Human Behavior Analysis With Social Media Platform Influence

Mandeep Kumar Gupta
B-tech(CSE)
Galgotias University
Uttar Pradesh, India

Utkarsh Jaiswal
B-tech (CSE)
Galgotias University
Uttar Pradesh, India

Dr. Sudeshna Chakraborty
Associate Professor
Galgotias University
Uttar Pradesh, India

Abstract - Sentiment analysis has advanced, allowing us to better understand how individuals or a specific person behaves through their comments or tweets on numerous social media platforms. To comprehend the general feelings present in the data gathered from various social media, sentiment analysis or opinion mining is used. Because of the Internet and many social media platforms like Twitter, Facebook, Instagram, and others where people may share their opinions, people are more exposed to the outside world. Social media is become more valuable to the general people because to affordable and quick communication. Applications in science and business can be made of social media data. Sentiment analysis and behavior analysis together made it easier and simpler to collect essential or valuable data for a variety of applications, such as character analysis and depression testing, among others. The sentiment score for the text and emoticons that was collected throughout the analysis will also be used to guide the behavior analysis. This essay covers the use of sentiment analysis to analyze user behavior on social media data.

***Key Words*: Emotions, Emojis, Feature Extraction, Classification, Behaviour Analysis, Sentiment Analysis, Data Pre-processing, Natural Language Processing, and Emotions**

1. INTRODUCTION

1.1 Behavior Analysis

The science that aids in understanding human behaviour is behaviour analysis. It investigates the effects of biological, pharmacological, and experiential elements on human behaviour. Individuals engage in behaviour, and behaviour analysts place specific attention on researching the variables that consistently affect people's behaviour. This is the science whose findings have helped to address socially significant behaviors including drug use, good eating, workplace safety, education, and the treatment of pervasive developmental disabilities (e.g., autism)[1].

1.2 Sentiment Analysis

Sentiment analysis is the practice of examining people's feelings and attitudes in a variety of contexts. It looks into texts that contain the user's opinions, feelings, etc.[2]. in order to determine how they would react in certain situations. It's becoming more popular these days because many institutions and organizations use this method to comprehend the perspectives and ideas of different people. For instance, the consumption of a product can be studied by looking at how people respond to it. Using a number of classification algorithms, the information or reviews submitted by various users in a social media network like Twitter can be categorized[3]. The data taken from social media will be processed using a variety of Natural Language processing methods.

1.2 Historical Analysis

The process of looking at the past through historical analysis is to examine the evidence. It is frequently located in the documentary evidence. It mostly consists of information produced within an organization, whether manually or mechanically[4]. Press releases, log files, financial reports, project and product documentation, emails, and other forms of communication are among the sources used for historical analysis. The main drawbacks of this study are that historical Data require more storage space and that it cannot be kept for a longer length of

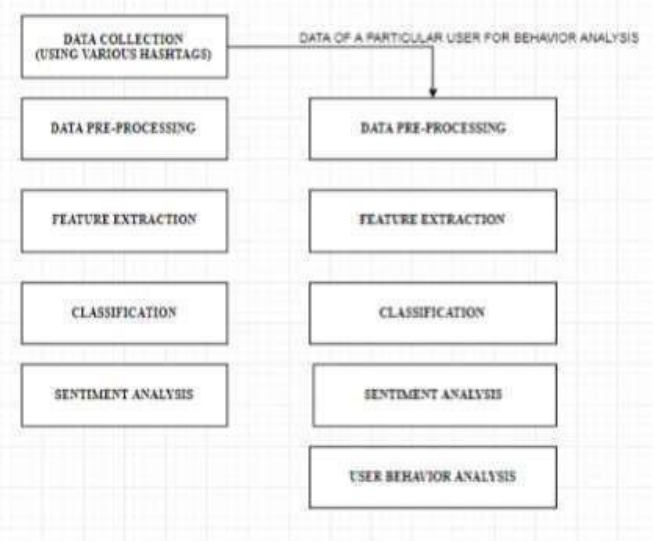

Fig 1 : Overall Approach

Using the keyword matching method, hashtags can be used to extract data from different social media platforms. For instance, #twitter can be used to retrieve data that matches the hashtag[5]. Since 140 characters is the maximum word count for data that can be posted online, such as tweets. Following data extraction, a number of pre-processing procedures will be completed, including the elimination of URLs, special symbols, full stops, stop words, and other ineffective or useless data. For this specific technique, many Natural Language tools or packages are required.

Using techniques like Term Frequency, Bag of Words, N-grams, etc., various features such as Sentiment features, Unigram features, Sarcastic features, Semantic features, etc. will be extracted from the tweets or data from social media[6]. A given tweet's retrieved features will be used to determine the polarity or score. Following the calculation of the score, the tweets will be classified using a variety of classification methods, including Naive Bayes, K-NN, Random Forest, etc. The accuracy of each class will then be determined[7].

2. PRELIMINARY REVIEW

- The Emotions Classification According to earlier studies, the majority of them occurred in the texts' Binary and Ternary Classification Algorithms employed are the main categories that this review falls within.
- The Features Employed
- Emojis

Classification Algorithms

When performing a multiclass classification in sentiment analysis, it is necessary to classify the words or emotions in order to determine user behaviour. The classification process can be carried out using a number of algorithms, such as Naive Bayes, K-NN, Random Forest, Support Vector Machine, etc.[8].

Decision Tree: Used to divide data into more manageable classes. Every node and leaf in this diagram represents a class of data, and each level represents a choice. Numerous types of variables, such as nominal (categorical and non-ordered), ordinal (categorical and ordered), and interval values, can be handled by decision trees (ordered values that can be averaged). They are categorical if they are made up of discrete categories without any determinable intrinsic value. Variables with an ordering adhere to it. Pure sets, or sets that don't need to be further subdivided, are a crucial component of decision trees. It is uncertainty in other sets. Vasile Paul Bresfelean (2007) used decision trees in the Weka environment to analyze and forecast student behaviour[9].

K-means: The number k, or the number of clusters to be created, serves as the algorithm's input. To match k patterns inside the hypervolume containing the specified pattern, choose k cluster centers at random. To receive each pattern, select the nearest cluster centers. Using the existing cluster memberships, recalculate the cluster canters. K-means was used by Liming Xue et al. (2015) to assess user behaviour[10].

Useful for identifying intriguing relationships between variables in sizable databases are association rules. It is employed to detect robust rules found in databases utilizing certain pertinent metrics. New rules are also produced as more data is analyzed using this rule-based methodology. The final objective is to assist a computer duplicate the feature extraction and abstract association capabilities of the human brain from fresh uncategorized data.

R. Geetharamani et al. (2015)[11] used association rule miningto forecast how visitors will access webpages.

Neural Networks: Neural Networks are capable of extracting meaning from voluminous or ambiguous data. Additionally, it can be used to extract patterns and identify trends that are too complicated for people or other computer tools to pick up on. A trained neural network can be compared to an authority in the field of information analysis. The benefits comprise: They are capable of

learning training or initial experience, The information it receives during the learning process can be organized or represented in a way that is uniquely its own. Parallel computations are possible, and specialized hardware is being developed. The performance of a network will be affected, among other things, if it is partially destroyed. Neural networks can be used to assess and forecast anomalous user behavior, according to Zheng Ruijuan et al. (2016)[12].

2.1 Features

The features that are extracted by the feature extraction technique include sentiment features, unigram features, punctuation features, sarcastic features, syntactic and stylistic features, top words, semantic features, pattern related features, hate speech features, etc. To calculate the sentiments, the extracted attributes from the dataset that were gathered based on the polarity of the sentiment can be used.

2.2 Emotions and Emojis

The examination of emoticons can also be used to calculate sentiment. Additionally, emoticons have precise textual meanings that can be used for sentiment research. These data will be obtained from several social media platforms, and the emotions can be analyzed similarly to texts. The emotions and emojis will be categorized during the procedure based on the sentiment polarities, and the emojis will also be considered emotions[13].

Table -1: Comparison of user behavior analysis model

Model	Core Algorithm
Classification algorithm model	Decision Tree
Clustering algorithm model	K-means
Association rule model	Association rules algorithm
Sequential pattern mining	Time series analysis
Neural network model	Network Model based on RBF
Factor analysis model	Principal component analysis/Factor analysis

3. OUTCOME OF SURVEY

Data gathered from social media, in this case twitter, can be used to determine people's emotions in specific situations. Binary, Ternary, and Multiclass Classifications are three different classification kinds that can be seen. However, in Ternary Classification, a new class called Neutral has been added, where the sentiments other than Positive and Negative are classified. In Binary Classification, the sentiments are separated into Positive and Negative[14]. Generally speaking, less accuracy can be demonstrated. Moreover, emotions can be included into the text and utilized to represent a range of emotions. Data is classified using a variety of techniques based on features, patterns, or sentiment analysis, such as Neural Networks, Association Rules, K-NN, K Means, Naive Bayes, Random Forest, Decision Tree, and Support Vector Machine. The efficiency of various categorization algorithms is evaluated using the four performance criteria Accuracy, Prediction, Recall, and F-Measure[15].

Neural network algorithms can be used to process or retrieve more precise and accurate classification or identification of user behavior because Spark is utilized for speedy data **p**rocessing and includes an in-memory database to handle and analyze data.

4. SUMMARY AND CONCLUSIONS

Sentiment analysis has been modified to include behavior analysis, which uses data uploaded by users in various social media platforms to determine or comprehend a user's behavior. Here, the information will be examined in order to comprehend the feelings or sentiments of a specific user. The attitudes or features will be classified using different classification algorithms following all of the pre-processing and feature extraction procedures. The data of a certain user will be categorized into several sentiment classes, such as positive, negative, and neutral, during the sentiment analysis phase. The user data will then be subjected to a behavior analysis technique based on the user's areas of interest following the sentiment analysis.

REFERENCES

1. Chhaya Chauhan, Smriti Sehgal "Sentiment Analysison Product Reviews," International Conference on Computing, Communication and Automation, 2017.
2. ZHU Nanli, ZOU Ping, LI Weiguo, CHENG Meng "Sentiment Analysis: A Literature Review," IEEE ISMOT, 2012.
3. Kiichi Tago, Qun Jin "Analyzing Influence of Emotional Tweets on User Relationships by Naive Bayes Classification and Statistical Tests," 10th International Conference on Service-Oriented Computing and Applications, IEEE, 2017.
4. Nasser Alsaedi, Pete Burnap "Feature Extraction and Analysis for Identifying Disruptive Events from Social Media," International Conference on Advances in Social Networks Analysis and Mining IEEE/ACM, 2015
5. Mondher Bouazizi, Tomoaki Ohtsuki "Sarcasm Detection in Twitter," IEEE, 2015
6. Hajime Watanabe, Mondher Bouazizi, Tomoaki Ohtsuki "Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection," IEEE Access, 2018
7. Mondher Bouazizi, Tomoaki Ohtsuki "Sentiment Analysis in Twitter: From Classification to Quantification of Sentiments within Tweets," IEEE, 2016
8. Mondher Bouazizi, Tomoaki Ohtsuki "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," IEEE Access, 2017
9. Alexandros Baltas, Andreas Kanavos, Athanasios K. Tsakalidis "An Apache Spark Implementation for Sentiment Analysis on Twitter Data," 2017
10. Mondher Bouazizi, Tomoaki Ohtsuki "A Large-Scale Sentiment Data Classification for Online Reviews under Apache Spark," 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks, 2018
11. Hao Wang, Jorge A. Castanon "Sentiment Expression via Emoticons on Social Media," International Conference on Big Data, IEEE 2017
12. Wies law Wolny, "Sentiment Analysis of Twitter data using Emoticons and Emoji ideograms," 2016
13. Pranali Borele, Dilipkumar A. Borikar, "An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques," 2016
14. Zheng Ruijuan, Spyros Sioutas, Athanasios Tsakalidis, Giannis Tzimas "Large Scale Sentiment Analysis on Twitter with Spark," EDBT/ICDT Joint Conference, 2016
15. Andreas Kanavos, Nikolaos Nodarakis, Spyros Sioutas, Athanasios Tsakalidis, Dimitrios Tsolis, Giannis Tzimas "Large Scale Implementations for Twitter Sentiment Classification," MPDI Algorithms, 2018