

Literature Study: Highway Traffic Management with Sentiment Analysis and Data Mining

Nurul Khairina^{1*}, Muhammad Khoiruddin Harahap²

¹Universitas Medan Area, Indonesia, ²Politeknik Ganesha Medan, Indonesia

¹nurulkhairina27@gmail.com, ²choir.harahap@yahoo.com]



*Corresponding Author

Article History:

Submitted: xx

Accepted: xx

Published: xx

Keywords:

Highway; Traffic; Management;
Sentiment Analysis; Data Mining

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

[In today's era, technology is growing rapidly, many of the latest technologies are in great demand by the Indonesian people, one of which is social media. Various social media such as Facebook, Twitter, Instagram, have become very popular applications for various ages, including teenagers, adults, and the elderly. Social media has a positive impact that can help people convey the latest information through posts on their respective accounts. Social media can disseminate information in a short time, this is why social media is an interesting application to research. The problem of road traffic congestion is strongly influenced by the number of vehicles that pass every day. A large number of private vehicles and public vehicles that pass greatly confuses the atmosphere of highway traffic. Congestion often occurs during working hours. Road congestion also often occurs when an unwanted incident occurs. Sentiment analysis algorithms and data mining algorithms can be combined to find information on traffic jams through social media such as Facebook, Twitter, Instagram, and other social media. The results show that sentiment analysis methods and data mining algorithms can be used to find information about current traffic jams through social media. The conclusion from this literature study can be seen that the K-Nearest Neighbor data mining algorithm is the best choice to overcome road traffic congestion, which will then be further developed in the form of highway traffic management modeling.]

INTRODUCTION

[Medan city is one of the largest provincial capitals in Indonesia. Medan City is a metropolitan city that has a fairly large population and reaches 2 million inhabitants per the year 2015 [1]. The number of private vehicles and public transportation often triggers traffic jams on the highway during working hours. According to Kartika Ratnasari, Medan is the 6th most congested city in Indonesia. The length of traffic jams in Medan in a year reaches 42 hours with a percentage of 20%. The percentage of time drivers are stuck in traffic jams reaches 25% during working hours and 23% outside working hours [2].

Natural disasters such as earthquakes, landslides, strong winds, tsunamis, and so on can also disrupt the flow of highway traffic. In addition, events such as demonstrations, student brawls, going home on National holidays, road repairs, Covid-19 prevention checks, zebra operations, traffic light damage, traffic accidents, and so on can also cause long traffic jams and take up time for drivers. According to Rozie Winata, the number of traffic accidents in 2019 has increased in 2018. Throughout 2019 there were 6,100 cases, while previously in 2018 there were only 5,990 cases. In addition, Kapolda also said that material losses due to traffic accidents in 2019 decreased, which was only Rp. 12,625,000,000, while in 2018 it previously reached Rp. 13,696,830,000 [3].

Minister of Transportation Budi Karya Sumadi said that traffic jams in urban areas had a major impact on the Indonesian economy. One of them inhibits economic growth[4]. In 2016, SITRAMP (Study on Integrated Transportation Master Plan) also estimated that losses due to congestion could reach Rp 12.8 trillion per year[5]. Until now, the Medan City Government continues to strive to overcome road traffic congestion, one of which is by building toll roads and flyovers[6]. In this study, the researcher plans to conduct a literature review of the sentiment analysis algorithm and data mining algorithm to find alternative solutions for handling traffic jams in terms of the field of informatics engineering which will also involve social media [7][8]. Several theories will be analyzed and tested to help the government overcome road traffic congestion.]

LITERATURE REVIEW

[As for previous research related to this research, Arvian Zanuardi & Hitapriya Suprayitno researched by applying Knowledge Discovery to the database to analyze the characteristics of traffic accidents in Surabaya [9]. Zuhri Muhs and Wahyu Herijanto conducted a study that focused on traffic management due to trams on the Surabaya highway with the

theory of Traffic Demand Management [10]. Laif Satoinong et al conducted a study that focused on performance analysis and traffic management at the ITS Roundabout and the Mulyosari Roundabout in Surabaya using the MKJI theory [11]. Popy Purnamasari Wahid Suyitno et al conducted a study by detecting traffic congestion in Jakarta with the Data Mining C4.5 algorithm and offering solutions to overcome traffic congestion [12]. Putu Amelia Vennanda Widyaswari et al conducted a study on the classification of traffic congestion in the city of Malang using the Neighbor Weighted K-Nearest Neighbor method [13]. Siti Nurwahyuni researched the form of sentiment analysis on the KRL Access Online Transportation Application using the Naïve Bayes method [14]. Myrna Ermawati and Joko Lianto Buliali conducted a study using a text-based approach to find out similar traffic incidents through social media Twitter[15].

METHOD

[This literature review study was conducted to determine whether sentiment analysis and data mining algorithms can prevent road traffic congestion through road traffic management. The literature review was conducted on three studies [13] [15] [16] by analyzing the algorithms and cases to be solved as can be seen in Table 1. below:

Tabel 1. Review Paper]

Author	Title	Algorithm	Case
Putu Amelia Vennanda Widyaswari, Indriati, Rizal Setya Perdana	Malang City Traffic Congestion Classification Through Twitter Media Using the Neighbor Weighted K-Nearest Neighbor (NW-KNN) Method	Neighbor Weighted K-Nearest Neighbor (NW-KNN)	Traffic congestion
Myrna Ermawati, Joko Lianto Buliali	Text Based Approach For Similar Traffic Incident Detection from Twitter	Text Based Approach	Traffic Incident
Amani AlFarasani, Tahani AlHarthi, Sarah AlHumoud	ATAM: Arabic Traffic Analysis Model for Twitter	Sentiment Analysis, support vector machines	Road Incident Prediction

RESULT

[The results of the literature review carried out by researchers can be in the form of comparative analysis of research results, research gaps, and the advantages and disadvantages of the three studies [13][15][16] that have been described in Table 1. The results of the literature review can be seen in Table 2 below:

Table 2
Comparative Analysis of Three Research on Sentiment Analysis and Data Mining

Algorithm	Object of Research	Research Result
Neighbor Weighted K-Nearest Neighbor (NW-KNN)	Traffic congestion	The average value of precision is 0.7336507, recall is 0.2210526, and accuracy is 0.665
Text Based	Traffic	Best average f-measure

Approach	Incident	70% was achieved by retrieval system that tested using jaccard coefficient
Sentiment Analysis, Support Vector Machines	Road Incident Prediction	The overall accuracy was 82% for all four classes

Table 3
 Research Gap from Three Research on Sentiment Analysis and Data Mining

Algorithm	Object of Research	Research Gap
Neighbor Weighted K-Nearest Neighbor (NW-KNN)	Traffic congestion	Optimizing the performance of the Neighbor Weighted K-Nearest Neighbor (NW-KNN) algorithm to overcome traffic jams
Text Based Approach	Traffic Incident	Optimizing traffic incident detection with semantic analysis
Sentiment Analysis, Support Vector Machines	Road Incident Prediction	Build a web service for live streaming and classifications.

After determining the results from the three articles [13][15][16], their strengths and weaknesses were reviewed, as shown in Table 4.

Table 4
 Strengths and Weaknesses from Three Research on Sentiment Analysis and Data Mining

Title	Strength	Weakness
Malang City Traffic Congestion Classification Through Twitter Media Using the Neighbor Weighted K-Nearest Neighbor (NW-KNN) Method	Able to detect congestion by providing information in the form of time, place, road congestion conditions, and weather conditions	In some trials there is still low accuracy, this is due to a large number of test data with incorrect conditions, such as test data that should be in the crash category are included in the non-jammed category.
Text Based Approach For Similar Traffic Incident Detection from Twitter	Able to filter sentences without symbols/punctuation from Twitter, can extract hashtags, conditions, locations, and objects	Unable to distinguish the presence of different words with the same meaning in a traffic incident, for example, the terms 'consecutive collision' and 'accident' streak'.

ATAM: Arabic Traffic Analysis Model for Twitter	Able to detect traffic incidents in the form of safe conditions, need attention, danger, and neutral	Still need Arabic speakers to measure the accuracy of word classification
---	--	---

DISCUSSION

[After conducting a literature study on sentiment analysis and data mining algorithms in the three studies listed in Table 2, Table 3, and Table 4, the researcher has a further research plan to build a highway traffic management model, as detailed in Figure 1 below:

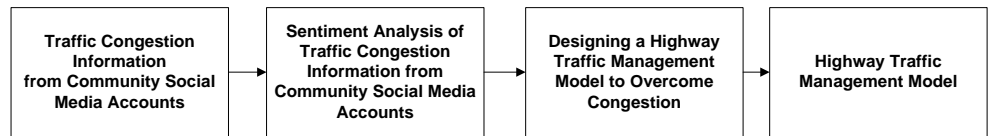


Figure 1. Planning for Further Research

CONCLUSION

[The literature study is an initial study to explore how sentiment analysis and data mining algorithms work, especially in traffic management, which will be useful for overcoming traffic congestion. After the researchers analyzed the three studies in Table 2, Table 3, and Table 4, it can be seen that the sentiment analysis method and data mining algorithm can be used to detect traffic congestion through social media. From the literature study, it can also be seen that the right data mining algorithm to use is K-Nearest Neighbor. For further research, researchers will build a traffic management model to avoid congestion, this is based on the results of literature studies on sentiment analysis methods and data mining algorithms.]

REFERENCES

1. P. Medan, "Kependudukan," 13 November 2013. [Online]. Available: <https://pemkomedan.go.id/hal-kependudukan.html>.
2. K. Ratnasari, "Daftar Kota Termacet di Indonesia," 13 August 2020. [Online]. Available: <https://artikel.rumah123.com/daftar-kota-termacet-di-indonesia-ternyata-bukan-jakarta-yang-terparah-54016>.
3. R. Winata, "Kasus Laka Lantas di Kota Medan," 6 January 2020. [Online]. Available: https://medanbisnisdaily.com/news/online/read/2020/01/06/97415/kasus_dan_korban_laka_lantas_di_sumut_meningkat/.
4. Y. A. Uly, "Kemacetan Menghambat Pertumbuhan," 5 August 2020. [Online]. Available: <https://money.kompas.com/read/2020/08/05/150600726/menhub--kemacetan-di-perkotaan-hambat-pertumbuhan-ekonomi->.
5. D. Subinarto, "Medan, Maet, dan Resiko Ekonomi," 13 October 2016. [Online]. Available: <http://www.umm.ac.id/en/opini/medan-macet-dan-resiko-ekonomi.html>.
6. W. A. Siregar, "Groundbreaking Tol Dalam Kota Medan Dimulai Februari 2020," 15 January 2020. [Online]. Available: <https://economy.okezone.com/read/2020/01/15/320/2153232/groundbreaking-tol-dalam-kota-medan-dimulai-februari-2020?page=1>.
7. Y. Tao, F. Zhang, C. Shi and Y. Chen, "Social Media Data-Based Sentiment Analysis of Tourists' Air Quality Perceptions," *Sustainability*, vol. 11, no. 5070, pp. 1-23, 2019.
8. A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys*, vol. 49, no. 2, pp. 28-41, 2016.

-
9. A. Zanuardi and H. Suprayitno, "Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database," *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, vol. 2, no. 1, pp. 45-55, 2018.
 10. Z. Muhs and W. Herijanto, "Manajemen Lalu Lintas Akibat Trem Di Jalan Raya Darmo Surabaya," *Jurnal Teknik Pomits*, vol. 3, no. 1, pp. 37-42, 2014.
 11. L. Satoinong, Mardijono, M. Donny, N. Ray and L. S. B. Wibowo, "Analisis Kinerja Dan Manajemen Lalu Lintas pada Bundaran ITS dan Bundaran Mulyosari Kota Surabaya," *Jurnal Perencanaan dan Rekayasa Sipil* , vol. 2, no. 1, pp. 16-22, 2019.
 12. P. P. W. Suyitno, R. E. Indrajit and M. Fauzi, "Penerapan Data Mining dalam Menangani Kemacetan di Jakarta," *Ikraith-Informatika*, vol. 1, no. 2, pp. 53-60, 2017.
 13. P. A. V. Widyaswari, Indriati and R. S. Perdana, "Klasifikasi Kemacetan Lalu Lintas Kota Malang Melalui Media Twitter Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NW-KNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* , vol. 3, no. 2, pp. 1464-1470, 2019.
 14. S. Nurwahyuni, "Analisis Sentiment Aplikasi Transportasi Online KRL Access Menggunakan Metode Naive Bayes," *Jurnal Swabumi*, vol. 7, no. 1, pp. 31-38, 2019.
 15. M. Ermawati and J. L. Buliali, "Text Based Approach For Similar Traffic Incident Detection from Twitter," *Lontar Komputer*, vol. 9, no. 2, pp. 63-71, 2018.
 16. A. AlFarasani, T. AlHarthi and S. AlHumoud, "ATAM: Arabic Traffic Analysis Model for Twitter," *International Journal of Advanced Computer Science and Applications (IJACSA)* , vol. 10, no. 3, pp. 328-336, 2019.

↓