2023 IEEE 26th International Conference on
Intelligent Transportation Systems (ITSC)
24-28 September 2023. Bilbao, Bizkaia, Spain

Enhancing decision-making in transportation management: A comparative study of text classification models

Eduardo Carneiro[1], Tânia Fontes[1], Rosaldo J. F. Rossetti[2] and Zafeiris Kokkinogenis[2]

2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3503-9946-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITSC57777.2023.10422056

***Abstract*— Machine learning algorithms offer the capability to analyze large volumes of real-time data, providing transport authorities with valuable insights into traffic conditions, congestion hotspots, and incident detection from diverse data sources. However, these algorithms face challenges related to data quality and reliability. We conducted a comparative analysis of machine-learning models that can be used to identify and filter transportation content from social media or other sources that can provide small and concise text. The filtrated result can then feed models and/or tools used to improve and automate traffic control, operational management, and tactical management decision-making. We consider factors such as run time, generalization capacity, and performance metrics as criteria to assess their suitability for different decision levels. The analysis is supported by a dataset consisting of Twitter content. The predictions from three groups of algorithms are evaluated: traditional machine learning algorithms (Support Vector Machines, Logistic Regression, and Random Forest), a fine-tuned Google BERT model, and Google BERT models without training (BERT-base and BERT-large). The tests are performed using New York, London, and Melbourne data. The findings of this research aim to assist decision-makers in making informed choices when selecting the most appropriate method to filtrate information subsequently used for models that contribute to different traffic management tasks.**

I. INTRODUCTION

Social media has rapidly emerged as a dominant communication and real-time information dissemination method in the digital age. Its widespread usage has opened up avenues for applications ranging from financial markets to transportation. The latter is a vital aspect of modern society, with millions of people commuting daily. In the United States, for instance, there were around 225.8 million licensed drivers in 2021, and approximately 86% of the population used private vehicles as their primary mode of transportation [1]. With such a large population on the move, it is essential to have access to real-time transportation information in order to make well-informed travel decisions.

Multiple public transportation companies already use social media to communicate with their passengers and communities [2]. The authors of these studies believe that this is key to improving the quality of their services. To the best of our knowledge, [3] was the first work to propose Twitter as an *artificial traffic sensor*, where tweets are used to sense traffic-related events. [4] suggests a social media processing pipeline focusing on subjective user opinions of mobility networks and public transportation services. However, the availability of real-time transportation information on social media has grown substantially in recent years, with users sharing information about traffic congestion, public transportation delays, accidents, and road closures, among other issues. The majority of this social media content is not directly relevant to transportation and therefore needs to be filtered out [5].

Several approaches have been explored and tested to harness the potential of social media content for evaluating transportation networks. However, neglecting the filtering process can impact the accuracy of results, particularly when conducting sentiment analysis or topic modeling. Furthermore, the absence of filtering can significantly increase the execution time of the algorithms, which is particularly critical for time-sensitive tasks.

This study focuses on the initial phase commonly encountered in social media analysis approaches, known as information filtering, which involves the extraction of the relevant content to the specific problem at hand without needing to train the models. The main objective is to explore the application of different information filtering techniques, specifically, text classification methods coupled with the exclusion of irrelevant content, to analyze social media data effectively. The study seeks to enhance the understanding of how these models perform, across various dimensions and decision-making scenarios, in extracting relevant insights from social media, enabling better decision-making in transportation network management. Key factors such as performance metrics and execution time will be considered to assess the effectiveness of the models.

The remainder of this manuscript is structured as follows. First, a literature review is performed, identifying the most common ways authors extract or filter transportation-related content. Then, the data used in this study is presented, followed by the applied models, with descriptions of the implementation done for each one. Next, the main results are outlined and discussed, with a reflection being made regarding which model might be the best for each one of the three tasks in the study. The document ends with remarks on the main conclusions and an overview of future directions.

II. LITERATURE REVIEW

Text mining techniques have been widely applied in the transportation domain [6] In the context of text mining, information filtering assumes a pivotal role in streamlining

[1]E. Carneiro and T. Fontes are with INESC TEC, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal ({eduardo.d.carneiro, tania.d.fontes}@inesctec.pt). Correspondence should be addressed to Tânia Fontes.

[2]R. Rossetti and Z. Kokkinogenis are with LIACC, Department of Informatics Engineering, Faculty of Engineering, University of Porto, Porto, Portugal ({rossetti, kokkinogenis}@fe.up.pt).

the analysis by focusing on the most relevant data while mitigating noise and irrelevant information.

The process of information filtering can be approached through diverse techniques, each one with its own strengths and weaknesses. While in some works the filtration process is conducted through search-based methods, which can be further categorized into filtration based on search parameters or filtration based on users or pages (e.g. [7]), in others it is done by text classification (e.g. [8]).

The filtration by search parameters is done during the extraction phase and consists of selecting content containing one or more entries from a list of words and/or hashtags [9], [10]. The number of words can vary, depending on the level of precision required. While Cebeci et al. [11] only use one word, Anastasia et al. [12] uses four. These words or hashtags can be general transportation words, such as "bus", "street" or "crash" [7] or more specific ones, usually related to public transportation companies [13] and/or private ones [14]. Points of interest points, like road names, zones [15] or bus stops/lines [16], [17] are also used.

Filtration by users or pages can be done in two distinct ways. The first consists in identifying accounts of interest and extracting all their content [18]. These accounts of interest can be public transportation companies or users/entities that post daily information about traffic, like congestion levels. The other option is to identify these accounts and extract the comments made on their posts or tag them [19]. Candelieri et al. [20] use both options for the account "@atm_informa" from Milan.

As mentioned earlier, filtration can also be achieved through the process of filtering. Text classification can be defined as a natural language processing (NLP) task that involves categorizing or assigning predefined labels or classes to textual data. In some studies [21]authors use topic modeling to group content according to multiple subjects being discussed after the initial filtering, done either by search or machine learning. Chen et al. [14] took a different approach and after doing initial filtering by search, they created multiple text classifiers based on Google BERT, to understand if the content was related to a particular travel mode or not (e.g. subway related). To the best of our knowledge, first attempts of using Google BERT is reported by Osório et al. [19] and Murços et al. [22].

Numerous studies have addressed the need for information filtering, specifically focusing on content sourced from various social media platforms, including Twitter [19], Facebook [21], Instagram [23], Flickr [23] and Weibo [24]. Although English is the most common language authors deal with, other languages can also be found [25], [26].

III. Data and Methods

Social media content was used to study the results and performances of different text classifiers when used to identify transportation-related text. In the following sections, we present the data employed for building and testing the models, followed by a description of the preprocessing steps taken. Furthermore, we delve into the models utilized during this study, along with a detailed explanation of the chosen evaluation methodology.

A. Data

With over 500 million tweets being written daily [1], Twitter provides a vast amount of data that can be leveraged for real-time evaluations. Hence, in this study, messages from this social network were used to assess a classification model.

A tweet is a concise message that allows users to express their thoughts and opinions, share news, or provide any other type of information within a 280-character limit. These private or public messages can contain text, photos, videos, links, and hashtags to enhance engagement and discoverability. Private tweets are restricted to the user's followers, while public tweets can be viewed, shared (retweeted), liked, or replied to by anyone. About 1% of the extracted tweets contain geolocation information, which can have coordinates or a location name.

The character limit of 280 characters for each tweet encourages users to be concise and creative in their expressions. As a result, many users utilize Twitter as a platform to chronicle their daily experiences, discussing a wide range of topics throughout the day.

In this study, data from three major cities English-speaking cities were used: Melbourne, London, and New York (Table I). The data used was collected between the May 16^{th} and July 6^{th} of 2017 [8].

TABLE I
Data characterization

		New York	Melbourne	London
DOMAIN				
- Area (km^2)		738.4	9992	1572
- Main public transport		S, B	T, TR, B	S, B, T
POPULATION (in 2020)				
- Population (M)		8.3	5.8	8.8
- Population (people.km^2)		10,194	453	5,598
- Main language		EN	EN	EN
- Most spoken languages		EN, SP, CH	EN, CH, C	EN, B, P
- Average people's age		36.9	36	35.9
- Social media networks		T, F, I	F, I, TT	F, I, T
MESSAGES				
Bounding box	South-West	-74.255641 40.495865	144.593742 -38.433859	-0.510365 51.286702
	North-East	-73.699793 40.91533	45.512529 -37.511274	0.334043 51.691824
- Messages available (M)		9.2	0.84	5.8

Languages: B: Bengali, C: Cantonese, CH: Chinese, EN: English, P: Polish, SP: Spanish.
Public Transports: B: Bus, S: Subway, T: Train, TR: Tram.
Social Networks: F: Facebook, I: Instagram, TT: TikTok, T: Twitter.

B. Pre-processing

The preprocessing of messages for classification involved several steps to transform the raw text into a format suitable for input into a model, in particular:

1) Replacing: This step consisted in substituting contractions and numbers by their integer occurrences.

[1]Twitter Usage Statistics, www.internetlivestats.com/twitter-statistics, Accessed: 2022-05-04

2) Cleaning: This step involved the elimination of unwanted or defective data from the text. This included removing HTML tags, special characters, hashtags, URLs, non-ASCII characters, user mentions, stop words, and punctuation irrelevant to the classification task.
3) Normalizing: This step aimed to standardize the text by converting it to a consistent format, which included converting all text to lowercase and removing extra whitespace.
4) Lemmatization: This step grouped together the inflected forms of a word (only performed for the verbs).

By following these preprocessing steps, the text is transformed into a representation that can be effectively classified. This enables the models to understand and extract meaningful features from the text data, leading to accurate classification results.

C. Models

Predictions were considered from three main groups of algorithms: (i) traditional algorithms; (ii) fine-tuned Google BERT models; and (iii) Google BERT models without training involved.

Traditional algorithms consist of well-established supervised machine-learning algorithms commonly used for classification tasks. In the context of the study, the traditional algorithms include Support Vector Machines (SVM), Logistic Regression (LR), and Random Forest (RF). SVM finds an optimal hyperplane that separates data points of different classes in the feature space. LR is a statistical algorithm used for binary or multi-class classification. LR can also be used for unsupervised learning tasks, but it is primarily utilized for supervised learning. It models the relationship between the independent variables and the probability of a certain class using the logistic function. RF is an ensemble learning algorithm that combines multiple decision trees to make predictions. Each tree is trained on a random subset of the training data, and the final prediction is made by aggregating the predictions of individual trees. The three models were trained and tested using a combination of bag-of-words and paragraph2vec embeddings [8].

In the fine-tuned Google BERT models, the last layers of the models can be retrained and employed to classify the selected content based on the specific dataset used for training, pertaining to the problem at hand. This conventional method of utilizing Google BERT has been extensively explored, but its application in transportation text classification problems remains relatively unexplored. However, the use of Google BERT models without additional training is relatively uncommon when addressing these types of problems. This approach involves leveraging the embeddings obtained from Google BERT and working with dictionaries associated with each class to develop a classification model that does not necessitate further training of the pre-trained Google BERT model and classifies the content according to the Cosine Similarity. Table II presents the dictionary of transport-related words used [8], [22]. For the fine-tuning task, the chosen model was the Google BERT Base, and for the classification without additional training Google BERT Large and also Google BERT Base.

TABLE II
DICTIONARY OF TRANSPORTATION-RELATED WORDS

accident	bus	highway	street	truck
avenue	buses	metro	streets	trucks
bicycle	cab	moto	subway	van
bicycles	car	motorcycle	taxi	vans
bike	cars	motorcycles	taxis	walk
bikes	driver	road	traffic	uber
boulevard	drivers	station	train	underground

D. Evaluation

The model was evaluated using two distinct datasets. These datasets were manually annotated in previous research [22]. The first dataset (Experience 1) consisted of 1,000 tweets specifically collected for the city of New York. These tweets were categorized to provide a ground truth for evaluating the model's performance on New York City-specific data. The second dataset (Experience 2) included 3,000 classified tweets collected from three cities: New York, London, and Melbourne. These tweets served as a broader evaluation set, allowing us to assess the model's performance across different cities and geographies.

By evaluating the model on both datasets, we aimed to assess its effectiveness in accurately classifying tweets for the city of New York and its ability to generalize across different locations. This approach provides a comprehensive evaluation of the model's performance in real-world scenarios and helps us understand its applicability in various geographical contexts. The metrics used to evaluate the performance of each algorithm were accuracy, precision, recall, and the F1-score. The research [27] showed that these are the most common approaches and therefore we believe they are the most adequate to be used during this comparison. The duration of the execution for each one was also timed. To ensure reliable results, k-fold cross-validation, with k=5, was applied to the machine learning approaches and the time of execution is the average result of five different executions for each one of the models.

This evaluation has been carried out using a commodity laptop with an NVIDIA GeForce GTX 1650 with Max-Q design, 16 GB of RAM, and 1 TB storage SSD disk.

IV. RESULTS & DISCUSSION

Table III presents the overall results obtained for the two conducted experiences. Experience 1 refers to the tests executed using the dataset with content only from New York City. Experience 2 includes data from three cities: New York City, Melbourne, and London.

Starting by looking at the table, it is possible to see that the common machine learning algorithms, SVM, LR, and RF have very similar results for the performance metrics. As for the training and prediction time, for the SVM and the LR the values are almost the same, while the RF takes almost 80% longer. From one experience to the other, the

TABLE III
COMPARISON BETWEEN DIFFERENT APPROACHES FOR TEXT CLASSIFICATION

	SVM	LR	RF	BERT-base	BERT-large	BERT fine-tuned
Experience 1						
Execution Time (seconds):						
- Training time	0.390±0.008	0.390±0.005	0.628±0.030	-	-	6184.476±54.345
- Prediction time	0.021±0.001	0.019±0.003	0.027±0.001	40.641±4.019	91.585±0.614	761.912 ±5.637
Performance:						
- Accuracy	0.670±0.04	0.671±0.03	0.668±0.05	0.67	0.54	0.945±0.03
- Precision	0.665±0.03	0.669±0.02	0.674±0.06	0.70	0.66	0.934±0.02
- Recall	0.684±0.06	0.674±0.06	0.648±0.09	0.67	0.54	0.956±0.04
- F1-Score	0.674±0.04	0.671±0.03	0.660±0.07	0.66	0.44	0.950±0.01
Experience 2						
Execution Time (seconds):						
- Training time	1.039±0.018	0.968±0.017	1.846±0.025	-	-	18377.758±35.345
- Prediction time	0.096±0.005	0.063±0.006	0.071±0.007	108.365±0.457	266.906±2.665	2293.379±4.857
Performance:						
- Accuracy	0.663±0.03	0.654±0.02	0.679±0.02	0.67	0.55	0.962 ±0.01
- Precision	0.642±0.03	0.647±0.02	0.679±0.02	0.69	0.67	0.945 ±0.02
- Recall	0.737±0.02	0.682±0.04	0.681±0.03	0.67	0.55	0.956 ±0.04
- F1-Score	0.686±0.02	0.663±0.02	0.679±0.02	0.65	0.44	0.947 ±0.02
Test Sentences Classification						
Sentence T1	related	related	related	related	related	related
Sentence T2	unrelated	unrelated	unrelated	related	related	related
Sentence nT1	related	related	unrelated	unrelated	related	unrelated
Sentence nT2	related	related	related	unrelated	related	related
Comparison						
Implementation complexity	1	1	1	2	3	4
Adaptation to Other Problems:						
- Different language	3	3	3	1	1	2
- Different cities	2	2	2	2	2	1
- New transports	3	3	3	1	1	2
- Specific transport mode	3	3	3	1	1	2

Test Sentences Classifcation: T: Transports related sentence, nT: non-Transports related sentence.

results are almost the same, with the first two algorithms having slightly worst results, but not in a significant way. The execution times are pretty much linear.

Regarding the BERT models without training, they do not have training time, and the prediction times are much longer than the ones from the previous algorithms. These prediction times were not completely linear for the base model, but the difference is irrelevant unless the dataset has extremely large proportions. From the base model to the large, the prediction time more than doubles, but this is not applicable to the results since the large model performs worse in every single metric. The results from both experiences are once again almost the same, without any difference relevant enough for conclusions to be drawn from.

Finally, for the BERT-base fine-tuned, the times both for training and testing are much longer than for the remaining models, but the model performs much better, achieving more 30% of accuracy and precision. The results were already high for experience 1 so there was not much space to improve in experience 2, in which the results are only slightly better.

Table III also has a **Test Sentences Classification** section that presents four different sentences and the classification results for each algorithm. The first 2 sentences are related to transports and represented in the table as T1 and T2, with T1 being ”Company trucks are delayed again due to the snow” and T2 being ”another accident involving bikes on fifth avenue”. The other 2 sentences represent tweets unrelated to transports and are represented as nT1 and nT2, with nT1 being ”What a sunny day it is today!” and nT2 being ”I can't miss another train”. Surprisingly Bert-Base was the only model that predicted all the sentences class correctly, with the fined tuned version coming close to it but failing to for the last sentence that was related to a gym train, not a vehicle train.

To complement this classification demonstration Table IV provides a confusion matrix of the classification results obtained for what is considered the best model according to the metrics, which is Bert Base fine-tuned. Analyzing this matrix, the model, even when fine-tuned, displays strengths in identifying clear-cut cases of transport-related discourse but struggles with metaphorical or indirect references, which however can also be particularly challenging for some humans. It also misclassifies a contextually non-transport-related statement due to the presence of a transport-related word, which is the word ”train”. These results signify the need for better context awareness and understanding of metaphoric language in machine learning models. It also emphasizes the importance of continual fine-tuning and training with diverse datasets to improve linguistic and contextual understanding, especially in tasks as intricate as text classification.

In the last part of the table, we compare the different models by ranking 1 (best) to 6 (worst). If models are considered similar, they can have the same ranking. The implementation complexity refers to both how difficult it is to implement each solution and how difficult and time-

TABLE IV
EXAMPLES OF TEXT CLASSIFICATION OF TRANSPORT AND NON-TRANSPORT TWEETS USING THE CONFUSION MATRIX (BERT BASE FINE-TUNED).

		Predict class	
		Positive	Negative
Actual class	Positive	*True-Positive* "My truck is stuck again due to the heavy snow, I'm so tired of days like this"	*False-Negative* "Smooth ride or bumpy road, it's all about shifting gears."
	Negative	*False-Positive* "I can not miss another train today!"	*True-Negative* "NFL season is back again tonight, LET'S GO BENGALS"

consuming it is to run it, and here we believe this complexity is the same for the first three solutions, followed by the models without training and lastly the fine-tuning, which requires a deeper understanding of the possible approaches.

In the "Other Problems Adaptation" four different categories can be found, and the ranking demonstrated what we believe to be the most accessible models to due the new tasks. For different languages, the BERT base/large approaches are the best since it will only be necessary to translate the used dictionary for the chosen language. The Bert fine-tuned approach will need a new training dataset just like the most common algorithms (SVM, LR, RF) but since it was trained in different languages, it will not require changes to the embedding generation. For different cities where English is the primary language, no changes need to be made to any model, so the approach with the best metrics (fine-tuned) is considered the superior choice. For new transports, the easiest choice is the one that doesn't require new training, just new additions to the dictionary, followed by the one that will give better results. Lastly, for a specific transport mode, the most simple choice is to use the BERT approach without training and reduce the dictionary entries only to reflect the desired mean of transportation, followed by the training approaches. This last point order would be inverted if the objective is a specific company, which would be a problem for the non-trained approaches.

V. APPLICATION TO TRANSPORTATION DECISION-MAKING

Machine learning algorithms can analyze vast amounts of data in near real-time. This enables transport authorities to gain valuable insights into traffic conditions, congestion hotspots, and incident detection promptly. With this

information transportation authorities can optimize traffic signal timings, lane configurations, and road markings based on real-time conditions. Besides text messages, machine learning models can analyze images and videos to detect and classify incidents such as accidents or road hazards. This allows for quicker incident response, minimizing disruptions, and improving road safety. These models can also analyze historical data and patterns to make predictions and forecasts about future traffic conditions. This information can assist in proactive decision-making and planning for different traffic management strategies.

However, machine learning algorithms also have some disadvantages. These models heavily rely on data quality and reliability. If the data used for training the models are incomplete, biased, or inaccurate, it can impact the effectiveness and performance of the models. Also, some models can be complex and difficult to interpret, and computationally demanding. This can make it challenging to understand how the models arrive at their decisions or recommendations, potentially limiting the transparency and trustworthiness of the system and requiring significant processing power and storage resources. Therefore, implementing and maintaining the necessary infrastructure to support machine learning-based systems can be costly.

Thus, when considering the use of machine learning methods to filter information used for traffic control, and operational and tactical management, factors such as time to train and predict, performance, complexity, and adaptability should be considered. Different machine learning methods have different strengths and limitations, and their suitability depends on the specific characteristics of the traffic management task. Therefore, selecting the appropriate method for this filtering depends on the specific requirements and objectives of the task. Therefore based on the evaluation performed in section IV the suitability of each method was assessed by traffic operation type.

While traffic control aims to ensure smooth and safe traffic flow, minimize congestion and delays, prevent accidents, and optimize the use of available roadway capacity, operational traffic management aims to improve traffic flow, enhance safety, and optimize transportation systems. On the other hand, tactical traffic management is rather focused on long-term planning and system optimization.

Machine learning techniques can be applied in each phase of traffic management to analyze data, make predictions, optimize operations, and enhance decision-making processes, namely for:

- **Traffic control**: Machine learning can be used in traffic control to improve signal timings, detect and predict traffic congestion, monitor traffic conditions through surveillance systems, and enhance incident detection and response, which is a short-term task and therefore needs a filtering option that can retrieve results fast. The best option for this seems to be the traditional machine learning approaches that can give predictions fast and for a task like this do not require new training often. If there is a lack of data for the initial training, the BERT models without training are also a good choice.
- **Operational traffic management**: Machine learning can assist in operational traffic management by analyzing real-time traffic data, predicting traffic patterns, optimizing signal timings and lane configurations, and providing accurate traveler information. This task is very similar to the previous one, so the same choice can be applied.
- **Tactical traffic management**: Machine learning can support tactical traffic management by analyzing historical traffic data, predicting future traffic demands,

optimizing route planning, and assisting in capacity planning and infrastructure development. This is a medium to long-term task and therefore, time is not a problem but retrieving as much useful information as possible is crucial. The fine-tuning BERT, although it is very resource depend is the best choice for something like this.

For near real-time traffic control operations, where immediate response is required, lightweight and fast algorithms may be preferred to ensure quick decision-making. On the other hand, for long-term operational traffic management, more accurate algorithms can be used to analyze historical and real-time data for optimizing traffic flow and managing congestion. It is important to highlight that although these are the algorithms/models we chose, similar algorithms or models could also be applied for each category described.

VI. Conclusion

In this work, we compared three different groups of text classification solutions applied in the context of transportation. The results allowed us to conclude that the use of social media data and machine learning methods in traffic management should be carefully considered based on the specific requirements and objectives of each control level, whether it is near real-time control, operational traffic management, or tactical traffic management. Suggestions were provided about how to make adequate decisions of the algorithms when choosing how they will perform the content filtration. This process allows to save time, money, and resources.

ACKNOWLEDGMENT

This work was supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF) within project DynamiCITY (NORTE-01-0145-FEDER-000073), and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project e-LOG (EXPL/ECI-TRA/0679/2021) and research unit LIACC (FCT/UID/CEC/0027/2020). Tânia Fontes also thanks FCT for the Post-Doctoral scholarship SFRH/BPD/109426/2015 and the application reference 2022.07805.CEECIND.

References

[1] TRB, "Highway statistics 2021," Policy and Governmental Affairs, Office of Highway Policy Information, Tech. Rep., 2021.

[2] S. Bregman, *Uses of Social Media in Public Transportation*, ser. Synthesis of Transit Practice. Washington, DC: TRB, 2012.

[3] S. Carvalho, L. Sarmento, and R. J. F. Rossetti, "Real-Time Sensing of Traffic Information in Twitter Messages," *Proceedings of the IEEE ITSC 2010 Workshop on Artificial Transportation Systems and Simulation (ATSS2010), Madeira Island, Portugal*, pp. 19–22, 2010.

[4] Z. Kokkinogenis, J. Filgueiras, S. Carvalho, L. Sarmento, and R. J. F. Rossetti, "Mobility network evaluation in the user perspective: Real-time sensing of traffic information in twitter messages," in *Advances in Artificial Transportation Systems and Simulation*. Cambridge, MA: Academic Press, 2015, pp. 219–234.

[5] A. El Abaddi, L. Backstrom, S. Chakrabarti, A. Jaimes, J. Leskovec, and A. Tomkins, "Social media: Source of information or bunch of noise," in *Proc. 20th Int. Conf. Companion on World Wide Web*. New York, NY: ACM, 2011, p. 327–328.

[6] F. Rebelo, C. Soares, and R. J. F. Rossetti, "Twitterjam: Identification of mobility patterns in urban centers based on tweets," in *IEEE 1st International Smart Cities Conference (ISC2)*. IEEE, 2015, pp. 1–6.

[7] F. Ali, D. Kwak, P. Khan, S. M. Islam, K. H. Kim, and K. S. Kwak, "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 33–48, 2017.

[8] J. Pereira, A. Pasquali, P. Saleiro, and R. Rossetti, "Transportation in social media: an automatic classifier for travel-related tweets," in *Progress in Artificial Intelligence: 18th EPIA Conf. on Artificial Intelligence*, ser. LNCS, vol. 10423. Cham: Springer, 2017, pp. 355–366.

[9] M. Dahbi, R. Saadane, and S. Mbarki, "Social media sentiment monitoring in smart cities an application to moroccan dialects," in *4th Int. Conf. on Smart City Applications (SCA' 19)*, 2019.

[10] G. Kulkarni, L. Abellera, and A. Panangadan, "Unsupervised classification of online community input to advance transportation services," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018, pp. 261–267.

[11] H. İbrahim Cebeci, S. Güner, Y. Arslan, and E. Aydemir, "Barriers and drivers for biking: What can policymakers learn from social media analytics?" *Journal of Transport & Health*, vol. 28, p. 101542, 2023.

[12] S. Anastasia and I. Budi, "Twitter sentiment analysis of online transportation service providers," in *Int. Conf. on Advanced Computer Science and Information Systems*. IEEE, 2017, pp. 359–365.

[13] V. Effendy, "Sentiment Analysis on Twitter about the Use of City Public Transportation Using Support Vector Machine Method," *International Journal on Information and Communication Technology (IJoICT)*, vol. 2, no. 1, p. 57, 2016.

[14] X. Chen, Z. Wang, and X. Di, "Sentiment analysis on multimodal transportation during the covid-19 using social media data," *Information*, vol. 14, no. 2, 2023.

[15] H. N. Chua, A. W. Q. Liao, Y. C. Low, A. S. H. Lee, and M. A. Ismail, "Challenges of mining twitter data for analyzing service performance: A case study of transportation service in malaysia," in *Int. Conf. on Business Information Systems*. Springer, 2021, pp. 227–239.

[16] J. Tomas Mendez, H. Lobel, D. Parra, and J. Carlos Herrera, "Using twitter to infer user satisfaction with public transport: The case of Santiago, Chile," *IEEE Access*, vol. 7, pp. 60 255–60 263, 2019.

[17] N. N. Haghighi, X. C. Liu, R. Wei, W. Li, and H. Shao, "Using Twitter data for transit performance assessment: a framework for evaluating transit riders' opinions about quality of service," *Public Transport*, vol. 10, no. 2, pp. 363–377, 2018.

[18] B. P. Santos, P. H. L. Rettore, H. S. Ramos, L. F. M. Vieira, and A. A. F. Loureiro, "Enriching traffic information with a spatiotemporal model based on social media," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, June 2018, pp. 00 464–00 469.

[19] J. Osorio-Arjona, J. Horak, R. Svoboda, and Y. Garcia-Ruiz, "Social media semantic perceptions on Madrid Metro system: Using twitter data to link complaints to space," *Sustainable Cities and Society*, vol. 64, p. 102530, 2021.

[20] A. Candelieri and F. Archetti, "Detecting events and sentiment on twitter for improving urban mobility," *CEUR Workshop Proceedings*, vol. 1351, pp. 106–115, 2015.

[21] F. Ali, D. Kwak, P. Khan, S. El-sappagh, and A. Ali, "Transportation sentiment analysis using word embedding and ontology-based topic modeling," *Knowledge-Based Systems*, vol. 174, pp. 27–42, 2019.

[22] F. Murços, T. Fontes, and R. J. F. Rossetti, "Are bert embeddings able to infer travel patterns from twitter efficiently using a unigram approach?" in *IEEE International Smart Cities Conference (ISC2)*, 2021, pp. 1–7.

[23] E. Steiger, T. Ellersiek, and A. Zipf, "Explorative public transport flow analysis from uncertain social media data," *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, p. 1–7, 2014.

[24] H. Chang, L. Li, J. Huang, Q. Zhang, and K.-S. Chin, "Tracking traffic congestion and accidents using social media data: A case study of shanghai," *Accid. Anal. Prev.*, vol. 169, p. 106618, 2022.

[25] L. Hong, G. Convertino, and E. Chi, "Language matters in twitter: A large scale study," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, 2011, pp. 518–521.

[26] J. Pereira, A. Pasquali, P. Saleiro, R. Rossetti, and N. Cacho, "Characterizing geo-located tweets in brazilian megacities," in *2017 International smart cities conference (ISC2)*. IEEE, 2017, pp. 1–6.

[27] M. Hossin and S. M.N, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, pp. 01–11, March 2015.