REVA UNIVERSITY
Bengaluru, India

A Project Report on

Emotion Analysis based Voice of Citizen using Twitter data

Submitted in Partial Fulfilment for Award of Degree of

Master of Business Administration

In Business Analytics

Submitted By

Surendra Tanniru

R20MBA09

Under the Guidance of

Dr. JB Simha

CTO, ABIBA Technologies

REVA Academy for Corporate Excellence - RACE

REVA University

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064

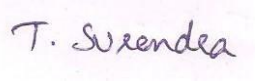
race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Surendra Tanniru** hereby declare that I have completed the project work towards the second year of Master of Business Administration in Business Analytics at, REVA University on the topic entitled **Emotion Analysis based Voice of Citizen using Twitter data** under the supervision of **Dr. JB Simha.** This report embodies the original work done by me in full fulfilment of the requirements for the award of degree for the academic year **2022.**



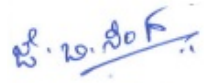
Place: Bengaluru
Date: 22nd October, 2022

Name of the Student: Surendra Tanniru
Signature of Student:

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to certify that the project work entitled **Emotion Analysis based Voice of Citizen using Twitter data** carried out by **Surendra Tanniru with SRN R20MBA09,** a bonafide student of REVA University, is submitting the second-year project report in fulfilment for the award of **MBA in Business Analytics** during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with a similarity score of less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said degree.



Signature of the Guide
Name of the Guide
Dr. JB Simha

Signature of the Director
Name of the Director
Dr. Shinu Abhi

External Viva
Names of the Examiners

1. Rajib Bhattacharya, Director – Data & Analytics, Cargill
2. Akshay Kulkarni, Manager – Data Science and AI, Publices Sapient

Place: Bengaluru
Date: 22-October-2022

REVA UNIVERSITY
Bengaluru, India

Acknowledgement

I am highly indebted to **Dr. Shinu Abhi,** Director, and Corporate Training for their guidance and constant supervision as well as for providing necessary information regarding the project and for their support in completing the project.

I would like to thank my project guide **Dr. JB Simha** for his valuable guidance provided to understand the concept and execute this project. It is gratitude towards all other mentors for their valuable guidance and suggestion in learning various data science aspects and for their support. I am thankful for my classmates for their aspiring guidance, invaluable constructive criticism, and friendly advice during the project work.

I would like to acknowledge the support provided by Hon'ble Chancellor, **Dr. P Shyama Raju**, Vice Chancellor **Dr. M. Dhananjaya**, and Registrar **Dr. N. Ramesh**. It is sincere thanks to all members of the program office of RACE who are supportive of all requirements from the program office.

It is my sincere gratitude towards my parents, and my family for their kind co-operation and encouragement which helped me in the completion of this project.

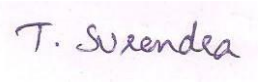
Place: Bengaluru
Date: 22-October-2022

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **Emotion Analysis based Voice of Citizen using Twitter data** was scanned for similarity detection. Process and outcome are given below.

Software Used: Turnitin
Date of Report Generation: 15th August, 2022
Similarity Index in %: 14 %
Total word count: 3648
Name of the Guide: Dr. JB Simha



Place: Bengaluru
Date: 17-October-2022
Verified by: M N Dincy Dechamma

Name of the Student: Surendra Tanniru
Signature of Student

Signature
Dr. Shinu Abhi,
Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	NRC	National Research Council
2	NLP	Natural Language Processing
3	WRI	World Resources Institute

List of Figures

List of Tables

Abstract

After the Big data boom, the number of social media network is increasing at a faster pace. Moreover, the impact that these social media messages and comments create socially and politically is enormous. They influence people mentally and psychologically. Out of many social media platforms, Twitter comes into focus which influences the public mostly because of its platform where many people express their feelings, behaviours, and emotions in the form of tweets which causes a lot of impact.

Traffic is one of the biggest problems that India is experiencing for many years. It might be because of the infrastructure, population, or the increasing number of vehicles on the road. People use Social media to convey their feelings and emotions that they have during the time they spent in traffic.

Many researchers tried to analyze the sentiments from Twitter data. But many focus on positive and negative sentiments that a tweet expresses. The study is extended to identify the emotions apart from the sentiments that people express on Twitter. There are different types of lexicons available to identify the Sentiments in textual data. Similarly, some lexicons can be used to identify the emotions in the textual data. Emotion lexicon from National Research council (NRC) and text2emotion are used in this study to identify the emotions like joy, fear, anger, surprise, trust, anticipation, disgust, and sadness.

This analysis can help the officials to identify the highly congested areas and take necessary actions to divert the traffic and manage the congestion. This also helps in reducing the time taken to travel to different places even if someone has to go through congested areas.

Keywords — Emotion, Emotion recognition, NRC lexicon, Twitter, Traffic, Text Analysis, Natural language processing (NLP)

Contents

Chapter 1: Introduction

A BBC journalist in 2016 wrote 'In Bangalore the hottest topic of conversation is traffic'. Many claims that walking is faster than driving in the city (BBC News, 2016).

According to a report by map-making company TomTom; Bangalore has the worst traffic congestion. Bangalore ranked first as the most traffic-congested city in the world out of 416 cities across 57 countries. The report also estimates that on average a person spends 71% more time in traffic than in any other city. It also says that people in the city are spending 243 hours i.e. around 10 days in traffic during peak hours. There is a report on top 10 Indian cities with high traffic congestion. Mumbai grabbed the 4th spot with 65% congestion. Pune was at the 5th spot with 59% and the capital New Delhi at the 8th spot with a congestion level of 56%. After the pandemic, the traffic in the city sees a dip in traffic congestion and has 49% less congestion during peak hours when compared to pre-Covid levels (Bangalore has the worst traffic in the world, 2020).

As per the report by TomTom, Bangalore ranked first as the most traffic-congested city in the world out of 416 cities in 2020. It ranked sixth in 2019 and even after the pandemic it still stands at 10th position in its global 25 indexes (Bangalore Traffic Report | TomTom traffic Index, 2019).

According to a report by World Resource Institute (WRI), an average citizen in India spends more than 240 hours stuck in jams every year. It also estimates that traffic delays cost the city 65bn rupees ($950m; £760m) a year (Mulukutla, 2015).

As per the WRI report, in Bangalore, the traffic moved at the speed of 35km (22 miles) an hour in 2005 whereas, in 2014, it had slowed down to 9.2km (5.7 miles). During the pre-Covid days, the speed at peak times is just 4-5km on the city's key Outer Ring Road.

There are many reasons for Bangalore having huge traffic congestion in many areas. Mostly it is because of the fast and unplanned growth of the city in the last 20 years. Bengaluru is home to many IT companies and is often described as the Silicon Valley of India. The city sees a global footprint of many IT companies and industries opening their

offices in and around Bangalore early 1990s itself. This has brought a huge number of people from all over India shifting their living to Bangalore making it densely populated. The population in Bangalore increased from 5.6 million in 2001 to 8.7 million in 2011. It is estimated that the current population of the city is around 12 million.

There is also another problem that comes with the number of vehicles hitting the road. There are about 10.04 million private vehicles in the city, which includes 7 million two-wheelers and more than 2.2 million cars. More than a million vehicles enter the city every day and 3,500 vehicles are getting registered in the city daily (Dev, 2021). Figure No. 1.1 shows the number of vehicles registered per year until 2019.

Figure No. 1.1: Total Vehicles registered per year

Even the CMO of Karnataka passed an order to decongest traffic bottlenecks at ten significant hotspots like Mahadevapura, Outer Ring Road, Silk Board Junction, Hebbal, and Whitefield Road.

Even the Social media platform Twitter is outraged by the tweets about Bangalore traffic. With the help of NLP, we were able to analyze the Sentiments from the text/tweet. Recently, the text analysis has gone beyond by just analyzing the sentiments but also identifying the topics and emotions. Further, analysis of emotions in text, from news to social media posts, will improve the understanding of not just how people convey emotions through language but also how emotions shape our behavior. This allows us to capitalize on

substantial previously unattainable opportunities in commerce, public health, government policy, social sciences, and art.

Tonality in a message is actually worth 38% of the content of your message and has a significant influence on the meaning of your communication and the reactions you will get. The Lexicon-based approach relies on the words in the text and the sentiment they carry.

This study is used to show a correlation between patterns of emotional expression in tweets. Identify the intensity of emotions of a citizen who expressed on Twitter regarding Bangalore Traffic. It also involves the following.

1. Extract Tweets from Twitter using Twitter API
2. Removal of special characters
3. Data cleaning/Text pre-processing
4. Tokenization
5. lemmatization
6. Stop word removal
7. Applying NRC lexicon
8. Applying text2emotion
9. Identifying the emotions from tweets

Chapter 2: Literature Review

This study is driven to develop and identify public opinion based on the emotions that are extracted from the Twitter data. There are many studies and research done in identifying the sentiments from social media platforms. Many focus on sentiments rather than emotions in the textual data.

Few studies are done on emotions by extending their study from sentiments toward emotions. Further, there are different types of approaches for each study ranging from lexical-based to machine learning based approaches. The lexical-based studies use the NRC emotion lexicon, DepecheMood, Topic-based DepecheMood, and EmoSenticNet.

The sentiment analysis done by Saif M. Mohammad using Twitter data used NRC Hashtag Sentiment Lexicon and Sentiment140 Lexicon to recognise the sentiments in the Twitter data. The study has almost 775,000 tweets which are used to create a large word–sentiment association lexicon (Mohammad, 2013).

In (Mihalcea, 2004), Strapparava provided a system to identify the emotions in headlines of news resources. The research took more than 1250 news headlines from different websites. They have annotated the data with emotions like joy, fear, anger, sadness, disgust, and surprise.

A similar analysis of emotion detection is done on Covid pandemic impact in (Amrita Mathur, 2020) and in (Huzaib Avez Sayyed, 2021). The former study identifies the sentiments and emotions on twitter data using AFINN and NRC lexicons whereas the latter examines the methods of machine and deep learning in textual data.

There are studies done on traffic analysis using deep learning methods. In (Shounak Kundu, 2020), the study is done based on LSTM and predicts the future traffic using time series analysis. Whereas, in (Ameena Padiath, 2009), the traffic analysis is done on Indian traffic using Artificial Neural Network by taking the video content of the traffic.

NRC emotion lexicon and DepecheMood lexicon are created using crowdsourcing. The words in the NRC emotion lexicon are selected from some of the most frequent nouns. The lexicon behind the classification of emotions is NRC Affect Intensity Lexicon. The lexicon has >14,000 unigrams (words) and ~25,000 word senses for eight emotions (Joy, Anger, Fear, Sadness, Disgust, Anticipation, Trust, and Surprise) and two Sentiments (positive and negative). On the other hand, the DepecheMood lexicon contains 25.3k words. Part-of-Speech tagging is applied to each word by identifying adjectives, nouns; verbs or adverbs, and a word-emotion matrix is created that contains 37k entries which are used as an emotion lexicon.

Similarly, text2emotion is the python package that helps in identifying the emotions from the content. This package will process any textual message and recognize the emotions (Angry, Fear, Sad, Happy, and Surprise) embedded in it.

In (Felipe Bravo-Marquez, 2016) Felipe and Saif Mohammad tried to create a lexicon based on word-emotion associations using Twitter data. The study identifies the multi-label classification of words extracted from unlabelled tweets such as unigrams, Brown clusters, POS tags, and word2vec embeddings which are used in classifying tweets into emotional categories.

In (Ema Kusen, 2017), extensive research is done on social media messages to identify emotions and compare different lexicon based emotion recognition techniques like NRC, DepecheMood, and EmoSenticNet. The study compares the results they obtained to the results provided by crowdsourcing.

Chapter 3: Problem Statement

Traffic is one of the biggest problems that India is experiencing. And, when it comes to Bangalore which is termed the silicon valley of India, It is the world's most traffic-congested city. There is always a topic that runs on social media platforms or in the news that talks about the traffic in Bangalore.

Many research papers and reports claim that Bangalore is facing traffic-related issues and the hottest topic of conversation in the city is 'Traffic'. There might be many reasons that vary from infrastructure, population, or the increasing number of vehicles on the road.

According to a report by World Resource Institute, an average citizen in India spends more than 240 hours stuck in jams every year. It also estimates that traffic delays cost the city 65 billion rupees a year.

People often use Social media to convey their feelings and emotions that they have during the time they spent in traffic. Twitter is one of the biggest social media platforms that influence people socially and politically.

This study is to recognize the Voice of Citizen by analyzing the emotions from the tweets that are extracted using Twitter API.

Chapter 4: Objectives of the Study

The goal of this study is to provide

1. Voice of Citizen on Bangalore Traffic by utilizing the NLP techniques and identifying the emotion behind the tweet.

2. To identify the word-emotion association that relies on the words in the text. NRC emotion lexicon and text2emotion are the two different approaches used to recognize the emotions in the Twitter data.

3. Identify the intensity of emotions of a citizen who expressed on Twitter regarding Bangalore Traffic.

This study will help in understanding the most traffic-congested locations of the city and highlight them. This can be utilized by government officials to take necessary action to divert the traffic thus reducing the total time spent by an individual in the traffic.

This study is used to show a correlation between patterns of emotional expression in tweets. This is further used to provide a dashboard to show the tweet classification and frequencies based on emotions identified.

Chapter 5: Project Methodology

The scope of this study is to identify the tweets about Bangalore traffic and analyze the data and recognize the eight emotions (Joy, Anger, Fear, Sadness, Disgust, Anticipation, Trust, and Surprise) from the data. This is further used to provide a dashboard to show the tweet classification and frequencies based on emotions identified. Figure No. 5.1 represents the system design.

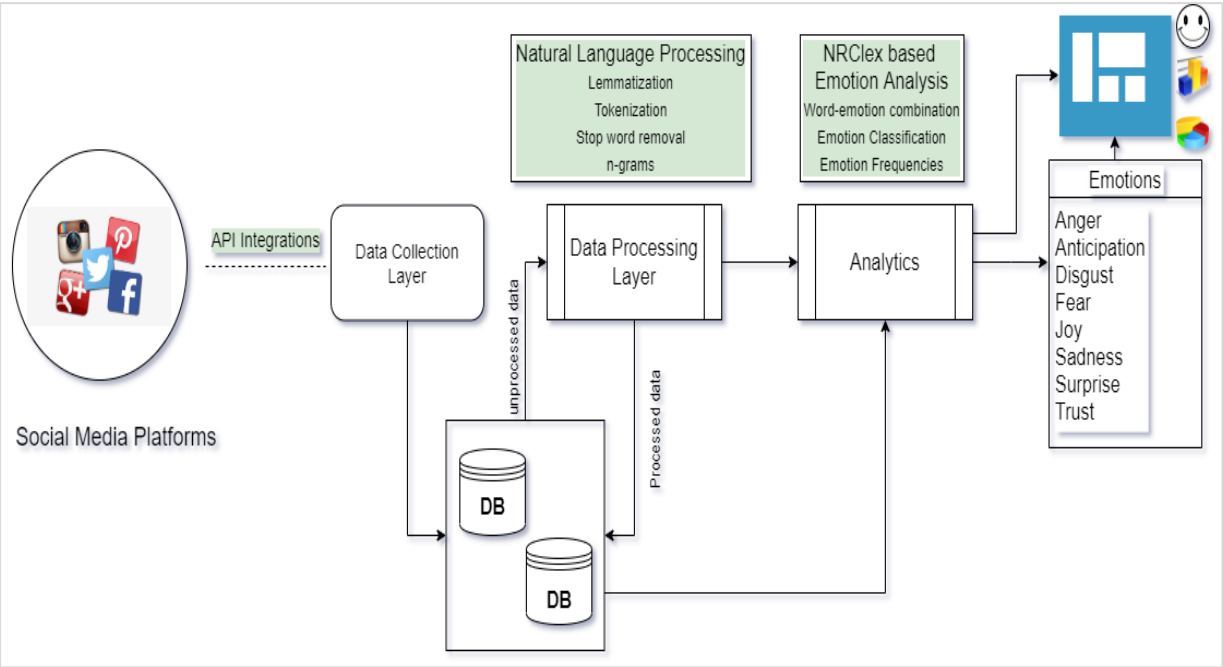


Figure No. 5.1: System Design

Data is collected from Twitter using Twitter-API. The data can be obtained using many search keywords. Once the data is extracted, it is pre-processed to remove hyperlinks and punctuations, symbols, emoticons, etc. using a regular expression to get the clean data. Figure No.5.2 represents the data flow diagram used in this study.

Tokenization: Once the data is cleaned and manually labelled, we read the dataset and create tokens (words) by splitting the words from the sentence

Normalization: Normalization helps in reducing the number of unique tokens present in the text, removing the variations in a text, and also cleaning the text by removing redundant information. Here to make the processing easy and we need almost everything from the tweet, we converted the data into lower words

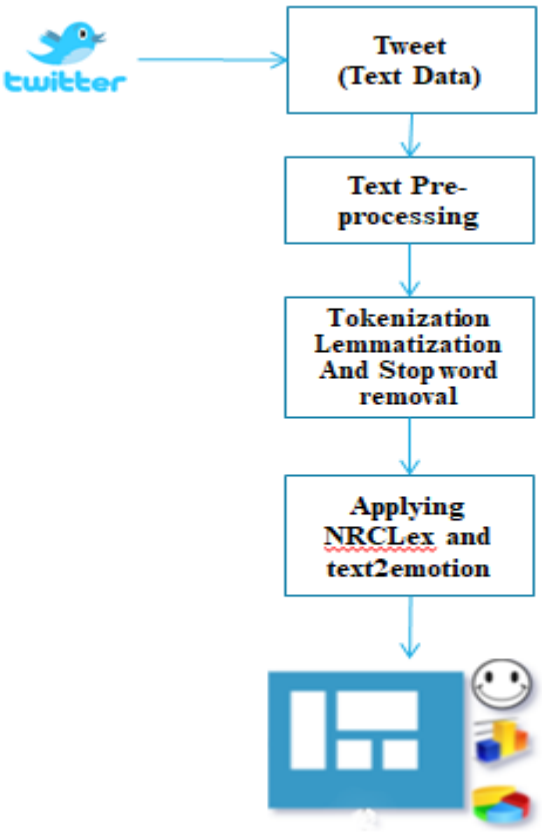


Figure No. 5.2: Data Flow Diagram

Lemmatization: The tokenized words then need to be converted to their base root form to make sense of each word.

Removal of Stop Words: Stop words are a set of commonly used words in any language. Words like ‘is’, ‘the’, ‘and’ are treated as stop words. In Natural language processing, analysis is based on the important words by removing stop words. For this study, Natural language tool kit (NLTK) and SPACY stop word libraries are used to remove the stop words. Further, the new corpus of stop words are added to the list and appended to the NLTK library. The stop words are removed from the word dictionary.

Emotion Lexicon: NRC emotion Lexicon is used on the word dictionary to identify the emotions from the word-emotion association provided by the NRC lexicon. Similarly, the words are also passed toward text2motion to identify the emotions of each tweet. Emotion classification and frequencies of emotions are identified from the analysis.

Dashboard: Further the emotions and their classification and frequency of occurrence based on Tweet are represented in a Dashboard for easy understanding.

Chapter 6: Data Collection and Pre-processing

Data Collection:

Data is collected from Twitter using Twitter API. Anyone can use Twitter and the sources of information that a Tweet can include media, individuals, officials, and others. Mixed sources of information provide a more well-rounded perspective of the impacts of the particular event and the actions being taken to deal with that event.

Twitter can accept different forms of inputs like pictures, videos, emoji's, and much content to convey a message. Twitter allows users to communicate directly in real time which helps to identify the events or incidents happening in and around.

There are various components like Username, Timestamp, Hashtag, hyperlinks, Retweets, Favorites, and location available from a tweet that can be extracted.

The data set contains 254 tweets about Bangalore traffic. Figure No. 6.1 to 6.3 are the data extracted from twitter and its components. Figure No. 6.4 and 6.5 shows the pre-processed data which inserted into database for further analysis.

```
['@CMofKarnataka @BSBommai The only way to solve traffic congestion and providing way to silk board need parallel roa… https://
t.co/Mr3HUT1MeZ',
 "@ChristinMP_TOI Sir sorry it's not like that:-)\nPl check how this 1 of d world class flyover… https://t.co/C8Z4vITnTH"]
```

Figure No. 6.1: Typical Tweet after extraction

	user	tweet	location	time
0	CodeKrafter	Getting #stuck in #traffic on every single #we...	India	2022-07-10 12:11:02
1	EarthDate2109	@kritikatwtss You still will find a place to s...	Bengaluru, India	2022-07-10 11:48:34
2	VBS_seeker	@jointcptraffic 🙏 Please do visit #SilkBoard ...		2022-07-08 15:19:19
3	VBS_seeker	@jointcptraffic Hope the new plan will deconge...		2022-07-07 17:08:15
4	Badass_Superdad	@JMawaali @whosthatmiss This traffic looks ver...	Silkboard Flyover	2022-07-07 16:15:06
5	sajhm13	Bad drains proving last straw for #SilkBoard j...	Bengaluru, India	2022-07-07 14:50:07
6	Clemenza2020	@vilakudy @minicnair Flat 50% discount it seem...		2022-07-07 03:08:49

Figure No 6.2: Tweet that contains username, location, and timestamp

	tweet_id	text	retweet_count	created_at
0	1546104710595477506	Getting #stuck in #traffic on every single #we...	0	Sun Jul 10 12:11:02 +0000 2022
1	1545427316494729216	@jointcptraffic 🙏 Please do visit #SilkBoard ...	0	Fri Jul 08 15:19:19 +0000 2022
2	1545092340749963264	@jointcptraffic Hope the new plan will deconge...	0	Thu Jul 07 17:08:15 +0000 2022
3	1545057580074307589	Bad drains proving last straw for #SilkBoard j...	3	Thu Jul 07 14:50:07 +0000 2022
4	1543608727936180224	#SilkBoard traffic 😑	0	Sun Jul 03 14:52:54 +0000 2022

Figure No. 6.3: Tweet that contains Tweet ID, retweet count, Location

Data Pre-processing:

For further analysis, the data obtained from the above mentioned sources had to be pre-processed to make it suitable for easier analysis and to provide relatable insights. Pre-processing of data involves the removal of hyperlinks and punctuations, symbols, emoticons, etc. using a regular expression to get the clean data

	Tweet
2	silkboard traffic will be the death of me
3	The capital of traffic in Bangalore Silk Board is missed in this reels I strongly condemn this baava SilkBoard
4	Suryapsingh32 satishacharya Fix your traffic first before talking about Hyderabad We can cross entire Telangana
5	Those who always crack silkboard traffic jokes they never lived in a city Called MUMBAI
6	In Bangalore who is richer than CEO The average traffic police
7	VandanaJain This is precisely what Bangalore people would not be caught dead doing You can move here find a j
8	I need help narendramodi PMOIndia Bangalore commissioner police Bangalore traffic commissioner police
9	vinodblru rajdugar dyamannavar anil lulla SWRRLY southernrailway ChristinMP TOI captsanthoshkc
10	BangaloreTrafficPolice bangaloretraffic In Bangalore most places NO Entry sign is either placed so high or hidd
11	Just returning back from Bangalore post a 3 weeks Live Project stint Mobility is a very big issue out there beyond
12	BangaloreMirror Intresting hope it s implemented on ground responsibly Hope Bangalore traffic officers rise to th

Figure No. 6.4: Tweet after Pre-processing

The data is then inserted into the database for further analysis.

	Tweet character varying
1	Reminds me of Silkboard 2018 a hilarious short film about Bangalore traffic
2	Getting stuck in traffic on every single weekends in front of pheonixmall We are having our on silkboard mome
3	kritikatwtss You still will find a place to stay But the greatest hazard is the traffic snarl Pre Covid was wors
4	jointcptraffic Please do visit SilkBoard Jn during the week end You will find Traffic jam from Madivala
5	jointcptraffic Hope the new plan will decongest Hebbal Jn I would now like to draw your attention to give soluti
6	JMawaali whosthatmiss This traffic looks very similar to Bannerghatta Silkboard junction amp Koramangala
7	Bad drains proving last straw for SilkBoard junction s traffic Bengaluru Bangalore NammaBengaluroo envyk blr
8	vilakudy minicnair Flat 50 discount it seems Caused mega traffic jam in TVM Silkboard type jam as per my in laws
9	I finally understood the meaning of this meme after experiencing evening traffic from Belandur to Silkboard
10	DgpKarnataka CPBlr jointcptraffic btppubliceye DcpComdCentre Please rectify Public eye mobile app If that is
11	SilkBoard traffic
12	akm1410 I don t see any improvement in Koramangala or Silk board traffic so No don t agree Also we still aren

Figure No. 6.5: Tweets inserted into the database

Chapter 7: Data Modelling and Analysis

Each tweet is converted to word level to identify the Tokens. Further, each token is lemmatized to convert to its base form to make out the meaning of the word. We also removed the stop words which are not required for lexicon analysis.

Figure No. 7.1 is a Word Frequency plot used to identify the most frequent words. This helps us in identifying which location or area is the most tweeted for traffic congestion.

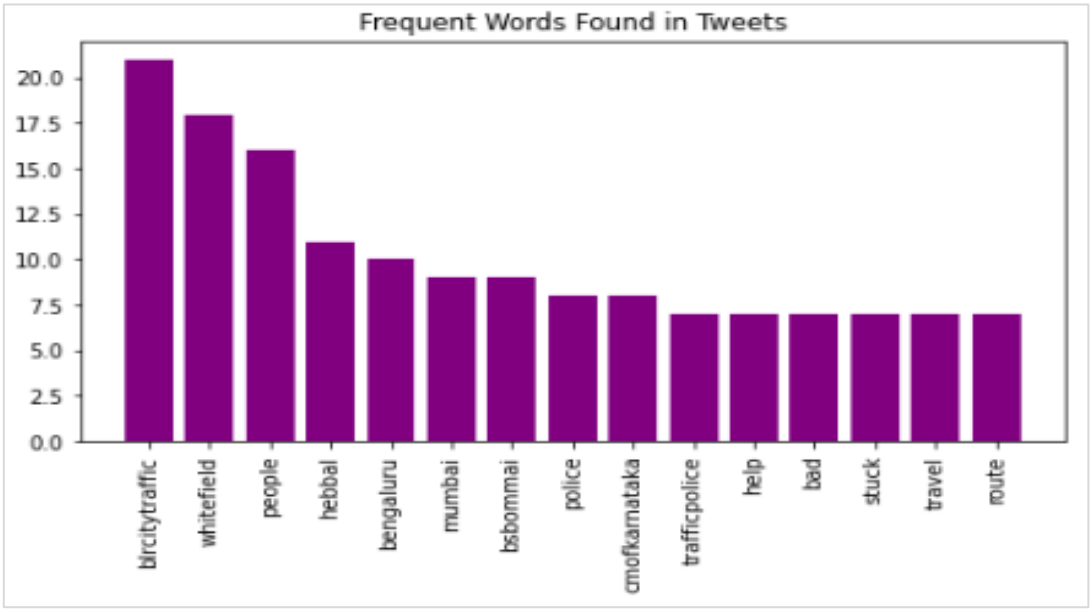


Figure No. 7.1: Word Frequency plot

From Figure No. 7.1, findings that can be observed is that Whitefield and Hebbal localities are more congested areas than Silk board which is most famous for its traffic.

Similarly, a word cloud is plotted for the same as shown in Figure No. 7.2. The word cloud is created using the same word frequency dictionary. NLTK library is used to sort the dictionary and create a word cloud.

Figure No. 7.2: Word Cloud based on frequent words in all Tweets

The tweets once cleaned for any unwanted words, removing hashtags, hyperlink, and stop words are then used to recognize the emotion associated to each word using NRCLex, NRC word-emotion association lexicon, and text2emotion.

For NRCLex and NRC word-emotion association lexicons, it is identified how many words of each tweet are matched to the lexicons that are available in the NRCLex and calculates the frequency of occurrences of each word-emotion association. And then, calculate the number of Tokens in each tweet, the number of lexicon tokens that are present in the emotion-word lexicon, and the lexicon ratio which is the number of lexicon tokens/ number of Tokens. Based on these the average lexicon value is calculated which is treated as the emotion score of each tweet.

A similar method is followed for text2emotion. The tweet is tokenized and lemmatized and stop words are removed from it to identify the emotion of each word. Further, the higher frequency of word-emotion association is selected as the overall emotion of the tweet.

Figure No.7.3 depicts a tree map that shows the total number of words that are associated to the lexicons in the NRCLex.

Figure No. 7.3: Tree Map for Emotion classification from all Tokens

The total words from the tweet are matched to the total lexicon tokens from the NRCLex lexicon and created an emotion classification to identify how the words from the data are classified based on the emotions. Figure No. 7.4 below shows the classification of emotions based on tokens from the Tweet.

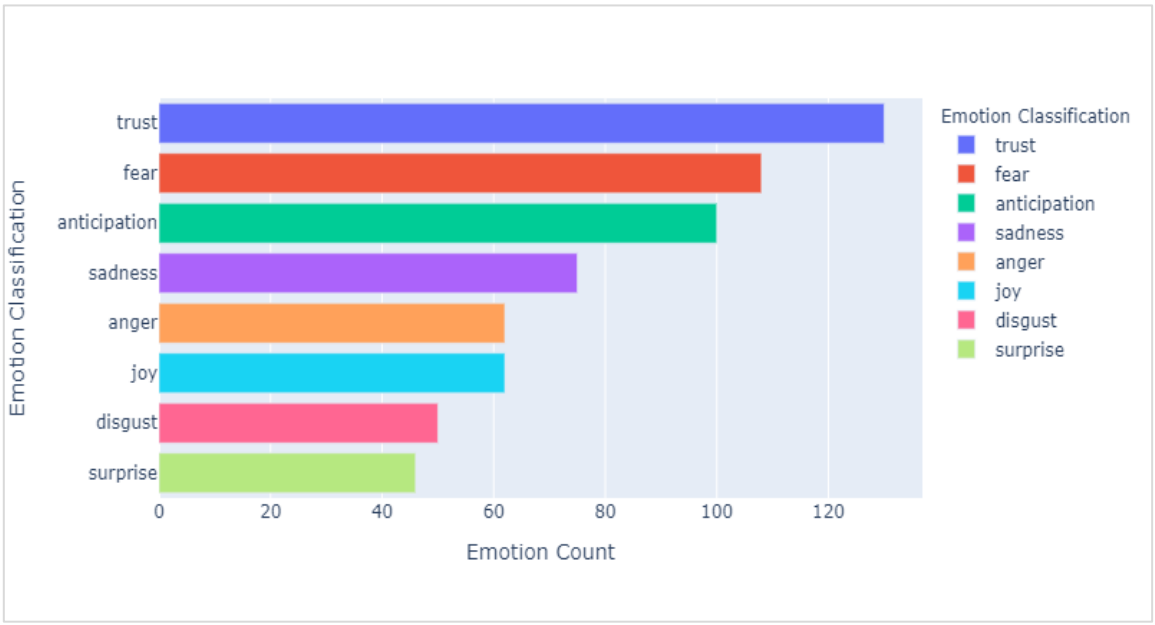


Figure No. 7.4: Emotion Classification based on Tokens

A total of 2622 tokens are available after normalizing and removing stop words from about 254 tweets. Out of which 633 words are associated with one of the eight emotions available in the NRCLex lexicon.

From Figure No. 7.4, it is observed that more than 20% of the words are associated with the emotion of Trust. And, around 17% and 15% of the words from tweets are associated with emotions Fear and anticipation respectively. From this, it is understood that many tweets are either related to Trust or Fear about the Bangalore traffic.

The combined emotions of Anger, Fear, Sad, Disgust, and Anticipation accounted for almost 63% of words from the tweets which convey that people are not happy with the traffic in Bangalore.

Similarly, the Frequency of words from the data is calculated based on how many times a word from the tweet has appeared in the lexicon words. For every lexicon word matched with the token from the tweet, the emotion is calculated and the frequency is calculated based on the total number of words of the tweet and the number of times the token matched with the lexicon word from the NRCLex word-emotion association lexicon. Figure No. 7.5 shows the emotion frequencies based on tokens

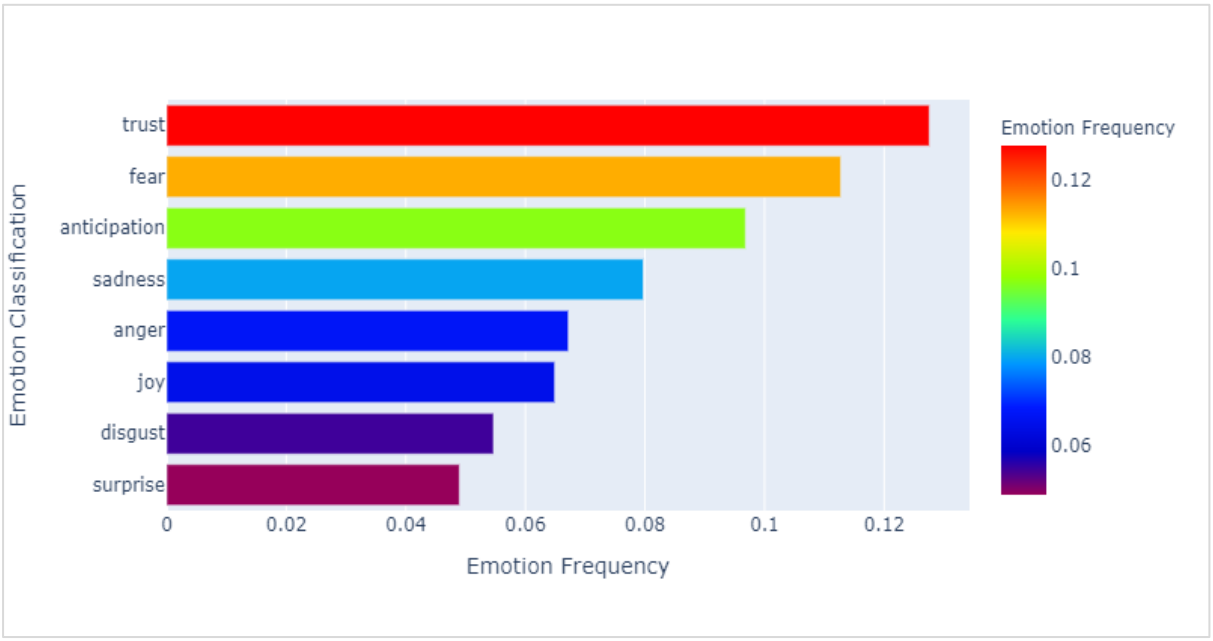


Figure No. 7.5: Emotion Frequencies based on Tokens

Further, overall Emotion is calculated in each Tweet by taking the higher frequency of emotions recognized in each tweet.

Figure No. 7.6 represents the total number of tweets and their associated emotions calculated based on the frequency of occurrences of the emotion from the word-emotion association identified from each word of the tweet.

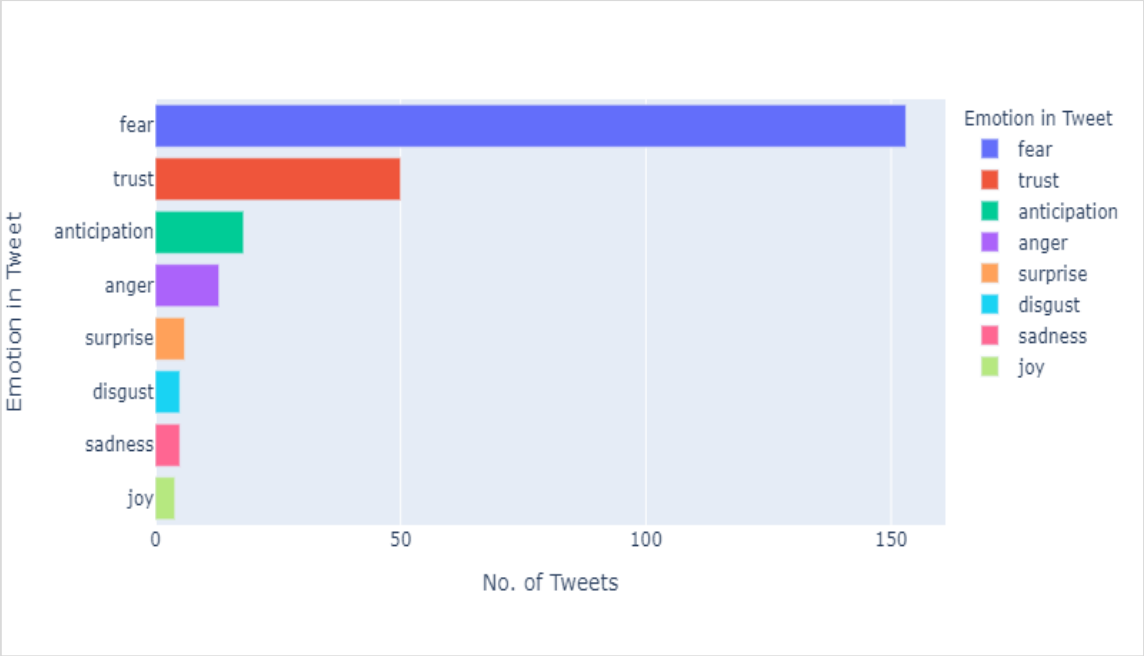


Figure No. 7.6: Classification of Emotions in Tweets using NRCLex

From Figure No. 7.6, it is observed that almost 65% of the tweets have emotions as Fear. Most of them tweeted that they get fear to travel in Bangalore traffic because of the time they have to spend in the traffic.

Also, it is observed that there are almost 20% of the tweets have emotions as Trust and almost 8% of tweets have emotion as Anticipation. This is because of the new rules and infrastructure development that the government or the traffic police are taking action into. There are less than 4% of tweets that has Emotion as Joy.

A similar analysis is done for the text2emotion library. The Emotion is identified for each tweet based on the higher frequency of word-emotion association from each tweet. The text2emotion contains only 5 emotions Anger, Fear, Sad, Happy, and Surprise.

Further, the emotion count is plotted for all the tweets from the data. Figure No. 7.7 represents the total number of tweets and their associated emotions calculated based on the frequency of occurrences of the emotion from the word-emotion association identified from each word of the tweet.

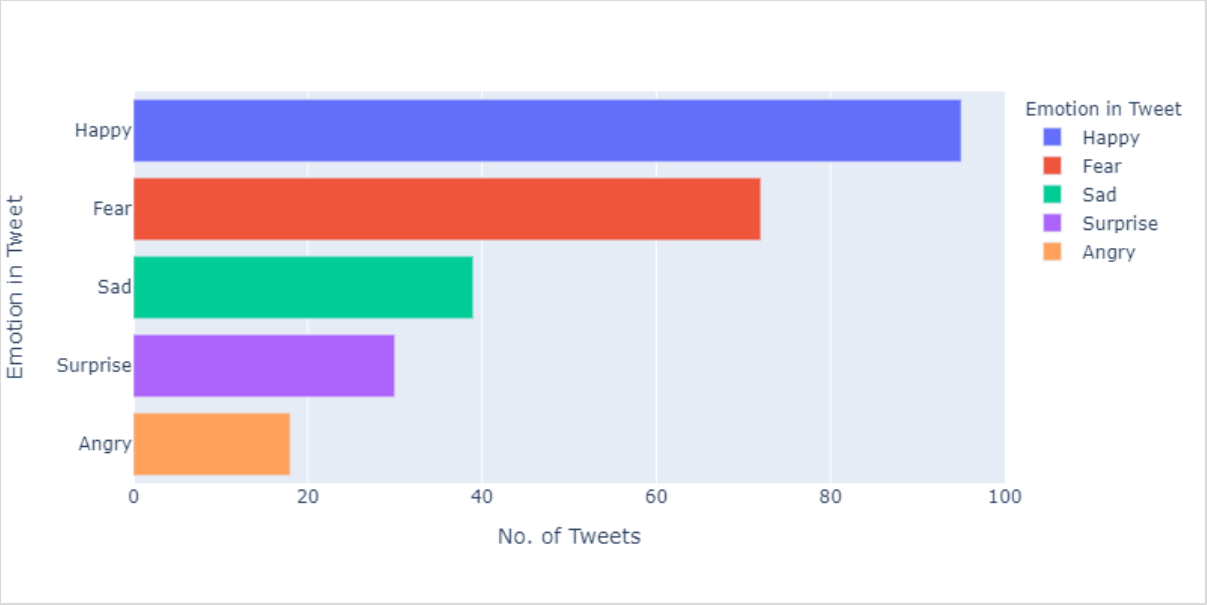


Figure No. 7.7: Classification of Emotions in Tweets using text2emotion

From Figure No. 7.7, it is observed that almost 37% of the tweets have emotions as Happy. And 29% of tweets have emotions as Fear.

Table No. 7.1 shows the comparison of emotions between NRCLex and text2emotion on the total number of tweets extracted via Twitter API. The comparison explains that most of the words in the NRC lexicon are associated with Fear followed by Trust whereas, in text2emotion, the lexicons are associated with Happy.

Total Emotions on Tweets		
Emotion	NRCLex	text2emotion
Fear	153	72
Trust	50	
Joy	4	95
Sad	5	39
Anticipation	18	
Surprise	6	30
Angry	13	18
Disgust	5	

Table No. 7.1 Comparison of emotions between NRCLex and text2emotion

Metabase Dashboard:

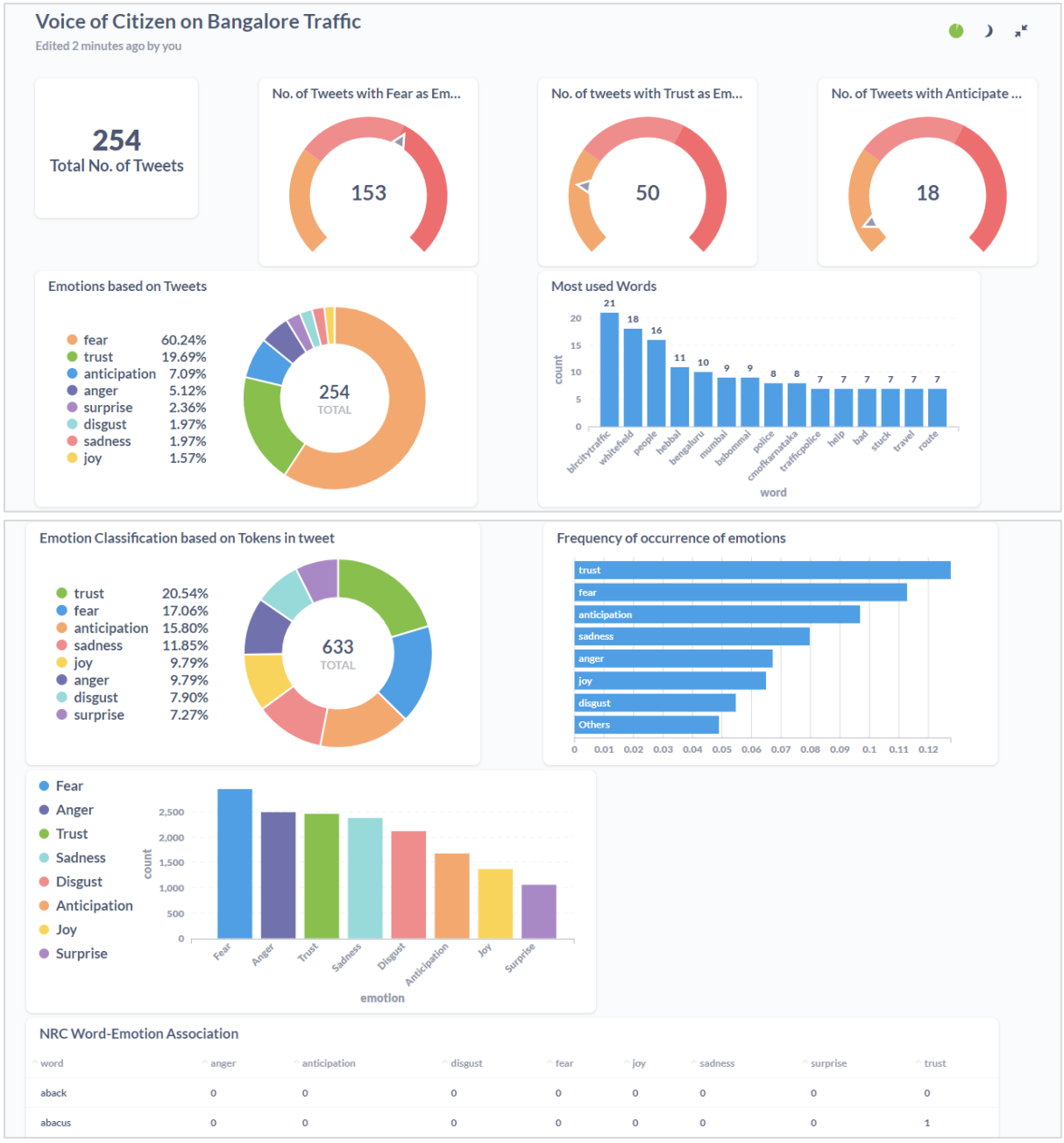


Figure No. 7.8: Metabase Dashboard on Voice of Citizen

Chapter 8: Conclusions and Recommendations for future work

From this study, it is analyzed that most of the tweets carry an emotion of Fear followed by Trust. This tells that the traffic congestion is very high in Bangalore which needs attention by government officials and traffic police.

It is also noted that most of the words in the NRC lexicon are associated with Fear and Anger followed by Trust. This also makes the analysis inclined towards those emotions. When compared to text2emotion, the word-emotion associations in text2emotion are mostly towards Happy. So, the number of words available in the lexicon does make a necessary impact on the analysis.

The Dashboard can be utilized by government officials and traffic police to analyze the tweets and identify the most traffic congested areas and take necessary action in diverting the traffic or make necessary infrastructure development to reduce the congestion.

Bibliography

Ameena Padiath, L. V. (2009). Prediction of Traffic Density for Congestion Analysis under Indian Traffic Conditions. *12th International IEEE Conference on Intelligent Transportation Systems*, 7.

Amrita Mathur, P. K. (2020). Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19. *Fifth International Conference on Communication and Electronics Systems (ICCES 2020)*, 5.

Bangalore has the worst traffic in the world. (2020, January 29). Retrieved from Live mint: https://www.livemint.com/news/india/bengaluru-has-the-worst-traffic-in-the-world-report-11580278531554.html

Bangalore Traffic Report | TomTom traffic Index. (2019, December 10). Retrieved from TomTom: https://www.tomtom.com/traffic-index/bengaluru-traffic/

BBC News. (2016, December 7). Retrieved from Why is Bangalore stuck in traffic jams?: https://www.bbc.com/news/world-asia-india-38155635

Dev, A. (2021, September 04). *35% increase in private vehicles on Bengaluru roads: Officials*. Retrieved from Hindustan Times: https://www.hindustantimes.com/india-news/35-increase-in-private-vehicles-on-bengaluru-roads-officials-101630695800757.html

Ema Kusen ˇ 1, G. C. (2017). Identifying Emotions in Social Media:Comparison of Word-emotion Lexicons. *5th International Conference on Future Internet of Things and Cloud Workshops*, 6.

Ema Kusen, G. C. (2017). Identifying Emotions in Social Media : Comparison of Word-emotion Lexicons. *5th International Conference on Future Internet of Things and Cloud Workshops*, 6.

Felipe Bravo-Marquez, S. M. (2016). Determining Word–Emotion Associations from Tweets by Multi-Label Classification. *2016 IEEE/WIC/ACM International Conference on Web Intelligence*, 4.

Huzaib Avez Sayyed, D. S. (2021). Study and Analysis of Emotion Classification on Textual Data. *6th International Conference on Communication and Electronics Systems*, 6.

Mihalcea, S. a. (2004). WordNet-Affect: an Affective Extension of WordNe. *ITC-irst, Istituto per la Ricerca Scientifica e Tecnologica*, 4.

Mohammad, S. M. (2013). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. *National Research Council Canada*, 7.

Mulukutla, A. M. (2015). *When it Comes to Reducing Car Congestion, India's Cities Can Learn from its Businesses.* Bangalore: World Resources Institute.

Shounak Kundu, M. S. (2020). Traffic Forecasting with Deep Learning . *IEEE Region 10 Symposium*, 6.

Appendix

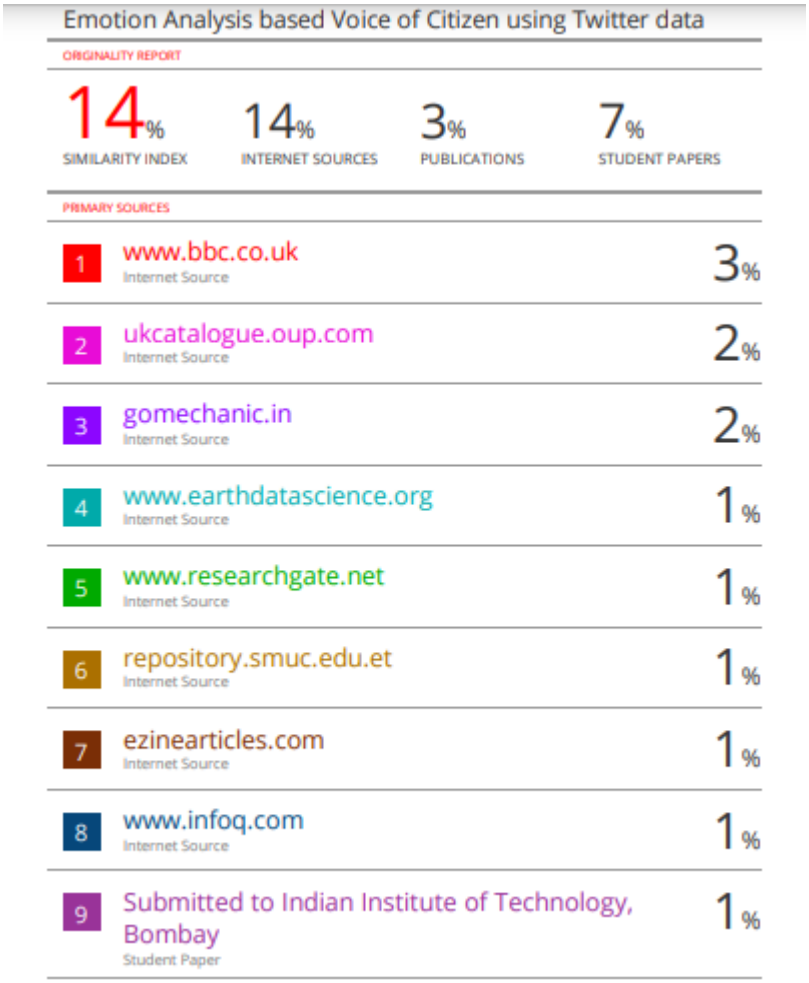
Plagiarism Report[1]

Similarity Index – 14%

Internet Sources – 14%

Publications – 3%

Student Papers – 7%



Emotion Analysis based Voice of Citizen using Twitter data

ORIGINALITY REPORT

14%	14%	3%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.bbc.co.uk Internet Source	3%
2	ukcatalogue.oup.com Internet Source	2%
3	gomechanic.in Internet Source	2%
4	www.earthdatascience.org Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	repository.smuc.edu.et Internet Source	1%
7	ezinearticles.com Internet Source	1%
8	www.infoq.com Internet Source	1%
9	Submitted to Indian Institute of Technology, Bombay Student Paper	1%

[1] Turnitn report to be attached from the University.

10	doi.org Internet Source	1%
11	Submitted to Bahrain Polytechnic Student Paper	<1%
12	Lecture Notes in Computer Science, 2013. Publication	<1%
13	medium.com Internet Source	<1%
14	www.cs.toronto.edu Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		