ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.6

Students Feedback for Mining Their Opinions Using Supervised Learning Algorithm

Sarpreet Kaur[1], Rasleen Deol[2]
Research Scholar of M.Tech[1], Assistant Professor M. Tech[2]
Department of CSE
Global Institute of Management & Emergent Technologies, Amritsar, Punjab, India

Abstract:
With the advancement in internet technology, people became keener to express and share their opinions on web regarding day-to-day activities and global issues too. Growth of social networking has also contributed eminently to these activities, thereby providing us a platform to share views across the world. Industries are working rigorously on opinion mining (also known as sentiment analysis), to extract and analyze public mood and views.

Keywords: Data Mining, Clustering, Classification, SVM, Neural network, GA, Opinion Mining

1. INTRODUCTION

1.1 DATA MINING:-

There are large amount of databases available with the increase in the Information Technology. Data is present in huge amount which comprises of various fields. For the purpose of future decision makings, the data needs to be stored and manipulated. For this, various databases have been developed and researches have been carried out for their managements. The process of extraction of useful information and patterns from large amount of stored data is known as data mining. Data Mining is the practice of Examining large pre-Existing Databases in order to generate new information. It is process of processing large volumes of data (usually data store in a database).Used for searching for patterns and relationship with data. The process of extraction of useful information and patterns from large amount of stored data is known as data mining. . Various types of data is analyzed with the help of certain data mining tools. There are certain applications such as the customer retention, education system, production control, healthcare, manufacturing engineering, decision making, and so on related to this technique [1].

DATA MINING PARAMETERS

- Association
- Sequence of path Analysis
- Classification
- Clustering
- Forecasting
- ASSOCIATION: - It defines the relationship between the events that connected with each other.
- SEQUENCE OR PATH ANALYSIS:-Analysis patterns where one event leads to the other event.
- CLASSIFICATION: - It is used to predict the target of class in each case of data.
- CLUSTERING:-Finding and documenting the groups of facts.
- FORECASTING (predictive Analysis):-Discovering the patterns in data that can lead to reasonable prediction about the future.

COMPONENTS OF DATA MINING

- Database and Data warehouse
- Database and Data warehouse Server
- Knowledge Base(KDD)
- Data mining Engine
- Pattern Evaluation Method
- Graphical User Interface

The relevant data is fetched from the database, data warehouse, information repository and the servers. The fetching of data is based on the user's data mining request. The search is conducted with the help of knowledge gathered from parts and further calculations are done on the basis of interesting patterns. The data mining modules interact with the pattern evaluation which can help in focusing the depiction of interesting patterns and graphical user interface. The users as well as data mining systems allows the user interface which communicates between the users and the data mining systems for providing a user interface within the system. Following are the categories for the data mining tasks. There are special kinds of functionalities within the data mining. These are utilized for specifying certain kind of patterns which can help in identifying various tasks of the data mining process. There are two categories in which the data mining tasks can be classified. They are the descriptive as well as the predictive. The tasks which characterize the general purpose properties of the data within the database are known as the descriptive mining tasks. The tasks which perform the inference on the present data for making predictions are known as the predictive mining tasks. The various data mining functionalities are given below [2]: 1. Characterization and Discrimination: Data characterization is a summarization of data of the class under study and data discrimination is a comparison of the target class with one or a set of comparative classes.

Class/concept descriptions are derived using these two functionalities.

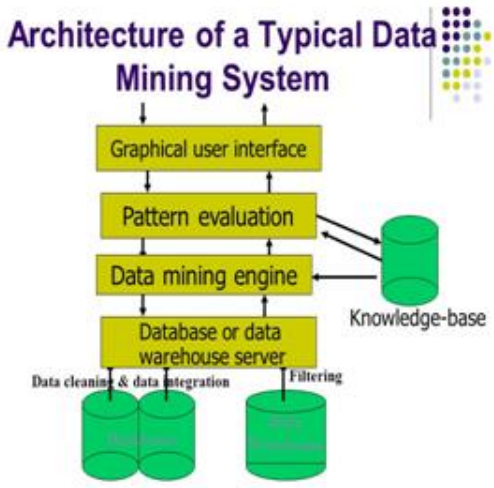
2. Frequent patterns are the patterns that occur frequently in data. Association rule mining is the process of finding interesting interrelations, incessant patterns or federation among sets of items in the negotiation, connected databases or other statistics repositories [3].

3. Classification and Regression:-Classification is a data mining (machine learning) technique used to predict group membership for data instances. Regression analysis is a methodology that is mainly used for numeric prediction.

4.Cluster Analysis:-Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters).

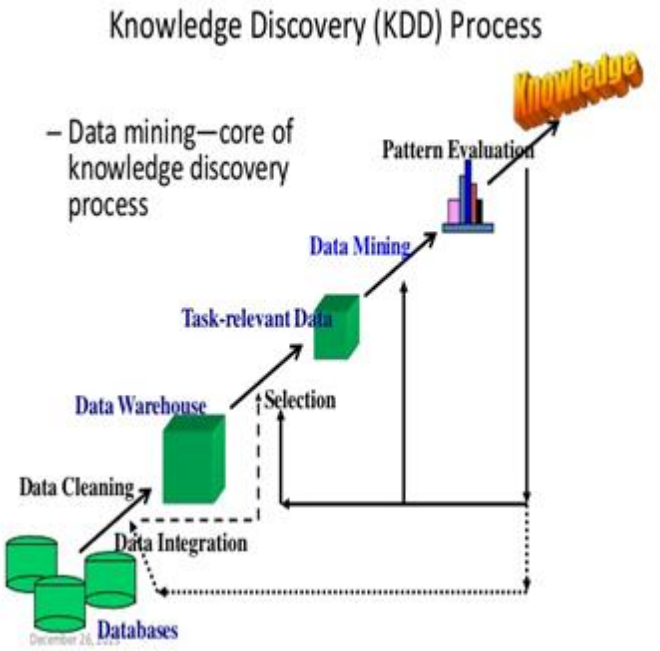
5. Outlier Analysis:-Some objects in a data set do not comply with the general behavior or model of the data. These data objects are outliers and analysis of outlier data is known as outlier analysis [4].

The large amount of data which needs certain powerful data analysis tools are thus put for the here which is also known as the data rich but information poor condition. There is an increase in the growth of data, its gathering as well as storing it in huge databases. It is no more in the hands of humans to do it easily or without the help of analysis tools. There are certain data archives created here which can be visited when the data is required. The insightful, interesting and novel patterns of data are discovered from large-scale data sets using the data mining. The knowledge discovery in databases process is a very important step in data mining. The data mining and KDD are often termed as synonyms. There are databases, data warehouses, internet, information repositories involved within the data sources. The end goal here is to extract information from the data set and transform it into an understandable structure which can be helpful in further use which can help provide the data mining process to evolve. Any kind of data repository can be presented in the through this technique. There are various types of algorithms and techniques which are utilized for various typed of data. There are different databases in which the data mining can be used. The object-relational databases, relational database, data warehouses and multimedia data bases and so on which can be involved here [5].



1.2 KNOWLEDGE DISCOVERY FROM DATA (KDD) PROCESS

There are certain steps which are to be followed for the KDD process of extracting knowledge from the data present and further emphasizing it on high-level applications for achieving particular data mining methods. There are various fields in which the data mining is very essential. The most important role for presenting the frequently utilized object sets is utilized with the help of determining the correlations between the various types of fields which are present



in the data base. The association rule is another important factor which can be utilized here for the identification of frequently used object sets in KDD process. The retail stores utilize this association rule concept for the purpose of managing the marketing, advertising and handling problems which are present in this field [6].

EXAMPLE:-

Data Mining in an Organization As it is already known that there is a huge growth in the information technology on daily basis and the databases are being created by organizations for managing the data. There are organizations which belong to the telecommunications, banking, marketing, transportation, and so on. It is vital to explore the complete databases efficiently for defining the valuable data. For identifying the information in huge databases, the data mining method is used. The data which shows well defined relationship between the variables is created using the KDD process. The useful patterns are discovered from the database with the help of automatic discovery process which is mainly the KDD method. It is also known as discovering required data from huge data bases. The association rule mining is one of the important rules which are developed for data mining. For the purpose of decision making, these rules are proposed on applications such as market based, banking, and so on [7].

1.3 CLASSIFICATION IN DATA MINING

Classification is the process of finding a model or functions that describe and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived is based on the analysis of set of training data.

CLASSIFICATION METHODS FOR DATA MINING ARE: -

The commonly used methods for data mining classification tasks can be classified into the following groups.

1. Decision tree induction methods.
2. Rule-based methods
3. Genetic Algorithm
4. Neural networks
5. Bayesian network
6. Support vector machines.

1.4CLUSTERING IN DATA MINING

Clustering is a process of partitioning a set of data (objects) into a set of meaningful sub-classes, called clusters. Clustering is the process of making a group of abstract objects into classes of similar objects.

Clustering for Data Mining

1. Scalability
2. Ability to deal with different kinds of attributes
3. Discovery of clusters with attribute shape
4. High dimensionality
5. Ability to deal with noisy data
6. Interpretability

Clustering Methods: -

Following are some methods of clustering:-

1. Partitioning Method
2. Hierarchical Method
3. Density-based Method
4. Grid-Based Method
5. Model-Based Method
6. Constraint-based Method

2.1 OPINION MINING

It is the type of natural language processing for tracking the mood of the public about a particular product. Opinion mining is also called Sentiment Analysis because it is the Process of determining the emotional tone behind a series of words. Opinion mining can be characterized as a sub-discipline of computational linguistics that spotlights on extracting people's opinion from the web. The current expansion of the web encourages clients to contribute and communicate by means of websites, recordings, and interpersonal interaction sites, and so on. Every one of these stages gives a tremendous amount of valuable information that we are interested to break down. Given a bit of text, opinion-mining systems examine [8]:

• Which part is opinion expressing?
• Who wrote the opinion;
• What is being commented.

Sentiment analysis, then again, is a piece of text in order to determine the writer's attitude like (weakly positive, mildly positive, strongly positive, and so forth.) of a bit of text – as such:

• **What is the opinion of the writer?**

It is seen recently that there is an increase of data availability, the alleged data deluge, controlled by an increased amount of electronic action performed, (for example, utilizing informal communities online) and the progressive pervasive reach of IT in all devices. The first of these trends is the purported "open data" movement, characterized by the way that the whole way across Europe and the US, governments are progressively publishing their data repositories for other people to access and utilize it [9]. Another pattern concerns the incomprehensible amount of data is made accessible by citizens through "participatory sensing": standard play a proactive role in publishing comments and grumbling on line, and progressively utilize innovation to record additional information, for example, photos or audio recordings, commonly through cell phones. Moreover sensors are getting to be distinctly embedded in regular non-ICT, for example, autos or the urban landscape, so that usable data are automatically gathered at a fast pace. At last these data, and in addition government data, are gathered by citizens as well as now made accessible to citizens so that new information gets to be distinctly accessible. At analytical level there are a few technological innovations that help comprehending the vast amount of data availability. The limits of human attention, combined with the current straightforward interfaces accessible for perusing exchange and comments, regularly prompts low levels of engagement and flaming wars, driving to polarization of arguments and enhanced risks of conflicts. To address this challenge, opinion mining contrasts from unadulterated data and text mining insofar it manages subjective statement. In this sense, it is a particular development of a discipline managing unstructured information extraction (IE) that was already fundamentally working with objective data, for example, natural disasters or bibliographic information. The explosion of client created content widens the application scope of general opinion mining tools, which are turning out to be more pervasive and accessible to the majority of citizens.

1.3.1 Opinion Mining Applications

1. Argument mapping software helps arranging legitimately these policy statements, by explicitating the logical links between them. Under the research field tools like Compendium, Debate graph have been developed to give a logical structure to a number of policy statement and to link arguments with the evidence to back it up.

II. LITERATURE REVIEW

Author and tittle	Year	Description	Outcome
Shoiab Ahmed, et.al," A Novel Approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using Web Data"	2015	The web data is collected utilizing web crawler applied with different preprocessing techniques which include removal of prevent words from online reviews, then stemming is performed utilizing Porter Stemmer algorithm, and after that reviews are tagged utilizing Stanford POS tagger [12]. The proposed approach is experimented on movie and product web domains and obtained higher success rate in terms of accuracy measured by various tools like Kappa statistics with an accuracy of 77.7% and has lower error rates.	The results demonstrate that the proposed novel approach has higher efficacy and it can be successfully used in Opinion Mining for the task of decision making by any web user.
Malhar Anjaria, et.al,'' Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning"	2014	In this paper, the novel approach of exploiting the user influence factor is introduced in order to predict the outcome of an election result. A hybrid approach of extracting opinion is proposed utilizing direct and indirect features of Twitter data based on Support Vector Machines (SVM), Naive Bayes, Maximum Entropy and Artificial Neural Networks based supervised classifiers [13].	Experimental results demonstrate that Support Vector Machines outperform every single other classifier with maximum successful prediction accuracy of 88% in case of US Presidential Elections held in November 2012 and maximum prediction accuracy of 58% in case of Karnataka State Assembly Elections held in May 2013.
LI Bing, et.al," A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data"	2014	Recently, some efforts have been made to mine social media for the analysis of public behavior. In the real life social media environment, the structure of the data is usually not clear and it does not directly generate enough information to completely represent any selected target [14]. However, the vast majority of these works were unable to accurately extract clear indications of general public opinion from the ambiguous social media data. They additionally lacked the capacity to summarize multi-characteristics from the scattered mass of social data and use it to compile useful models, likewise lacked any efficient mechanism for managing the Vast data.	This paper proposes a novel matrix-based algorithm to determine the defined multilayered Twitter data.
Dhanalakshmi V, et.al," Opinion mining from student feedback data using supervised learning algorithms"	2016	This paper explores opinion mining utilizing supervised learning algorithms to discover the polarity of the student feedback based on pre-defined features of teaching and learning. The study conducted involves the utilization of a combination of machine learning and natural language processing techniques on student feedback data gathered from module evaluation survey results of Middle East College, Oman [15]. In addition to giving a step by step explanation of the process of implementation of opinion mining from student comments utilizing the open source data analytics tool Rapid Miner.	The results are compared to locate the better performance with respect to various evaluation criteria for the different algorithms.
Gaurav Dubey, et.al," User Reviews Data Analysis using Opinion Mining on Web"	2015	The web world is thriving with e-commerce these days and the need for online reviews has become pivotal. The outcomes guide the consumer and help them in settling on decisions regarding various available products which otherwise would bemuse them. In any case, one issue hampering this decision making problem is to sift through the huge jumbled piles of reviews available on the boundless web [16].	The automatic summarization and classification is different for different domains and varies with the testing situations..
Ram Chatterjee, et.al," Tactics of Twitter Data Extraction for	2015	In our paper, the various methods available through which the twitter data can be extracted are analyzed. An overall picture is provided of how each method is different from another for extracting tweets. Windows 10 has been selected as a keyword	In this research we have defined the method to extract twitter data utilizing various sentiment tools. Each sentiment

Opinion Mining"		to compare the opinions as extracted by different sentiment tools. Twitter, a prevalent micro blogging helps us to extract a great many tweets over various distinct domains [17].The choice of sentimental tool to be used is entirely depend on user and his/her need. For example, if large corpus is needed then we will use "Topsy" or 'Trackur', or in the event that we need tweets into an excel sheet then 'Tweet Archivist' is used. In future, we will be utilizing the extracted tweets utilizing sentiment tools to construct corpus. The collected corpus will be used to perform sentiment analysis utilizing classifiers. Our classifier will be able to determine the tweets as positive, negative or neutral. This helps the end user to frame a decisive opinion on the query search.	tool is different from other. In future, we will be utilizing the extracted tweets utilizing sentiment tools to construct corpus. The collected corpus will be used to perform sentiment analysis utilizing classifiers.
Pooja Kherwa, et.al," An approach towards comprehensive sentimental data analysis and opinion mining"	2014	This paper is the outcome of our research in gathering opinion and review data and processing the data utilizing the rules of natural language and grammar to discover what exactly was being talked about in the user's review and the sentiments that people are expressing. Our approach diligently scans every review (categorized by aspects) alongside various graphical visualizations [18]. The paper aims to provide summarized positive and negative features. It is basically an approach to make use of the boundless measure of developing user domination on the Internet in reviewing critical business decisions, policies, services etc by dissecting and picturing the data in a user-friendly manner, to-the-point.	A novel application of this approach is helping out product manufacturers or the government in gaging response. The paper aims to improve our system as discussed in the prior section, and furthermore preparing many pilot tests to further enhance the summarization results of the system to develop in more detail.
Lokmanyathilak Govindan et.al," A Framework for Fast-Feedback Opinion Mining on Twitter Data Streams"	2015	This paper focuses on the computational infrastructure for fast-feedback opinion mining. This calls for a versatile platform to handle all the possible problems arisen from mining data streams of a social networking site. Specifically, the trouble of getting customer feedbacks faced by companies that produce free software is considered [19].	These data streams are filtered and analyzed and fast feedback is obtained through opinion mining. The framework is based upon Apache Hadoop to deal with huge volume of data streamed from Twitter. The experiments have indicated 84% accuracy in the sentimental analysis.
Saurabh S.Wani, et.al," Analysis of Data Retrieval and Opinion Mining System"	2015	This paper presents analysis of text based data retrieval system and opinion mining on social sites. Collected data from various sources like nearby machine, email accounts, social networking accounts of respective user. Multiple users can use this system by giving log in credentials [20]. This paper explains significance of Inverse Document Frequency and Term Frequency in Lucene scoring formula. The vast majority of the people express their correct reviews on social networking websites than whatever others. This paper discusses an approach to classify each tweet from Twitter into different categories.	This paper additionally presents techniques to improve performance of Lucene by modifying certain parameters of document scoring formula. Lucene performance additionally can be improved by modifying algorithm for incremental indexing.

III. CONCLUSION

The data mining is the approach which is applied to extract useful information from the raw data. The opinion mining is the technique of data mining which is used to analyze opinions of users for further analysis. In this paper, various opinion mining approaches have been reviewed and discussed. In future, the efficient opinion mining technique will be proposed and results will be analyzed in terms of various parameters.

IV. REFERENCES

[1]. Klenner, M., Fahrni, A., & Petrakis, S.," PolArt: A robust tool for sentiment analysis", 2009, Computational Linguistics, pp. 235–238

[2]. Kozinets, R.V.," E-tribalized marketing: The strategic implications of virtual communities of consumption", 1999, European Management Journal, 17(3), 252-264

[3]. Miao, Q., Li, Q., & Dai, R.," AMAZING: A sentiment mining and retrieval system", 2009, Expert Systems with Applications, 36, 3, Part 2, 7192-7198

[4]. Miao, Q., Li, Q., & Zeng, D.," Fine-grained opinion mining by integrating multiple review sources", 2010, Journal of the American Society for Information Science, 61(11), 2288-2299

[5]. Mudambi, S.M., & Schuff, D," What makes a helpful online review? A study of customer reviews on Amazon.Com", 2010, MIS quarterly, 34(1), 185-200

[6]. Prabowo, R., & Thelwall, M.," Sentiment analysis: A combined approach", 2009, Journal of Informetrics, 3(2), 143-157

[7]. Senecal, S., & Nantel, J," The influence of online product recommendations on consumers' online choices", 2004, Journal of Retailing, 80(2), 159-169

[8]. Sharma, A., & Dey, S.," A comparative study of feature selection and machine learning techniques for sentiment analysis", 2012, ACM

[9]. Sher, P.J., & Lee, S.H.," Consumer skepticism and online reviews: An elaboration likelihood model perspective", 2009, Social Behavior and Personality: An international journal, 37(1), 137-143

[10]. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M.," Lexicon-based methods for sentiment analysis", 2011, Computational Linguistics, 37(2), 267-307

[11]. Tan, S., & Zhang, J.," An empirical study of sentiment analysis for chinese documents", 2008, Expert Systems with Applications, 34(4), 2622-2629

[12]. Shoiab Ahmed, Shoiab Ahmed," A Novel Approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using Web Data", 2015, IEEE

[13]. Malhar Anjaria, Ram Mahana Reddy Guddeti," Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning", 2014, IEEE

[14]. LI Bing, Keith C.C. Chan," A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data", 2014, IEEE

[15]. Dhanalakshmi V., Dhivya Bino, Saravanan A. M.," Opinion mining from student feedback data using supervised learning algorithms", 2016, IEEE

[16]. Gaurav Dubey, Ajay Rana, Naveen Kumar Shukla," User Reviews Data Analysis using Opinion Mining on Web", 2015, IEEE

[17]. Ram Chatterjee, Monika Goyal," Tactics of Twitter Data Extraction for Opinion Mining", 2015, IEEE

[18]. Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Prashast Kumar Singh," An approach towards comprehensive sentimental data analysis and opinion mining", 2014, IEEE

[19]. Lokmanyathilak Govindan Sankar Selvan, Teng-Sheng Moh," A Framework for Fast-Feedback Opinion Mining on Twitter Data Streams", 2015, IEEE

[20]. Saurabh S.Wani, Prof. Y. N. Patil," Analysis of Data Retrieval and Opinion Mining System", 2015, IEEE

[21]. Khan, Khairullah, "Mining opinion components from unstructured reviews: A review," Journal of King Saud University – Computer and Information Sciences, Vol. 26, 2014.

[22]. Dirk T. Tempelaar, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context", Journal of Computers in Human Behavior, Vol 47, pp: 157–167, 2015.

[23] .Kumar Ravi, Vadlamani Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications", Published in Knowledge-Based Systems Vol. 89 , 14–46, 2015.

[24] .S.Vijayarani, "Preprocessing Techniques for Text Mining - An Overview", International Journal of Computer Science & Communication Networks, Vol 5(1),7-16.

[25]. Trisha Patel, "Sentiment Analysis of Parents Feedback for Educational Institutes", International Journal of Innovative and Emerging Research in Engineering, Volume 2, Issue 3, 2015.