Hypes Vibes: Sentiment Analysis Using Social Media

Sanjay S. Kadam
Bharati Vidyapeeth of Engineering,
Belapur CBD, Navi Mumbai, Maharashtra, Pin code-400614

Bhavik Khatau
Bharati Vidyapeeth of Engineering,
Belapur CBD, Navi Mumbai,Maharashtra, Pin code-400614

Deepak Maurya
Bharati Vidyapeeth College of Engineering,
Belapur CBD, Navi Mumbai,Maharashtra, Pin code-400614

Yash Koli
Bharati Vidyapeeth College of Engineering,
Belapur CBD, Navi Mumbai,Maharashtra, Pin code-400614

ABSTRACT
There is various type of information uploaded and shared on social media in the form of text,videos, photos and audio. Social media is rich with raw and unprocessed data and theimprovement in technology, especially in machine learning and artificial intelligence, allow
the data to be processed and converted it into a useful data that they can benefit most businessorganization.
Most of the previous study applied sentiment analysis into a product or movie review to betterunderstand their customer and make the necessary decision to improve their product or
services.
Our Project focuses to provide a better understanding of the application of sentiment analysisin social media platform with better accuracy. Sentiment analysis is an approach that uses
Natural Language Processing (NLP) to extract, convert and interpret opinion from a text andclassify them into positive, negative, or natural sentiment.
Keywords: *Machine Learning Algorithms, Tweeter Data,NaturalLanguage Processing*

I. INTRODUCTION

Most of the previous study applied sentiment analysis into a product or movie review to better understand their customer and make the necessary decision to improve their product or services. Our Project focuses to provide a better understanding of the application of sentiment analysis in social media platform with better accuracy. Sentiment analysis is an approach that uses Natural Language Processing (NLP) to extract, convert and interpret opinion from a text and classify them into positive, negative, or natural sentiment.

II. Literature Review

Social MediaSentiment Analysison Twitter Datasets,2020.

With the help oftraining dataset, thesystem is trained bythe classifiers usingSVM. Random Forestand Decision Tree are used to evaluate theperformances basedon the sentiment analysis domain.This data can have auseful resourceunderstanding thesentiments of people.It is also observedthat Decision Treeand RandomForest algorithmhave moreaccuracy thanSVM technique forthe presentanalysis.It isnow understandable thatpeople's opinion is how muchimportant for any product-based company orgovernment.[1]

Deep Learning forAutomatedSentiment Analysisof social media,2019.Sentiment analysis isdone at different levelsusing commoncomputationaltechniques likeUnigrams, lemmas,negation and so on. If you havethousands offeedbacks per month,it is impossible forone person to readall of theseresponses. Whilesentiment analysis isuseful, but it isnot acompletereplacement forreading surveyresponses.

Sentimentanalysis is auseful tool toanalyzeuser-generatedcontent on socialmedia sites.However, the complexity anddynamic nature of socialmedia data makes it difficulttoaccurately identifysentiments.[2]

SentimentAnalysis insocial media,2019.The application ofsentiment analysisranges from businessand marketing,politic,health to public action.Sentiment analysis isnotlimited to oneapplication, but itprovides a vastapplication in differentareas to assist indecision making.If you havethousands offeedbacks permonth, it isimpossible for oneperson to read all theseresponses. Whilesentiment analysisis useful, but it isnot a completereplacement forreadingsurveyresponses.The conductedsystematic literaturereviewprovidesinformation onstudies on sentimentanalysis insocialmedia.The paper makes thefollowing threecontributions.First, we show what is the method used in analyzing sentiment in social media. There is various method introduced by research, still, the most common method uses in Lexicon based method is Sent Wordnet and TF-IDF while for machine learning is Naïve Bayes and SVM.[3]

Sentiment Analysis for Social Media,2015.Formally stated their task and interpreted how to mathematically incorporate social context and topical context into the basic prediction model.The major challenges were 1.Incremental Approach 2.Parallel Computing For Massive Data, 3.Sarcasm,4.Grammatically Incorrect Words.The main goal of this paper is to give an overview of latest updates in sentiment analysis and classification methods, and it includes the brief discussion on the challenges of sentiment analysis for which the work needs to be done.If the challenges are overcome the project would be of great use.[4]

A Study on Sentiment Analysis Techniques of Twitter Data,2019.In this research paper they have compared machine learning classifiers such as Naïve Bayes, Maximum Entropy and Support Vector Machine and Document - level sentiment analysis. One of the greatest difficulties was in determining the best approach for detecting sentiments in Twitter data because comparing various approaches is a highly challenging task when there is a lack of agreed benchmarks.It was expected that ensemble Twitter sentiment-analysis methods would perform better than supervised machine learning algorithms, However, hybrid methods also performed well and obtained reasonable classification accuracy scores, since they were able to take advantage of both machine learning classifiers and lexicon-based Twitter sentiment-analysis approaches. Comparison of different techniques were studied.[5]

Using Twitter Data to Predict the Performance of Bollywood Movies,2015.This paper uses sentiment analysis and prediction algorithms to analyze the performance of Indian movies based on data obtained from social media sites. The authors used Twitter4j Java API for extracting the tweets through authenticating connection with Twitter web sites and stored the extracted data in MySQL database and used the data for sentiment analysis.The current study suffers from the limitation of not having enough computing resources to crawl the data. For predicting box office collection, there is no correct availability of ticket price information, total number of seats per screen and total number of shows per day on all screens.For the movie Singham Returns over method of prediction gave a box office collection as 84 crores and the actual collection turned out to be 88 crores, which was pretty accurate.After going through the entire research paper, it is quite clear that it can predict the movie performance even before it is released.[6]

III. OBJECTIVE

- This project aims to do sentiment analysis on tweets of upcoming movies thus finding thelocation of favorable crowd
- This project aims to do sentiment analysis on tweets of newly launched products.
- This project aims to give an idea about who is going to form government in upcoming elections.
- This project aims to give an idea about how satisfied are the people with their leaders.
- This project aims to do sentiment analysis on tweets ofdifferent people thus giving the opinion oftweets as positive, negative, or neutral.

IV. EXPERIMENTAL SETUP

4.1 Frontend
tech stack = HTML, CSS, JavaScript.
HTML (the Hypertext Markup Language) and CSS (Cascading Style Sheets) are two of the core technologies for building Web pages. HTML provides the structure of the page, CSS the (visual and aural) layout, for a variety of devices.

4.2 Backend
tech stack = Flask, MySQL.
Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for

object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

MySQL (/ˌmaɪˌɛsˌkjuːˈɛl/) is an open-source relational database management system (RDBMS).Its name is a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured Query Language. A relational database organizes data into one or more data tables in which data types may be related to each other; these relations help structure the data. SQL is a language programmers use to create, modify and extract data from the relational database, as well as control user access to the database. In addition to relational databases and SQL, an RDBMS like MySQL works with an operating system to implement a relational database in a computer's storage system, manages users, allows for network access, and facilitates testing database integrity and creation of backups.

V. PROPOSED SYSTEM

5.1 System Architecture

System analysis is the process of defining the architecture, components, and data of a system to satisfy specified requirements. Design is a method of studying a system by examining its component parts and their interactions. Before implementation began the system was analyzed and designed.
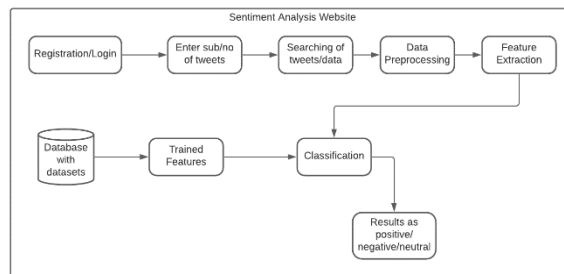

Figure 2: System Architecture diagram

5.2 Activity diagram

UML models are basically three types of diagrams, namely, structure diagrams, interaction diagrams, and behavior diagrams. An activity diagram is a behavioral diagram i.e., it depicts the behavior of a system. An activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed. We can depict both sequential processing and concurrent processing of activities using an activity diagram. They are used in business and process modeling where their primary use is to depict the dynamic aspects of a system
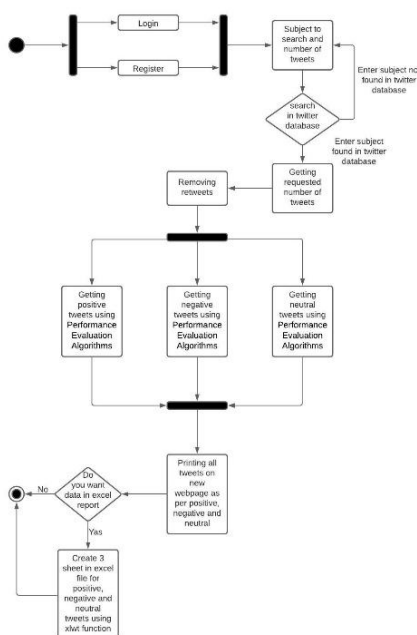

Figure 3: Activity diagram

5.3 Interaction design and Sequence Design

From the term Interaction, it is clear that the diagram is used to describe some types of interactions among the different elements in the model. This interaction is a part of the dynamic behavior of the system. This interactive behavior is represented by two diagrams known as the Sequence diagram and the Collaboration diagram. The basic purpose of both diagrams is similar. The sequence diagram emphasizes the time sequence of messages and the collaboration diagram emphasizes the structural organization of the objects that send and receive messages.
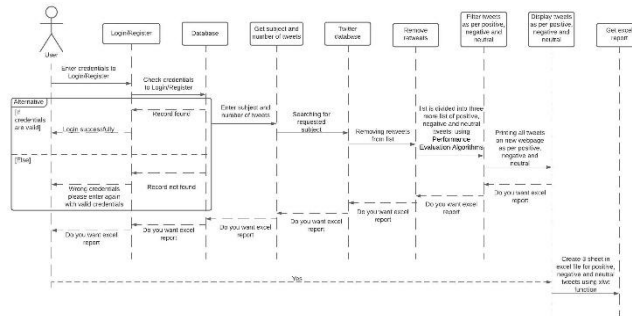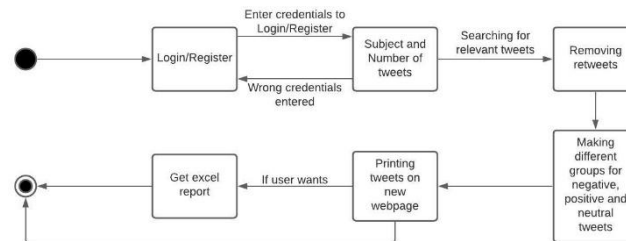

Figure 4: Interaction Diagram

Figure 5: Sequence Diagram

5.4 Component Diagram

Component diagrams are used to model the physical aspects of a system. Physical aspects are the elements such as executables, libraries, files, documents, etc. which reside in a node. Component diagrams are used to visualize the organization and relationships among components in a system. These diagrams are also used to make executable systems. It does not describe the functionality of the system but it describes the components used to make those functionalities.
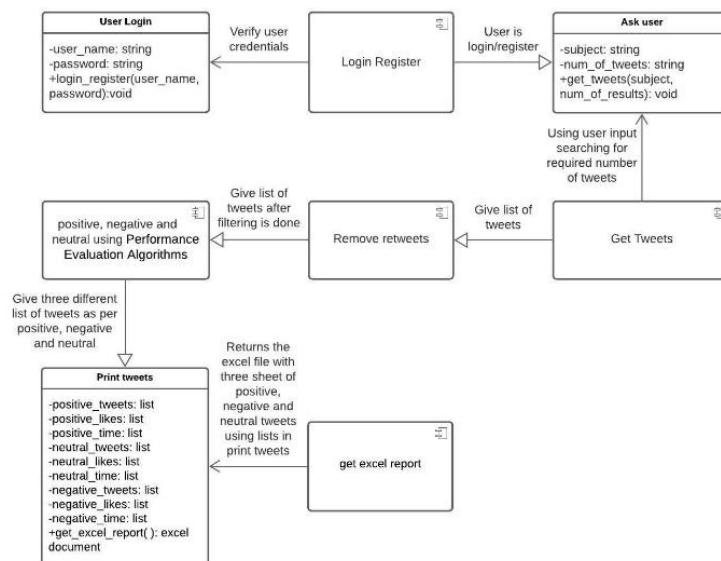

Figure 6: Component Diagram

4.5 Deployment Diagram

Deployment diagrams are used to visualize the topology of the physical components of a system, where the software components are deployed. Deployment diagrams are used to describe the static deployment view of a system. Deployment diagrams consist of nodes and their relationships. Deployment diagrams show how they are deployed in hardware. However, these two diagrams are special diagrams used to focus on software and hardware components.
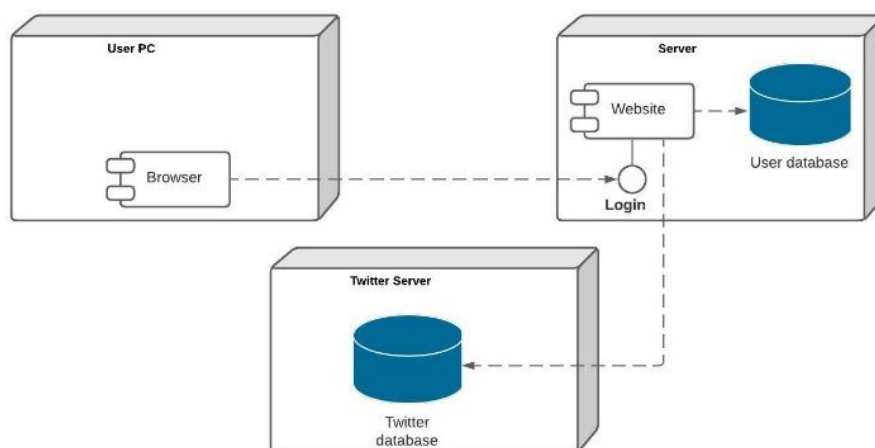
.

Figure 7: Deployment Diagram

VI.   CONCLUSION

Masses of users share their feelings on social media, making it a valuable platform for tracking and exploring public sentiment. Social media is one of the biggest platforms where massive instant messages are published every day which makes it an ideal source for capturing the opinion towards various curious topics, such as products, goods, politicians, or celebrities, etc.

SO, we are looking forward for making a project based on the discussed topics.

VII.   FUTURE SCOPE

The futureof this project is to improved design; implementation and documentation in such a way that anyone can use this project for better perform. I will develop the site more dynamically. In future I will add the few modules for better improvement of the project such as, real-time chat bot option. In future I will also add mobile version app of this website. We will also try to improve their accuracy. And we will try to predict the persons feeling such as happy sad etc. We will add the database where the user can see its previous analysis.

REFERENCES

[1].   Poche, Flora, Claus Ester, and Christine Strauss. (2018) "Social Media Metrics and Sentiment Analysis to Evaluate the Effectiveness of Social Media Posts." Procedia Computer Science 130: 660- 666,2012.
[2].   Joyce, Brandon, and Jing Deng. (2017) "Sentiment Analysis of Tweets for the 2016 US Presidential Election", in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE,2014.
[3].   Iroko, Victoria, Maria Sharmila, Khaleel Malik, and Riza BatistaNavarro. (2018) "Analysing Sentiments Expressed on Twitter by UK Energy Company Consumers", in Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 95-98): IEEE.2015.
[4].   Rahman, S. A. El, F. A. Altai, and W. A. Altieri. (2019, 3-4 April 2019). "Sentiment Analysis of Twitter Data", in the 2019 International Conference on Computer and Information Sciences (ICCIS).2016.
[5].   Garg, Neha, and Rankle Rani (2017) "Analysis and visualization of Twitter data using k-means clustering." International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 670- 675.2011.
[6].   Mishra, Prerna, Ranjana Rajnish, and Pankaj Kumar. (2016) "Sentiment analysis of Twitter data: Case study on digital India." International Conference on Information Technology (Incite)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds. IEEE, 148 – 153.2010.2014.
[7].   Devika, R., and S. Revathi. (2018) "Survey on clustering techniques in Twitter data." Second International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 1073- 1077.2017.
[8].   Colne Ric, Niko, and Jane DeMar. (2018) "Emotion recognition on twitter: Comparative study and training a unison model." IEEE Transactions on Affective Computing, 1949 - 3045.2015.
[9].   Tripathi, Pragya, Santosh Kr Vishwakarma, and Ajay Lala. 2015 "Sentiment analysis of English tweets using rapid miner." International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 668 – 672.2014.2016.
[10].   Anandarajan, Murunga, Chelsey Hill, and Thomas Nolan. (2019) "Sentiment Analysis of Movie Reviews Using R." Practical Text Analytics. Springer, Cham, 193-220.2015.

[11]. Kumar, Akashi, and Geetanjali Garg. "Sentiment analysis of multimodal twitter data." Multimedia Tools and Applications 78.17 (2019): 24103-24119.2020.

[12]. Rosenthal, Sara, Nour Farrah, and Presley Nakao. "SemEval-2017 task 4: Sentiment analysis in Twitter." Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval2017).2017.

[13]. Amolik, Akshaya, et al. "Twitter sentiment analysis of movie reviews using machine learning techniques." International Journal of Engineering and Technology 7.6 (2016)

[14]. Kharche, S. R., and Lokesh Bijole. "Review on Sentiment Analysis of Twitter Data." International Journal of Computer Science and Applications 8 (2015)

[15]. S. Hota and S. Pathak. Kenn classifier-based approach for multiclass sentiment analysis of twitter data. In International Journal of Engineering Technology, pages 1372–1375. SPC, 2018.

[16]. Agarwal, B. Xie, I. Varsha, O. Ram bow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Median2011, pp. 30- 38,2016.

[17]. Neethu M, S and Rajshree R," Sentiment Analysis in Twitter using Machine Learning Techniques" 4th ICCCNT 2013, at Tiruchengode, India. IEEE – 31661.2014.

[18]. Bhumika Gupta, Monika Negi, Kanika Vishwa, "Study of Twitter Sentiment Analysis Mistreatment Machine Learning Algorithms on Python" International Journal of laptop Applications 0975- 8887, 2017.

[19]. E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg!" Icwsm, vol. 11, no. 538-541, pp. 164, 2011.

[20]. J. Bollen, A. Pepe, and H. Mao. Modelling public mood and emotion: Twitter sentiment and socio-economic phenomena. Proc. of ICWSM, 2011

[21]. Hodeghatta, U.R. (2013), "Sentiment analysis of Hollywood movies on Twitter", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1401-1404,2014.

[22]. AndranikTumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In Proceedings of the International Conference on Web and Social Media (ICWSM'10) 10, 1 (2010), 178–185,2010.

[23]. Jose, R.; Chooralil, V.S. Prediction of election result by enhanced sentiment analysis on twitter data using word sense disambiguation. In Proceedings of the 2015 International Conference on Control Communication & Computing India (ICCC), Trivandrum, India, 19– 21 November 2015; pp. 638–641,2014.

[24]. Ibrahim, M.; Abdallah, O.; Wicaksono, A.F.; Adriani, M. Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14–17 November 2015; pp. 1348– 1353,2013.

[25]. M. Anjaria and R. M. R. Guddeti, "Influence factor-based opinion mining of Twitter data using supervised learning," in 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), pp. 1-8. 2014.

[26]. E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg," Icwsm, vol. 11, no. 538-541, p. 164, 2011.

[27]. A. Go, R. Bhavani, and L. Huang. "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford 1, no. 12, 2009.

[28]. M. Bandgar, D. S. Sheela (2016), "Analysis of real time social tweets for opinion mining", International Journal of Applied Engineering Research, 11(2), pp. 1404-1407,2012.