day server Twitter receive data tweet with a very large amount, thus, we can perform data mining used for a particular purpose. One is for the visualization of traffic jams in the city. Naive Bayes classifier is an approach that refers to the Bayes theorem, by combining the K-Nearest Neighbor prior knowledge to new knowledge. So that is one of the classification algorithms are simple but have high accuracy. (Sandi Dawn Rodiyansyah, Edi Winarko 2012). Classification of data mining is a technique used to predict and are used to classify data into different classes by considering some constraints. Data classification problems have many applications in various fields of data mining. This is because the problem aims at studying the relationship between a set of variables and variable features most important targets. (Sayali D. Jadhav, HP Channe 2014).

Therefore, in this study will prove the ability of Naive Bayes classifier and K-Nearest Neighbor to classify tweets containing information on traffic congestion in Bandar Lampung. with one of the techniques used in the processing of such data by using Classification Naïve Bayes and K-Nearest Neighbor.

2. Literature

To support this study, the researchers took several sources of previous similar studies in lakunan by previous researchers regarding the classification Naïve Bayes and K-Nearest Neighbor.

2.1 Classification Posts Twitter Bandung City Traffic Congestion Using Naive Bayesian Classification

Research conducted by Sandi Dawn Rodiyansyah and Edi Winarko on "*Classification of Posts Twitter Bandung City Traffic Congestion Using Naive Bayesian* Classification" (2012). Naive Bayesian classifiers classification techniques. This approach is referred to the Bayes theorem which is a statistical chance to combine the principle of prior knowledge to new knowledge. This principle is then used to solve classification problems naive Bayesian classifier is a classification algorithm that is simple but has the capability and high accuracy.

2.2 Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques

Research conducted by Sayali D, Jadhav and HP Channe about "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques" (2014). Data classification is the process of organizing data into categories / groups in such a way that the same group of data objects more similar and data objects from different groups are very different. classification algorithm assigns each instance

for a particular class so that misclassification will be slight. It is used to extract a model that accurately defines a class of important data in a given dataset.

2.3 SMS Classification Based on Naive Bayes Classifier and Apriori Algorithm Frequent Item Set Research conducted by Ishtiaq Ahmed, Donghai Guan and Tae Choong Chung about "SMS Classification Based on Naive Bayes classifier and Apriori Algorithm frequent item set" (2014). Naïve Bayes is a simple probabilistic classifier based on Bayes' theorem with strong naive independence assumption. These assumptions are treated each and every word as a single, independent and mutually exclusive. This model can be described as "Model Independent Feature". As the complexity in learning Bayesian Classifier is colossal, there must be some way to reduce the complexity and thus the

classification was introduced

2.4 Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)

Research conducted by Ishtiaq Ahmed, Donghai Guan and Tae Choong Chung about " Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms) "(2016). Classification is an act to provide groups in every state. Each country contains a set of attributes, one of which is the class attribute. This method needs to find a model that can explain the class attribute as a function of the input attributes. A decision tree is one of the most popular methods of classification because it is easy to be interpreted by humans. Here, each branching stating conditions to be fulfilled and the tips of the tree declare a class data. C4.5 decision tree algorithm is the most famous, but the algorithm is not able to handle the data that has a large scale.

3. Results And Discussion

After all databases have been collected, the next step is to use rapid application miner 5 to classification Naïve Baiyes and K-Nearest Neighbor test the accuracy of probability models with the data testing and experimenting with the use of data already collected.

- -

3.1 Testing with Device Rapid Miner 5

Below is a database table that can tweet

Table 1. Database Tables.									
NO	Sample	Results Sample	Year	Remarks					
1	2	2	Year 2011	Solid					
2	1	1	In 2012	No Solid					
3	2	2	In 2013	Solid					
4	1	1	2015	Not congested					
5	5	4	2016	Compact					

Testing With Naives Bayes classification. In tests using other software in this research is to use Rapid Miner 3.1 The method used to classify the data of this study is to use bayes classifier naive method.





Figure 1. Display Rapid Miner.

Figure 2 The Process Started Validation

- 1. In this test, the data used is clean data that have been collected are stored in the data table. Then the read of the database to retrieve data. The next process is the final stage is the stage of testing using validation process. For greater clarity, the entire testing process with software Rapid Miner 1 can be seen in the picture below.
- 2. Classification process with Naive Bayes classifier Rapid Miner 1 shows the classification process with Rapid Miner.
- 3. Naïve Bayes classifier is used to classify the data of this study in order to obtain this accuracy at rapid miner.
- 4. Naives Bayes classification process after connecting to the database.



Figure 3. Validation Process in Rapid Miner



Figure 5 Connected With The database

C Operators A 1 0 0	Process II S XML II	 St Parameters II O Context II
Petro Resources Resources Resources Resources		Constructions of Constructions Constructions of Constructions Constructions of Constructions Construction of Validations Construction of Validations Construction of Validations Constructions Constructions
Samples (unit) (G DB (C DB) (C D	Trations () Lap ()	Compatibility inset 531002
Contraction of the local division of the loc	9 potential problema	and the second s

Figure 4 Naives Bayes classification process.

		100	v process* > - RapidMiner 5.3.000 @ Rar	mdan				
Ele Est Process	Toola New Reb							
C 📬 🖩 🖡	2 9 🔊 A 🖉 🛛							
i tun	pleBet (Patrieve database) 🚿	SingkOstro.dor	(Nate Bayes) 11	Example Def (R.ocal Repository/database)				
Table (Bolthes)	Tartilar Association		27 veronterreserts 6	reconsector a	P.A. P.A. P.D. D.D.			
Criterion Selector	A . D 7 Satalas new							
61430					3 (BCB			
receil	(e) Table Vew () Platiview	(@ Table Wew () Plot Wew						
UC (optimistic)	accuracy: 60.02%	accuracy 60.07%						
UC Disententet		The padal	thue bidak padal	class precisies				
co parente	pred padat	3	2	60.00%				
	pred bdsk padat	0	0	0.00%				
					0 0 0 0 0			
BLOD X H o	10				System Wondar 💷 😳 🖸 🖾			
	a				Cystere mondar 2 2 0 6			
Log 35 20 0 40 15 2017 15 152 40 15 2017 15 152 40 15 2017 15 152 40 15 2017 15 151 40 15 2017 15 151 40 15 2017 15 151 40 15 2017 15 152 40 15 20 15 20 15 40 15 20 15 20 15	THING Roading sample set. 1945 RSC No Bename given for in 1945 RSC No Bename given for in 1945 RSC Source mission 1945 RSC Source mission 194	wal file, using altitual for logging resultar antidy affect 3 x wall file, using altitual for logging resultar						

Figure 6 Results Of Classification Naives Bayes

5. can be seen that the result of the naves Bayes methods using rapid miner as shown below with an accuracy of 60%.

3.2 Tests By Classification Nearest Neighbor

In tests using other software in this research is to use Rapid Miner 3.1 The method used to classify the data of this study using the method of classification Nearest Neighbor and use the data that is used naives Bayes.

Elle Edi Brones Tools New Help		Ein Est Bocass Ioss Non Pala	
E Opening X (1 + E)	El Parameters II O Context	■ Coperators II 12 = 11 (2 ² Process II) (2) (ML) II)	E Parameters X O Context X
0 - Fat 0 > 1 + + + # Pross - 0 - 1 + 2 3 8		🖉 = nale bayes 🛛 👰 🦻 💭 👘 🕶 = 🛊 🚆 Process + 🖏 Yalidaton + 🖉 = 🛄	\$1 1 1 1 4 + 2 5 5 7 5 8 +
e g (Prozz (2004))) e g (Prozz (2004)) e g (2004) e	topologi topologi https://www.communications/ mouthing/ and mouthing/ and mouthing/ a	B G toots 2 B G toots 2 C S Construction of Reparation (Construction) C S Construction (Constructi	Image: Statistics (1) indexter(s)
	ecolog 201553 •	Resolute 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Compatible level 131000 - Maile Communit S XValidation (Populator Com)
u data book /kunau-fa - vi 2 (ST) / SD / Arabima / Bug / S	Synapsis	data book Revision of Diverting A Problems A Diverting A Diverties	Synoph
- 🔰 tasat Ruman Re-ut, Shent ni 12 Ala 👔 Na polalena faund		A federita proterns	
Hosspe Pies Licobe	The root operator which is the outer most operator of every process.	Mossage Press Press Press Mossage Op Parameter responsive retry accesses a reposition by raining (c) court Riv. To use of the values of the value o	Location This operator performs a cross-validation in order to to training of the standard the standard of the

Figure 7 Display Rapid Miner blank.

Figure 8 The process started Validation

In this test, the data used is clean data that have been collected are stored in the data table. Then the read of the database to retrieve data. The next process is the final stage is the stage of testing using validation process. For greater clarity, the entire testing process with software Rapid Miner 3.1 can be seen in the picture below.

	<rew process*=""> - RapidMiner 53.000 @ Ramdan</rew>	- 0 - X	2 *	<new process=""> > napidwiner 3.3.000 @ name</new>	
Elle Edit Bracess Tools Yew Help			Elle Edit Discess Jouls yew Help		
	1 I I I I		1 4 E B 9 A 4 B	🛐 💐 🛞	
		(The second data and	E Operators X X = E	ess X FXML X	Bi Parameters # (A) Context #
C CPERITS A C C C C C C C C C C C C C C C C C C	2 M. A	grannen Goner	Ø 0 > I	- 2 EPocess + % Validation +	0.02234.199098.
P to Modelina (2)	Threes .	2.Ft + F 2 2. ft + 4 1 4 1 4 1 4 1	C C Nodeling (5)		📆 Validation (X. Validation)
B G Cassification and Regression (2) Item Process		E Hoss	E G Lab Wodeling (1)	in the Archivelouter	Potermanez
		logerbosky int *	W Bild Distering and Segmentation (4)	(=) (= g/ =) (=
- V Nave Bajes (Kersel)	(Retrines funct)	toglia 🚳	Allegra Kanel	9 m 0 m 1 v	
	(m m)	southe	Belleans (tast) Belleansta		number shalidations 13
	0 A				samping type statiled samping '
		random deed 2001			use local random seed
	0 <u>8</u>	send mail rever *			
		ercoding SYSTEM *	Rousenerven		
(Boundary V V a B.)			2 4 - 0 0 0 -		
8.4.9.0.0.0			() Samples ment		
a Samples			(ECG (CG (CG) (CG)		Compatibility level 63.000 🕆
e 🗟 De			8 G GR Bardan Po		Hat M Comment M
E Q dda fander fo		O Help II Convent II	- O LEAST DESIGNATION		97
B D processes (fumion-fro)					The X-Validation (Tapithine Cent)
- B b83 Randar Pt - v(01417 225 All -		Process	A P100	ens X BLog X	Same
- Booki (Rente-Re), 21217 1100 P	The v	and a second sec	Dr One p	deritél protein	
	lighting in	olaskar	- Perma	Nessage Files In Sector with amazing a sector to some (5 cm Da - A to societ to avaid	Location This operator performs a
M change in	Nessage Files	Location The root operator which is	n D	a channel and annear a channel of and form at . A sector of party	admitted the statistic of
A Parameter hopos	itory entry' accesses a repository by name (/Local Re 🕘 No quick fix available	Retrieve tweet Retrieve tweet Provide and the outer most operator of revery process.			
C D Handatory input n	nissing af port Validation training 📀 Choose among 2 available quic	S Valdzton taning	Figuro 1(Noorost Noighh	or Classification
0			rigure n	r incarest ineight	

Figure 9 Inserting Validation

Process

Nearest Neighbor classification process by a classifier Rapid Miner 3.1 shows the classification process with Rapid Miner 3.1 Nearest Neighbor method classifier is used to classify the data of this study in order to obtain this accuracy at rapid miner. Naives Bayes classification process after connecting to the database. Can be seen that the result of the Nearest Neighbor method using rapid miner as shown below with an accuracy of 80%.

of the second						1		-0164	hores a surbruside portor di al	carbiera -	
1 📦 🖩 🖓 🥬 🔿 🗛						in Lit Process	Tote New Helb				
E Operators X 12 0 0	Process X EXHL X			Paramete	es II 🕐 Conted II		39 04 01				
🖉 🗸 👔 petam	Ø • 0 • 1 ∰ Pozza • Ø • □ \$ Ø Ø Ø			8-200004-		Immeliat Paters dataset Z 200/ClassRedex 64% Z Sectional Context Reports (International Context Reports (In					Repostories X 2 o 0
Attitutes (1) Performance (Attitute Cou					# Process	() Table / Plot View (C Ted View () Aenolations			*	4.84.4000
🕀 🇐 Clustering (5)	[Betrinen faced]	Walishafiren		iosected	int	Criterion Selector	(a) Multiclase Classification P	edumance () Amotators		X	a
S Cluster Count Performance Gunter Distance Performance	the line of the li	(10 mil)		as a	(precision	Table View () Pictules				B @ Local Repository Annumina
Cluster Density Performany	<u>e 6</u>	ay "		en logfie		recall AUC (optimizatio)	accuracy: 30.00%				H C Ala Artist Pij
S Map Clustering on Labels		40		na resuttie		AUC .		true partor	Inue listak partial	class precision	
- S References		0 2				No. geometry	pret polat)	1	75.00%	
S. Extract Performance		- NO		random seed	2001		pred tidat podat	0	1	100.00%	
- R Performance (User-Based)					(0311 (604	100.00%	00.00%		
Performance (Min-Mai)				pero man	never						
- % Performance to Data				encoding	RVSTEN						
4 V											17 (Territoria) (T
Repositories # 25 0 D						Dates 11 Mar					Derben Menter 2 2 4 8
8 A - 9 4 4 4 4											
E 👌 Samples moter						Fab 16 2017 1016 1	3 PN RIFO Process Invided secon	unturk antari de s			0
8 (B CB						Fab 18, 2017 12:18:1	0 PH NFO Reading wainple set.				1
B @ Local Repository (Landah-Po)						Fab 10, 2017 10 10 10	9 PH FUTO, HIGHLING AVAILUPA DAL. 3 PH FUTO, NA MARGINA, given far re	out the using attend for register resulted			
🗄 🛄 dala (raman Fil)				Help 📈	Comment II	Fab 18,0017 10 18:5	2 PRI INFO. Process starts				
- I LISE daman for at the TT hill				100		Feb 18 2017 12:18:5	2 Pill NPC Saving results.				
bik) Handas Hu - st. Starti 8 State				Process		Fab 10,0017 10:105 Eab 10,0017 10:105	2 PB RPO Process Strahed autors 7 PB RPO has Bananta down for or	inflaty after 0 s.			I March March
- Dookf marrier Po - et. 215/17 (1.00 m						Fab 10, 2017 10:201	7 PH NFO, Process stats				
- 👔 data book Sunder Pa - et. 01417 1.0	A Problems X B Log X			Sympole		Fab 78, 2017 10:22 1 Fab 18, 2017 10:20 1	7 PH NFO: Loading Initial data. 7 PH NFO: Staving results.				20. 118
WHITE A CONTRACTOR	Die polential problem					Fait 19, 2017 19:20-17 PM RPD Process Research successfully after Ex-				U	
	Message	Fors	Location	The root i	operator which is				1 0 1		
	A Parameter repository entry' accesses a repository by name (il. ocal Re.	🙆 No quick for available	Retreve tweet	The outer	most operator of	Figu	ire 12	the resu	Its of cla	ssificatio	n Neares
				EVEL U							

Figure 11 Connected With The database



4. Conclusions And Recommendations

4.1 Conclusion

Based on the analysis discussion by using Comparative Method Naves Bayes and K-NN In Analysis Postings Twitter Traffic Congestion Bandar Lampung we can conclude:

- 1. The test results accuracy by using software Rapid Miner 3.1 Naive Bayes classification method produces a value classification accuracy of 60.00% . while Nearest Neighbor produce the smallest value amounted to 80.00% accuracy.
- 2. In the classification by using the software Rapid Miner 3.1 it can be concluded that the value of Nearest Neighbor classification method accuracy greater than Naive Bayes classification.
- 3. In recent times research until the testing classification Naive Bayes and Nearest Neighbor Classification using rapid miner has not been an overall 3.1 and the data obtained is not optimal.

4.2 Suggestions

Based on the results of research and discussion in the previous chapter there are some suggestions that can be given to the author for subsequent research:

a. Keep testing whether the results of this classification is appropriate or not to maximize the accuracy results will be obtained

b. To maximize the accuracy results will be obtained Hopefully with research analysis this can be exploited by a user associated with the well.

References

- [1] Password Dawn Rodiyansyah * 1, Edi Winarko 2. Classification Posts Twitter Bandung City Traffic Congestion Using Naive Bayesian Classification IJCCS, Vol.6, No.1, January 2012, pp. 91 ~ 100 ISSN: 1978-1520 taken from https://journal.ugm.ac.id/ijccs/article/download/2144/1923.
- [2] Sayali D. Jadhav 1, HP Channe. Comparative Study of K-NN, Naive Bayes and Decision Tree Classification. Techniques international Journal of Science and Research (IJSR) ISSN (Online): 2319-7064. Index Copernicus Value (2013): 6:14 | Impact Factor (2014): 5,611 taken from https://www.ijsr.net/archive/v5i1/NOV153131.pdf.
- [3] Ishtiaq Ahmed, Donghai Guan, and Tae Choong Chung. SMS Classification Based on Naive Bayes classifier and Apriori algorithm frequent itemset. International Journal of Machine Learning and Computing, Vol. 4, No. 2, April 2014 taken from www.icmlc.org/icmlc2014/012_icmlc2014.pdf.
- [4] Leni Marlina 1, Muslim 2, Andysah Putera Utama Siahaan 3. Comparison Data Mining Classification (Naïve Bayes and C4.5 Algorithms). International Journal of Engineering Trends and Technology (IJETT) - Volume 38 Number 7- August 2016 taken from www.ijettjournal.org/2016/volume-38/number-7/IJETT-V38P268.pdf
- [5] Siti Asmiatun. Application of Collision Detection Algorithms and Bayesian For Striking Distance DECA T strategy to the NPC (Non Player Character) Using Unity 3d. Transformatika Journal, Volume 14, Number 1, July 2016 Taken From Http://Journals.Usm.Ac.Id/Index.Php / Transformatika / Article / Download /382/243.
- [6] Lukmanul Hakim and Akhmad Fauzy. Determination of Traffic Accidents Relationship Patterns Method Using Association Rules With Apriori Algorithm. University Research Colloquium 2015 ISSN 2407-9189 vol 3, No. 2 taken from https://publikasiilmiah.ums.ac.id/bitstream/handle/11617/5177/8.Lukmanul%20Hakim.pdf?se quence=1.
- [7] Puspa Tantriyana Anggitasari. Klasifikasi Spammers On Twitter Based on User Behavior Using C3.0 algorithm. Taken from http://code.cs.ipb.ac.id/abrari/RevisiPPKI/uploads/ a0dfdae19b18c87dd2dc01690c526398/SKRIPSI_TANTRI.pdf.

- [8] Anoraga Nandari, Sukadi. News Portal Website Creation Jetis. IJNS Indonesian Journal on Networking and Security - Volume 3 No. 3 July - 2014 taken from https://ijns.org/journal/index.php/ijns/article/view/272/266
- [9] Hendri Risman, Nugroho Didik and YUSTINA Retno WU, APPLICATION K-Nearest Neighbor method At Determinants Scholarship Application. Students In Stmik Sinar Nusantara Surakarta ISSN: 2338-4018 is taken fromhttp://p3m.sinus.ac.id/jurnal/index. php / TIKomSiN / article / download / 199/176