

Social Media Analytics with Big Data

¹Heta Chetan Shah, ²Durva Haresh Patel, ³Pankaj Rathod

^{1,2}Student, Shri Bhagubhai Mafatlal Polytechnic, Department of Information Technology, Mumbai, India

³Professor, Shri Bhagubhai Mafatlal Polytechnic, Department of Information Technology, Mumbai, India

Email IDs: shahheta2209@gmail.com, durva.patel2003@gmail.com, pankaj.rathod@sbmp.ac.in

Abstract - Due to the rapid increase in the use of internet and rapid growing businesses, education over internet, Big Data is becoming the fastest growing research area. Opportunities have been provided to every individual. All the data generated by the users on social media (their daily insights, activity, background details, browsing information) is collected in huge amount which is known as 'Big data'. The review about different topics related to big data is discussed in brief i.e., about sentiment analysis and its algorithm.

Keywords: Big Data, social media, data analytics, sentiment analysis.

I. INTRODUCTION

Internet and social media provide real time information, opinions and reviews from all over the world because of the rapid growth in technologies worldwide. People from all over the world uses internet for the purpose of business, marketing, education, etc. and therefore a large amount of data referred to as big data is generated. Generally, the data sets which are included in big data are so enormous and complex that it is burdensome for common software to collect, process, analysis and act on the data. Big data plays a crucial role in social media. It is because of big data we can track all social media comments, insights, tweets, posts, likes and all other activities. [1]

This paper is divided into following sections:

I. Introduction, II. Big data, III. Big social media data, IV. Sentiment analysis, V. Survey, VI. Algorithms, VII. Future Scope for sentiment analysis and followed by Conclusion and References.

II. BIG DATA

Big Data and its analysis have led to several interesting applications in the fields of medicine, social media, and ecommerce.

Big data are stored data lakes. A consolidated storehouse that allows users to store all the structured and unstructured data at any quantity is known as data lakes. Unlike data warehouses, data lakes can store non-relational databases

(NoSQL). For example, data coming from websites, mobile apps and social media. [2,4,5]

Applications:

- Tracking Customer Spending Habit, Shopping Behavior: In online systems big data is useful for analyzing consumer behaviour and their likes and dislikes.
- Smart Traffic System: Data collected through the cameras besides the roads can be used to analyze the real time traffic, for example: Google maps.
- Auto Driving Car: A car can be driven without any human interpretation, for example: Tesla
- Education Sector: Big data can be used in online education system to search information about candidate's interested courses.
- Energy Sector: Big data can be used in meter reading and it can estimate power load every 15 minutes and sends the data to the server.
- Media and Entertainment Sector: Here big data is use to keep track on user activity like the time spent on a particular site like Netflix, Amazon and provides recommendation to the user.[12]

III. BIG SOCIAL MEDIA DATA

Social media in context with big data refers to all kinds of photos, videos and messages that are uploaded or transmitted via social media. The maximum amount of data that is created on the internet is due to social data. Every time we have an interaction on through any social media sites or applications data is created. Fig 1. Shows the processing of social media data. It is carried out in following steps, first the data is collected from social media sources, then integrated at a place and then cleaning is performed. Then the next step is to process the data then the data is supplied to the end user. Any online shopping sites work by understanding its customer which can be done through social media analytics. We get our ads on the browser similar to the things we browse over the internet. [1,3]

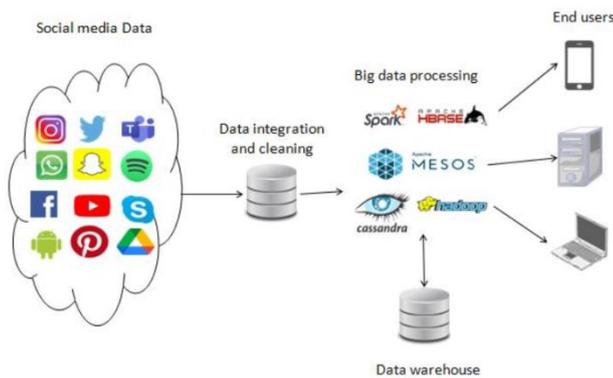


Figure 1: Big social media data processing

IV. SENTIMENT ANALYSIS

The big data phenomenon applied to social media is growing area of study known as ‘sentiment analyses’. The process of sentiment analysis is to extract information about a consumer’s outlook of a product, service or brand. Sentiment analysis studies the subjective information in an expression, that is, the opinions, appraisals, emotions, or attitudes towards a topic, person or entity. The expressions studied in sentiment analysis are classified as positive, negative, or neutral.[2]

A short portrayal of sentiment analysis is demonstrated below with the help of python programming.[11]

Before implementing the below program, we need to install textblob library in command prompt: -

```
pip install -U textblob
```

Python code for sentiment analysis:

```
fromtextblob import TextBlob
text = "\nI am happy after taking diploma."
text1 = "\nI am not happy after taking diploma."
text2 = "\nPost SSC education and Diploma would work the same."

blob = TextBlob(text)
blob1 = TextBlob(text1)
blob2 = TextBlob(text2)

string = [blob,blob1,blob2]

fori in string:
#Sentiment function of textblob returns two properties namely polarity and subjectivity.
    if (i.sentiment.polarity<0):
        print(i)
        print(i.sentiment)
        print("Negative\n")
```

```
elif (i.sentiment.polarity == 0):
    print(i)
    print(i.sentiment)
    print("Neutral\n")
else:
    print(i)
    print(i.sentiment)
    print("Positive\n")
```

“Textblob is python library which is available for free (open-source)

Textblob is used for processing textual data.

This library performs tasks like noun phrase extraction, sentiment analysis, translation, etc on textual data.” [11]

Output:

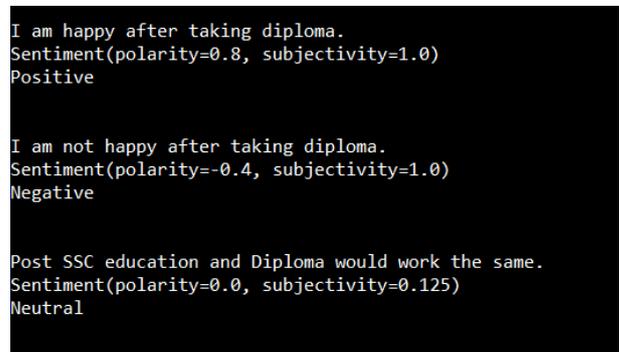


Figure 2: Output of the above program

V. SURVEY

Based on the above program we performed a survey by creating a Google form which gives review above diploma or taking post SSC. We have mentioned three statements as in the above program which represents positive, negative or neutral statement and took responses from people.[8]

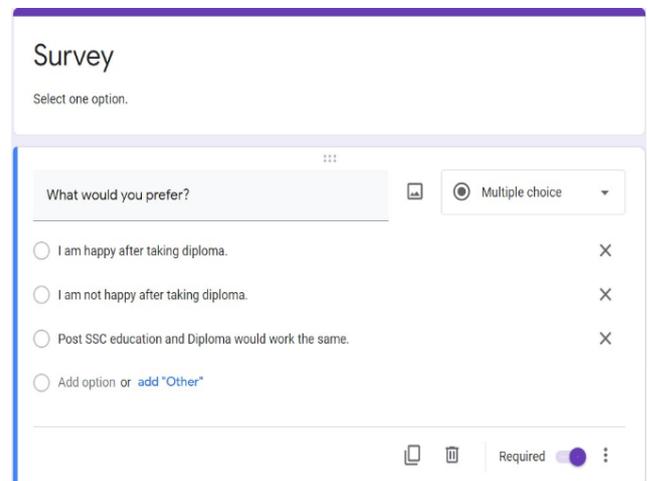


Figure 3: Google form

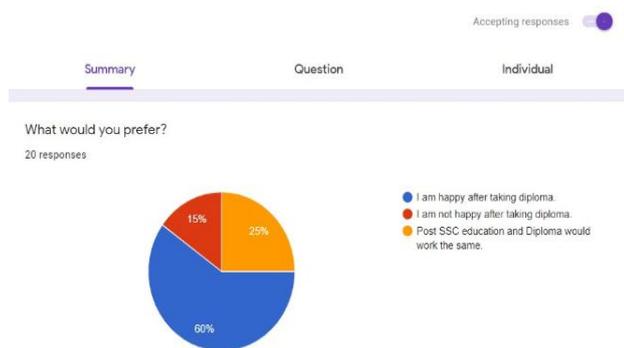


Figure 4: Responses in the form of pie diagram

A1	Timestamp	What would you prefer?
2	11/12/2021 20:49:46	I am happy after taking diploma.
3	11/12/2021 20:51:26	I am not happy after taking diploma.
4	11/12/2021 21:02:23	I am happy after taking diploma.
5	11/12/2021 21:02:41	Post SSC education and Diploma would work the same.
6	11/12/2021 21:03:52	I am happy after taking diploma.
7	11/12/2021 21:04:44	Post SSC education and Diploma would work the same.
8	11/12/2021 21:54:13	Post SSC education and Diploma would work the same.
9	11/12/2021 22:04:04	I am happy after taking diploma.
10	11/12/2021 22:31:05	I am happy after taking diploma.
11	11/12/2021 23:42:31	Post SSC education and Diploma would work the same.
12	11/13/2021 9:49:44	I am happy after taking diploma.
13	11/13/2021 10:04:05	I am not happy after taking diploma.
14	11/13/2021 10:06:52	I am happy after taking diploma.
15	11/13/2021 10:10:29	I am happy after taking diploma.

Figure 5: Responses in form of excel sheet

VI. ALGORITHMS FOR SENTIMENT ANALYSIS

Natural Language Processing (NLP) methods and algorithms are used in sentiment analysis to know the sentiment of a text or emoticon. Some algorithms or approaches used are mentioned below:

A) Rule-based approach

In rule-based approach algorithm based on clearly defined description of an opinion is used. It includes identifying subjectivity, polarity, or the subject of opinion.[4]

Here's how it works:

- There are two lists of words.
 - The first list includes only positive words.
 - The second List includes only negative words.
- The algorithm scans the entire text and finds the word which matches the criteria above.
- The next step is calculation; the algorithm calculates which type of words is more suitable in the text.
- If there are more positive words present in the text then the text is said to have a positive polarity and if there are a greater number of negative words then it is said to have a negative polarity.[4]

B) Automatic Sentiment Analysis

The rule-based approach cannot be considered as a real tool, automated sentiment analysis comes into picture. This approach goes into the detail of the text and delivers appropriate results. This approach uses machine learning instead of clearly defined rules to get precise meaning of the text. This increases the precision and accuracy of the processes and you can curate the information on various criteria without making the process too complex.

The automatic approach includes supervised machine learning classification algorithms. Sentiment analysis could be considered as one of the most sophisticated examples of how to use classification to maximum effect. Unsupervised machine learning algorithms are also used to explore data.[2,4]

C) Approaches used in automatic sentiment analysis

Before discussing about approaches in automatic sentiment let discuss in short again what is automatic sentiment analysis. Automatic sentiment analysis doesn't depend on any predefined set of rules but it depends on machine learning and its techniques. So, it automatically classifies the sentiment of the text or emoticon that whether it is positive, negative or neutral.[2]

a) Emotion-based approach

In this approach the selection of a tweet or comment is done only if it contains at least one emoticon. The sentiment of the text is known based according to the emoticon.

Example:

If the emoticon is :) - The feeling expressed is happy so it shows positive sentiment.

If the emoticon is :(- The feeling expressed is sad so it shows negative sentiment.[2]

b) Word-based approach

In this approach the selection of tweet or comment is automatic i.e., the words that express sentiment like good, bad, excellent or terrible. So, from the above-mentioned words we can estimate whether it is positive or negative.[2]

c) Hybrid approach

In the hybrid approach both emoticons and words are used as classification criteria. Based on emoticons and words it estimates whether the tweet or comment is positive or negative.[2]

VII. FUTURE SCOPE

By the amount of data that is being generate day by day it is estimated that data will be growing rapidly in future. The volume of data will increase to about 175ZB (zetta bytes) by 2025. Cloud storage will be utilized to store half of this data. The demand for Data scientists and CDOs (Chief Data Officers) will increase, as without these professionals who collect and analyze data using analytics and reporting tools, big data will be useless. As data is growing into huge volumes every day, it also put forth a challenge in protecting it from intrusions and cyber attacks. Hence, privacy will be a matter of concern. 'Fast Data' and 'Actional Data will come into picture'. Fast data allows for processing in real time streams. It could bring value to organizations which takes actions and makes decision when data arrives. Actionable big data is the bridge between business value and big data.[13]

VIII. CONCLUSION

Big data is a game changer, various organizations use the strategic information extracted from big data for the betterment of the organization. Huge profit can be gained from the slightest of change in the efficiency and smallest savings this is the reason why most of the organization and business are moving towards big data. Big data makes it easy for the social marketer to focus on a particular buyer of consumer and recommend the product in which they are interested which would result in a great success.[14]

REFERENCES

- [1] Norjihan Abdul Ghani, Suraya Hamid, Ibrahim AbakerTargioHashem, Ejaz Ahmed, "Big Social Media Data Analytics: A Survey, Computers in Human Behavior" (2018).

- [2] Ana C.E.S.Lima, Leandro N. de Castro, "Automatic Sentiment Analysis of Twitter Messages", 2012 Forth International Conference on computational Aspects of Social Networks (CASonN).
- [3] Vikas Dhawan and Nadir Zanini, "Big Data and Social media analytics", Research Matters 2014: A Cambridge Assessment publication.
- [4] <https://theappsolutions.com/blog/development/sentiment-analysis/>
- [5] Ahmed Oussous, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, Samir Belfkih, "Big Data Technologies: A Survey", Journal of King Saud University - Computer and Information Sciences 2017.
- [6] <https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications>
- [7] <https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-Big-Data.html>
- [8] S.Uma Maheshwari, S.S Dhenakaran, "Aspect based Fuzzy Logic Sentiment Analysis on Social media" Publisher- IEEE.
- [9] <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>
- [10] https://www.researchgate.net/figure/An-example-of-aspect-based-sentiment-analysis-ABSA-Sentiment-expressions-such-as_fig1_335075154
- [11] <https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/>
- [12] <https://www.geeksforgeeks.org/applications-of-big-data/>
- [13] <https://www.itransition.com/blog/the-future-of-big-data>
- [14] <https://www.educba.com/what-is-big-data-analytics/>

Citation of this Article:

Heta Chetan Shah, Durva Haresh Patel, Pankaj Rathod, "Social Media Analytics with Big Data" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 5, Issue 12, pp 60-63, December 2021. Article DOI <https://doi.org/10.47001/IRJIET/2021.512012>
