

Mini Project Report:

Submitted in partial fulfillment of the requirements
of the degree
**BACHELOR OF ENGINEERING IN COMPUTER
ENGINEERING**

By

Gautam Wadhwani/65

Neeraj Chawla/11

Tanmay Thakare/62

Yash Kewlani/32

Supervisor

Prof. Sujata Khedkar



Department of Computer Engineering

Vivekanand Education Society's Institute of Technology

HAMC, Collector's Colony, Chembur,

Mumbai-400074

University of Mumbai

(AY 2021-22)

CERTIFICATE

This is to certify that the Mini Project entitled “ **Brand Analytics using Twitter Data** ” is a bonafide work of **Gautam Wadhwani (65), Neeraj Chawla(11), Tanmay Thakare(62), Yash Kumar Kewlani(32)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of “**Bachelor of Engineering**” in “**Computer Engineering**” .

(Prof. Sujata Khedkar)

Supervisor

(Prof. _____)

Head of Department

(Prof. _____)

Principal

Mini Project Approval

This Mini Project entitled “**Brand Analytics using Twitter Data** ” by **Gautam Wadhwani (65), Neeraj Chawla(11), Tanmay Thakare(62), Yash Kumar Kewlani(32)** is approved for the degree of **Bachelor of Engineering in Computer Engineering**.

Examiners

1.....
(Internal Examiner Name & Sign)

2.....
(External Examiner name & Sign)

Date:

Place

Contents

Abstract	i
Acknowledgments	ii
List of Figures	iii
1 Introduction	1
1.1 Introduction	
1.2 Motivation	
1.3 Problem Statement & Objectives	
1.4 Organization of the Report	
2 Literature Survey	11
2.1 Survey of Existing System	
2.2 Literature Survey Chart	
2.3 Mini Project Contribution	
3 Proposed System (eg New Approach of Data Summarization)	18
3.1 Introduction	
3.2 Architecture/ Framework	
3.3 Algorithm and Process Design	
3.4 Details of Hardware & Software	
3.5 Experiment and Results	
3.6 Conclusion and Future work.	
References	32

Abstract:

Twitter is one of the most running social media platforms. Twitter is widely used in our day to day life. We often use tweets to analyse the data for Data Analytics(DA) which creates a great scope for the businesses to enhance in their business as well as they provide better help to customers. The project works by using Vader which is an algorithm for sentiment analysis using python for two automobiles brands which are Mahindra XUV and TATA Safari. Comparison is done between both brands which helps to analyse which one is better . The ML algorithms such as Logistic Regression , Naive Bias ,SVM and Random Forest are used and analysis of which algorithm is best is done. Preprocessing of tweets and use of word cloud and doing exploratory analysis of the tweets.

Acknowledgement:

We thank our project guide Dr Sujata Khedkar and our colleagues who had put their whole soul into the project. We would like to thank Mr Sanjay Mirchandani also for giving advice and reviewing our project.

List of Figures:

- 1)DFD Level 0
- 2)DFD Level 1
- 3)DFD Level 2
- 4)Tweets Extraction
- 5)Vader Algorithm
- 6)Exploratory Analysis
- 7)Machine Learning Algorithms
- 8)Literature review

Chapter 1:Introduction

1.1 Introduction:

Social media has become a part of our lifestyle. We discuss a lot of things on it which includes commenting, expressing views,etc. This social media data can be used to study and analyze people's view on a particular brand.

- Automobile industry is a widely popular industry and is rapidly growing.
- Providing analysis based on social media data on particular automobile model will help the buyers and also the company owners.
- We will be using twitter data for analysis as it is one of the trusted growing social network and also allows researchers to access data.

1.2 Motivation:

Data Analytics and Machine learning are the most largely growing industries in our era. With the support of our Mentor , which ignites a great motivation for the team and also all of us united well as a team , everyone was eager to learn and wanted to do their best . Twitter helps us for getting data and also the analysis part about the project is quite fascinating.

1.3 Problem Statement and Objective:

Problem statement:

- To provide customers with the information about automobile brands, its popularity among people, analysis based on views of people on the brand. This will help the customer to make a right choice while purchasing it.
- To provide information to company owner about the views of people on their brand and people's expectations from their brand. This will help the owners to save their product from becoming flop.

Objectives

- To analyze people's perception about the brand of Automobile company and its product.
- To provide analysis of particular automobile brand using various Machine Learning algorithm and present the statistics.
- To compare between two automobile brands
Mahindra XUV 700 and Tata Safari
- To provide pre launch and post launch views of people on the brand.
- Our Project covers two sustainable goals,[8] Decent Work and Economic Growth,[9] Industry, Innovation and Infrastructure.

1.4 Organization of Report:

Chapter 1:Includes the introduction to the report, the problem statement and our reasons for selecting this project.

Chapter 2: Includes various literature surveys related to twitter analysis and

algorithms for analysing the data with also preprocessing the tweets.
Chapter 3: includes the detailed explanation of our project with the various tests we conducted and the conclusions we derived from it. Lastly, we give the results of the project.

Chapter 2: Literature Survey

2.1 Survey of existing system:

Efendi et al.[1], the authors of this paper used twitter data to analyse the popularity of automobile brands in Indonesia. The study involves both the cars and motorcycles industry. After determining the popularity scores of the brands, they compared them with the sales report released on various automobile websites to see if there is a correlation between the popularity and sales of the brand. They found that people's interest towards a brand makes it more of a deciding factor for them to buy a Japanese car and motorcycle. Whereas features come second.

Limitations: Twitter data showed inaccuracies in reflecting the real world. The result of the analysis is also inconsistent. Sentiment analysis also affected the result greatly in a negative way and reduced the Pearson correlation score.

Shukri et al.[2], They Compared three Automobile Brands(Audi,BMW,Mercedes) using twitter API and other social media sites.In that they collected nearly 3000 tweets, 1000 each brand. In that they categorized it in three types of tweets that are positive, negative and neutral, based on the users reviews about that brand.

In their research, they found out that the positive percentage of Audi was more than other brands and also the negative score of Audi was less as compared to other brands.

Limitations: The survey was limited to Europe only, also regional languages twitter data were not taken for survey. Also which part of Europe they have covered was not mentioned.

Brain et al.[7], Their main aim was to understand how twitter affects the tourism field. They made an in-depth analysis of tweets by dividing into different types of user groups and analysed using network mapping. Their exploratory study aimed to contribute to tourism and hospitality industry research and thus benefiting the academics by introducing Twitter as a data source and social media analytics as a tool for big data analysis.Cruise related tweets were collected and social media analytics were used to extract valuable intelligence regarding Twitter usage and groups.

Limitations: Methodological issues, specifically validity and reliability, have not yet been tested for user generated content. Also this study remained largely exploratory.

Rathore et al.[8], the authors compared individually the pre launch and post launch sentiments of three different brands and showcased their popularity through charts and diagrams. They compared users' emotions before and after the launch of three new products in the market—a pizza, a car and a smartphone—for possible inputs for NPD. Their analysis showed that for the car and the phone, firms need to focus on user attitudes towards product attributes, whereas for pizza, firms should concentrate on physiological changes and promotional sides.

Limitations: They did not analyse cross cultural data. Also the study did not include comparison between the brands under the same domain.

Vidya et al.[6], This paper measures the brand reputation based on customer satisfaction

through customer's sentiment analysis from Twitter data. For this they have taken five mobile phone providers. the sentiments are classified and compared using three different algorithms: Naïve Bayes, Support Vector Machine, and Decision Tree classifier method. They measure customer satisfaction on five products: 3G, 4G, Short Messaging, Voice and Internet services. This paper also discusses some correlated business insights in a telecommunication services industry. The results showed that SVM gives better performance than other two classifiers (Naïve Bayes and Decision Tree) in time processing and accuracy.

Limitations: The search was limited to a particular region only. They compared only a few features of mobile phones.

Azgar et al.[5], Analysis of sentiments is the fastest growing field that utilizes text mining, computational linguistics and natural language processing, linguistic mining of text and calculation to extricate valuable data to assist in decision making. According to the classification of the polarity, the paper concluded that Audi has 87% of the positive tweets compared to 74% for BMW, 84% for Honda, 70% for Toyota and 81% for Mercedes. What's more, the results demonstrate that Audi has negative polarity 18% against 10% for BMW, 20% for Mercedes, 15% for Honda and 25% for Toyota. All the analysis was done using Naive Bayes.

Limitations: Naive Bayes is not as efficient as the other algorithms of machine learning such as Random forest , SVM or Logistic Regression.

Fumiki et al.[3], This paper has explained two keyword count analysis algorithms they are BM 25 and TF-IDF. Both focused on how many times a word occurred in a document which helps to know the relatability of the document with respect to keywords. From the paper we concluded that BM 25 was more efficient than TF-IDF. This thing could be used in tweets such a way that we could know whether the tweets are relatable or not .

Limitations: TF-IDF was less effective .

Wisdom et al.[10], This paper helped us to learn how to use twitter API keys for extraction of data, like how to collect and access data, after that storing the collected data. Then preparing the data for pre processing like tokenizing the tweet, removing stop-words. They explained the algorithms for analyzing the data and creating bigrams and finding out most used mentions and hashtags. And various techniques of analyzing streaming data and visualized it on the world map.

Limitations:- It is not a beginner friendly paper it took time to understand what was the author was trying to say.

2.2 Literature Survey chart:

Sr. No.	Title of paper	Details	Description	Limitations
1.	Study of Automotive Brands Popularity in Indonesia Using Twitter Data	Authors: Stevent Efendi, Alva Erwin, Kho I Eng Swiss German University, Indonesia Published on:	Their objective was to prove whether it is possible to predict popularity of automobile brands in Indonesia using twitter data and comparing them with the	The result of the analysis is inconsistent. Sentiment analysis affected the result greatly in negative way and reduced the Pearson

		25/4/2016	sales report.[1]	correlation score.
2.	Twitter Sentiment Analysis: A Case Study in the Automobile Industry	Authors: Sarah E. Shukri; Rawan I. Yaghi; Ibrahim Aljarah; Hamad Alsawalqah Published in: 2015 IEEE Jordan Conference on AEECT	They Compared three Automobile Brands(Audi,BMW,Mercedes) using twitter API. In that they categorized it in three types of tweets that is positive, negative and neutral type based on the users reviews about that brand.[2]	The survey was limited to Europe only, also regional languages twitter data were not taken for survey.
3.	Study of algorithm For the analysis of count for keywords	Authors : Fumiaki Saitoh, Fumiya Shiozawa, and Syohei Ishizu. Department of Industrial and Systems Engineering, College of Science and Engineering	The document file has explained the comparison between the algorithms like BM25 and TF-IDF.[3]	The TF-IDF formula is less efficient than the BM25 method.
4.	Internet Marketing by Automobile Industry: Special Reference of Indian Counterparts	Authors : Teena Bagga, Programme Coordinator, Amity Business School and Deepak Gupta, B.Tech, Mechanical and Automation Engg. MBA	Internet Marketing is being widely used as a tool in the Automotive sector for imparting mass scale knowledge of their products and for large scale promotions.[4]	Increased use of promotional strategies as a result of increasing competition.
5.	Using Twitter Data for Cruise Tourism Marketing and Research	Author: Seunghyun “Brian” Park and Chihyung “Michael” Ok and Bongsug “Kevin” Chae.	Their main aim was to understand how twitter affect the tourism field. They made an in-depth analysis of tweets by dividing into types of user groups.	This study remained largely exploratory. Methodological issues, specifically validity and reliability, was not yet been tested for user generated content.

6.	Pre-andpost-launch emotions in new product development: Insights from twitter analytics of three product	Authors:Ashish Kumar Rathore, P. Vigneswara Ilavarasan	In this, they compared individually the pre launch and post launch sentiments of three different brands and showcased their popularity through charts and diagrams.	They did not analysed cross cultural data. Also the study did not include comparison between the brands under same domain.
7.	Analysis of Sampling Algorithms for Twitter	Deepan Palgunay, Vikas Joshiz, Venkatesan Chakaravarthy, Ravi Kothariz and L V Subramaniam ySchool of ECE, Purdue University, Indiana, USA	This paper had sentiment analysis algorithm .This paper had kept the count of tweets and number of words in a tweet which showed whether the word is positive ,negative or neutral.	They did not built a multi lingual based sentiment classification to acquire tweets in different languages and perform analysis.
8.	Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers, Procedia Computer Science	Author:Nur Azizah Vidya, Mohamad Ivan Fanany, Indra Budior	This paper measures the brand reputation based on customer satisfaction through customer's sentiment analysis from Twitter data.For this they have taken five mobile phone providers .	The search was limited to a particular region only. They compared only few features of mobile phone.
9.	Sentiment Analysis on Automobile Brands Using Twitter Data	Zain Asghar ¹ , Tahir Ali ² , Imran Ahmad ³ (&), Sridevi Tharanidharan ⁴ , Shamim Kamal Abdul Nazar ⁵ , and Shahid Kamal ⁶ ¹ University of Central Punjab, Lahore, Pakistan ² Gulf University of Science and Technology, Kuwait City, Kuwait ³ Riphah International University, Lahore,	This paper the Algorithm Naive Bayes theorem was mentioned,which concluded Positive polarity of Audi was highest than of Honda,BMW,Toyota and Mercedes. This helped the customers to plan for the car to be selected.	The negative calculation made the people feel that the brand is totally useless which creates Psychological impact to the people.

		Pakistan		
10.	An introduction to Twitter Data Analysis in Python	Vivek Wisdom, Rajat Gupta Artigence Inc {Vivek, Rajat}@artigence.in	This paper helped us to learn how to use twitter API keys for extraction of data.	It is not a beginner friendly paper it took time to understand what was the author was trying to say

2.3 Mini Project Contribution:

In the project we have used twitter data for analysing sentiments of people for two automobile brands which are TATA Safari and Mahindra XUV 700 . We have used various processes of data pre-processing, data filtering and exploratory analytics and finally the output is in the form of different graphs such as comparison bar graphs ,word clouds of different shapes, community detection ,greed optimization , emotional analysis. The project was achieved by reading 15 insightful papers and some resourceful videos.

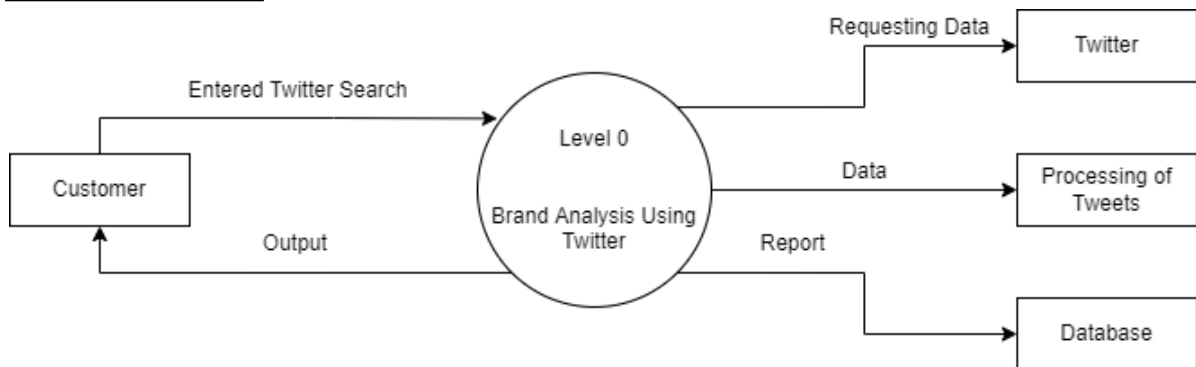
Chapter 3:Proposed System

3.1 Introduction:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. One can figure out which product of which brand is better by just using data visualization. A code for extracting tweets is used. The project uses Vader algorithm for sentiment analysis and R programming for the graphs and visualization for the tweets . Where word clouds, bar graphs, histograms and many more designs are done.

3.2 Architecture/Framework:

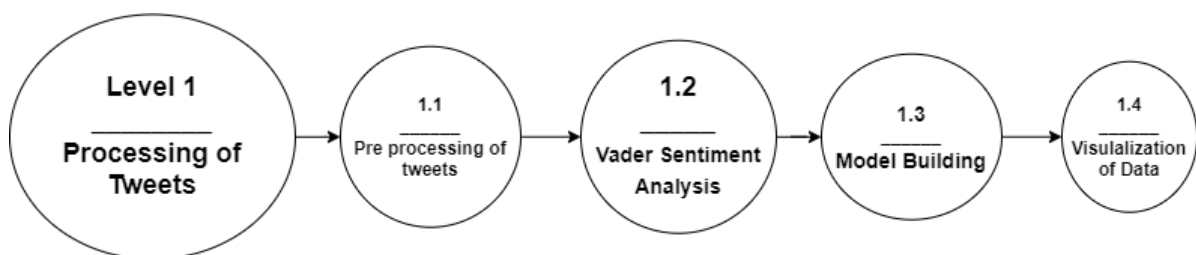
DFD Level 0:



DFD level 0

This is about how to extract tweets and process them, for example removing emojis , non-ascii characters ,etc. After the processing the final report will be going in the database and it will show as an output to the customer.

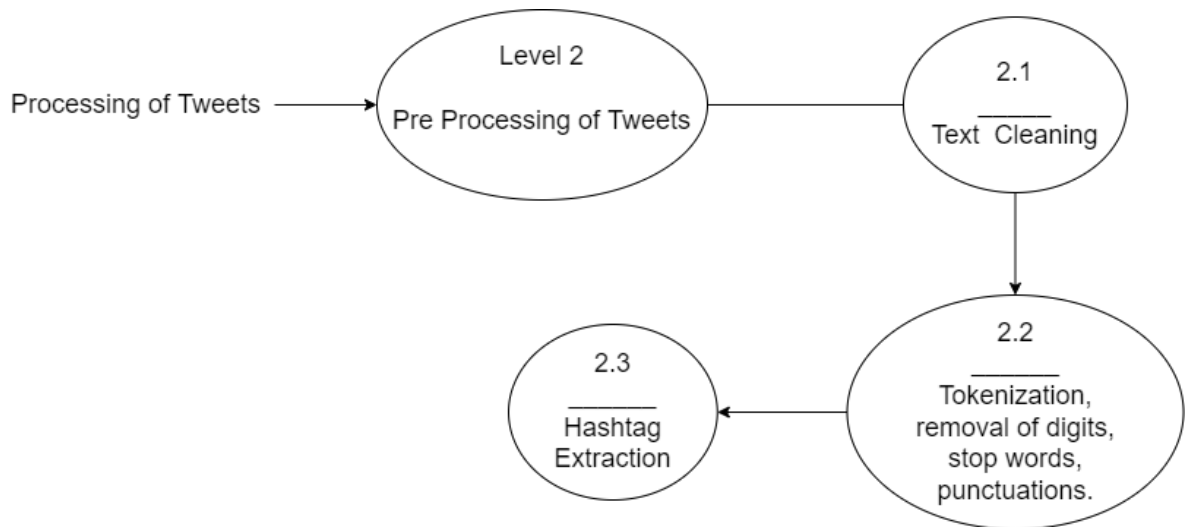
DFD Level 1:



DFD level 1

Here the sentiment analysis will be done using a vader algorithm which will tell people reaction on the tweets of brands whether they are positive or negative, and then finally visualization using bar graphs and spreadsheets will be done.

DFD Level 2:



DFD level 2

It explains the pre-processing part in detail which includes text cleaning ,tokenization, removal of digits ,stopwords and punctuation marks.

Data Extraction:

Developer's account was created on twitter and API keys were obtained. In order to authorize our app to access Twitter on our behalf, we used the OAuth Interface. Tweepy is used as a tool to access twitter data with python. The search was made for particularly two automobile brand.1) Mahindra XUV700 and 2) Tata Safari.

The tweets were extracted along with the username and location. The extracted data was stored in a dataframe using pandas. The data frame consists of three columns: user, location and tweets. This dataframe was downloaded as a csv file.

Preprocessing:

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format. The tweets contain many unwanted symbols and words. this unwanted data has to be cleaned. In data cleaning, punctuation, numbers and special characters are removed many tweets contain url links and twitter accounts in the tweets. These are also removed. Thus Data preprocessing prepares raw data for further processing.

Sentiment analysis:

The processed data is then sent for sentiment analysis. We have particularly used Vader sentiment analysis.

VADER SENTIMENT ANALYSIS

VADER stands for Valence Aware Dictionary and sentiment Reasoner. Vader performs well for the analysis of sentiments expressed in social media. These sentiments must be present in the form of comments, tweets, retweets, or post descriptions, and it works well on texts from other domains also. VADER is a lexicon and rule-based analysis tool. Using Vader we provide scores called polarity_score to every tweet. The polarity scores include positive, negative neutral and compound. The Compound score is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1 (most extreme negative) and +1 (most extreme positive).

Based on the polarity scores various graphs and diagrams were plotted which makes comparison between the two car models for different polarity scores.

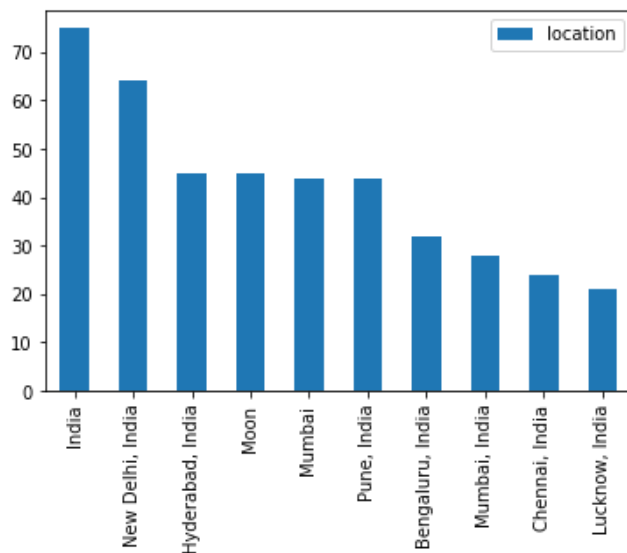
Then based on the compound score we labeled the tweets as negative, neutral and positive in a new column named “analysis”.

(1)

From the graph (1) above, we infer that mahindra XUV 700 has greater mean positive, mean negative and mean Compound as compared to that of Tata Safari. However the mean neutral for Tata Safari is greater than Mahindra XUV 700.

Also individually for both car model has greater mean neutral followed by mean compound, mean positive and lastly mean negative.

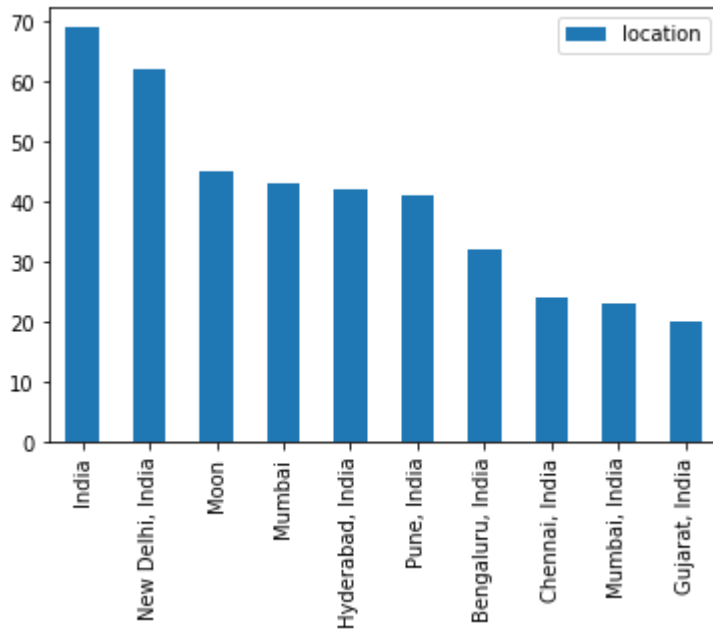
Also graphs showing no. of tweets at different locations was plotted.



(2.a) Top 10 locations for tweets for both brands

Graph 2.a shows number of tweets(both brands) as per top 10 locations.

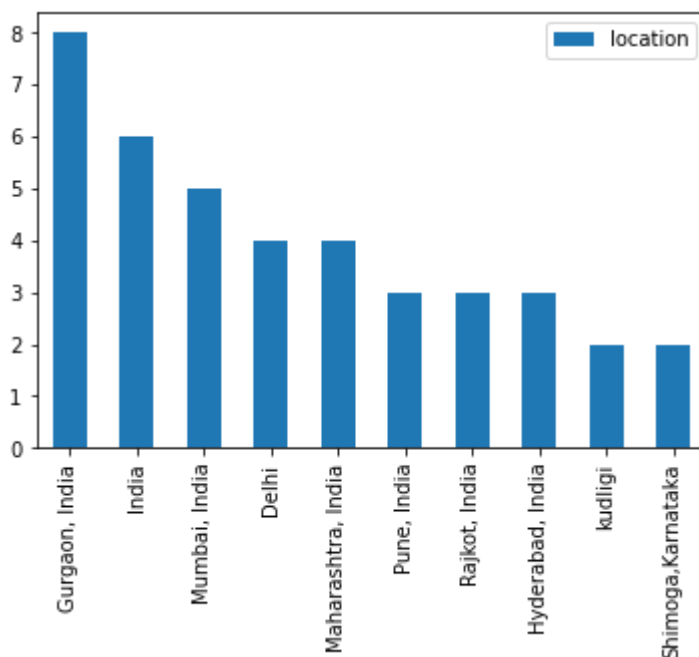
we see that maximum number of tweets are from New Delhi, India followed by Hyderabad, Mumbai and Pune.



(2.b)Top 10 location for Mahindra XUV 700

Graph 2.b shows no. of Mahindra XUV 700 tweets for top 10 locations.

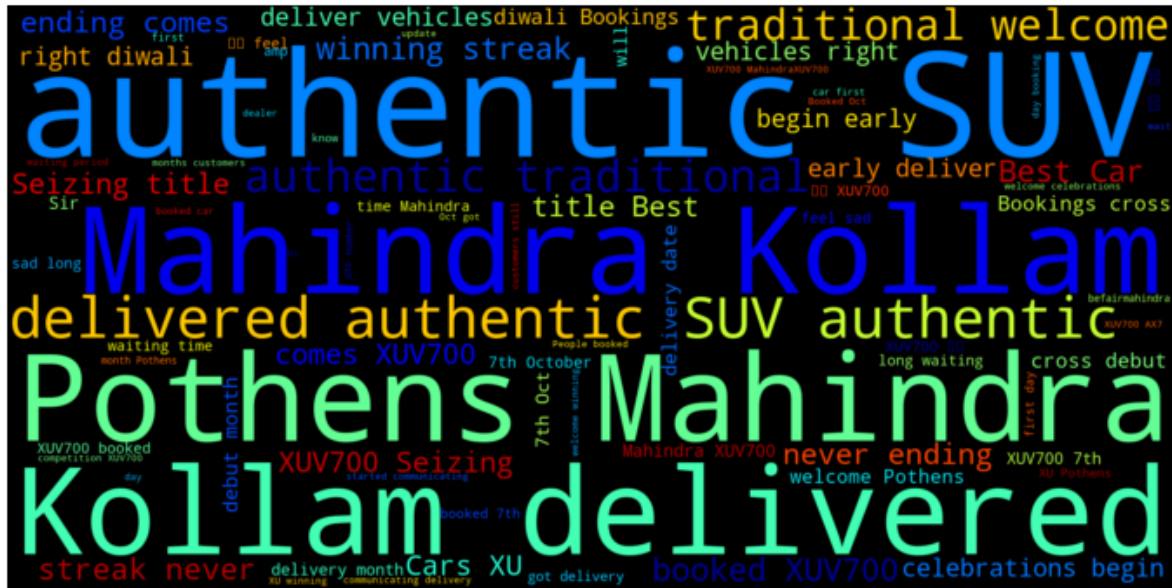
We see that largest amount of tweets are from New Delhi followed by Mumbai, Hyderabad and Pune.



(2.c)Top 10 location of Tata Safari

Graph 2.c shows number of Tata Safari tweets for top 10 locations.

We see that largest number of tweets are from Gurgaon followed by Mumbai, Delhi and Pune.



(2f) Wordcloud of Mahindra XUV using Vader

Word cloud 2 is of Mahindra XUV700. We see that the word SUV is largest which means SUV has occurred highest number of times followed by authentic, Mahindra, delivered etc.

3.3 Algorithm and Process Design:

ML Algorithms:

The dataset for this part consist only of tweets and label(analysis) To execute various algorithms we need to split data. Data splitting allows unbiased evaluation of the model's performance on fresh data that was not previously seen by the model. Data is splitted as into 80% train and 20% test. For this we will be using Sklearn and Model_selection. Sklearn is a Python lography that offers various features for data processing that can be used for classification, clustering, and model selection. Model_selection is a method for setting a blueprint to analyze data and then using it to measure new data. Then we use Counter Vectorizer. This vectorizer will transform our data into vectors of features. After this various ML models were implemented.

We performed four Machine Learning Models:

- 1) Support Vector Machine
- 2) Logistic Regression
- 3) Random Forest
- 4) Naive Bayes

- 1) Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate.

- 2) Logistic regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent

variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no). Mathematically, a logistic regression model predicts $P(Y=1)$ as a function of X . It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

3) Random Forest

Random Forest is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

4) Naive Bayes

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

For each model above we find accuracy, recall, F1 score and precision.

Algorithm	Accuracy	Recall		F1 Score		Precision	
		Positive	Negative	Positive	Negative	Positive	Negative
Naive Bias	0.821918	0.96	0.66	0.84	0.75	0.75	0.86
Random Forest	0.871233	0.94	0.90	0.89	0.86	0.84	0.83
SVM	0.879452	0.92	0.96	0.90	0.85	0.87	0.77
Logistic Regression	0.893151	0.93	0.90	0.91	0.89	0.90	0.88

The above table gives positive and negative Recall, F1 Score and Precision along with Accuracy for various Models.

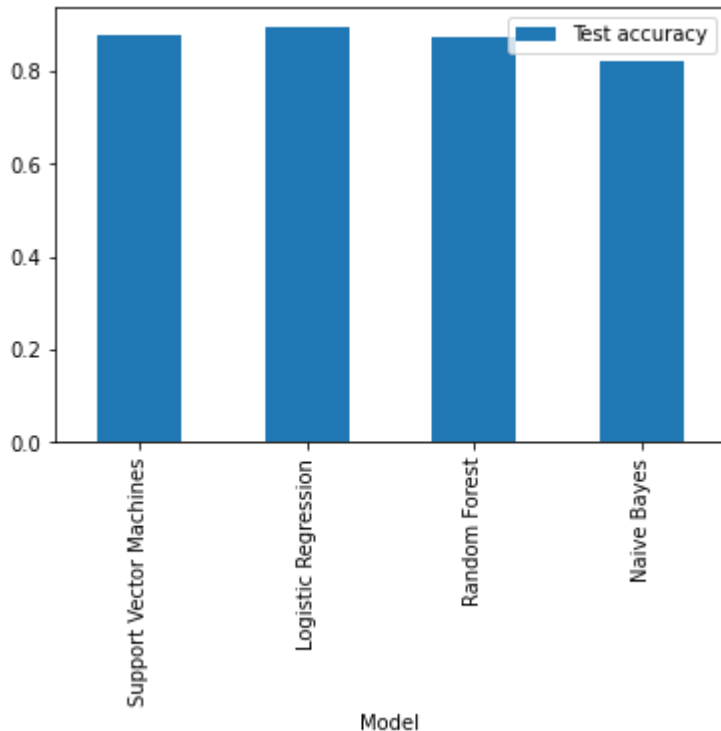


Fig 3) Algorithm efficiency bar graph.

From the Graph Model we see that the Test accuracy for Logistic Regression(89.31%) is highest among the four followed by SVM(87.94%), Random Forest(87.12%) and lastly Naive bayes(82.19%).

Thus among the four Logistic Regression is the better Algorithm than others.

3.4 Details of Hardware and Software:

1. Google Colab:

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

Some Key Features of Colaboratory are:

- Write and execute code in Python
- Document the code which supports the mathematical equations
- Create new notebooks
- Upload the existing notebooks
- Share the notebooks with the google link
- Import data from Google Drive

- Save notebooks from/to Google Drive
- Import/Publish notebooks from GitHub
- Import external datasets e.g. from Kaggle
- Integrate PyTorch, TensorFlow, Keras, OpenCV
- Free Cloud service with free GPU and TPU

2. R Studios:

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

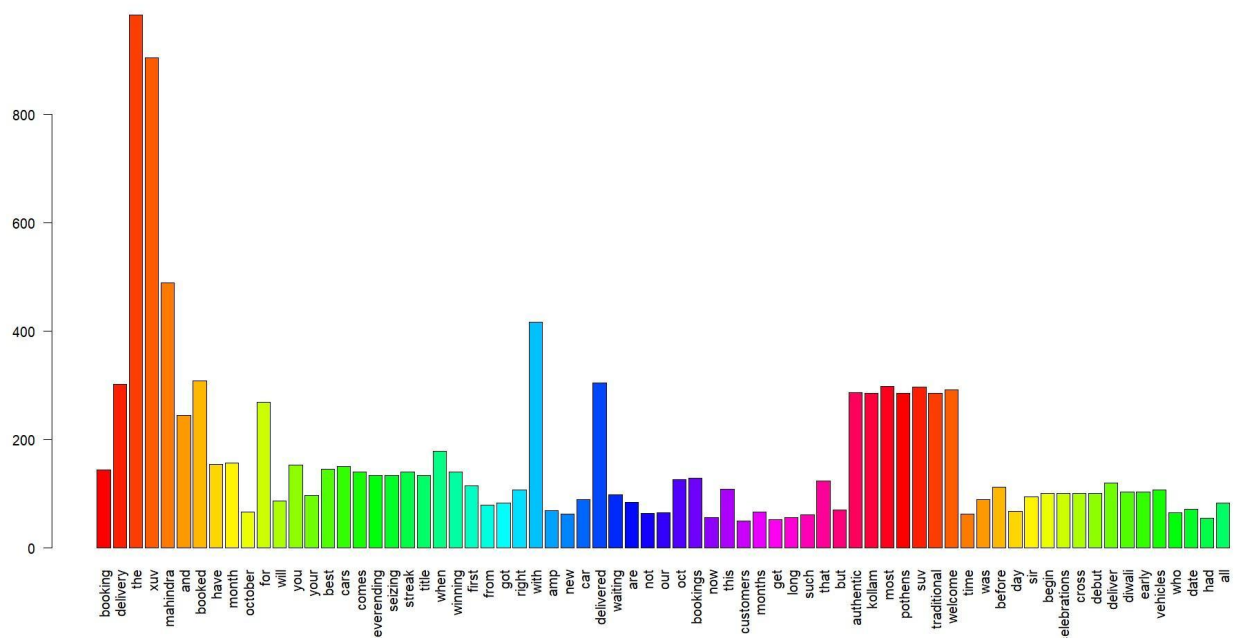
Some Key Features of R Studios are:

- Access RStudio locally
- Syntax highlighting, code completion, and smart indentation
- Execute R code directly from the source editor
- Quickly jump to function definitions
- View content changes in real-time with the Visual Markdown Editor
- Easily manage multiple working directories using projects
- Integrated R help and documentation
- Interactive debugger to diagnose and fix errors
- Extensive package development tools

3.5 Experiments and Results:

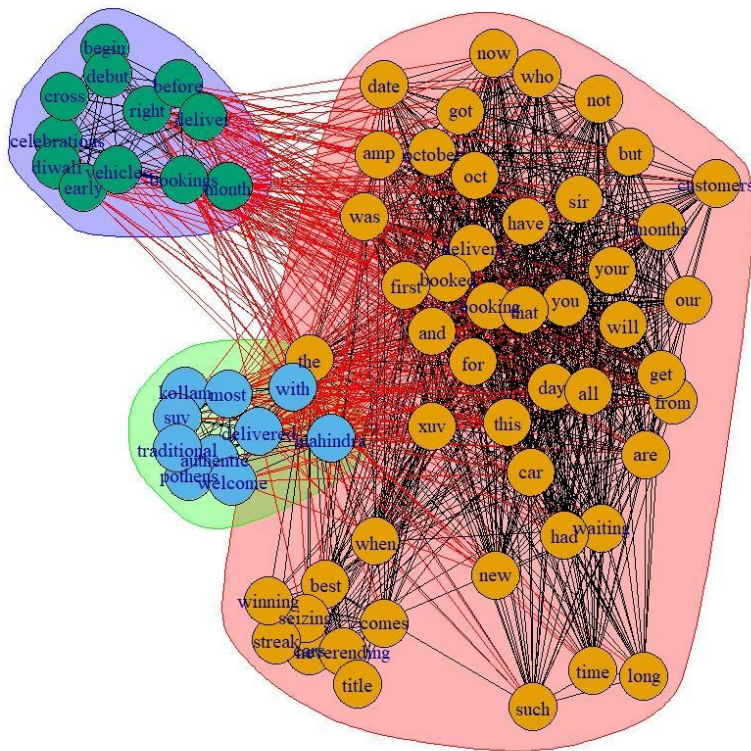
The experiment was to analyse tweets of Tata Safari and Mahindra XUV 700 the results of visualizations are as follows:

Mahindra XUV 700:



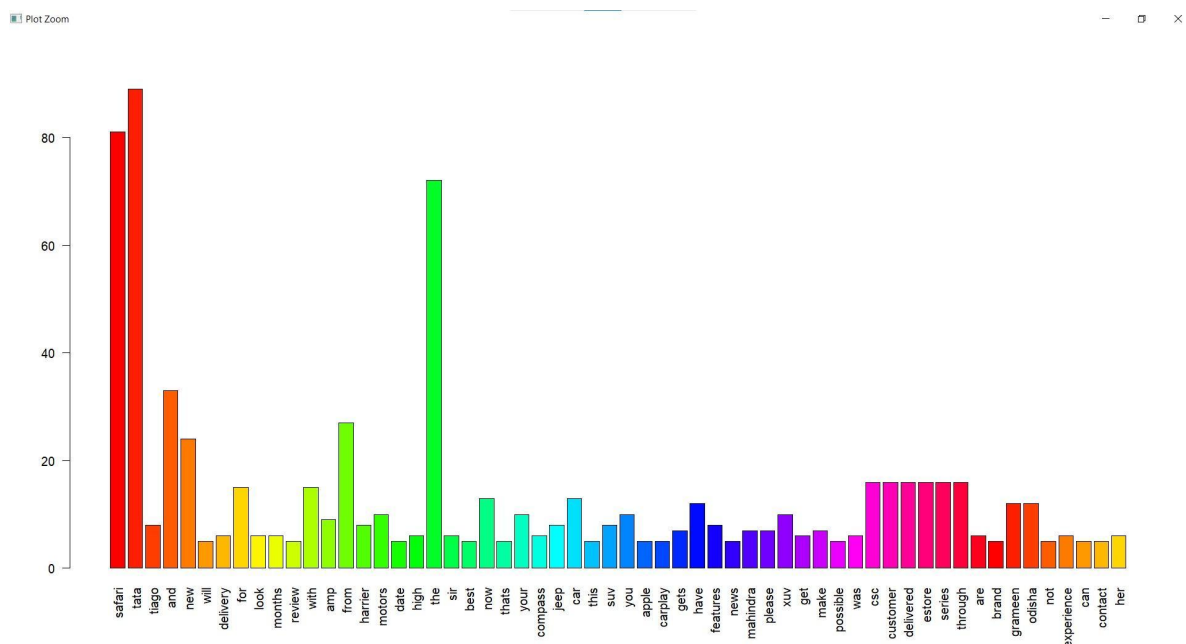
Fig(1.1) Bar Graph for Frequency vs Words in tweets.

From the given graph we can see that words like Delivery , delivered ,authentic and pothens are with high frequencies.The better understanding of this graph is in the word cloud.



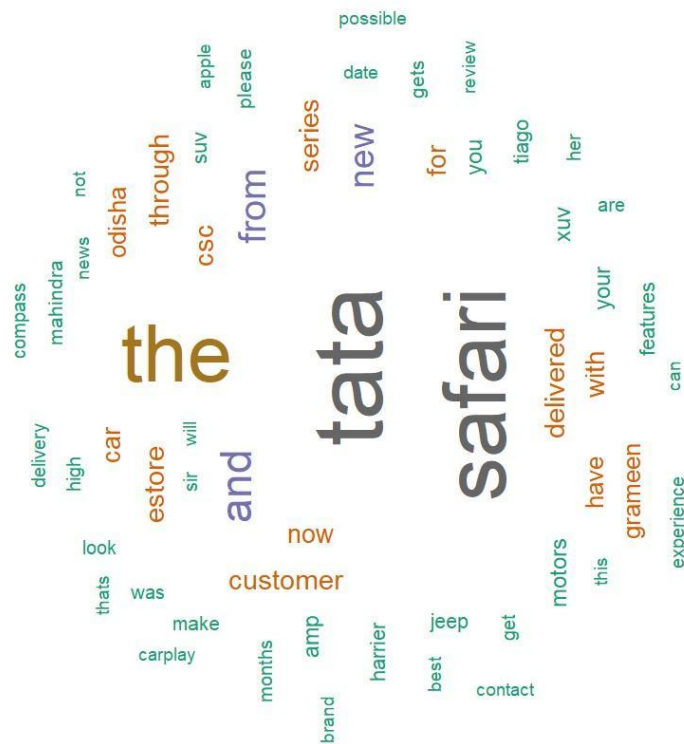
(Fig 1.33) Prop Optimization Detection

2. TATA Safari



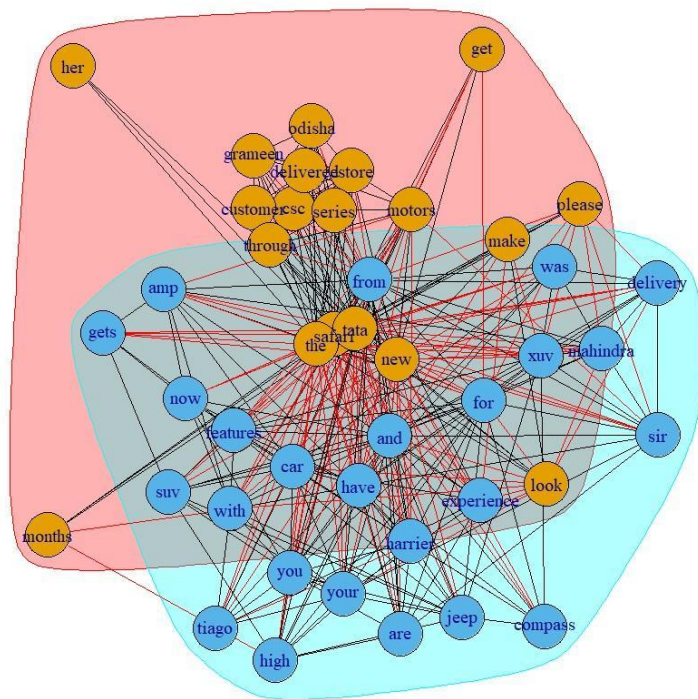
(Fig 2.1) Bar Graph for Frequency vs Words in tweets.

Bar Graph From the above figure we can see that the word Tata , Safari and car are used in high frequency

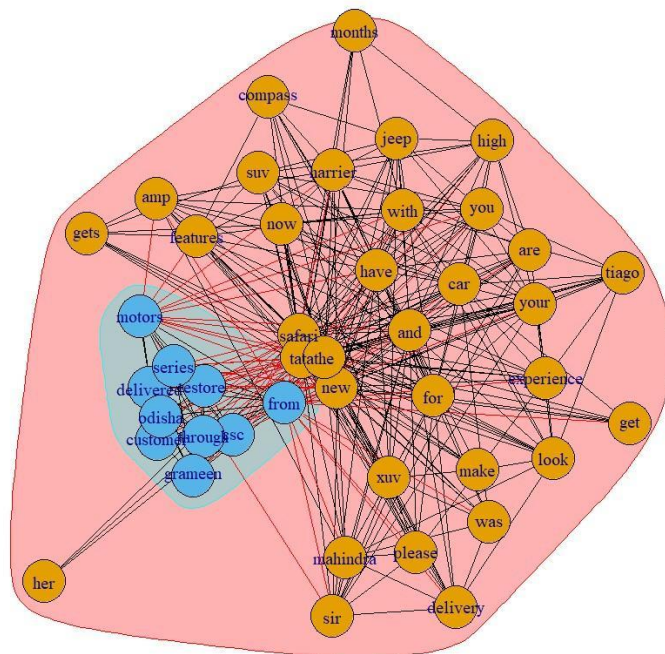


(Fig 2.1) Word Cloud

The figure shows the words tata and safari and few more words such as customer ,grameen ,delivered ,estore,series,csc,etc are used in high to moderate frequencies respectively.



(Fig 2.31)Community Detection



(Fig 3.2)Greed Optimization

Brand	Compound	Positive	Negative	Neutral
Mahindra XUV 700	<u>0.191035</u>	<u>0.113603</u>	<u>0.041314</u>	<u>0.845079</u>
TATA Safari	<u>0.171880</u>	<u>0.091983</u>	<u>0.027933</u>	<u>0.880084</u>

References:

- [1] Efendi, S., Erwin, A., & Eng, K. I. (2016). "Study of Automotive Brands Popularity in Indonesia Using Twitter Data.", *Journal of Applied Information, Communication and Technology*, 3(1), 23-33.
- [2] S. E. Shukri, R. I. Yaghi, I. Aljarah and H. Alsawalqah, "Twitter sentiment analysis: A case study in the automotive industry," 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2015, pp. 1-5.
- [3] Fumiaki Saitoh, Fumiya Shiozawa, Syohei Ishizu, "Knowledge Extraction About Brand Image Using Information Retrieval Method", "HCI International 2016 -- Posters' Extended Abstracts", "2016", "Springer International Publishing", pages 291-295".
- [4] Teena Bagga and Deepak Gupta, "Internet Marketing by Automobile Industry: Special Reference of Indian Counterparts", *International Journal of Computer Applications*, volume 97, 2014, pages 9-16.
- [5] Zain Asghar, Tahir Ali, Imran Ahmad, Sridevi Tharanidharan, Shamim Kamal Abdul Nazar, Shahid Kamal, "Sentiment Analysis on Automobile Brands Using Twitter Data Intelligent Technologies and Applications", 2019, Volume 932
- [6] Nur Azizah Vidya, Mohamad Ivan Fanany, Indra Budi, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers, *Procedia Computer Science*", Volume 72, 2015, Pages 519-526.
- [7] Seunghyun "Brian" Park and Chihyung "Michael" Ok and Bongsug "Kevin" Chae, "Using Twitter Data for Cruise Tourism Marketing and Research", *Journal of Travel & Tourism Marketing*, volume 33, Pages 885-898, (2016).
- [8] Ashish Kumar Rathore, P. Vigneswara Ilavarasan, "Pre- and post-launch emotions in new product development: Insights from twitter analytics of three products", *International Journal of Information Management*, Volume 50, 2020, Pages 111-127.
- [9] Pimprikar, R., Ramachadran, S. and Senthilkumar, K., "Use of machine learning algorithms and twitter sentiment analysis for stock market prediction." *Int J Pure Appl Math*, 2017 115(6), pages 521-526..
- [10] Wisdom, V. and Gupta, R., "An introduction to twitter data analysis in python.", "Artigence Inc. 2016
- [11] Palguna, D.S., Joshi, V., Chakaravarthy, V., Kothari, R. and Subramaniam, L.V., "Analysis of sampling algorithms for twitter.", In *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015, June.