International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Social Media Mining: Sentiment Analysis on Twitter Data

Rashi Bhattad, Mansi Satpute, Ujjwal Mishra

Rashi Bhattad, PICT, Pune
Mansi Satpute, PICT, Pune
Ujjwal Mishra, PICT, Pune

Abstract –

"Social Media Mining", Which essentially means a sentiment analysis done on people on social by using various statistics and analyzing algorithms of the pattern of people's activity on social media sites. We were able to get the data from various social media sites and the Machine Learning algorithms were implemented on them for keyword analysis. Which is basically analyzing the patterns about the general populations' behavior regarding a specific topic via keyword searches, mentions or tweets done on the social media sites. We had to get the raw data from these social media sites and then process it in such a way that it was viable for performing the process of Machine Learning on it. We have made use of K-Nearest Neighbor (KNN) algorithm to train the Machine, as well as Natural Language Processing (NLP) for enabling the Machines to understand the Human language. And Natural Language ToolKit (NLKT) for sentiment analysis to obtain insight of the audience on social media.

Key Words: **Social Media Mining, Sentimental Analysis, Natural Language Processing.**

1. INTRODUCTION

In today's world social media sites such as Google, Facebook, YouTube, Twitter, etc. play an important role in every individual's life. People on these sites upload and download data according to their needs. All these social media sites are filled with trillion bytes of data which can be recognized on different aspects of social media and human interactions. To better understand these interactions on the web, social media mining can be done to get a better understanding of the latest trends.

1.1 Literature Survey

The first stage is to find different data sources, geared around answering a specific question and figuring out a way to compile various forms of raw data from multiple sources for this data can be used as input for data mining. The next stage is to explore the data by using visualization tools to improve and filter the initial idea and to transform the data to answer our queries. Moving forward the next stage is to model the data by creating and applying multiple analytical algorithms that will find patterns and draw assumptions based on data presented. Data modelling can be done by three ways which are descriptive, predictive, and prescriptive. Many models are created and by using machine learning algorithms, the most valid and practical models are found. The final stage is to deploy the best models by making calculated decision made by humans, and operational decision, made by machines, which will answer the question. Lastly, with continuous monitoring and measuring of the models, the success of the models' outcomes is evaluated. The four stages are altered slightly to progressively increase the effectiveness of the specific data mining application in question.

1.2 Relevance

Web-based Media Mining includes web-based media, network investigation, and information mining to give a helpful stage to understudies and task administrators to comprehend the extent of online media mining. It presents the issues emerging from web-based media information and presents essential ideas, impending problems, and pragmatic calculations for network investigation and information mining. With the assistance of the Machine Learning course, we can apply ideas, standards, and techniques in different situations of online media mining.

1.2 Scopes and Objective

The scope and objectives of social media mining are to: Understand social aspects of the Web with Social Theories, Social media and Mining. Learn to collect, clean, and representable social media data. To measure essential properties of social media and simulate social media models. Find and analyze communities in social media. Understand how information propagates in social media. Understanding friendships in social media, performing recommendations, and analyzing behavior. Study or ask interesting research issues. Startup ideas/research challenges. Learn representative algorithms and tools

2. THEORETICAL DESCRIPTION

The most common way of mining social information includes a mix of measurable Machine Learning, science, and statistics. The initial step is to accumulate and handle social information from various web-based media sources. Aside from online media stages like Twitter, or YouTube, information diggers likewise remove information from

different sites, news locales, gatherings, or other public pages where clients cooperate and leave remarks. All of this data should then be handled prior to continuing to the subsequent stage. Whenever information is gathered and handled, what follows is the use of different information digging procedures which consider more straightforward ID of normal examples and the connection of different elements in huge datasets. A portion of the more generally utilized online media information mining methods incorporate arrangement, affiliation, following examples, prescient investigation, catchphrase extraction, feeling examination, and market/pattern investigation.

Besides, web-based media information mining additionally utilizes various online media information mining programming answers for upgrade the most common way of mining. Probably the most popular information mining programming arrangements incorporate the accompanying: Microsoft SharePoint, Sisense, IBM Cognos, RapidMiner, and Dundas BI. Given that a more inside and out assessment of information is required, information diggers might choose to utilize AI in the process too.

The last advance in the mining system is to make a visual portrayal of the experiences acquired from the entire interaction to convey the data to the designated crowd. This is generally finished by utilizing web-based media investigation or an assortment of information perception devices, like Infogram, ChartBlocks, Tableau, and Datawrapper, to give some examples.

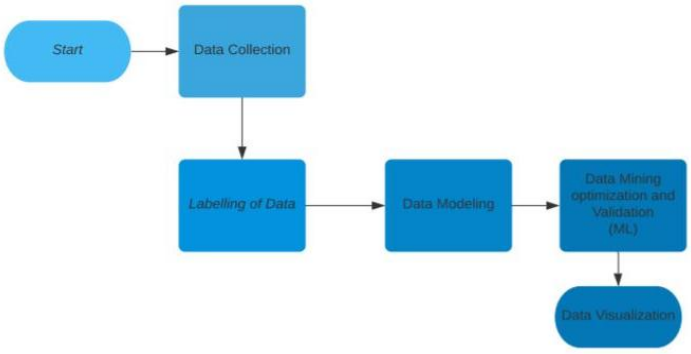

Fig 1- Block Representation

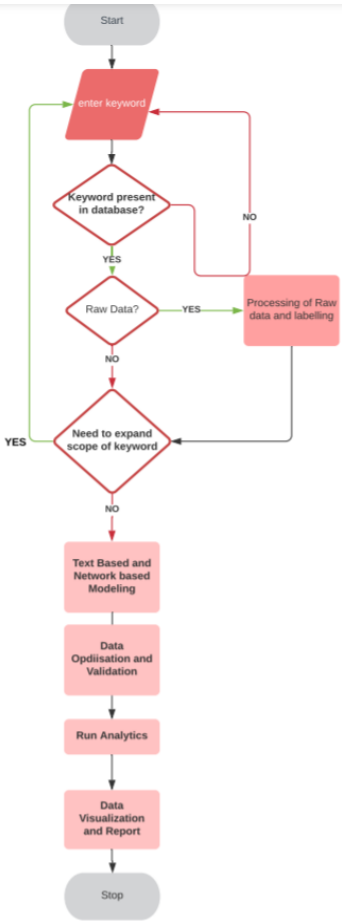

Fig.2-Algorithm Implemented

3. SYSTEM DESIGN

System designing consists of the following steps:-

3.1. Data Collection

This step involves collection of information from social media platforms, aggregating and analyzing it to identify patterns and trends. It harvests various types of data that is either publicly available or is generated on a daily basis on social media platforms. Data collection itself involves various steps such as- data selection, data cleaning, constructing data, data integration, formatting data.

3.2. Data Labelling

As the data collected is raw, data labelling involves assembling similar type of data and adding a meaning to it order to train a machine learning model so that it can recognize patterns in future. Machine learning algorithms learn from this tagged data and recognize repetitive patterns and trends. Also, labelling is also used to identify key features present in the data while minimizing human involvement.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

3.3. Data Modelling

Data modeling is the process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures. In modelling, various methods are selected and applied, and their parameters are measured to optimum values. The tasks involved in modelling include- selecting modeling technique, generating test design, building model, accessing model.

3.4 Data Mining optimization and validation

Optimization in data mining is reducing the cost of a function in order to optimize some performance measures. The required function is minimized with respect to the parameter on the training set and the error encountered at this point is suffuse to be low.

In supervised machine learning the data is divided into train set and test set. Using this train set the model is trained to perform predictions and analyze trends and then the model further becomes the trained model. Validation is the process where in a trained model is evaluated with a test set. The main purpose of validation is to test the generalization ability of the trained model.

3.5. Data Visualization

Visualization of data is the last step in the flow of the mining process. But it does not mean that its use is limited to there only. Data visualization is very important, and it can be done in many ways by using Pie charts, bar graphs, Data Clustering, histograms Wordcloud, etc. All this can be performed using predefined libraries in python such as Matplotlib Seaborn, Bokeh, Altair, Plotly, and ggplot. We could visualize the data and represent it in the form of Bar graphs for negative words and Positive words as hashtag on the x-axis and count for the words on the y-axis.

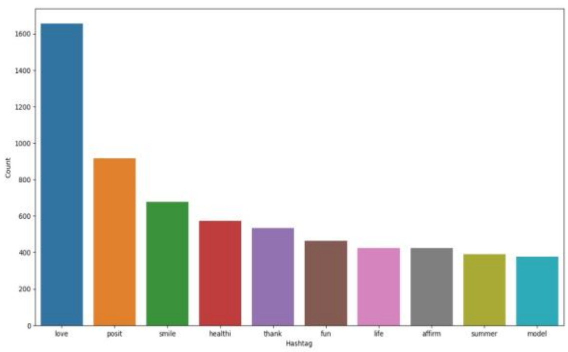

Fig.3 – Graphical Visualization of Positive words

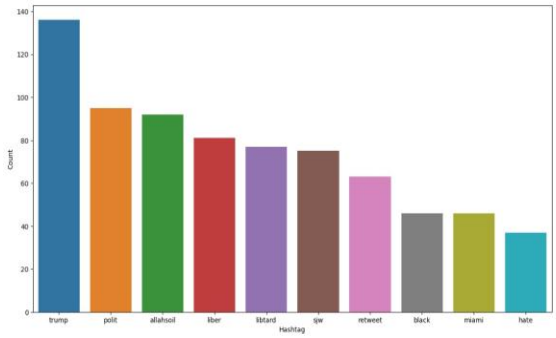

Fig.4 – Graphical Visualization of Negative words

4. RESULTS

We obtained our results in the form of wordcloud for sentiment analysis done by keyword search. We obtained a wordcloud for frequent words that had appeared in the data multiple times, out of which we could separate the words on the basis of their sentiment and segregate it as Positive words and Negative words.

Fig.5 – Frequently Appeared Words

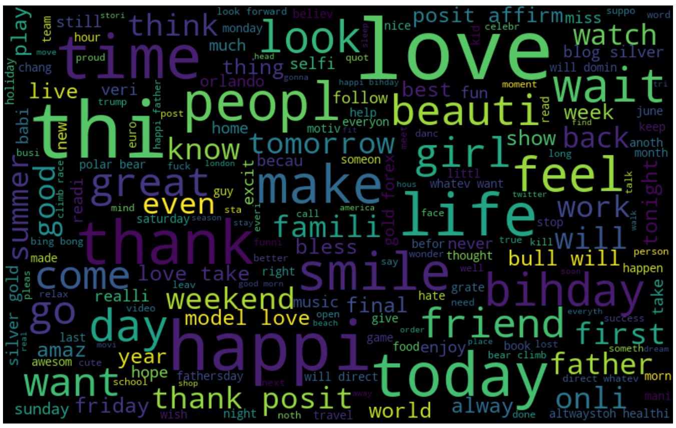

Fig.6 – Positive Wordcloud

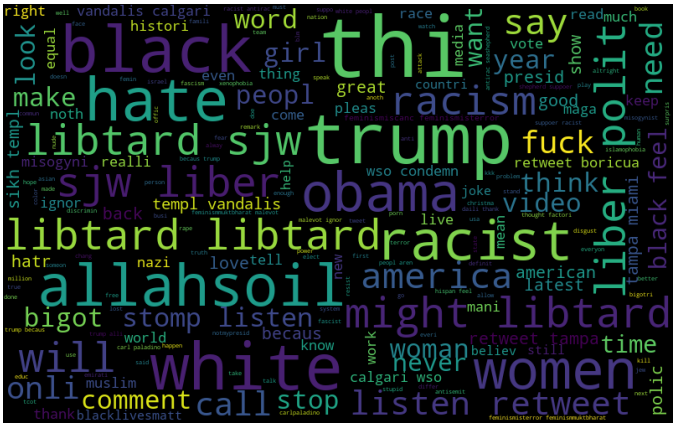

Fig.7 – Negative Wordcloud

5. CONCLUSION

Social media mining basically aims at predicting human behavior by analyzing and mining his social data and is by far the most common strategy used for marketing and business application. Using a wide range of concepts from computer science, machine learning and statistics, patterns and trends are represented, analyzed and sent for further application of use.

In this technical paper, we examined the importance of social network analysis and its applications in various fields. In this scientific study, we concentrated on Twitter and have utilized the Python programming language. A programme that performs sentiment analysis on the data using keyword search. The findings were displayed on using several wordclouds. We realized that the neutral sentences are disproportionately large high, indicating that Twitter has to be improved analysis of feelings.

Applications of social media mining include business development, marketing, social science research, health services and educational purposes.

This concept of mining has reduced the human effort to deal with hundreds of descriptive attributes, automatic analysis of databases and the recognition of important trends and behavioral patterns. Hence, we can conclude by saying that data mining bounds the tools to formally represent, measure and model meaningful patterns from large scale social media data. With the digitization and globalization, social media mining plays an important role in every line of business.

6. FUTURE SCOPE

The way users interact with social media continues to improve and evolve, leading to changes in analytics. The future trends of social media and analytics continue to have extensive effects on how the world communicates and how businesses gather and interpret data. It would help in providing personalized ads to the user via various social media platforms. It would help Business intelligence uses software and services to analyze data and turn it into useful information to help make decisions about marketing, goals, and procedures within the company. With social media analytics, marketers can collect data directly from consumers, including their conversations about products, services, industries, and brands, and use it to make decisions that relate to the future of the company.

7. References

Books:

[1] **"Social Media Mining: An Introduction"** - https://books.google.co.in/books?hl=en&lr=&id=H9FkAwAAQBAJ&oi=fnd&pg=PR15&dq=social+media+mining&ots=cXZ8prCMR9&sig=dooHbnAVK1RQyR20XSgicrZX9PQ#v=onepage&q&f=false

Transaction/Journals:

[2] **"A Study on Social Network Analysis through Data Mining Techniques":**https://ijarcce.com/wp-content/uploads/2016/11/IJARCCE-ICRITCSA-10.pdf

[3] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010).

[4] Scott, J. (2011). Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1), 21-26.

International/National Conference Proceedings:

[5] Social Network Analysis and Mining (SNAM)- by Reda Alhajj https://www.springer.com/journal/13278

[6] Fortuna P, Nunes S (2018) A survey on automatic detection of hate speech in text. Computer Survey 51(4):1–30. https://doi.org/10.1145/3232676

Technical Reports:

[7] Swati Agarwal and Ashish Sureka. 2015. Using KNN and SVM based one-class classifier for detecting online radicalization on Twitter. In Proceedings of the International Conference on Distributed Computing and Internet Technology. Springer, 431--442.

[8] Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17, 26-32.

Website:

[9] Social Media Data Mining: Understanding What It Is and How Businesses Can Use It: www.sandiego.edu/blogs/business/detail.php?_focus=76022

.