

**SOCIAL NETWORKS USER BEHAVIOR AND CLASSIFICATION  
TECHNIQUES**D.Sridhar<sup>1</sup>, Dr.V.Kathiresan<sup>2</sup><sup>1</sup>Research Scholar & Assistant Professor, Department Of Computer Science,  
Dr.SNSRCAS, Coimbatore Tamilnadu, India<sup>2</sup> HoD, Department of Computer Applications (PG), Dr.SNSRCAS, Coimbatore, Tamil Nadu, India

**Abstract** — Today human life is mainly fulfilled with digital devices like mobile phones, iphone, laptop, digital TV etc., in this internet plays a vital role. Social networks such as facebook, twitter, LinkedIn, whatsapp are making themselves to be independent. In this busy world, using digital devices we can connect anytime anywhere in the world. Digital device technologies still growing day by day with new trends and ideas. Now day's internet is used like as a cable connection in homes, buildings and famous places. The technologies like mobile shopping, Tele shopping, online marketing, online booking, e- learning and Online Social Networks (OSN) like this step by step has improved according to the modern life. It is very beneficial to the user but it also leads to internet addiction and depression, variation in mental health status of day to day life.

**Keywords**- Web usage mining, Behavior Analysis, Mental Health Status, Machine Learning Algorithms

**I. INTRODUCTION:**

Data in Web Usage Mining, can be obtained in server logs, browser logs, proxy logs, or collected from an organization's database. These data collections vary in terms of the location of the data source, the kinds of data available, the segment of population from which the data was obtained, and techniques of implementation [16]

The web mining additionally enables organizations in such a way that companies, banks etc can keep check on the fake installments. Data mining can be used to investigate this kind of information and data. This data can be helped to grow better progressed and defensive techniques which could be used in the prevention of all the fraudulent activities. Web data mining innovation is not only concerned about simply assembling the information but rather it is likewise providing a great deal of concerns identified with the security of the crucial information. On the internet abundant of personal data is available and to secure it, web mining played a major role [11].

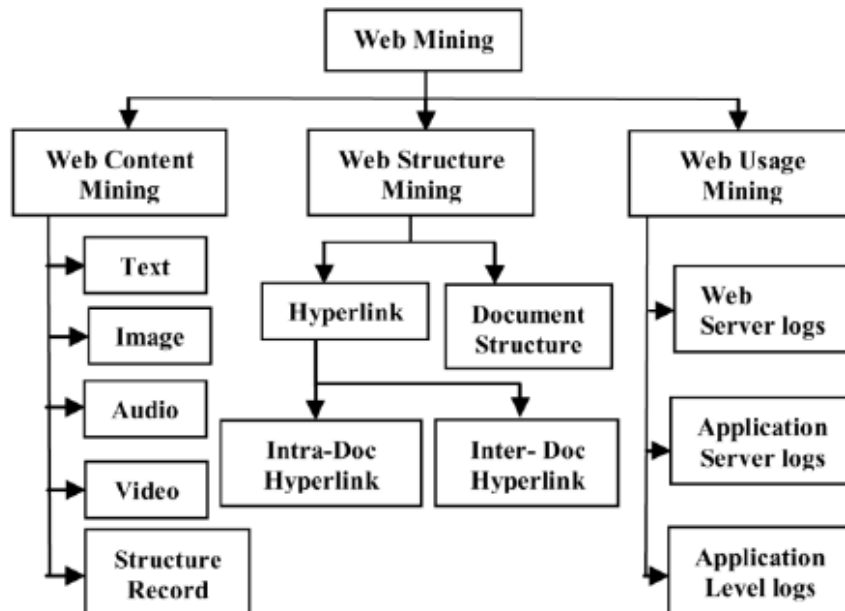
The three basic categories of web mining are:

- Web Content Mining
- Web Structure Mining
- Web Usage Mining

**Web Content mining** is utilized to investigate the information gathered via web crawlers. It is mainly concerned with the finding of important information or data from the unstructured; semi structured or structured content from the web documents. Unstructured documents contain images, text, audio and video. HTML documents are comprised by the semi structured documents and the structured documents include tables and lists [12].

**Web Structure mining** is utilized to taking in the information, associated with example of specific site and utilization mining are utilized to investigate information identified with a specific client programs are and also the information to assemble the types of the clients. This data can be assembled by means of Web mining. A utilizing customary information mining defined techniques are the association, classification and clustering, and then examining of the successive patterns [12].

**Web Usage Mining** is an information mining strategy that mines the data by breaking down the log records that contains the client get to designs. Web Usage Mining mines the optional information which is exhibit in log records and got from the associations of the clients with the web. Web use Mining systems are connected on the information display in web server logs, program logs, treats, client profiles, bookmarks, mouse clicks and so forth. This data is regularly assembled consequently get to web log through the Web server [13].



**Fig 1: Web Data Mining Structure**

## II. ONLINE SOCIAL NETWORKS:

Now-a-days, online social networks (OSNs) are extremely popular. There are more than a hundred websites such as Facebook, Twitter, LinkedIn and Youtube offering a range of services for users who depend on the networks for sharing their interests, photos, videos and blogs. However, the main purpose of online social networks is sharing information.[14] Table I shows the most popular social networking sites according to a ranking from ebizMBA (<http://www.ebizmba.com>).

| S.No. | Website     | Estimated Unique Monthly Visitors |
|-------|-------------|-----------------------------------|
| 1     | Facebook    | 1,500,000,000                     |
| 2     | YouTube     | 1,499,000,000                     |
| 3     | Twitter     | 400,000,000                       |
| 4     | Instagram   | 275,000,000                       |
| 5     | LinkedIn    | 250,000,000                       |
| 6     | Reddit      | 125,000,000                       |
| 7     | VK          | 120,000,000                       |
| 8     | Tumblr      | 110,000,000                       |
| 9     | Pinterest   | 105,000,000                       |
| 10    | Google Plus | 100,000,000                       |

**TABLE 1 THE MOST POPULAR SOCIAL NETWORKING SITES. (JULY 2017 EBIZMBA)**

## TYPE OF ONLINE INFORMATION

Online social networks have a wide range of data formats used for such features as sending messages, pictures and videos. The activities of the users on those websites can be used to analyze their behavior. There are two types of data in online: static and dynamic.[14]

## III. ANALYSIS OF USER BEHAVIOR

There is a large amount of data on online social networks. Reviewing that data may help to investigate user behavior in terms of how and why users interact. At present, there are many research studies on the various methods used to analyze user behavior as follows:

#### *A. Content Analysis*

Content analysis is an analysis of information in the form of unstructured texts such as sentences, paragraphs, documents and web pages to understand the human sense of expression and are implemented to meet the requirements. In language sentences, there are two related words like syntax and semantic. Syntax is a structure, grammar and order of the elements in a language statement, while semantic is the meaning of those sentences. These two words can be classified into two types of content analysis in terms of being quantified and qualified.

#### *B. Sentiment Analysis*

Sentiment Analysis, also known as opinion mining is a process to analyze the feeling the user expresses on topics posted, items “liked” and opinions on Internet and labeled to classify polarities as positive, negative or neutral which helps to determine a user’s attitudes.

#### *C. Temporal Behavior*

Temporal behavior is an analysis of fetching technology (posting, tweeting) and pull technology (feeding) activities on social network streams in the form of the number of posts over time such as the time of day, the day of the week, monthly, yearly, etc. to understand characteristics and temporal aspects.

#### *D. Latent User Attribute Analysis*

Latent user attribute analysis is a method to classify user attributes including age, gender, race, nationality, regional origin, political orientation and other features that are not be revealed from informal communication such as tweets on Twitter or posting status on Facebook. The discovery of the hidden attribute of users’ messages valuable to distinguish user and referred to user attributes relative to properties of network structure such as follower and following frequency in Twitter or communication behavior such as response frequency (percentage of reply tweet from user), retweet frequency (percentage of retweets of tweet messages) and tweet frequency (percentage of user tweet).

#### *E. Personality Analysis*

Personality analysis is an analysis of user activities on OSNs that are related to their personalities by extracting profile features to observe the relationship between the personality of the user and user’s profile. The profile features are separated into two categories. First, profile features depend on user actions such as the number of posted photos, events and the number of “liked” objects. Second, the profile features depend on the user’s action with their friends such as the number of times that users are tagged in photos and the size and density of friend networks. Five factor models (FFM) is a widespread models in the study of personality consists of openness to experience, conscientiousness, extraversion, agreeableness and neuroticism are used to described user personality and psyche (human mind, conscious and unconscious).

#### *F. Location Based Analysis*

Location based analysis is a method to investigate a relationship between friendships and user movements. The location features such as check-in data or the status of users that tagged places are used to identify the location of a user. The user interaction in social networks can be focused on various factors such as moving closer to a friend’s home, the influence of friends that affect the movement of the user, the move to an area where friends have checked in, limitations of a relationship to predict the movement of the user and using temporal and geographic data to analyze the periodicity of user movement. [14]

### **IV. BEHAVIOR ANALYSIS IN SOCIAL MEDIA**

By analyzing behaviors observed on social media, we can categorize these behaviors into individual and collective behavior. Individual behavior is exhibited by a single user, whereas collective behavior is observed when a group of users behave together. User activities on social media generate behavioral data, which is massive, expansive, and indicative of user preferences, interests, opinions, and relationships.[15]

#### **Individual Behavior Analysis**

Individual behavior can be considered one of the following:

- User-User Behavior. This type of behavior is observed between two users. For example, befriending and following in social media are examples of such behavior.
- User-Entity Behavior. This type of behavior is exhibited with respect to entities on social media (for example, user-generated content). For instance, liking a post on Facebook or posting a tweet on Twitter are examples of user-entity behavior.
- User-Community Behavior. This is the type of behavior that users exhibit with respect to communities. Joining and leaving communities are examples of this type of behavior. [15]

### **Collective Behavior Analysis**

Collective behavior analysis can be performed by either analyzing the individuals that exhibit the collective behavior independently, or by analyzing the individuals that exhibit the collective behavior as one group.[15]

## **V. MENTAL HEALTH**

Mental health is an essential component of the human health. Currently, like depression, mental health problems have been one of the leading causes for global burden of disease [1], [2]. Therefore, preventing mental health problems can be helpful to improve human well-being. Early detection is a basis of the prevention of mental health problems [3]. However, because of several reasons such as lacking mental health knowledge, and stigmatizing attitudes towards mental patients, people with mental health problems are not motivated to seek professional help [4], [5]. More importantly, traditional methods for detecting individual mental health problems (e.g. self-report techniques and clinical diagnosis) cannot identify individual mental health status in real-time, which may lead to delayed reporting and can have negative impacts on personal mental health.

More importantly, traditional methods for detecting individual mental health problems (e.g. self-report techniques and clinical diagnosis) cannot identify individual mental health status in real-time, which may lead to delayed reporting and can have negative impacts on personal mental health [6].

The development of Internet and information technology gives us an opportunity to find new method for detecting mental health problems. First, based on information technology, Internet behaviors can be collected and processed in a non-intrusive, accurate and efficient manner. Given that the relationship between Internet behaviors and psychological features (e.g. personality) has been confirmed in previous studies, which implies the possibility of detecting mental health problems through Internet behaviors analysis. Amichai-Hamburger and Ben-Artzi (2000) found that there exists relationship between Internet behaviors and personality [7]. Gosling et al. (2011) collected digital records of human behaviors on social media and proved the accuracy of predicting personality by perceiving Internet behaviors [8]. Furthermore, Kosinski and colleagues (2013) established computational models for predicting Facebook user's psychological profile and personal preference [9]. Wu et al. (2015) argued that computer-based personality judgments are more accurate than those made by humans [10].

## **VI. COMMON MACHINE LEARNING ALGORITHMS FOR BEHAVIOR ANALYSIS**

Machine learning algorithms can be divided into 3 broad categories: supervised learning, unsupervised learning, and reinforcement learning.

### ***Naïve Bayes Classifier Algorithm:***

It would be difficult and practically impossible to classify a web page, a document, an email or any other lengthy text notes manually. This is where Naïve Bayes Classifier machine learning algorithm comes to the rescue. A classifier is a function that allocates a population's element value from one of the available categories. For instance, Spam Filtering is a popular application of Naïve Bayes algorithm. Spam filter here, is a classifier that assigns a label "Spam" or "Not Spam" to all the emails.

Applications of Naïve Bayes Classifier: Sentiment Analysis- It is used at Facebook to analyse status updates expressing positive or negative emotions. Document Categorization- Google uses document classification to index documents and find relevancy scores i.e. the PageRank. Email Spam Filtering-Google Mail uses Naïve Bayes algorithm to classify your emails as Spam or Not Spam

### ***Support Vector Machine Learning Algorithm***

Support Vector Machine is a supervised machine learning algorithm for classification or regression problems where the dataset teaches SVM about the classes so that SVM can classify any new data. It works by classifying the data into different classes by finding a line (hyperplane) which separates the training data set into classes. As there are many such linear hyperplanes, SVM algorithm tries to maximize the distance between the various classes that are involved and this is referred as margin maximization. If the line that maximizes the distance between the classes is identified, the probability to generalize well to unseen data is increased.

### ***Decision Trees:***

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance-event outcomes, resource costs, and utility.

### ***Linear Regression Machine Learning Algorithm:***

Linear Regression algorithm shows the relationship between 2 variables and how the change in one variable impacts the other. The algorithm shows the impact on the dependent variable on changing the independent variable. The independent

variables are referred as explanatory variables, as they explain the factors the impact the dependent variable. Dependent variable is often referred to as the factor of interest or predictor.

#### **Decision Tree Machine Learning Algorithm:**

A decision tree is a graphical representation that makes use of branching methodology to exemplify all possible outcomes of a decision, based on certain conditions. In a decision tree, the internal node represents a test on the attribute, each branch of the tree represents the outcome of the test and the leaf node represents a particular class label i.e. the decision made after computing all of the attributes. The classification rules are represented through the path from root to the leaf node.

### **CONCLUSION**

Social Networks user behavior can be analyzed with machine learning algorithm by applying various attributes like text, Profile images, Similes, Image post, admin for the Social Network page, advertisement liked, Search keywords, and check-ins to identify the level of mental health status.

### **REFERENCES**

- [1] M. Prince, V. Patel, S. Saxena, "No health without mental health", The lancet, vol. 370, pp. 859-877, 2007.
- [2] T. Vos, AD. Flaxman, M. Naghavi, "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990C2010: a systematic analysis for the Global Burden of Disease Study 2010", The lancet, vol. 380, pp. 2163-2196, 2012
- [3] PY. Collins, V. Patel, SS. Joestl, "Grand challenges in global mental health", Nature, vol. 475, pp. 27-30, 2011.
- [4] AF. Jorm, "Mental health literacy: empowering the community to take action for better mental health", American Psychologist, vol. 67, pp. 231-243, 2012.
- [5] Changye Zhu and Baobin Li, Ang Li, Tingshao Zhu "Predicting Depression from Internet Behaviors by Time-frequency Features" IEEE/WIC/ACM International Conference on Web Intelligence, 2016
- [6] S. Clement, O. Schauman, T. Graham, "What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies", Psychological Medicine, vol. 45, pp.11-27, 2015.
- [7] YA. Hamburger, E. Ben-Artzi, "The relationship between extraversion and neuroticism and the different uses of the Internet", Computers in Human Behavior, vol. 16, pp. 441-449, 2000.
- [8] SD. Gosling, AA. Augustine, S. Vazire, "Manifestations of personality in online social networks: self-reported Facebook-related behaviors and observable profile information", Cyberpsychology, Behavior, and Social Networking, vol. 14, pp. 483-488, 2011.
- [9] M. Kosinski, D. Stillwell, T. Graepel, "Private traits and attributes are predictable from digital records of human behavior", Proceedings of the National Academy of Sciences of the United States of America, vol. 110, pp. 5802-5805, 2013.
- [10] Y. Wu, M. Kosinski, D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans", Proceedings of the National Academy of Sciences, vol. 112, pp. 1036-1040, 2015
- [11] <http://www.web-datamining.net/>
- [12] Anurag kumar, Vaishali Ahirwar, Ravi Kumar Singh, "A Study on Prediction of User Behavior Based on Web Server Log Files in Web Usage Mining", International Journal Of Engineering And Computer Science, Volume 6 Issue 2,ppno: 20233-2026,2017.
- [13] Ashwini Ladekar, Dhanashree, Raikar, Pooja Pawar, "Web Log Based Analysis of User's Browsing Behavior", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 3 Issue 11,ppno: 3895-3899 2014.
- [14] Sawita Yousukkee, "Survey of Analysis of User Behavior in Online Social Network" The 2016 Management and Innovation Technology International Conference (MITiCON-2016), IEEE, 2016
- [15] Reza Zafarani and Huan Liu, "Behavior Analysis in Social Media" IEEE Intelligent Systems, July/August 2014
- [16] P.Nithya, Dr. P.Sumathi, "A Survey on Web Usage Mining: Theory and Applications" Int.J.Computer Technology & Applications, Vol 3, 2012
- [17] <https://www.dezyre.com>