

# Knowledge Discovery From Social Media Data For Education Enhancement

Ms.Neha Dubey<sup>#1</sup>, Ms.Aakankshi<sup>#2</sup>, Ms.Neelam Borde<sup>#3</sup>

<sup>1</sup>nehadub186@gmail.com

<sup>2</sup>aakankshiar@gmail.com

<sup>3</sup>neelamb1994@gmail.com

<sup>#123</sup>Department of Computer Engineering  
JSPM's

Imperial College of Engineering & Research  
Wagholi,Pune-412207



## ABSTRACT

Social media has gained a lot of popularity amongst the students. Students share their feelings and their day to day experiences on it in a very informal and casual manner. Experiences and problems of students revealed through social media needs human interaction or human analysis. Knowledge from such uninstrumented environments can present valuable data to report student problem. But mining knowledge from such data can be a very challenging task. The huge amount of data requires automated data analysis techniques. In this paper, a work-flow is developed which combines both qualitative investigation and large-scale data mining scheme. The data posted by students are collected and analysed. It is found that certain issues like heavy study load, hectic schedule and lack of sleep are encountered by the students. Hence these issues are classified using Naive Bayes Multi-label Classifier algorithm. This classification can help in understanding the student's problem in a very efficient way.

**Keyword:** Data mining, social media, text mining.

## ARTICLE INFO

### Article History

Received: 24<sup>th</sup> September 2015

Received in revised form :

25<sup>th</sup> September 2015

Accepted : 28<sup>th</sup> September 2015

**Published online :**

**28<sup>th</sup> September 2015**

## I. INTRODUCTION

Social media:

Social media is a group of internet-based applications that build on the ideological and technological foundations of Web. It has become part and parcel of life. It is not only popular and affordable but mostly used platform where people share their experiences with the world. Social media shatters all the geographical boundaries and allows communicating with the whole world.

Nowadays, there are numerous social media sites like twitter, facebook, photo bucket etc. These sites provide people a way to express their thoughts and feelings in front of huge amount mass. Social media enables us to be connected and interact with each other anywhere and anytime – allowing us to observe human behaviour in an unprecedented scale. This provides golden opportunities to understand individuals at scale and to mine human behavioural patterns otherwise impossible.

Social media sites also provide a way to advertise and share with people hence they are now used in various fields like politics and educational systems. Social media having endless advantages also comes with few disadvantages like overuse of these sites by people, posting objectionable things or using it for harming others.

Social media comes in various categories like some sites are for sharing of data in purely text format while some are for sharing pictures and videos.

More people are becoming interested in and relying on the social media for information, breaking news and other diverse subject matters. They find out what other people's views are about certain product/service, film, school etc.

Organizations are now conscious of the significance of the opinion of consumers which they post on social sites to the development of their products or services. Moreover, personalities make efforts to protect their image and are being conscious of how they are perceived on these sites.

With the rise of social media, the web has become very vibrant and lively. Hence more and more people are actively participating in these sites. Social media has become an ever increasing field in today's world.

## II. SOCIAL MEDIA MINING

Social media mining is the process of representing, analyzing, and extracting meaningful patterns from data in social media, resulting from social interactions of people. It is a field which encompasses techniques from computer science, data mining, social network analysis, network science, sociology and mathematics. Mining is also called as "Knowledge Discovery" i.e to discover or gain knowledge from raw data. In social media mining we mine the knowledge from various posts generated by people. These are then analysed and used for producing the results.

In this process information is collected, analysed, classified according to need and then final results are produced. Hence the mining helps in understanding individuals better, which can be used to design better computing systems tailored to individuals' needs that will serve them and society better.

## III. CHALLENGES IN SOCIAL MEDIA MINING

Social media mining comes with various challenges which poses certain difficulty in mining process.

The first challenge is its enormous size.

Second is the user generated data which comprises of noise and un-structured data.

Third is to classify the data which is of use and dump the remaining data.

## IV. RELATED WORK

### 1. Text mining

Text mining is the process of formatting the given text (parsing, with addition of some linguistic features, addition of some subsequent data), forming the structured database and evaluation and interpretation of output. The good quality of text mining refers to the combination of relevance, interestingness. The real time application is to scan the set of documents which is natural language and document set for the predictive classification.

How to do text mining?

1. Information retrieval: Collecting or identification of set of text documents, taken from the social media which is posted by the user.

2. Natural language processing: In this we are recognize the part of speech which is tagged, syntactic parsing.

3. Name entity recognition: In this we are identifying the named text like:-

the name of the peoples, place, organization, gestures and the postures symbols, etc.

4. Pattern identified entities: Some characteristics like the telephone no., e-mail id, and address can be identified by using the regular expression or by matching of data by using concern data sets.

5. Conference: Identifications of grammatical noun phrases using the posted dataset.

6. Sentiment analysis: In this process, we are identifying the user's mood, emotion, opinion, sentiments; by using this interpreter helps to identifying the concepts and the opinion holder and opinion objects.

7. Qualitative text analysis: The semantic or the grammatical relationship between the words is extracted from the posts made by the users on the social media to find the meaningful text.

## V. PROPOSED WORK

This paper targets only the student's generated data on the sites. Students regularly post their experiences, thoughts, ideas on these sites. This raw and unformatted data is taken, classified into various categories and according to those categories results are produced. Sentiment analysis can be referred to as discovery and recognition of positive or negative expression of opinion by students. It deals with the feelings of users. Opinions expressed on social media by users are often convincing and these indicators can be used to form the basis of choices and decisions made by people. Consequentially it has become necessary to analyse sentiment expressed by users using data mining techniques in order to generate a meaningful framework that can be used as decision support tool. This can be made possible by employing algorithms and techniques to ascertain sentiment.

The motive behind opinion investigation and sentiment analysis of students is basically to recognize potential drift in the educational system as it concerns the attitudes, observations, sentiment and the expectations of students.

It is important to translate sentiment and experiences expressed to useful knowledge by way of mining and analysis.

### Mining Twitter Data

Twitter is one of the most popular site. This social media site is public i.e the data or the content of Twitter is very concise. It allows only 140 characters to be read or tweet by the users. Twitter provides an application interface which can be used to stream the data. Various analysis methods can be used to mine the data. Analysis methods include qualitative content analysis, quantitative analysis, etc. But mining social media data is a very difficult task. Many methodological difficulties arise during analysing the huge amount of dataset i.e. textual dataset.

## Cluster

Clustering and classification are both the fundamentals of data mining. Classification are also known as the supervised machine learning and clustering also known as the unsupervised machine learning technique.

Clustering is the method of finding the similar data set and collect into the same group which is more similar to each other. The main base of clustering is to grouping the data based on the given data description. based on it is classified into following types:

### 1. Hierarchical based clustering:

It is also called as the connectivity based clustering. In this the no. Of objects are collected to form the clusters based on the distance between them.

### 2. Distribution based clustering:

The statistics are based on the distributed clustering. In this the objects are belongs to same distribution.

### 3. Density based clustering:

In this the clusters are defines based on the remainder of datasets.

### 4. Text pre-processing:

When the user post the comments on the social media he no. Of words are repeated. Meaningless words, gestures and postures that contains the noise to the text. In this it is used before classification.

Examples:

- a) Strict partitioning clustering: in this each cluster must be within one cluster.
- b) Strict clustering with outlier: which not belongs to cluster is said to be the outlier.
- c) Overlapping clustering: the objects are belongs to one or more clusters.
- d) Hierarchical clustering: the objects are belongs to the child as well as the parent cluster.

## VI. CONCLUSION

The conclusion is based on mining social media data which helps in recognizing the student's problems. Mining social media data is helpful to researchers in learning analytics, educational data removal, and learning skill. It gives a way to examining social medium statistics that conquer the main restrictions of both physical qualitative analysis and huge scale computational study of user produced textual content.

## VII. REFERENCES

- [1] Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan. "Mining Social Media Data for Understanding Students' Learning Experiences" IEEE Transactions on Learning Technologies, ID, DOI 10.1109/TLT.2013.2296520.
- [2] "Using the Twitter Search API | Twitter Developers <https://dev.twitter.com/docs/using-search> , 2013.
- [3] Ferguson, "The State of Learning Analytics in 2012: A Review and Future Challenges," Technical Report KMI-2012-01, Knowl-edge Media Inst. 2012.
- [4] IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 45, NO. 1, FEBRUARY 2015 Integrating Human Behavior Modeling and Data Mining Techniques to Predict Human Errors in Numerical Typing Cheng-Jhe Lin, Changxu Wu, Member, IEEE and Wanpracha A. Chaovalitwongse, Senior Member, IEEE.