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

R Wijaya[1], U Rizal[1]

[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.
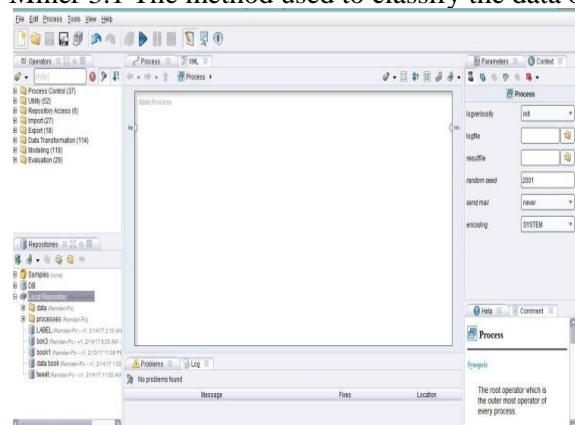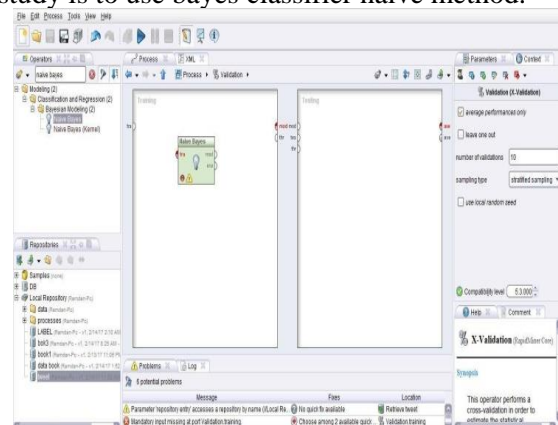

Figure 1. Display Rapid Miner.

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.
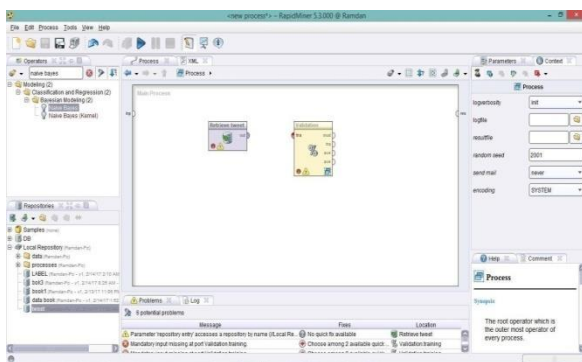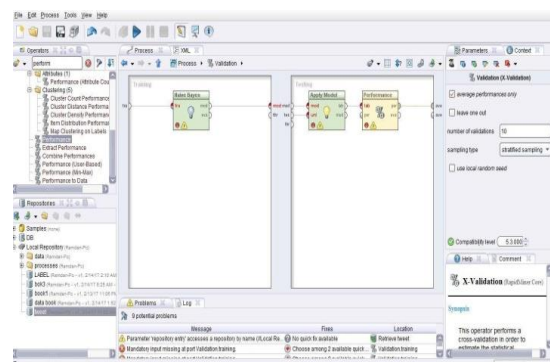

Figure 3. Validation Process in Rapid Miner

Figure 4 Naives Bayes classification process.
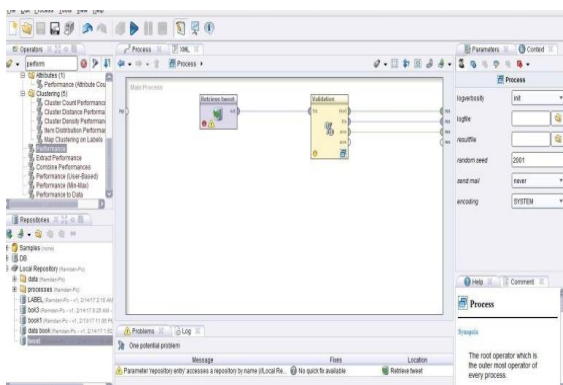

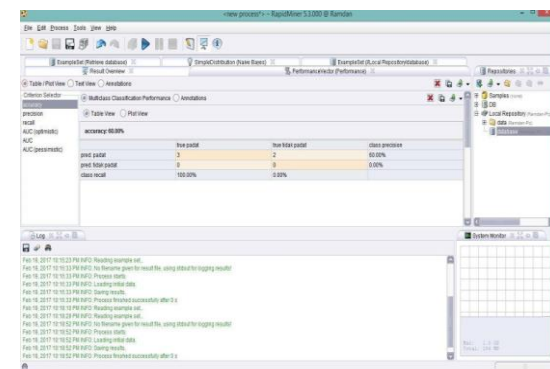**Figure 5** Connected With The database

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. ,
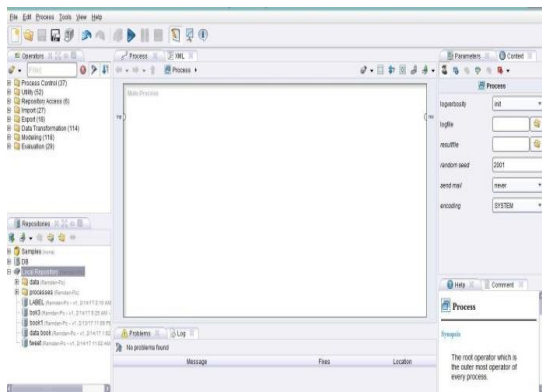
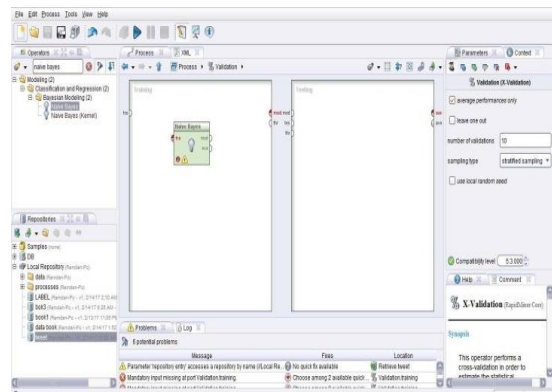**Figure 7** Display Rapid Miner blank.

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.
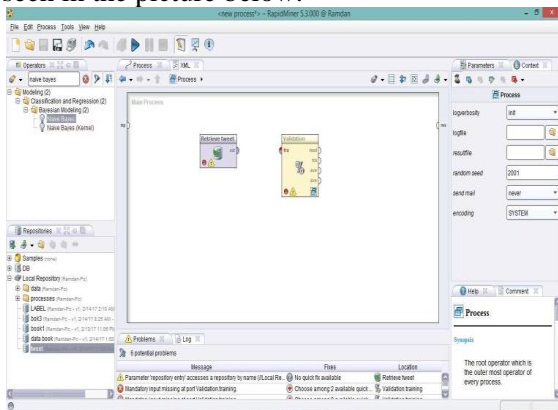

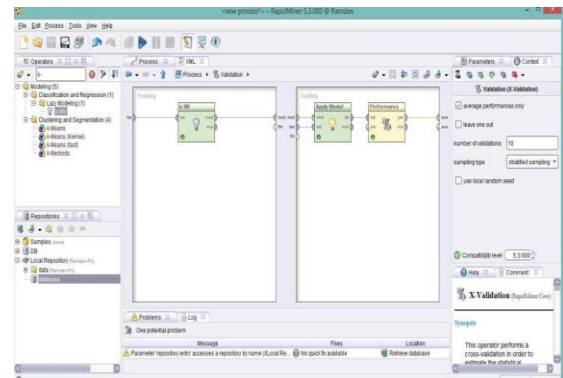**Figure 9** Inserting Validation

Figure 10 Nearest Neighbor Classification 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%.
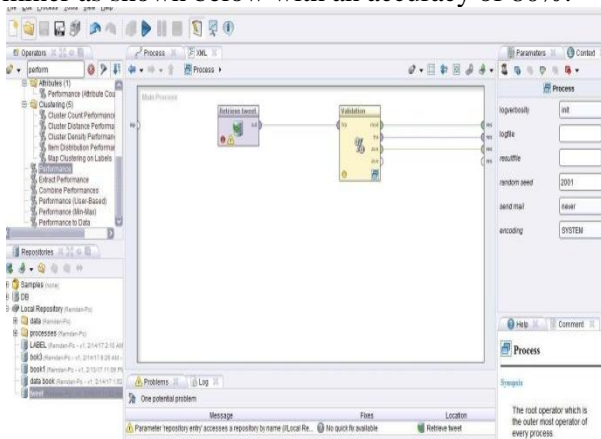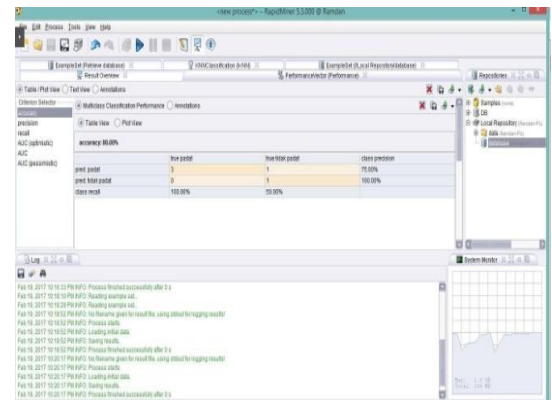

Figure 11 Connected With The database

Figure 12 the results of classification Nearest Neighbor.

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?sequence=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