Classification of Tweets Causing Deadlocks in Jakarta Streets with the Help of Algorithm C4.5

Qurrotul Aini [1,*], Jehad A.H Hammad [2], Taslim Taher [3] and Mohammed Ikhlayel [4]

[1] Dept. of Information Systems, UIN Syarif Hidayatullah Jakarta, Indonesia
[2] Department of Computer Information Systems Al-Quds Open University, Palestine
[3] Department of Computer Science & EngineeringInternational University of Business Agriculture and Technology, Dhaka, Bangladesh
[4] Dept. of Information Technology and Communications Al-Quds Open University, Palestine
[1] qurrotul.aini@uinjkt.ac.id*; [2] jhammad@qou.edu; [3] taslim.cse@iubat.edu; [4] miklil@qou.edu
* corresponding author

(Received: August 5, 2021; Revised: September 18, 2021; Accepted: October 8, 2021; Available online: December 31, 2021)

Abstract

Congestion seems to be a daily occurrence in the Indonesian city of Jakarta. As a consequence, the rider has access to essential information regarding traffic conditions at all times, which is beneficial. Through social media platforms such as Twitter, this information is readily available to the public. On the other hand, the information offered on Twitter is still uncategorized text. DKI Jakarta, as a consequence, developed a congestion classification system that included data mining techniques, a classification approach based on the decision tree technique, and C4.5 as a component. This C4.5 method transforms a large amount of information into a decision tree that shows the rules. Geocoding will be utilized to illustrate the locations that have been gathered, and a data split with a confusion matrix will be used to assess how well the categorization process has worked. According to the study's results, the average accuracy rate is 99.08 percent, the average precision rate is 99.46 percent, and the average recall rate is 97.99 percent.

Keywords: C4.5; Sentiment Analysis; Deadlock; Classification; Traffic Prediction.

1. Introduction

As a result, it is essential for motorists to be informed about the upcoming traffic conditions. At the moment, information on the status of a road may be accessed through social networking sites that are extensively utilized for communication and sharing of events. Certain major social networking services, such as Twitter, provide APIs for data retrieval from social networks [2]. This is deemed inefficient, however, since the existing data is displayed individually. It is required to combine data from several sources that have the same information in order to further classify the road conditions based on the current data [3]. The C4.5 algorithm was employed in this work to categorize traffic congestion in DKI Jakarta, which is a data mining approach. The C4.5 method is a mechanism for converting all very big facts into decision trees [4]. The C4.5 method is deemed the best suitable for twitter data categorization since it achieves the maximum classification accuracy when compared to other algorithms such as Naive Bayes [5].

This article explains the technique of utilizing the Nave Bayes Classification algorithm to identify twitter data including information on traffic congestion [6]. The system created in this work is divided into four components: downloading (data retrieval), pre-processing of tweets, tweet categorization, and display of classification findings [7]. The Tweet data collected will be classified as "stuck," "smooth," and "unknown." The "hour" and "fluent" classes are used to tweet data including information on traffic jams, while the "unknown" class is used to twitter data that does not fall into the "hour" or "fluent" groups. This study discovered that the software developed utilizing a naive bayesian classifier approach had the lowest accuracy rate of 78 percent for 100 sample data and the highest accuracy rate of 91.60 percent for 13106 sample data. The same findings were achieved when RapidMiner 5.1 was used, with the lowest accuracy being 72% on 100 sample data and the highest being 93.59% on 13106 sample data.

The purpose of this project is to develop a visualization application for obtaining traffic congestion information from an Android smartphone's map. The objective of this project is to create an application that will make it simpler for users to get information regarding congestion. During the technique, tweet data is extracted and then classified into the "jammed" or "smooth" class using the K Nearest Neighbor Classification algorithm, which is based on the nearest neighbor classification algorithm. The study's conclusion indicated that respondents believed the application was effective and really beneficial.

The purpose of this research is to discuss the method of utilizing the Naive Bayes Classification algorithm to categorize twitter data including information on traffic congestion. The system created in this work consists of four components: data retrieval (downloading), tweet pre-processing, tweet classification, and display of classification findings. The Tweet data collected will be classified as "stuck," "smooth," and "unknown." The "hour" and "fluent" classes are used to tweet data including information on traffic jams, while the "unknown" class is used to twitter data that does not fall into the "hour" or "fluent" groups. This study discovered that the software developed utilizing a naïve Bayesian classifier approach had the lowest accuracy rate of 78 percent for 100 sample data and the highest accuracy rate of 91.60 percent for 13106 sample data. The same findings were achieved when RapidMiner 5.1 was used, with the lowest accuracy being 72% on 100 sample data and the highest being 93.59% on 13106 sample data.

The purpose of this project is to develop a visualization application for obtaining traffic congestion information from an Android smartphone's map. The objective of this project is to create an application that will make it simpler for users to get information regarding congestion. During the procedure, tweet data is extracted and then categorized into the "jammed" or "stuck one" class using the K Nearest Neighbor Classification algorithm. The study's conclusion indicated that respondents believed the application was effective and really beneficial.

Using the categorization and extraction of tweets, the goal of this study is to extract information about online transactions in Indonesia that may be used for further investigation. The findings of this investigation led to the creation of an application known as SaFE-F. This application operates by searching for tweets that contain specific keywords (Search), filtering content that is relevant to online transaction activities (Filters), extracting information about online transaction activities (Extraction), and storing the extracted information in a more structured format (Storage). This study examines three classification algorithms, namely Naive Bayes, C4.5, and IBk, and compares their performance. After doing this research, it was discovered that the C4.5 algorithm performs much better in categorizing Twitter data than the Naive Bayes and IBk algorithms. Specifically, the goal of this study is to investigate a range of classification algorithms, such as decision trees, k-nearest neighbor, logistic regression, naive bayes, the C4.5 algorithm and support vector machines, as well as linear classifiers. This study compared the classification success rates of seven algorithms based on the influence of discrete and continuous characteristics, as well as the size of the dataset, on the AUC. The findings of this research reveal that since the classification process becomes more reliable, it is possible to conduct comparisons between multiple classification methods using a large number of datasets. It is known from this study that Decision Tree and Algorithm C4.5 work efficiently on all datasets. The study strategy, based on the findings, will classify traffic congestion data gathered from Twitter using the C4.5 algorithm. In this research, the congestion class is classified into three subtypes: smooth, dense, and jammed.

2. Literature Review

2.1. Traffic Jam

Cars exceeding the capacity of the allotted traffic lane produce congestion, which results in a bottleneck in traffic flow (road). The occurrence of such incidents is common in major cities, such as Jakarta [8]. There are three things that are often associated with congestion data. To begin, there is the day of the week; normally, the congestion circumstances on Monday are different from those on Tuesday and other days of the week. The second factor is time; the length of the day and night has an impact on the amount of traffic congestion. The third factor is the location and direction of the congestion.

2.2. Data Mining and C4.5 Algorithm

In data mining, also known as knowledge extraction, the process of inputting and analyzing extremely large data sets with the assistance of a computer is called knowledge extraction. It is a unifying interdisciplinary engineering field of machine learning, concept patterns, statistics, databases, visualization, and neural networks classification in one of the Data Mining Techniques [9-11]. Classification and prediction are two forms of data analysis that can be used to extract models that describe important data classes or to predict future data trends [12]. This is not a hotel category label (class), a predictive model of continuous content function [13,14]. Data mining is a process used to find useful information and knowledge, obtained from the data owned [15]. There are many approaches for data mining, including regression and clustering. We may be certain that there are several classes, and the objective is to develop a system for classifying new observations into any of the existing classes [16]. There are several techniques for data mining, including classification, regression, and clustering. Several were canceled, including the following:

- Classification is a statement of the grouping process that occurs inside an attribute and is projected onto a class label.
- In contrast to categorization, regression is a kind of regression that predicts a numerical value, such as 68 degrees Fahrenheit.
- Composting, grouping, categorizing, and grouping items with similar attributes.

Along with data completion, data data, data transformation, and data mining, the following steps of the data mining process will be discussed briefly:

1) Selection of Data

   To process data mining, it is necessary to separate the data collected from a collection of operational data. The chosen data will be saved separately from the prior operating data in multiple distinct files.

2) Cleaning of Data

   Essentially, all data, whether gathered from the experimental database or elsewhere, contains imperfection in the form of missing, inaccurate, or mistake data. As a result, data mining is used to eliminate noise from the data.

3) Transformation of Data

   Numerous data mining approaches need the use of certain data formats throughout the data mining process. Data transformation is the process of transforming data into a format appropriate for data mining.

4) Exploration of Data

   The process of identifying patterns in data using certain methodologies.

5) Evaluation of Patterns

The outcomes of data mining methods are planned patterns and forecasts that are used to conduct an assessment procedure to evaluate if the current target has ceased to exist.

The C4.5 method is a decision tree approach used in classification, in which the data set is converted into a decision tree including the decision rules [17]. This method is used to analyze previously collected data and then build patterns or models. A decision tree is a hierarchical structure made up of two kinds of nodes: leaf nodes and decision nodes. Leaf nodes denote classes, whereas decision nodes denote branching and sub branches for each output based on the derived value [18].

For the purposes of this research, the C4.5 algorithm will be utilized to map traffic congestion based on Twitter data classification. Using the C4.5 algorithm, the purpose of this project is to classify text data pertaining to traffic congestion and visually represent it as a map in order to communicate information about traffic congestion. Each tweet's location data will be geocoded so that it may be linked to a digital map [19,20].

3. Research Model

3.1. System overview

In this study, a data mining classification technique will be applied using the algorithm C4.5. The approach is advantageous for constructing a decision tree that contains a rule or decision rule. This study is described as follows:

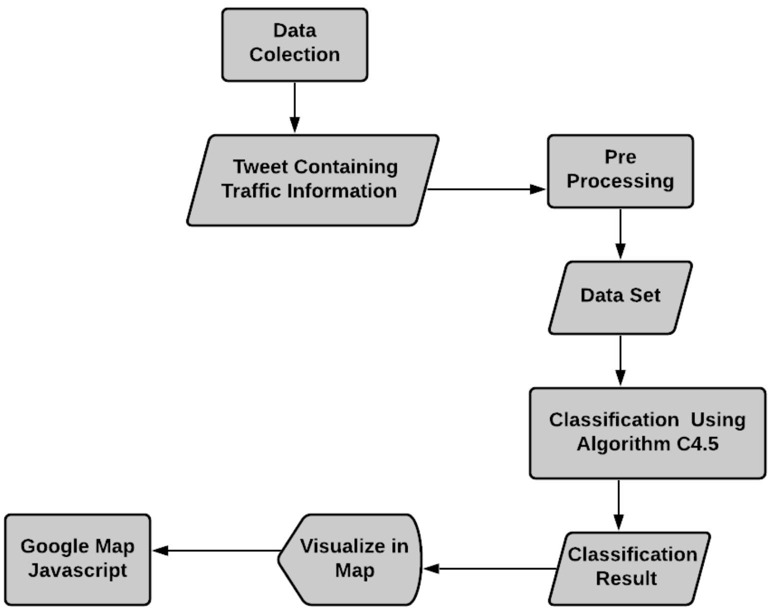

Figure. 1. Research Model

3.2. Twitter data

3.2.1. Twitter Account

The data for this study was gathered from Twitter accounts that offer updates on traffic conditions in Jakarta, including @TMCPoldaMetro. This tweet data will be analyzed and a judgment or rule will be generated. An example of a tweet from the @TMCPoldaMetro account is as follows:

Figure. 2. Example of an @TMCPoldaMetro account tweet

3.2.2. Road Information

Roads are constructed for the purpose of facilitating movement by road and are meant for vehicular traffic. This road also serves as a connector between two additional roads. The roadway is alluded to in this instance. Each highway is identified with a street name. This is the name given to a route in order for it to be identified on a map. Typically, a road has a certain distance and is crossed by cars using regular one-way circumstances referred to as banned conditions and two-way. When the @TMCPoldaMetro account updates traffic conditions, it often includes the direction of traffic, as seen below:

Figure. 3. Sample tweet

3.2.3. Label Condition

In general, the traffic conditions on a roadway may be described as smooth, crowded, or jammed. Flow is the state of a vehicle flowing smoothly and without impediments. Tweets from monitored accounts are often used to communicate traffic conditions and are classed as follows:

1) Smooth: Without impediments
2) Solid: Almost no impediments
3) A traffic gridlock has developed as a result of an obstruction.

Congested traffic circumstances occur when the number of cars on the road is greater than the existing number. Typically, the observed twitter accounts are informed with busy, bustling, packed, or creeping solidworks. While the scenario is one in which the vehicle comes to a complete halt for a little or extended period of time, the observed twitter account often notifies of the traffic jam's status with the words stuck or jammed.

3.2.4. Time

Twitter accounts monitored by transit authorities, such as @TMCPoldaMetro, give real-time information on traffic conditions. Along with information gleaned from tweets, the account maintains CCTV such as @via, which is positioned on numerous highways in Jakarta. We can deduce the date, month, year, and hour of the post from the monitored Twitter account.

3.3. Data Collection

To get tweet data from the monitored account, you must use the official Twitter API. It is an open source library created by Jefferson Henrique, a backend, frontend, and mobile developer, and is available for download at http://jeffersonhenrique.com/ or https://github.com/JeffersonHenrique. This library utilizes the Twitter REST API to gather the following tweets:

1) Establish a connection to the Twitter API using the API Key and Consumer Key that were registered.
2) Retrieve Restricted Tweets from the Twitter API.
3) Captures the following information on observed accounts: id, permalink, username, text, date, retweets, favorites, mentions, hashtags, and geolocation.

The TwitManager will utilize the following search parameters: username, since, up to, and max tweets. By entering the username of the account being monitored, you may choose the date range for setting the old tweet data and the date range for generating the most current tweet data from Twitter. The acquired tweet data is in the CSV format and has not been placed in the database. To ease the subsequent procedure, build a twitter data database and define the properties already included in the CSV, then import the collected CSV data into the tweet data database. The following is a graphical representation of the tweet data database.

Table. 1. Database Implementation

Id	Date	Tweet	Time Range	Day
1	2020/08/03 16:41	16.41 Traffic flow situation from the Tomang Underpass and Jl. Lt. Gen. S Parman headed for the Tangerang Toll Road, the Kebon Jeruk exit was observed to be crowded.	16:01-17:00	Monday
2	2020/06/01 16:28	16.28 Traffic flow situation on Tl. Cempaka Putih this afternoon to Pulo Gadung, Cawang, Senen, and Tj. Priok is observed to be busy.	16:01-17:00	Monday
3	2020/06/11 08:26	08.26 The current situation of the traffic flow at the Old Kodim TL is monitored smoothly.	08:01-09:00	Thursday
4	2020/06/08 20:50	20.50 The traffic flow situation of Pejaten Traffic Light, South Jakarta is monitored smoothly	20:01-21:00	Monday
5	2020/06/08 20:15	20.15 Traffic light situation at Fatmawati South Jakarta Traffic Light is observed to be busy and smooth	20:01-21:00	Monday

3.4. Data Selection

Done in order to extract key terms from the given data. Typically, the classifications below are the keywords used to define the bottleneck, namely:

Table. 2. Keyword Classification

Keywords	Description
Smooth	Without impediments
Solid	Almost no impediments
Jam	With impediments

The data was collected from the monitoring twitter account (@TMCPoldaMetro) between September 1, 2020 and September 30, 2021. According to the statistics, the following information about congestion situations with locations is available:

Table. 3. Data Location Sample

Location
Tanah Abang - Jl. KH Mas Mansyur
Jl. KH Mas Mansyur - Tanah Abang
Rasuna Said - Mampang
Mampang - Rasuna Said
Jl. MT Haryono - Pancoran
Pancoran - Jl. MT Haryono
Kalibata - Pancoran
Pancoran - Kalibata
Dukuh Atas - Bundaran HI
Casablanca - Kp. Malay

As an example, the rationale for selecting an existing place is that it has the highest frequency of occurrence and the location that appears in the account garners attention.

3.5. Pre-Processing

After the data selection procedure is complete, the data preparation stage begins. Preprocessing data is a step in the processing of data. The data to be processed includes day, time, location, and traffic conditions. Preprocessing was performed in four phases in this study: case folding, cleaning, data transformation, and tokenization. The following stages constitute preprocessing:

1) Lowercase the input; any characters other than letters will be removed.
2) Cleaning data from tweets, including removing any words other than the date, time, location, and traffic conditions.
3) Convert text containing hour information to a time range or a date to the appropriate day category.
4) Converts space-separated texts to an array to enable database input.

Table. 4. Preparation of Illustrations

Process Data	Data
Actual Data	17.15 The traffic situation between Jl. KH Mas Mansyur. 06 Desember 2020
Following the Case Folding Process	Jl. KH Mas Mansyur, the traffic situation is still favorable at 17.15. 06 Desember 2020
Following the Cleaning Process	17.15 Jl. KH Mas Mansyur crowded December 06, 2020
Following the Transformation Process	17.15-18.15 Jl. KH Mas Mansyurcrowded smoothly sunday
Following the Tokenization Process	[1] 17.15-18.15 [2] Jl. KH Mas Mansyur [3] crowded smoothly [4] sunday

Following the last run, the data will become a stable dataset. The first array will be inserted into the column "Time" (time), the second array will be inserted into the column "Location", the third array will be inserted into the column "Class", and the fourth array will be inserted into the column "Day", resulting in the following final result of the preprocessing data:

Table. 5. Illustration of final data set

No	Day	Time Range	Location	Class

1	Sunday	17.15-18.15	Jl. KH Mas Mansyur	Solid

3.6. Maps

After all categorization operations are completed, the site will get congestion statistics based on its latitude and longitude. The following is a map of the mapping location in the form of a traffic jam map derived from Twitter traffic statistics.

4. Result

4.1. Using a Map to Show Traffic Congestion

The system was tested and successfully displayed the traffic conditions on the route between points A and B, ranging from smooth to solid to congested.

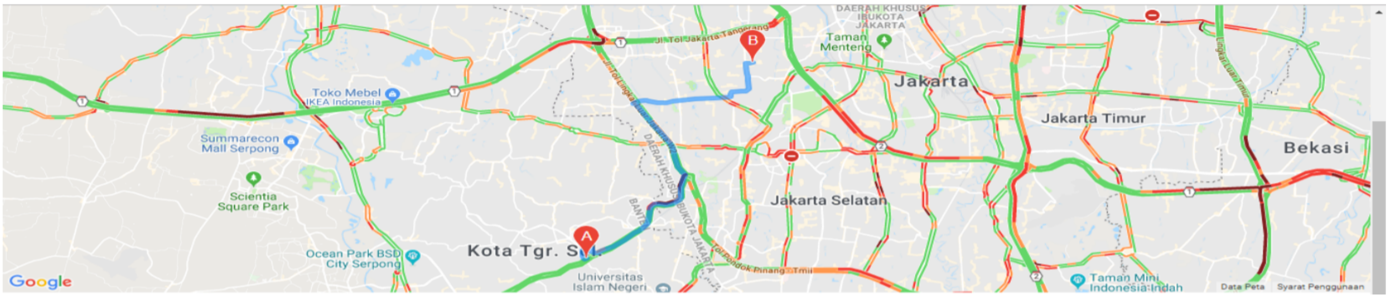

Figure. 4. The web application screenshot displaying the start and finish

The data collected at the Bintaro XChange sub-district site on Saturday between 16:01 and 17:00 a.m. indicate red traffic conditions.

4.2. Classification Using C4.5 Algorithm Analysis

4.2.1. Data preparation

The next step will be to do preprocessing on the 1482 tweets. Performance testing data is divided into two categories: training data and test data. Training data is the data that is used to teach the subject. To calculate the entropy and maximum gain values for each characteristic, the C4.5 training data approach will be utilized in conjunction with the data. Testing data will be used to calculate the entropy and maximum gain values for each attribute.

4.2.2. Process Testing

This exam is intended to assess the performance of a C4.5 Algorithm, which is used to categorize data into predetermined categories, in a numerical setting. The test gives training data that will be used to create a data division table in the future. By supplying output values for accuracy, precision, and recall calculations, performance is improved. This research will repeat the testing procedure five times, each time using a different subset of the test results. In the first section, 50% of the test data is used for training, while 50% of the 1482 outcome data sets are used for testing.

Table. 6. Performance Results of Training Data 50%

Status Data (1482 Data)	Training Data (742 Data)	Testing Data (740 Data)
Smooth (128 Data)	64	64
Solid (1119 Data)	560	559
Jam (235 Data)	118	117

Smooth data is currently 128 data, Solid data is 1119 data, and Jam data is 235 data, based on the precision value of 99.82%, recall value of 99.15%, and accuracy value of 99.59%.

4.2.3. Test Results Summary

Table. 7. Calculation of Max Gain on Algorithm C4.5

No	Attribute Gain Ratio Max	Attribute	Value Attribute	Total Case	Total Case Smooth	Total Case Solid	Total Case Jam	Entropy	Informati on Gain	Split Info	Gain Ratio
1	Smooth	Total	Total	742	65	564	113	0.511			0
2	Smooth	Smooth	no	677	0	564	113	0.3253	0.2142	0.4228	0.5
3	Smooth	Smooth	yes	65	65	0	0	0	0.2142	0.4228	0.5
4	Smooth	Solid	no	179	65	1	113	0.4957	0.3914	0.7971	0.491
5	Smooth	Solid	yes	563	0	563	0	0	0.3914	0.7971	0.491
6	Smooth	Jam	no	630	65	564	1	0.2479	0.3005	0.6122	0.4909
7	Smooth	Jam	yes	112	0	1	112	0	0.3005	0.6112	0.4909
1	Jam	Total	no	677	0	563	113	0.3253			0
2	Jam	Solid	yes	114	0	564	113	0.0363	0.3192	0.654	0.4881
3	Jam	Solid	no	563	0	0	0	0	0.3192	0.654	0.4881

4	Jam	Jam	yes	565	0	564	1	0.0094	0.3715	0.6472	0.4906
5	Jam	Jam	no	112	0	0	112	0	0.3715	0.6472	0.4906
1	Solid	Total	yes	565	0	564	1	0.0094			0
2	Solid	Solid	no	2	0	1	1	0.5	0.0076	0.0339	0.2242
3	Solid	Solid	yes	563	0	563	0	0	0.0076	0.0339	0.2242

It is created for the first iteration of the Smooth condition because the Smooth attribute has the highest gain value, followed by the second iteration of the Jam condition because the Jam attribute has the second highest gain value after Smooth, and finally for the Solid condition because the Solid attribute has the second highest gain value after Smooth.

Table. 8. Summary of each test

Testing	**Data Partison (%)**		**Tabel Confusion Matrix**		**Predicted Class**		
	Training Data	**Testing Data**			**Smooth**	**Solid**	**Jam**
1st	50	50	**Actual Class**	Smooth	64	0	0
				Solid	0	559	0
				Jam	0	3	115
2nd	45	55		Smooth	70	0	0
				Solid	0	61.5	0
				Jam	0	5	125
3th	40	60		Smooth	77	0	0

			Solid	0	671	0
			Jam	0	9	133
4th	35	65	Smooth	83	0	0
			Solid	0	727	0
			Jam	1	12	140
5th	30	70	Smooth	90	0	0
			Solid	0	783	0
			Jam	1	12	152

Table. 9. Summary of performance system

Trial	**Data Partition**		**Accuracy (%)**	**Precision (%)**	**Recall (%)**
	Training Data (%)	**Testing Data (%)**			
1st	50	50	99.60	99.82	99.15
2nd	45	55	99.39	99.73	96.72
3rd	40	60	98.99	99.56	97.56
4th	35	65	98.65	99.06	97.17
5th	30	70	98.75	99.13	97.37
Average (%)			**99.08**	**99.46**	**97.99**

4.2.4. Result Analysis

The precision, recall, and accuracy of each test were determined using the test results. 99.46 %, 97.99 %, and 99.08 %, respectively. All trials had a mean value of 99.82%, whereas all trials had a mean value of 99.06%. All trials have a recall rate of 99.15%, whereas all trials have a recall rate of 97.17%. The total value of all experiments is 99.60%, whereas the total value of all experiments is 98.65%. For the sake of graphical presentation, recall and accuracy values for each trial are plotted graphically as shown in Figure 5.

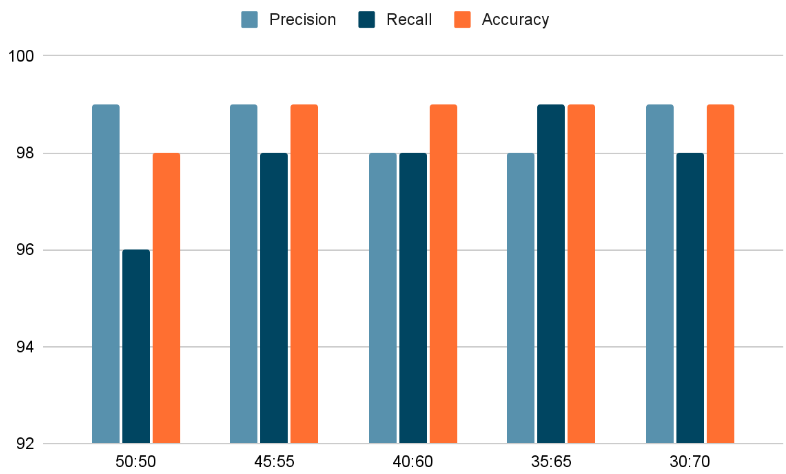

Figure. 5. Graph of performance evaluation

The x-axis in Figure 5 represents the data partitioning between training and testing data, while the y-axis represents the output created during the precision, recall, and accuracy calculations.

5. Conclusion

In order to construct a model that can predict the location and timing of congestion based on the findings of this research, various phases are required, including data gathering from Twitter, data clustering, data pre-processing, and data categorization, among others. Compared to other applications, the classification method of C4.5 has the highest level of accuracy (99.08%). According to comparisons with comparable research to identify congestion on Twitter, the average findings with the greatest accuracy are 96.12% on average. By incorporating modeling into our congestion detection program, customers may get real-time information on the location and timing of the congestion. When testing apps and comparing them, testing is done 25 times, resulting in 17 times the same circumstances that caused the jamming and eight times the same conditions that did not cause the jamming In this case, there is a drawback in that modeling will not function if there is no tweet to use as an example.

Furthermore, there are still many abbreviated words in text information, which can be improved by using spelling correction algorithms in pre-processing. To improve the accuracy of text information, images and videos in Twitter posts can be used in conjunction with information from other social media platforms like Facebook and Instagram, as well as information from social media platforms that are currently trending.

References

[1] M. I. Zulfa, R. Hartanto, and A. E. Permanasari, “Caching strategy for Web application – a systematic literature review,” Int. J. Web Inf. Syst., vol. 16, no. 5, pp. 545–569, Jan. 2020, doi: 10.1108/IJWIS-06-2020-0032.

[2] S. Yamamoto, K. Wakabayashi, T. Satoh, Y. Nozaki, and N. Kando, "Twitter user growth analysis based on diversities in posting activities," Int. J. Web Inf. Syst., vol. 13, no. 4, pp. 370–386, Jan. 2017, doi: 10.1108/IJWIS-04-2017-0034.

[3] Y. Wang and J. Zhang, "Exploring topics related to data mining on Wikipedia," Electron. Libr., vol. 35, no. 4, pp. 667–688, Jan. 2017, doi: 10.1108/EL-09-2016-0188.

[4] R. Varma, Y. Verma, P. Vijayvargiya, and P. P. Churi, "A systematic survey on deep learning and machine learning approaches of fake news detection in the pre- and post-COVID-19 pandemic," Int. J. Intell. Comput. Cybern., vol. 14, no. 4, pp. 617–646, Jan. 2021, doi: 10.1108/IJICC-04-2021-0069.

[5] D. Rossmann and S. W. H. Young, "Social media optimization: making library content shareable and engaging," Libr. Hi Tech, vol. 33, no. 4, pp. 526–544, Jan. 2015, doi: 10.1108/LHT-05-2015-0053.

[6] D. Ramachandran and P. Ramasubramanian, "Event detection from Twitter – a survey," Int. J. Web Inf. Syst., vol. 14, no. 3, pp. 262–280, Jan. 2018, doi: 10.1108/IJWIS-11-2017-0075.

[7] S. Okazaki, A. M. Díaz-Martín, M. Rozano, and H. D. Menéndez-Benito, "Using Twitter to engage with customers: a data mining approach," Internet Res., vol. 25, no. 3, pp. 416–434, Jan. 2015, doi: 10.1108/IntR-11-2013-0249.

[8] K. Nagaraj, B. Bhattacharjee, A. Sridhar, and S. GS, "Detection of phishing websites using a novel twofold ensemble model," J. Syst. Inf. Technol., vol. 20, no. 3, pp. 321–357, Jan. 2018, doi: 10.1108/JSIT-09-2017-0074.

[9] A. Kumar, M. A. Kaviani, E. Bottani, M. K. Dash, and E. K. Zavadskas, "Investigating the role of social media in polio prevention in India: a Delphi-DEMATEL approach," Kybernetes, vol. 47, no. 5, pp. 1053–1072, Jan. 2018, doi: 10.1108/K-06-2017-0210.

[10] C. Kaiser and F. Bodendorf, "Mining consumer dialog in online forums," Internet Res., vol. 22, no. 3, pp. 275–297, Jan. 2012, doi: 10.1108/10662241211235653.

[11] S. Joo, N. Choi, and T. H. Baek, "Library marketing via social media," Online Inf. Rev., vol. 42, no. 6, pp. 940–955, Jan. 2018, doi: 10.1108/OIR-10-2017-0288.

[12] F. A. Johannsen, "Process-related value propositions of enterprise social media use for the external communication with end consumers," Bus. Process Manag. J., vol. 24, no. 1, pp. 183–215, Jan. 2018, doi: 10.1108/BPMJ-03-2016-0049.

[13] G. Ignatow, N. Evangelopoulos, and K. Zougris, "Sentiment Analysis of Polarizing Topics in Social Media: News Site Readers' Comments on the Trayvon Martin Controversy," in Communication and Information Technologies Annual, vol. 11, Emerald Group Publishing Limited, 2016, pp. 259–284.

[14] N. Ferdinand and J. Tucker, "Digital Engagement Conference: A Road Map for the Future of Festivals, July 9‑11 2014, Bournemouth University, Talbot Campus, Poole, Dorset, United Kingdom," J. Tour. Futur., vol. 1, no. 2, pp. 162–166, Jan. 2015, doi: 10.1108/JTF-12-2014-0012.

[15] M. Dai, W. He, X. Tian, A. Giraldi, and F. Gu, "Working with communities on social media," Online Inf. Rev., vol. 41, no. 6, pp. 782–796, Jan. 2017, doi: 10.1108/OIR-01-2016-0002.

[16] N. Choi and S. Joo, "Understanding public libraries' challenges, motivators, and perceptions toward the use of social media for marketing," Libr. Hi Tech, vol. 39, no. 2, pp. 352–367, Jan. 2021, doi: 10.1108/LHT-11-2017-0237.

[17] T. Cawsey and J. Rowley, "Social media brand building strategies in B2B companies," Mark. Intell. Plan., vol. 34, no. 6, pp. 754–776, Jan. 2016, doi: 10.1108/MIP-04-2015-0079.

[18] [J. Bhattacharjya, A. Ellison, and S. Tripathi, "An exploration of logistics-related customer service provision on Twitter," Int. J. Phys. Distrib. Logist. Manag., vol. 46, no. 6/7, pp. 659–680, Jan. 2016, doi: 10.1108/IJPDLM-01-2015-0007.

[19] A. Al-Rawi, J. Groshek, and L. Zhang, "What the fake? Assessing the extent of networked political spamming and bots in the propagation of #fakenews on Twitter," Online Inf. Rev., vol. 43, no. 1, pp. 53–71, Jan. 2019, doi: 10.1108/OIR-02-2018-0065.

[20] B. Abu-Salih, P. Wongthongtham, and C. Yan Kit, "Twitter mining for ontology-based domain discovery incorporating machine learning," J. Knowl. Manag., vol. 22, no. 5, pp. 949–981, Jan. 2018, doi: 10.1108/JKM-11-2016-0489.