Sentiment Analysis on Social Media

Dilesh Tanna
Information Technology
A.P. Shah Institute of Technology
Thane,India
dileshtanna35@gmail.com

Manasi Dudhane
Information Technology
A.P. Shah Institute of Technology
Thane,India
mansidudhane31@gmail.com

Amrut Sardar
Information Technology
A.P. Shah Institute of Technology
Thane,India
redamrut96@gmail.com

Prof. Kiran Deshpande
Information Technology
A.P Shah Institute of Technology
Thane,India
kbdeshpande@apsit.edu.in

Prof. Neha Deshmukh
Information Technology
A.P Shah Institute of Technology
Thane,India
npdeshmukh@apsit.edu.in

***Abstract*—Social media consists of various kinds of emotions and sentiments of its users in the form of electronic media. To analyze the reactions or sentiments of the users on a certain post is also a challenging task. Our project aims to automate this task of analyzing the reactions and the posts and generate a report based on the outcome. A custom social media platform is proposed that would give the ability to the users to perform activities such as post, like, comment and share. The users can be enrolled into different groups such as business or university. This platform would have the ability to perform sentiment analysis on all the activities of the users in a group and create a report according to their reactions and their posts on the platform. Every activity would be rated based on the sentiments behind it on a scale of -1 and +1, 0 being a neutral sentiment. The appropriate admin would receive the reports which can be used for the further actions. The analysis would be performed on various factors such as the response of other users on a post. According to the user's posts, its home page would display other relevant posts. The analysis would facilitate the decision-making process for the admins (around other activities on the platform) and also help detect any need to give special attention to any user in a group, such as a student coping up with depression. Also, the proposed platform can be used to share content on other social media platforms as well. As a result, the users would have a single platform with the ability to do a lot more than any existing social media platform facilitates.**

Keywords—sentiment analysis, social media, emotion classification, sentiment computing.

I. INTRODUCTION

As in today's world, everyone shares their emotions online, through social media platforms and thus, the data generated by these platforms can be used for analysis of the sentiments expressed by the users on various different posts. Emotions such as anger, sadness, happiness, excitement can be extracted from the posts and further analyzed for reporting and decision-making purposes.

Existing social media platforms don't give us the privilege to track the activities of the users and analyze the user's behavior for future predictions such as what to post, when to post and whom to target.

The user would receive feeds based on their interests recorded by the system through their past behavior. A user would also be able to share a post on other supported social media platforms.

Sentiment analysis on a user's activities would generate a report for the admin of a group that the user belongs to; such as the Head of a department or a college counselor. Reports can be generated by the analyses done on the data of the users on the platform. According to the posts and the actions of the users, the data that is generated is analyzed by the sentiment model. The administrators may take further actions based on the reports that can be generated using the results of the sentiment analysis model. In a corporate world or a university, it can be considered important to keep track of the related population's sentimental behavior towards the institution as it gives a great amount of detail on how a user feels about being a part of that institution and in what way should the institution engage further with the concerned user. The right kind of analysis performed on the students might help determine their 'attitude' which the companies visiting the Institute for hiring the students can use to choose the right candidate. A social media platform with the ability to perform sentiment analysis and produce a report for the 'high level' users is not a mere content sharing platform anymore. It becomes a full-fledged authoritative tool which would facilitate the decision-making process.

II. LITERATURE SURVEY

[1] Have proposed VADER, which is a simple rule-based model for general sentiment analysis, and compare its effectiveness to common state-of-practice benchmarks including LIWC, ANEW, SentiWordNet, and machine learning techniques like Naive Bayes and Support Vector Machine (SVM) algorithms. VADER then combines these lexical features with general rules of grammatical and syntactical ways for expressing and emphasizing sentiment intensity. As comparing to the normal technique vader has the advantages. First, it is both quick and computationally economical. It takes a fraction of a second to analyze with VADER can take hours when using more complex models like SVM. Second, rules used by VADER are accessible. [1]

[2] Has presented that how to program a machine to analyze the different grammatical words, cultural variations, take out emotions, and get sentiment and meaning behind that words using machine learning techniques.
Author has made a comparison among Support Vector Machine, Naïve Bayes and Maximum Entropy classifiers regarding sentence level sentiment analysis for depression measurement. Author has examined the performance of our proposed methods on two datasets, twitter dataset and 20newsgroups. [2]

[3] Have discussed challenges and opportunity in Online Social Networks (OSN). They have highlighted two main challenges; first, it is very important for OSN users not only to share media content but also to receive the specific media they want to see more.
Second, to secure users who are followed by a greater number of users and for them who share too much of their personal information. Author has focused his attention on the relevant research challenges regarding semantics and security. Author has introduced the challenges of sentiment detection, and coping with phishing attacks, Sybil attacks, and spamming. Thus this paper has outlined a new research agenda in the field of OSN as well as it is a knowledgeable for the everyday users. [3]

[4] Have introduced the procedure to identify the intensity of the sentiments behind the phrase put on social media platform by a user and phrases with emotional contents will be filtered through a Machine Learning program such as pronouns and adjectives verbs, etc. They have introduced to a solution that notify to other persons such as doctors or relatives about the potential emotional behavior of the user. An authorized person will have the information about the specific user for considering as counter action. [4]

III. Existing System

The current system of social media platforms provides little or no freedom to the administrators of important accounts like businesses or colleges to analyze the sentiments related to their activities. The services available are very limited and charge a fortune for the right kind of analysis. Sentiment analysis is practiced in very few fields in the current technological timeline. Social media platforms use it to some extent. However, it doesn't provide a proper workaround it. The administrators don't have the privilege to perform analysis on users related to their circle such as students of a university or employees of a business.

IV. Problem Statement

User data is precious and would reap great results for an institution like a college or a business to make their decisions. Decision making can be a very risky job when it comes to taking huge decisions although, small decisions such as what to perform, when to perform and with/for whom to perform a certain task can collectively result into benefits. The existence of a system that facilitates this using the social media behavior of the point of interest of the institute can be a game changer when it comes to decision making. The electronic media serves a great role in data mining and making effective use of it right from the beginning. The ability to enroll a set of users in a group, share something to various platforms from one place and also analyzing the users in the group is something that can be considered a problem to solve.

V. Proposed System

A social media platform with the ability to perform sentiment analysis for 'high level users' is the desired outcome of this project. The platform would allow almost all activities that a normal social media platform does. The users would be able to share content and view other users' content and also express their personal views on the same. A user can be enrolled into different groups by an entity such as a business or a university. The user would also be able to share content from this platform to other social media platforms. Also, certain privileges are given to the admins using which the admins can make their posts compulsory. As a result, all the users for whom the post is made compulsory will have to respond to the post. This can used by universities where the admins want certain action on their posts in the form of feedback. The users can also share the post privately with a group of users on the platform. The other aspect of the project is the sentiment analysis part where all the activities of a user in a group would be analyzed.
The flow of the analysis is as follows:

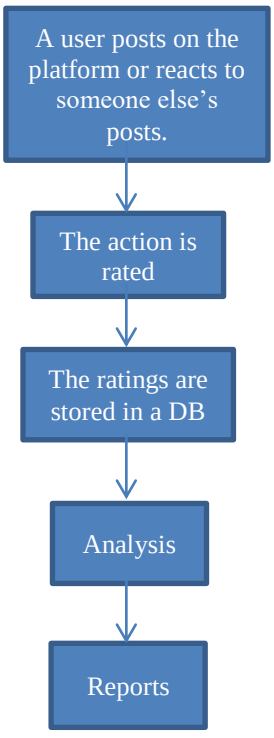

Fig.1 Flow of the data

The flow diagram above depicts a high level view of the proposed architecture. The user, as on any other social media platform, would perform an activity such as to post something.

User's reactions such as likes or comments on the other users' post will also be rated, such as if the comment is positive then the user who is commenting, its score will be increased, similarly if a user posted something depressing then the user will get negative score. For argument's sake, let us say that the user expressed their grief about the decision of their university to increase the minimum required attendance from 75% to 80%. This post, in a traditional platform would be stored in the DB as it is. However, in our proposed system, the post would be rated first and then stored in the DB along with the rating. The rating would be from -1 to +1. This rating and the post stored in the DB would be further used by the system to analyze it along with other related items in the DB and then generate a report for the high level user to study. Let's say that in our case, the report suggests that about 79% of the concerned population is unhappy with the new decision of the university. With these results, the admin may further decide to take an action or may simply ignore it.

The mechanism of the model is such that it analyses the likes, comments and posts of the users. A model which contains different lexicons and NLP libraries and has a set of grammar and positive and negative words is used. According to the analyses, reports are generated and displayed on the dashboard of the admin. Sentiment analyses is performed and the over-all score of the users according to their activities on the platform is calculated. Through these analyses, the admins can take necessary actions on it.

VI. Methodology

The sentiment analysis model consists of lexicon libraries.
A sentiment lexicon has a list of features, which are labeled regarding their sentiments as positive or negative. Here, the words are classified into binary classes (positive or negative). SentiWordNet and Senticnet are two databases in which words are associated with valence scores for sentiment intensity.

LIWC is a text analyses software which is designed for studying the various emotional, structural, cognitive and process components present in text samples. The dictionary has 4,500 words organized and used by LIWC and has 76 categories.

Machine learning approaches have also been proposed in this model. Naïve Bayes, Maximum Entropy and Support Vector Machine are some models proposed. These algorithms help in processing the analyses more effectively since using only the lexicon libraries for analysis is not as effective as it needs to be. So, model is trained to process the datasets and give the effective results.

The dataset includes attributes such as engagements of the users, feedbacks and their post. These attributes rule out the scores of the user and give results. The social media platform's home page is also affected by the score of the particular user. The home page changes based on these results.

Classify	Words
Positive	Love, happy, excited, good
Negative	Bad, ugly, sad, depress, abuse

Emoticons also depict an emotion along with acronyms such as 'LOL'. Sentiment-related acronyms are included in lexicon features list. The model also rates such instances effectively.

For example: A user post a message "I am happy! I adopted a puppy". This will get rated by the model and the results would be shown on the admin's dashboard and the admin has the privileges to view the user's scores. If a user posts a status as "I'm not feeling good, I'm depressed", then here the score of the user would be negative. The following activity diagram shows the flow of the analysis on the platform.

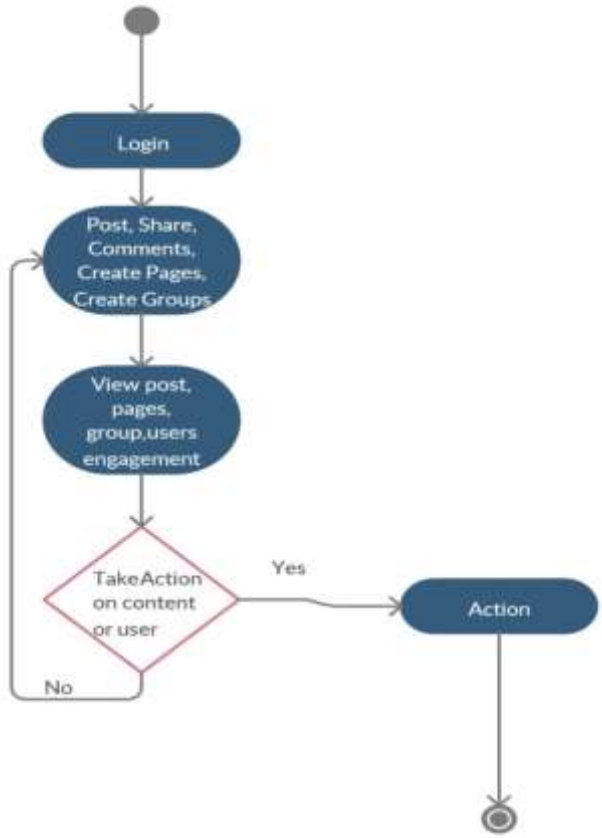

Fig.2 Activity Diagram

VII. Result and Analysis

The analysis carried on the actions of the users on social media through the sentiment model is depicted in the pie chart.

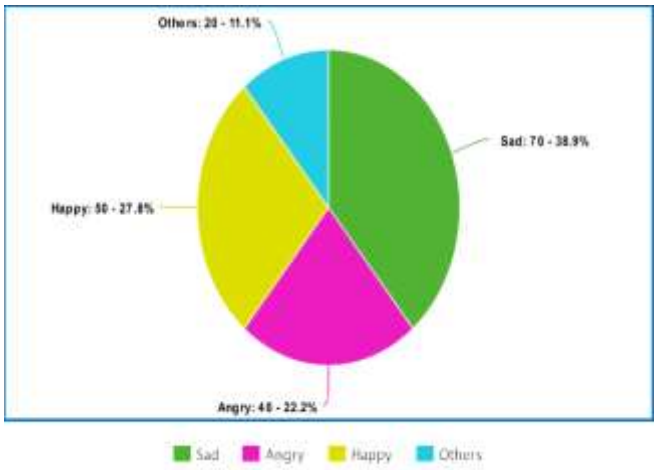

Fig.3 Result on Dashboard

Here, the report contains the graph and the scores of the user. The pie chart depicts the overall engagements of the users. There are mix emotions involved on the platform, such as sad, happy, angry, etc. Through these analyses, the admins of the platform can take certain actions on the content which is being shown on the media and also take feedback from the users. Admins can also identify the users whose reports are constantly negative.

The scores are given according to the user engagement on the platform by the trained model. The model gives scores from -1 to +1 as discussed above. The database stores the score of the user. The following table shows the scores of the user.

USER	SCORES
Dev	+0.78
Ayushi	-0.23
John	0
Manila	-0.54

The table shows the scores of the users in which Ayushi and Manila have negative scores, Dev has a positive score and John has a neutral score of 0.

VIII. TECHNOLOGY STACK

Programming languages, libraries and tools used:

1. MongoDB
2. Express.js
3. Node.js
4. React.js

Mongodb is an opensource relational database. The architecture consists of collections and documents instead of tables and rows such as in relational databases. Thanks to its architecture, this database is extremely scalable and can handle large volumes of data. This helps in making the tasks of storing user related data simpler.

Express is a lightweight server-side JavaScript framework. Express is used together with Node.js to simplify the development of the middle-tier. Node.js makes it easier to write modular, secure and fast applications. Express is a flexible Node.js web application framework that helps in middle tier application development.

Node.js is a JavaScript based runtime compiling and simplifies building performance centric and scalable web apps. Node.js has an ecosystem of open source libraries and components.

React is an Javascript library developed by Facebook. It has a concept of virtual dom, which helps build robust, dynamic Single Page Applications (SPAs).

Together, these technologies form the MERN stack. Using the MERN stack is the modern way of building robust, scalable apps.

CONCLUSION

The main motive behind our work is to study the emotions of the users through the social media platform and analyze their behavior and how it affects everyone on the platform in a positive and negative way. In a way, use of sentiment analysis helps in feedback analysis, competitor monitoring etc. Our solution would provide various groups like business or universities to analyze the sentiments of the users related to their circle. The analysis will help universities to make their own social media platform for the use of their institute, The analysis would not only help them improve their future activities but would also help in finding alarming situations, such as depressed users. The solution uses modern technologies which makes the application more scalable, fast, robust and secure. The users would still retain their privacy as their activities won't be tracked. Rather, they would only be analyzed. It also gives escalated privileges to the administrators such as analyzing the content on the platform, broadcasting messages, creating groups, etc.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to A.P Shah Institute of Technology for giving us opportunity to work under this project. We got an opportunity to learn about the sentiment analysis
We would like to thank our HOD Prof. Kiran Deshpande and Prof. Neha Deshmukh for arranging the necessary approvals and guiding us throughout the project work. In the end we will also like to thank our parents for their emotional support and motivation

REFERENCES

[1] Hutto C.J., Gilbert E., VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, AAAI, 2014.

[2] Anees Ul Hassan, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, Sungyoung Lee,'Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression', 2017 International Conference on Information and Communication Technology Convergence (ICTC).

[3] Persia, F., & D'Auria, D. (2017). A Survey of Online Social Networks: Challenges and Opportunities. 2017 IEEE

International Conference on Information Reuse and Integration (IRI).

[4] Rosa, R. L., Rodriguez, D. Z., Schwartz, G. M., de Campos Ribeiro, I., & Bressan, G. (2016). Monitoring system for potential users with depression using sentiment analysis. 2016 IEEE International Conference on Consumer Electronics (ICCE).

[5] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, "Lexicon-based methods for sentiment analysis", Comput. Linguistics, vol. 37, no. 2, pp. 267-307, 2011.

[6] A. kaushik, S. Naithani, ³A Study on Sentiment Analysis: Methods and Tools´ International journal of Science and Research, vol. 4, 2015, pp 287-291.

[7] Y. Rao, Q. Li, X. Mao, and L. Wenyin, ³Sentiment topic models for social emotion mining,´ Information Sciences, vol. 266, pp. 90- 100, 2014.

[8] D. Jiang, Xiangfeng Luo, Junyu Xuan and Zheng Xu, ³ Sentiment computing for the news event based on the social media Big Data´ IEEE Access pp. 1-6, 2016.

[9] D. Tang, B. Qin, and T. Liu, ³Learning semantic representations of users and products for document level sentiment classification,´ in Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, 2015.

[10] J. Xuan et al., "Uncertainty analysis for the keyword system of web events", IEEE Trans. Syst. Man Cybern. Syst., vol. 46, no. 6, pp. 829-842, 2016.

[11] Abinaya. R, Aishwaryaa. P, Baavana. S, ³Automatic Sentiment Analysis of User reviews´, IEEE International conference on Technological Innovations in ICT for Agriculture and Rural Development, 2016, pp 158-162.

[12] Y. Yamamoto, T. Kumamoto, A. Nadamoto, "Multidimensional sentiment calculation method for twitter based on emoticons", Int. J. Pervas. Comput. Commun., vol. 11, no. 2, pp. 212-232, 2015.