Opinion Mining for Travel Route Recommendation using Social Media Networks (Twitter)

C. Nalini T. Poovizhi

Abstract: *Most of the organizations use text analytics to uncover purposeful information from an unstructured text as a result of considering the linguistic communication process techniques area unit extremely difficult. They typically cause several issues because of the inconsistency in syntax and linguistics. Sentiment analysis based on the opinion of the users. On twitter, many people post about their experience on the traffic routes. This project discusses the prediction of text mining analysis. On that post collecting from the data set and we find out which path is the best path for the travellers and waiting for commuters. In this project we discuss the traffic mining tweets using the keywords predicting the positive and negative comment on the Twitter. Experimentation involves discussion and comparison of ensemble classifiers over tagged tweets. Finally, it will be finding the best accuracy.*

Index Terms: Sentiment analysis, Traffic, Twitter data, Route Recommendation

I. INTRODUCTION

A. Web Mining

Web could be an assortment of inter-related files on one or a lot of internet servers. Thus, data processing applied to the web has the potential to be quite useful. Web mining is mining of information associated with the planet wide net. This could be information really gift in websites or knowledge related to a internet activity. The primary is ancient trying of net pagan via content, whereas the second is a further search of pages found from a previous web mining tasks area unit usually divided into several classes. Web page mining examines the content of web page likewise as results of net searching. The substance incorporates message moreover as designs information. Web content mining is any separated into web content mining and list items mining. The essential is antiquated watching out of net agnostic by means of substance, though the second is any pursuit of pages found from a past inquiry.

B. Sentiment Analysis

Sentiment analysis also called as opinion mining, is that the field of concentrate that breaks down individuals' feelings, assumptions, assessments, evaluations, frames of mind, and feelings towards substances like stock, administrations, associations, people, issues, occasions, themes, and their qualities.

Opinion" is key influencers of our practices. Our convictions and view of the truth are molded on anyway others see the globe. At whatever point we need a decision a choice, to make look out the assessments of others. The assumption that shows up in the content comes in two flavors: unequivocal where the emotional sentence straightforwardly communicates a supposition ("It's a wonderful day"), and verifiable where the content infers a conclusion ("The headphone broke in two days").Most of the work done as such far spotlight on the first sort of assessment since it is the less demanding one to break down.

C. Opinion Mining

In this weblog, we will talk about the venturing stone to Opinion Mining – linguistically labeling a sentence. It'll encourage us to separate a sentence+ into its hidden linguistic structure – things, action words, descriptors and so on that may encourage us partner what was previously mentioned concerning what. When we tend to are sufficiently able to attempt and do that, we can separate supportive assessments that may encourage us to answer the inquiry uncover by our buyer over. Opinions are central to the majority human activities as a result of they are key influencers of our behaviors.

II. LITERATURE SURVEY

Yu-Ting Wen, Jinyoung Yeo, Wen-Chih Peng, Member, IEEE, and Seung-Won Hwang (2017) this paper, they need to propose a efficient Keyword-aware Representative Travel Route framework that uses information extraction from clients' verifiable[1]versatility records and social communications. Unequivocally, they designed a watchword extraction module to order the POI-related labels, for compelling coordinating with inquiry catchphrases.

Masahiko Itoh, Daisaku Yokoyama, Masashi Toyoda, Yoshimitsu Tomita, Satoshi Kawamura, and Masaru Kitsuregawa (2016) they did visual combination environment to investigate changes in streams of travelers on the capital of Japan metro and their circumstances and end results by misuse [2] more than four years' estimation of information extricated from the smart card framework and Twitter.

Opinion Mining for Travel Route Recommendation using Social Media Networks (Twitter)

Hang Do Minh (2016) the motivation behind this examination is to build up a Finnish traffic data framework that depends via web-based networking media information. The capability of exploitation interpersonal organization streams in rush hour gridlock data extraction has been incontestable in numerous [3] enormous urban communities; in any case, no examination has hence far explored the potential use in littler Networks like urban communities in the European nation.

Sandeep G Panchal, Prof. R.S.Apare (2017) they need to execute the ongoing traffic identification and investigation of the Twitter tweets returning from those territories inside the town. Android application to [4] bring up and suggest the graphical course organization of the traffic space.

Jie Bao , Yu Zheng , David Wilkie, Mohamed Mokbel (2015) they did these informal communities as area based interpersonal organizations (LBSNs).First classifications the recommender frameworks by the target of the suggestion, which may encapsulate areas, clients, exercises, or internet based life. Also, to classifications the recommender [5] frameworks by the strategies utilized, including content-based, interface investigation based, and helpful sifting based approachs. At last to classes the frameworks by the information sources utilized, just as client profiles, client online accounts, and client area narratives.

III. METHODOLOGY

A. Dataset

At this stage, the steps will be carried out for preparing for those tweets dataset for that next phase. First, any irrelevant tweets from the collected corpus are removed. All tweets with "traffic" keyword need not represent road traffic status. For example, it is free of cast to browse the website to know more about the traffic to your #website when sharing links https://t.co/1LId0EY2uF@BajpayeeManoj Awesome Trailer of #Traffic...No one other than you would have taken this message to the whole India...Kudos!! The first tweet represents website traffic, while the second tweet is with regard to a trailer of a movie called Traffic. Absolute dataset 875 and that dataset having two property tweets, opinion. In tweets, remarks on the traffic-related substance of human conclusion. What's more, conclusion dependent on positive and negative.Labelled data could be a group of samples that are tagged with one or additional labels. Labelling usually takes a collection of untagged data and augments every bit of that unlabelled data with important tags that are informative. Collecting traffic related data from the twitter, it would be tweets. On that tweets only finding the accuracy of this project. Based on the original tweet, an opinion i.e. "p" or "n" is added to the bag of words. Here "p" is considered as a positive and "n" is considered as a negative. Cluster as to finding the two attribute one is traffic related tweets another one is opinion of route recommendation that is positive and negative.

B. Pre-processing

Data preprocessing describes any kind of method performed on raw data to arrange it for an extra method procedure. Clustering is the way toward making a gathering of conceptual items into classes of comparable articles. In this section have to splitting word into the excel and preprocessing to the weka tool. K means algorithm finding the "p" and "n" opinion tweets. Furthermore, "p" is considered as a positive, "n" is considered as a negative. Then using the simple K means algorithm to the preprocessing section to finding the keywords from the traffic related tweets. K-means is one of the least complex unsupervised learning algorithms that tackle the outstanding grouping issue. In this section will helping the extracting keyword from data set. Using this algorithm finding the traffic related comments on twitter.

C. Sentiment Analysis

Sentiment Analysis is the process of determining whether a piece of writing (tweet, product/movie review, etc.) is positive, negative or neutral. This analysis finding the feedback about the route recommendation on using twitter. Transportation system plays a major role among people but ITS system helps getting opinions from the public. To complete the space, wisdom and opinions of the public are collected from the analysis of several information resources like blogs, forums the groups of social networking as well as international confidential resource, new magazine like wikis.

D. Traffic Route Recommendations

Real-time data streaming is that the method by that massive volumes of data are processed quickly specified a firm extracting the data from that data will react to dynamical conditions in real time. This data is utilized in totally different use cases such as finding trends associated with a selected keyword, measure complete sentiment, and gathering feedback regarding new product and services. To open the data stream to own Tweets delivered, you would like to send a association request to the API. Within the streaming model, this association exposes the pipeline for data to be delivered to you because it happens, and can exist for an indefinite period of your time. See the documentation for the particular APIs for information on establishing the association. Traffic congestion is a universal problem to contend with. This project extracting public opinion on traffic conditions from the tweets people make while on the road. Sentiment analyses have recently been developed as the internet usages have seen successful growth among people. Solving the difficulty technically, an endeavour has been made to produce a system for mining opinions concerning the traffic standing from tweets.Collection of traffic connected tweets victimisation the twitter API.Victimisation the text mining methodology. Analysis and also the extraction of data. Manual encryption of traffic information as Positive (p) and Negative (n). Building associate ensemble model victimisation the labelled dataset. Testing and validation of the model in classifying traffic connected tweets as "p" and "n". Below the process used to.

E. *Block diagram*
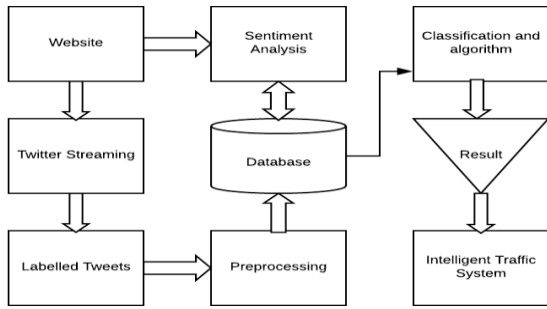
Figure 1.1 Architecture diagram

F. *Classification*

Classification could be a data mining technique that assigns classes to a collection of data so as to assist in additional correct predictions and analysis. Classification techniques are widely utilized in data mining to classify data among numerous Classes. Classification techniques are being employed in several businesses to simply determine the kind and cluster to that a selected tuple belongs. Classification facilitates the identification of organisms. Helps to establish the relationship among various groups of organisms. Helps to study the phylogeny and evolutionary history of organisms.

Classification is a procedure that is utilized for dividing the information into various classes as indicated by some compels or it groups everything in a dataset into one of a predefined set of classes or gatherings. It is a managed learning approach having known class categories. Several major kinds of classification algorithms including Naive Bayes classifier, Decision tree, Neural Network, Support vector machine (SVM). To predictTraffic congestion is a universal problem to contend with. This project extracting public opinion on traffic conditions from the tweets people make while on the road. Sentiment analyses have recently been developed as the internet usages have seen successful growth among people. Solving the difficulty technically, an Endeavour has been made to produce a system for mining opinions concerning the traffic standing from tweets. To discover the twitter data set through the weka tool. Algorithms are Decision tree, Neural Network, Support vector machine, Naive Bayes to finding the classification.That classification finding negative and positive

tweets from the traffic route recommendation. In this case decision tree finding the negative tweets more than another algorithm. Finding the best accuracy of the tweet from the traffic route recommendation.

G. *Navie Bayes*

It is an arrangement system supported Bayes' Theorem with a presumption of autonomy among indicators. In simple terms, a Naive Bayes classifier expects that the nearness of a particular element in an exceedingly class is inconsequential to the nearness of the other element. Naive Bayes classifiers mainly used in content order (because of higher lead to multiclass issues and autonomy rule) have a higher achievement rate when contrasted with various calculations. Subsequently, it's broadly used in Spam separating (distinguish spam email) and Sentiment Analysis (in social media analysis, to recognize positive and negative customer estimations).

H. *Decision Tree*

A decision tree might be a tree wherever every non-terminal hub speaks to a check or decision on the contemplated information thing. Selection of a specific branch relies on the consequence of the test. To classify a particular data item, we will in general start at the root hub and pursue the attestations down until we will in general achieve a terminal hub (or leaf). A decision is framed once a terminal hub is drawn closer. Decision trees may likewise be taken as an exceptional sort of a standard set, described by their various leveled association of guidelines.

I. *Support Vector Machine*

Support vector machines (SVMs) are a lot of related administered learning techniques utilized for order and relapse. Bolster vector machine (SVM) is an algorithmic guideline that attempts to discover a direct extractor (hyper plane) between the data purposes of two classifications in the multi-dimensional zone. SVMs are comparative temperament to tending to collaborations among alternatives and excess options.

TP Rate: It is the capacity which is utilized to locate the high true positive rate. The genuine positive rate is additionally called as affectability.

$$TPR = \frac{tn}{tn+fp}$$

Precision: Precision is given the correlation of a number of modules effectively classification to the number of whole modules classified fault-prone. It is the number of units accurately anticipated as broken.

$$Precision = \frac{tp}{tp+fp}$$

F-Measure: F-Measure is the one has the mix of both precisions and recall which is utilized to figure the score.

$$F\text{-Measure} = \frac{precision.call}{precision + recall}$$

J. *Neural Network*

Neural systems (NN) are those frameworks displayed dependent on the human brain working. Since the human brain comprises of uncountable neurons that are interconnected by synapses, a neural system could be a lot of associated information/yield units in which each affiliation joins a load identified with it. The network learns within the learning phase by adjusting the weights thus on are ready to predict the proper category label of the input.

IV. EXPERIMENTAL RESULT

The algorithm can either be connected straightforwardly to a dataset or called from your very own Java code. Weka contains instruments for information pre-preparing, classification, regression, clustering, association principles, and perception. It is likewise appropriate for growing new machine learning schemes. The trial examination of classification algorithm is done dependent on the execution proportions of arrangement systems.

Classification Techniques

IR Recall: In data recovery, review is the portion of the significant records that are effectively recovered.
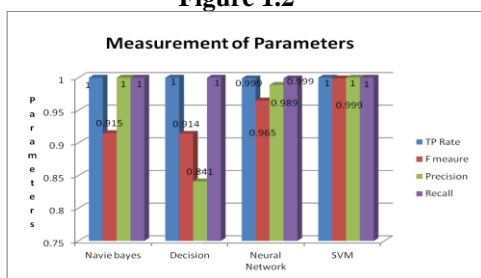
$$Recall = \frac{tp}{tp+fn}$$

Table1: classification techniques measurement

Algorithm	Correctly classified	Incorrectly classified	TP Rate	F-measure	Precision	Recall
Naive Bayes	739	136	1.000	0.915	1.000	1.000
Decision tree	736	139	1.000	0.914	0.841	1.000
Neural network	822	53	0.999	0.965	0.989	0.999
SVM	874	1	1.000	0.999	1.000	1.000
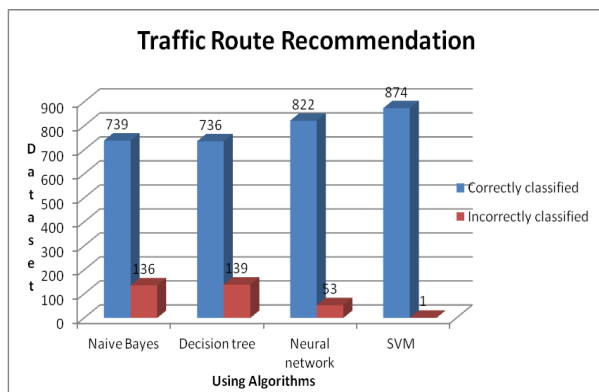
Figure 1.2

Figure 1.3

From the experimental result, Decision tree performs best in classifying process than neural network, Support vector machine (SVM), Naive bayes algorithm. The algorithm which has higher qualities from the order systems has picked as the best algorithm. In this classification, the decision tree has the greatest order esteems. So it is considered as the best classification calculation.

V. CONCLUSION

In this paper work has shown how the direction of extract tweets related to traffic status and opinions. A system is provided for mining opinions about the traffic status from tweets. TSA (Traffic Sentiment Analysis) system is also proposed to treat the traffic problems in newer angle. Ensemble model has been built by applying various classifier algorithms over the labeled tweets of training set.

Experimentation involves discussion and comparison of ensemble classifiers over tagged tweets. Twitter data set as inserted to be weka tool and exploratory outcome, Decision tree performs best in characterizing process than the neural system, Support vector machine (SVM), Naive Bayes calculation. The calculation which has higher characteristics from the request frameworks has picked as the best calculation. In this characterization, the Decision tree has the best request regards. So it is considered as the best characterization count. Finally, it will be finding the best accuracy of negative and positive tweets. With a help of classification and various algorithm to finding the travel route recommendation.

REFERENCES

1. Yu-Ting Wen, Jinyoung Yeo, Wen-Chih Peng, Member, IEEE, and Seung-Won Hwang "Efficient Keyword-Aware Representative Travel Route Recommendation" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 29, NO. 8, AUGUST 2017
2. Masahiko Itoh, Member, IEEE, Daisaku Yokoyama, Member, IEEE, Masashi Toyoda, Member, IEEE,Yoshimitsu Tomita, Satoshi Kawamura, and Masaru Kitsuregawa, Fellow, IEEE "Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network" (2016)
3. Hang Do Minh "Detection of Traffic Events from Finnish Social Media Data" (2016) University of Tampere School of Information Sciences Computer Science / Software Development, 48 pages,5 index and appendix pages November 2016
4. Sandeep G Panchal, Prof. R. S. Apare (2017) "Real Time Traffic Detection using Twitter Tweets Analysis" International Journal of Engineering Trends and Technology (IJETT) – Volume 47 Number 8 May 2017
5. Jie Bao · Yu Zheng · David Wilkie · Mohamed Mokbel Recommendations in Location-based Social Networks: A Survey.Published online: 6 February 2015
6. Muhammad Usman Shahid Khan, Student Member, IEEE, Osman Khalid, Ying Huang, Rajiv Ranjan, Fan Zhang, Senior Member, IEEE, Junwei Cao, Senior Member, IEEE, Bharadwaj Veeravalli, Senior Member, IEEE, Samee U. Khan, Senior Member, IEEE, Keqin Li, Fellow, IEEE, and Albert Y. Zomaya, Fellow, IEEE "MacroServ: A Route Recommendation Service for Large-Scale Evacuations" (2017) IEEE TRANSACTIONS ON journal name, manuscript ID
7. Yu-Ting Wen, Kae-Jer Cho, Wen-Chih Peng, Jinyoung Yeo, Seung-won Hwang ,National Chiao Tung University, Taiwan ,Pohang University of Science and Technology, Republic of Korea,Yonsei University, Republic of Korea" KSTR: Keyword-aware Skyline Travel Route Recommendation" 2015 IEEE International Conference on Data Mining
8. P.V.Varalakshmi, T.Sarayu, Thamarai Selvi.N.D."User Specific Xenogogy From Social Media" 2017 IEEE International Conference on Technological Innovations in Communication, Control and Automation (TICCA)
9. Zhao Jianqiang, Gui Xiaolin "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis"IEEE (2017)
10. M.S. Vivekanandan and Dr. C. Rajabhushanam, "Enabling Privacy Protection and Content Assurance in Geo-Social Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 49-55, April 2018.

AUTHORS PROFILE

Dr. C.NALINI received Ph.D and M.Tech from the Bharath University in 2004, 2007 respectively. Now she is working as a professor in the Department of CSE at Bharath Institute of Higher Education and Research. She has published more than 70 papers in scopus indexed journals and 100 papers published in international journal.

She has presented the paper in 45 national conferences and 33 international conferences, and received Radha Krishnan gold medal Award for outstanding individual achievement in 2014. She is a member of many professional bodies like ISTE, CSI, IEEE and IANG.

Second Author, Poovizhi.T, Pursuing M.Tech Final year in Bharath Institute of Higher Education and Research. She has Published in two paper in Scopus indexed journal and Published two papers in conference level.