A Study on Machine Learning Techniques for Sentiment Analysis of Users in Urban Cities for Transportation Using Social Media Data

Prof. G. G. Sayyad[1], Ms. Prachi Prakash Barbole[2], Ms. Sandhya Balu Dhainje[2], Mr. Gaurav Rajendra Rayate[2], Mr. Rohan Shivaji Rokade[2]

[1]Assistant Professor, [2]Student

Department of Computer Engineering, S. B. Patil College of Engineering, Savitribai Phule Pune University, Maharashtra, India

A B S T R A C T

Nowadays, there's increasing number of users on social media using various types of social media applications like FaceBook, Tweeter, WhatsApp, Instagram, SnapChat, and many more to be explored. These increasing numbers of user over social media account are drastically increasing the amount of data over their respective servers. This increasing data show social interaction of people over various topics.These interactions or discussions of people over social media lead to various activities that can be of good as well as bad cause; such good or bad activities can lead to traffic conjession in those respective cities and sometimes may lead to chaos or many more activities that can happen.

In order to have observance or control (by traffic authorities or security personnel's or governing bodies of city) over such activities that can be performed by discussion or interaction of people over social media that are causing conjession or peace harm in the city.These observations over such interactions or discussions over social media can help to reduce traffic conjession, chaos, peace harm reduction etc.

Keyword: Sentiment, Machine Learning, Analysis, Social Media, Support Vector Machine, Support Vector Regression, Decision Tree, Naive Bayes

I. INTRODUCTION

In the recent decade the usage of social media has increased drastically, this drastic increase in usage of social media has led to tremendous data collection and storage to respective social media servers.The social media companies have tremendous amount of data stored in their database servers, these servers consist of data like users contact details, photos, videos, audio files, personal information, feeds uploaded on the social media etc.

These social media companies use these data for analysis purpose and taking appropriate decisions regarding their respective businesses.The data of users collected through user social media account can also be used for transportation purpose in cities, village, states or nations world-wide. These data can help authorities handling city or security personnel's or governing bodies of that particular city or area.These data's can be gathered from different social media companies or through their API's (Application Programming Interface) which is accessible through some process of the same.

Considering this project we can observe that people using social media account does activities such as interacting with each other, having discussion about some events or activities to be performed, posting feed like photos, videos, articles, audio files etc.Through the above observation we come to know that above mentioned activities can tell us whether what is going to happen or what has happened in the city or area. These activities can be positive or negative.These positive or negative activities may sometimes lead to traffic conjession, chaos, strike, traffic blockage (Jam) etc. that leads to disturbance in traffic flow of the city and in such result it may cause GDP change of the city and peace loss.To achieve such project objectives we will be using different algorithms, techniques, method that will help to collect data, extract them, pre – process them, analyse them, identify sentiments out of it, classify them and then take appropriate decision related to traffic or city goodwillness.

II.   INTRODUCTION TO LITERATURE SURVEY

1. The paper with title name Mining Social Media for Open Innovation in Transportation Systems was published by Daniela Ulloa, Pedro Saleiro, Rosaldo J. F. Rossetti and Elis Regina Silva in the year 2016, in this paper they solved problem by User data was analysed using their social media account, through it collected sentiments of the users and took decision for the products to be sold or suggested them as per their sentiments, which was useful for the companies doing e-commerce business and for improving quality and better meet users.To access the impact of OSN as a user centred information provider this problem were solved using NPD process proposing a framework, Harvesting information, Search entity, Geological Portugal techniques. The future scope for their published paper was All those collected data using the techniques are monitored and further given to the companies which requires the sentiment and reviews about various kind of products and send suggestions to the users for further purchase [1].

2. The paper with title name Exploring the Potential of Twitter to Understand Traffic Events and their Locations in Greater Mumbai, India was published by Rahul Deb Das and Ross S. Purves in the year 2019, in this paper they solved problem by The use of twitter to extract and locate traffic related events in the area of Greater Mumbai. Built a pipeline capable of identifying and locating such events on map and allowing analyst to explore the nature and emergence of events. This in turn helped urban planners in their decision making process this problem were solved using Combination of Machine Learning simple rules,Hybrid geo referencing model techniques. The future scope for their published paper was this work may further used for other cities too, with their local regional language[2].

3. The paper with title name Scientific Text Sentiment Analysis using Machine Learning Techniques was published by Hassan Raza, M. Faizan, Ahsan Hamza , Ahmed Mushtaq, Naeem Akhtar in the year 2019, in this paper they solved problem by The use of dataset to extract scientific text sentiment using various techniques used below for analyze the sentiments of citation sentences extracted from different scientific research papers just because of linguistic differences in this domain this problem were solved using LR

(Logistic regression),DT (Decision tree),KNN (K – Nearest Neighbour), SVM (Support Vector Machine), NB (Navie Bayes), Classification algorithms, Evaluations matrix, Naviebayes F - score. The future scope for their published paper was This work may further used for other cities too, with their local regional language[3].

4. The paper with title name Sentiment Analysis: Machine Learning Approach was published by Dipak R. Kawade, Dr.Kavita S. Oza in the year 2017, in this paper they solved problem by the use of twitter to extract and locate sentiment related events in the area of defence and research. Built a pipeline capable of identifying and locating such events on twitter and allowing analyst to explore the nature and emergence of events. This in turn helped defence personnel's in their decision making process this problem were solved using Pre-processing techniques, Naive Bayes, Genetic Algorithms. The future scope for their published paper was in future big data analysis can be used to classify all emotions for large volume of tweets[4].

5. The paper with title name Optimization of sentiment analysis using machine learning classifiers was published by Jaspreet Singh, Gurvinder Singh and Rajinder Singh in the year 2017, in this paper they solved problem by the use of twitter to extract sentiment related behaviour of data or tweets in the area of e – commerce business. Built a pipeline capable of identifying sentiments and locating such events on twitter and allowing analyst to explore the nature of human behaviour on social media like tweeter. This in turn helped companies in their decision making process this problem were solved using Naïve Bayes used for sentiment classification, J48 algorithm used for sentiment prediction, BFTREE algorithm used for sentiment prediction, OneR algorithm used for sentiment prediction. The future scope for their published paper was The task of sentiment analysis has scope to improve pre-processing with word embedding's using deep neural networks and can also extend this study through convolution neural networks[5].

6. The paper with title name Sentiment analysis using machine learning was published by Anjali Gupta, AmitaDhankar, SurayanshDabas in the year 2018, in this paper they solved problem by The use of data available on social media to extract sentiment related behaviour of data in the area of opinion taking or decision making or suggestion giving process. Built a pipeline capable of identifying sentiments and suggesting or opinion giving such events on social media and allowing analyst to explore the nature of human behaviour on social media.This in turn helped companiesin their decision making process this problem were solved using Naïve Bayes, Maximum Entropy. The future scope for their published paper was the task of sentiment analysis has scope to be used in investigation purpose or evidence collection improve pre-processing with word embedding's using deep neural networks and can also extend this study through convolution neural networks[6].

7. The paper with title name Using Information from Heterogeneous Sources and Machine Learning in Intelligent Transportation Systems was published by A.L.C. Bazzan, J.C. Chamby-Diaz, R. S. Estevam, L. de A. Schmidt, M. Pasin, J. L. A. Samatelo and M. V. L. Ribeiro in the year 2019, in this paper they solved problem by The use of data available on social media to extract sentiment related behaviour of data in the area of city to identify the city traffic flow, amount of traffic on the day or too, conjession, potholes in the city road etc. This is observed and further decision is taken to solve such problems this problem were solved using Computer vision techniques for traffic flow analysis, ANN(Artificial Neural Network) in video images to pothole detection, Named Entity Recognition (NER), Gray-Level Co-Occurrence Matrix (GLCM), Light Detection and Ranging (LiDaR). The future scope for their published paper was The

application of such methods over data that relates to the same city in order to assess the full potential of the integrated framework, as well as extending the prediction tools[7].

8. The paper with title name Sentiment analysis in English text was published by ArwaAlshamsi, ReemBayari , Said Salloum in the year 2021, in this paper they solved problem by This research paper aims to obtain a dataset of tweets and apply different machine learning algorithms to analyze and classify texts this problem were solved using K-NN, Decision Tree, Random Forest, Random Tree, Naive Byes and ID3. The future scope for their published paper was Applying Sentiment analysis to new type of data such as image, videos and audios[8].

9. The paper with title name A Study of Sentiment Analysis: Concepts, Techniques, and Challenges was published by Ameen Abdullah QaidAqlan, B. Manjula and R. LakshmanNaik in the year 2020, in this paper they solved problem by The importance of opinions and comments on Web sites and how to extract them through certain techniques this problem were solved using Naive Bayes,SVM and Hadoop tools. The future scope for their published paper was Using big data techniques (Hadoop) in sentiment analysis field to give more effective and accurate results[9].

10. The paper with title name Sentiment analysis and opinion mining applied to scientific paper reviews was published by Brian Keith Norambuena in the year 2019, in this paper they solved problem by To determine the feelings, opinions, emotions, among other things, of people on something or someone. To do this, natural language techniques and machine learning algorithms are used this problem were solved using Naive Bayes, SVM and Hadoop tools. The future scope for their published paper was Future work could deal with the longitudinal evaluation of consistency between the review and the acceptance or rejection of the paper by each reviewer[10].

11. The paper with title name Sentiment Analysis of Twitter Data: A Survey of Techniques was published by Vishal A. Kharde,S.S. Sonawane in the year 2016, in this paper they solved problem by Sentiment analysis of twitter data which is helpful to analyse the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases this problem were solved using Naive Bayes, SVM andMax Entropy. The future scope for their published paper was The study of combining machine learning method into opinion lexicon method in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages[11].

12. The paper with title name A survey on sentiment analysis methods, applications, and challenges was published by Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni in the year 2022, in this paper they solved problem by The primary objective of this work is to investigate and complete classification methods with their advantage and disadvantages in sentiment analysis this problem were solved using Naive Bayes, SVM. The future scope for their published paper was Continuously expanding the comparison area with additional findings[12].

13. The paper with title name Sentiment analysis using product review data was published by Xing Fang and Justin Zhan in the year 2015, in this paper they solved problem by To tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis this problem were solved using Naive Bayes, SVM, Random forest. The future scope for their published paper was Experiments for both sentence-level categorization and review-level categorization are performed with promising outcomes[13].

14. LPR in modern transport systems identifies vehicles via computer vision. Our novel SR algorithm improves license plate legibility in traffic videos[14].

III. LIMITATIONS OF EXISTING WORK

- The system may not always be able to get accurate data from social media database.
- The system may not always be able to extract exact data collected from social media database.
- The system may not always be able to give appropriate decision based on the available data.
- Sometimes it may not get updated data.
- Sometimes the system might get confused between the positive or negative sentiments or good or bad sentiments.
- Sometimes it might give wrong decisions that are not precise or accurate.

IV. CONCLUSION

We have proposed a system that will collect dataset from social media database that will be further extracted, pre – processed, analysed and decision will be taken in favour of cities proper flow, maintaining peace overall city.[1][2]

This extracted data then identifies the sentimental words, posts, feeds, messages etc. and then they will be classified such that they will be in two groups i.e. positive or negative or good or bad further they will be analysed and decision will be taken accordingly like whether today the city will have conjession or any kind of chaos will be created or something else. [3][7][8][13]

V. REFERENCES

[1]. Daniela Ulloa, Pedro Saleiro, Rosaldo J. F. Rossetti and Elis Regina Silva, "Mining Social Media for Open Innovation in Transportation Systems", 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC) Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

[2]. Rahul Deb Das and Ross S. Purves, "Exploring the Potential of Twitter to Understand Traffic Events and Their Locations in Greater Mumbai, India", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[3]. Hassan Raza1 , M. Faizan2 , Ahsan Hamza3 , Ahmed Mushtaq4 , Naeem Akhtar5 School of Computer Sciences National College of Business Administration and Economics Lahore, Pakistan," Scientific Text Sentiment Analysis using Machine Learning Techniques", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 12, 2019

[4]. Dipak R. Kawade#1, Dr.Kavita S. Oza*2 # Department of ComputerScience, Sangola College, SangolaDist-Solapur (MS) India 1 dipakkavade@gmail.com * Department of ComputerScience, Shiveji University, Kolhapur (MS) India 2 skavita.oza@gmail.com, "Sentiment Analysis: Machine Learning Approach", ISSN (Print) : 2319-8613 ISSN (Online) : 0975-4024 Dipak R. Kawade et al. / International Journal of Engineering and Technology (IJET)

[5]. Jaspreet Singh*, Gurvinder Singh and Rajinder Singh, "Optimization of sentiment analysis using machine learning classifers", Singh et al. Hum. Cent.Comput. Inf. Sci. (2017) 7:32 DOI 10.1186/s13673-017-0116-3

[6]. Anjali Gupta1 , Amita Dhankar2 , Surayansh Dabas3 M.Tech student, Assistant Professor, B.Tech MSIT Delhi Department of C.S.E. UIET, M.D. UNIVERSITY, Rohtak, Haryana India, "SENTIMENT ANALYSIS USING MACHINE LEARNING: A REVIEW", JETIR1802163 Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org

[7]. A.L.C. Bazzan, J.C. Chamby-Diaz, and R. S. EstevamInstituto de Informatica, UFRGS ´ 91.501-970 P. Alegre, RS, Brazil Email: {bazzan, jccdiaz, rhuam.sena}@inf.ufrgs.br L. de A. Schmidt and M. Pasin Univ. Fed. de Santa Maria Av. Roraima 1000 97.105-900 Santa Maria, Brazil Email: {lschmidt, marcia}@inf.ufsm.br J. L. A. Samatelo and M. V. L. Ribeiro Univ. Fed. doEsp´ırito Santo Av. Fernando Ferrari, 514 29075-910 Vitoria, Brazil ´ Email: jorge.samatelo@ufes.br, matheusvlr@yahoo.com.br, "Using Information from Heterogeneous Sources and Machine Learning in Intelligent Transportation Systems", 978-1-7281-4914-1/19/$31.00 c 2019 IEEE

[8]. Xing Fang* and Justin Zhan," Sentiment analysis using product review data", Fang and Zhan Journal of Big Data (2015) 2:5 DOI 10.1186/s40537-015-0015-2

[9]. Arwa Alshamsi1 , Reem Bayari1 , Said Salloum, "Sentiment Analysis in English Texts", Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1683-1689 (2020) www.astesj.com Special Issue on Multidisciplinary Sciences and Engineering

[10]. Zhou Gui Zhou, "Research on Sentiment Analysis Model of Short Text Based on Deep Learning", Hindawi Scientific Programming Volume 2022, Article ID 2681533, 7 pages https://doi.org/10.1155/2022/2681

[11]. Brian Keith Norambuena* ,Exequiel Fuentes Lettura and Claudio Meneses Villegas," Sentiment analysis and opinion mining applied to scientific paper reviews", Intelligent Data Analysis 23 (2019) 191–214 191 DOI 10.3233/IDA-173807 IOS Press

[12]. Vishal A. Kharde Department of Computer Engg, Pune Institute of Computer Technology,Pune University of Pune (India) S.S. Sonawane Department of Computer Engg, Pune Institute of Computer Technology,Pune University of Pune (India), "Sentiment Analysis of Twitter Data: A Survey of Techniques", Sentiment Analysis of Twitter Data: A Survey of Techniques

[13]. Ameen Abdullah QaidAqlan, B. Manjula and R. LakshmanNaik, "A Study of Sentiment Analysis: Concepts, Techniques, and Challenges", Department of Computer Science, Kakatiya University, Warangal 506009, Telangana, India e-mail: ameenaqlan218@gmail.com R. LakshmanNaik Department of Information Technology, Kakatiya University, Warangal 506009, Telangana, India © Springer Nature Singapore Pte Ltd. 2019 N. Chaki et al. (eds.), Proceedings of International Conference on Computational Intelligence and Data Engineering, Lecture Notes on Data Engineering and Communications Technologies 28, https://doi.org/10.1007/978-981-13-6459-4_16

[14]. Dhakane, Vikas Nivrutti, and Jalinder Nivrutti Ekatpure. "Super Resolution of License Plates Using Generalized DAMRF Image Modeling."