# User Behavior Analysis on Social Media data using Sentiment Analysis or Opinion Mining

**Nirmal Varghese Babu[1], Fabeela Ali Rawther[2]**

*[1]Student, Department of Computer Science and Engineering, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India*
*[2]Assistant Professor, Department of Computer Science and Engineering, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India*

---

**Abstract -** Understanding the behaviour of people or a particular user using his comments or tweets in various social media is an advancement of the Sentiment Analysis. Sentiment Analysis or Opinion Mining is used to understand the overall sentiments present in the data collected from various social media. The people have more exposure to the outside world due to the existence of the Internet and Various Social Medias like Twitter, Facebook, Instagram etc. where they will be sharing their thoughts. Cheap and fast communication has made social media more valuable among the public. Social Media data can be used for various Scientific and commercial applications. The combination of Sentiment Analysis and Behavior Analysis made the extraction of needed or useful data more easy and simple for various applications which include Character analyzing, Depression Testing etc. Moreover, the behaviour analysis will be done based on the text and emoticon sentiment score obtained during the analysis. This paper describes the User Behavior Analysis on the Social Media Data using Sentiment Analysis.

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

## 1. INTRODUCTION

### 1.1 Behavior Analysis

Behaviour analysis[21],[24] is the science that helps to understand the behaviour of individuals. It studies how biological, pharmacological, and experiential factors influence the behaviour of humans. Behaviour is something that individuals do, behaviour analysts have a special emphasis on studying factors that reliably influence the behaviour of individuals. This is the science which has made discoveries that have proven useful in addressing socially important behaviour such as drug taking, healthy eating, workplace safety, education, and the treatment of pervasive developmental disabilities (e.g., autism).

### 1.2 Sentiment Analysis

Sentiment Analysis is the process of analyzing the sentiments and emotions of various people in various situations. It aims at a user's attitude toward various situations by investigating and extracting texts which involve the user's opinion, sentiments etc.[2] Nowadays, it's an emerging trend because a lot of organizations or institutions following this procedure to understand the views and opinions of various people. For example, the usage of a particular product can be analyzed by the way people respond to it[1]. Various classification algorithms can be used to classify the data or reviews posted by various users in a social media say, Twitter. Various Natural Language processing tools will be used to process the data extracted from social media.

### 1.2 Historical Analysis

Historical analysis is the method of examination of evidence to understand the past. It usually found in the evidence contained in documents. It mostly contains the data generated either manually or automatically within an enterprise. Press releases, log files, financial reports, project and product documentation, email and other communication are the sources of historical analysis. The main disadvantages of this analysis are that it cannot be retained for a longer period of time and historical Data need more storage space.

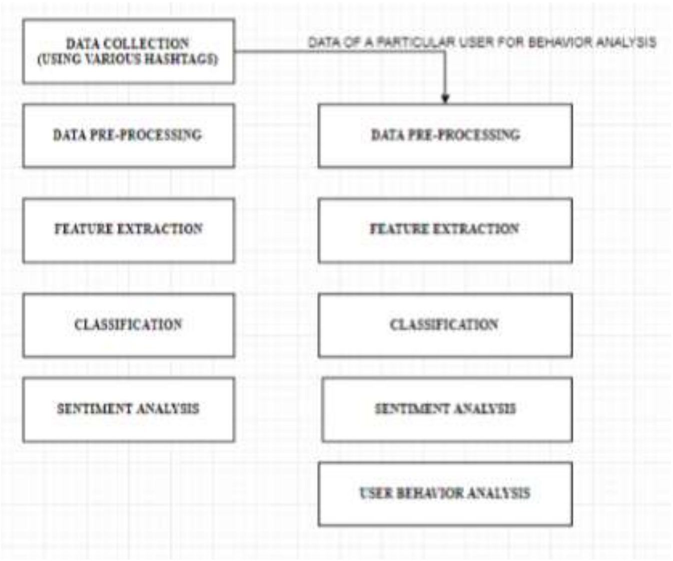


Fig 1 : Overall Approach

Hashtags can be used to extract the data from the various Social Media using the keyword Matching process where the data matching with the hashtags will be extracted, for example, #twitter. Since the maximum word count of data which can be posted online ie tweets is 140. After, the extraction of data, various pre-processing[3] steps which include removal of URL's, Special symbols, Full stops, Stop words etc. will be done where the invalid or not useful data will be removed. Various Natural Language tools or packages are called for this particular procedure.

Various features include Sentiment features, Unigram features, Sarcastic features, Semantic features etc will be extracted from the tweets or data from the social media using algorithms like Term Frequency, Bag of Words, N-grams etc. The polarity or the score will be calculated based on the extracted features available in a particular tweet. After finding the score, various classification algorithms like Naive Bayes, K-NN, Random Forest etc. will be used to classify the tweets based in the various class or sentiments then the accuracy of each class will be calculated.[17][18]

## 2. PRELIMINARY REVIEW

Based on the previous researches, most of them have taken place in the Binary and Ternary Classification of the texts. This review can be categorized mostly by:

– The Classification Algorithms used.
– The Features used.
– Emoticons and Emojis

### 2.1 Classification Algorithms

In Sentiment Analysis, the classification of the sentiments or words are required to identify the user behaviour or to perform the Multiclass Classification. There are several algorithms[19] which can be used to perform the classification which includes Naive Bayes, K-NN, Random Forest, Support Vector Machine etc.

Decision Tree: Used for splitting the data into smaller classes. Here, each level represents the decision and all nodes and leaves consists of a class of data. There are various types of variables accepted by decision trees: Nominal (categorical and non-ordered), ordinal (categorical and ordered) and interval values (ordered values that can be averaged). Categorical means that they consist of discrete categories with no inherent value that could be computed upon. Ordered variables follow an ordering. A key ingredient in decision trees has pure sets: sets that do not need to be split further. In other sets, it is uncertainty[32]. Vasile Paul Bresfelean(2007), analyzed and predicted Student's Behavior Using Decision Trees in Weka Environment.

K-means: Input to the algorithm consists of k, the number of clusters to be constructed[33]. Choose k cluster centres to coincide with k, randomly selected patterns inside the hypervolume containing the pattern set. Assign each pattern to the closest cluster centres. Recompute the cluster canters using the current cluster memberships. Liming Xue et all(2015), analyzed user behaviour using the K-means algorithm.

Association Rules: Used for discovering interesting relations between variables in large databases[31]. It is used to identify strong rules discovered in databases using some measures of interest. This rule-based approach also generates new rules as it analyzes more data. The ultimate goal is to help a machine mimic the human brain's feature extraction and abstract association capabilities from new uncategorized data. R. Geetharamani et all(2015), predicted users webpage access behaviour using association rule mining.

Neural Networks: Neural Networks[16],[20],[27]-[30] have the ability to derive meaning from complicated or imprecise data. It can also be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an expert in the category of information it has been given to analyse. The advantages include: They have the ability to learn to do tasks based on the data given for training or initial experience, It can create its own organisation or representation of the information it receives during learning time, The computations may be carried out in parallel and special hardware devices are being designed and manufactured, Partial destruction of a network leads to the corresponding degradation of performance etc. Zheng Ruijuan et all(2016), stated that Neural Networks can be used to analyze and predict abnormal user behaviour.

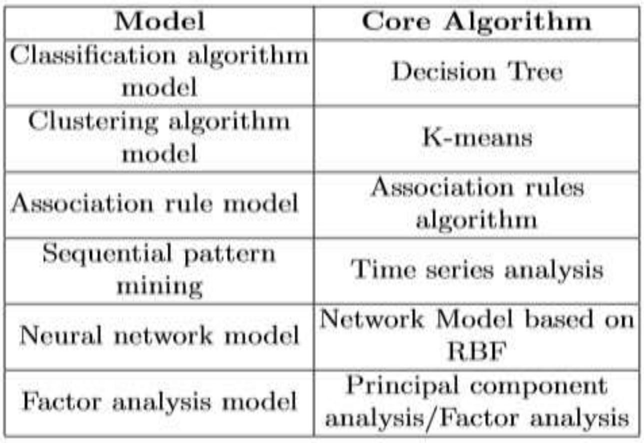
### 2.2 Features

The Features[4]-[13] that are extracted during the feature extraction procedure which may include Sentiment Features, Unigram Features, Punctuation Features, Sarcastic Features, Syntactic and Stylistic Features, Top Words, Semantic Features, Pattern related Features, Hate Speech Features etc. The Sentiments can be calculated using the extracted features collected from the dataset based on the sentiment polarity.

### 2.3 Emoticons and Emojis

Sentiment analysis can also be calculated using the analysis of Emoticons[14],[22]. The emoticons too have exact textual meaning where the sentiment analysis can be done based on these textual meaning. These data will be downloaded from various social media[15][23],[25],[26] and the emoticons can be processed as the same way of the texts. Moreover, the emojis will also be considered as emoticons where the emoticons and emojis will be classified during the process based on the sentiment polarities.

**Table -1:** Comparison of user behavior analysis model, Source*: Mimi Zhang et all(2015)



| 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

The Sentiments of people under certain circumstances can be identified using the data collected from social media, say twitter here. Three types of Classifications can be observed: Binary, Ternary and MultiClass Classification. The sentiments will be calculated into Positive and Negative in Binary Classification whereas in Ternary Classification, a new addition of class called Neutral, where the sentiments other than Positive and Negative will be classified. Mostly, less accuracy can be observed. Also, emoticons can also be a part of the text where it can be used to identify various sentiments. The classifications are done by various classification algorithms like Neural Networks, Association Rules, K-NN, K Means, Naive Bayes, Random Forest, Decision Tree, Support Vector Machine etc. using the features and patterns collected or from the sentiments collected from the data. Accuracy, Prediction, Recall and F-Measure are the 4 performance indicators used to calculate the efficiency of various classification algorithms.

More precise and accurate classification or identification of user behaviour can be processed or obtained using Neural Network Algorithm in Spark Architecture since Spark is used for fast processing of data since it has an In-memory database to handle and process data.

## 4. SUMMARY AND CONCLUSIONS

Behavior analysis is a modification of the already existing Sentiment Analysis where the behaviour of a particular user will be find out or understand **using** various tweets or data posted by a user in various social media. Here, the data will be analyzed to understand the sentiments or emotions of a particular user. After all the pre-processing and feature extraction steps, the sentiments or features will be classified based on various classification algorithms. In Sentiment Analysis procedure, the data of a particular user will be classified into various sentiment classes like Positive, Negative, Neutral etc. After the Sentiment analysis, the user data will be analyzed using the behaviour analysis procedure based on the fields or area of his/her interest.

## REFERENCES

1. Chhaya Chauhan, Smriti Sehgal "Sentiment Analysis on 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. Balakrishnan Gokulakrishnan, Pavalanathan Priyanthan, Thiruchittampalam Raghavan, Nadarajah Prasath, AShehan Perera "Opinion Mining and Sentiment Analysis on a Twitter Data Stream," The International Conference on Advances in ICT for Emerging Regions, 2012.
4. Huma Parveen, Prof. Shikha Pandey "Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm," IEEE, 2016
5. M. Tirupati, Suresh Pabboku, G. Narasimha, "Sentiment Analysis on Twitter using Streaming API," The International Advance Computing Conference, 2017.
6. 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.
7. 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
8. Mondher Bouazizi, Tomoaki Ohtsuki "Sarcasm Detection in Twitter," IEEE, 2015
9. 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
10. Mondher Bouazizi, Tomoaki Ohtsuki "Sentiment Analysis in Twitter: From Classification to Quantification of Sentiments within Tweets," IEEE, 2016
11. Mondher Bouazizi, Tomoaki Ohtsuki "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," IEEE Access, 2017
12. Alexandros Baltas, Andreas Kanavos, Athanasios K. Tsakalidis "An Apache Spark Implementation for Sentiment Analysis on Twitter Data," 2017
13. 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
14. Hao Wang, Jorge A. Castanon "Sentiment Expression via Emoticons on Social Media," International Conference on Big Data, IEEE 2017
15. Wies law Wolny, "Sentiment Analysis of Twitter data using Emoticons and Emoji ideograms," 2016
16. Pranali Borele, Dilipkumar A. Borikar, "An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques," 2016
17. Nikolaos Nodarakis, Spyros Sioutas, Athanasios Tsakalidis, Giannis Tzimas "Large Scale Sentiment Analysis on Twitter with Spark," EDBT/ICDT Joint Conference, 2016
18. Andreas Kanavos, Nikolaos Nodarakis, Spyros Sioutas, Athanasios Tsakalidis, Dimitrios Tsolis, Giannis Tzimas "Large Scale Implementations for Twitter Sentiment Classification," MPDI Algorithms, 2018
19. R. Ragupathy, Lakshmana Phaneendra Maguluri "Comparative analysis of machine learning algorithms on social media test," International Journal of Engineering Technology, 2018
20. Ahan M R, Honnesh Rohmetra, Ayush Mungad "Social Network Analysis using Data Segmentation and Neural Networks," International Research Journal of Engineering and Technology (IRJET), 2018
21. Mario Cannataro, Nicola Ielpo, and Barbara Calabrese "Using social networks data for behaviour and sentiment analysis," Springer International Publishing Switzerland, 2015
22. Georgios S. Solakidis, Konstantinos N. Vavliakis, Pericles A. Mitkas "Multilingual Sentiment Analysis using Emoticons and Keywords," IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014
23. N. AZMINA M. ZAMANI, SITI Z. Z. ABIDIN, NASIROH OMAR1, M. Z. Z.ABIDEN "Sentiment Analysis: Determining People's Emotions in Facebook ," Applied Computational Science
24. Saqib Iqbal, Ali Zulqurnain, Yaqoob Wani, Khalid Hussain "The survey of sentiment and opinion mining for behavior analysis of social media," International Journal of Computer Science Engineering Survey (IJCSES) Vol.6, No.5, 2015
25. Wies law Wolny "Emotion Analysis of Twitter Data That Use Emoticons and Emoji Ideograms," 25TH INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS DEVELOPMENT, 2016
26. Ga¨el Guibon, Magalie Ochs, Patrice Ballot "From Emojis to Sentiment Analysis," https://hal-amu.archives-ouvertes.fr/hal-01529708, 2017
27. Anindita A Khade "Performing Customer Behavior Analysis using Big Data Analytics," 7th International

Conference on Communication, Computing and Virtualization, 2016

28. Siva Subramanian Raju, Prabha Dhandayudham "Prediction of customer behaviour analysis using classification algorithms," AIP Conference Proceedings, 2018

29. Mimi ZHANG, Yan WANG, Jianping CHAI "Review of User Behavior Analysis Based on Big Data: Method and Application," International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII), 2015

30. Zheng Ruijuan, Chen Jing, Zhang Mingchuan, Zhu Junlong, Wu Qingtao "User abnormal behaviour analysis based on neural network clustering," The Journal of China Universities of Posts and Telecommunications, 2016

31. R GEETHARAMANI1, P REVATHY and SHOMONA G JACOB "Prediction of users webpage access behaviour using association rule mining," Indian Academy of Sciences, 2015

32. Vasile Paul Bresfelean, "Analysis and Predictions on Students' Behavior Using Decision Trees in Weka Environment," 29th International Conference on Information Technology Interfaces, 2007

33. Liming Xue, Weixin Luan "Improved K-means Algorithm in User Behavior Analysis," Ninth International Conference on Frontier of Computer Science and Technology, 2015