WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

"Application of Data Mining Techniques for better Decision Making in an Organization for Business Competitiveness"

Vishal Kumar

Masters of Business Administration, Amity University, Uttar Pradesh, India.

ABSTRACT

Data Mining is the technology that helps in finding some hidden useful previously unknown information from the large amount of data. This data can be structured (data marts, data warehouse) or unstructured data (web documents, text data) in nature. Web is a huge source of data; most of the structured data can be mined to extract useful information for enhancing business opportunities using data mining techniques like classification, clustering and association rule mining. In order to mine unstructured data different text mining techniques can be applied. This research work mainly focuses on mining user reviews and opinion regarding an item or product. Opinion mining and sentiment analysis is one of the upcoming concepts in field of data mining. Sentiment analysis or Opinion Mining is a deterministic technique for classifying and evaluating other people's opinions. Online item surveys have risen as a capable medium to produce electronic word of mouth. Several buyers post their purchasing experience about items and administrations in online reviews that might be utilized by potential shoppers to help them in settling on item decisions and buy choices. Consumer reviews communicated in online product reviews are vital for potential customers to settle on very much educated buy choices and for item makers to get a few bits of knowledge about their items' qualities and shortcomings. This research work contends that online product reviews are a rich source of promoting intelligence that can be extricated as consumer views by applying text processing and investigation strategies. Specifically, this research work proposes the utilization of sentiment analysis for extricating showcasing knowledge from online mobile phone reviews. The test results on mobile phone reviews demonstrate that sentiment analysis can be a compelling approach to infer showcasing knowledge and benchmarking data from online reviews. This research work focuses on in- depth study of the topic and discusses all concepts and terminologies of opinion mining. This research work also discusses the methods and techniques used for gathering reviews extracting the phrases based on the subjectivity and thereafter calculating the semantic orientation of the collected reviews. This research work proposed the methods that have mined the mobile phone reviews at different granularity, like at document level, and sentence level. This research has also performed aspect or feature level opinion mining and sentiment analysis on mobile phone reviews. This research work had made use of support vector machine learning technique in order to compute the polarity of a mobile phone review (positive or negative). This research work has made use of rule based approach along with NLP (Natural language processing) to generate summary of the reviews. This research work have also successfully identified the key features words that affect the overall orientation of product review.In this research work we have also proposed materialized view approach to find the frequent query sets posed by different users. To address this issue and to reduce the response time of frequently posed queries on data warehouse, materialized views are constructed. The resulting queries would help in providing response to future posed queries in reduced time. Faster processing of queries will help the managers to make timely and better decision in an organization.

INTRODUCTION

Motivation and scope: As the market is becoming competitive, organizations are looking for finding the new ways to increase their business. They are more focusing on increasing their number of customers and also finding ways to retain the old customers. With the introduction of web 2.0 it has become easier even for a non-technical person to connect with internet. . Now a day's people builds their perception and make decisions by analyzing the facts and reviews of other people either manually or computationally. Since everything is online now a day's , hence internet has become an integrated part of human lives and is thus used for exchanging all aspects of human life viz. sentiments, emotions, affection, support, opinions, trade, business etc. With the onset of social media there has been numerous platform such as blogs, discussion forums, reviews and social networks where an individual can post his or her reviews, feedbacks and list their likes and dislikes for a product's attributes or features or comparison of different products (same or different feature). As the cost of accessing internet is reducing, the easiness to access the web has been increased significantly. People are now a days more comfortable and have become more interactive on web. Things have become so easy that even a non-technical person can use internet medium to express his or views. People are posting their thoughts or views more freely on various topics. They are writing post purchase experience on various sites in the form of reviews. These reviews are gathered and are analyzed to evaluate the overall orientation of the collected reviews. With the increasing popularity of the Internet and with the development of new and easy tools for

content writing it has become easy to write blogs and reviews. This easiness to access the web has given a confidence even for a non-technical person. Besides this there are more user friendly tools available online which help user to interact with people who are participating actively and sharing their views and ideas online. The web usage has been increased drastically in today's time. Huge amount of data is available on web in both structured and unstructured form which can be explored and analyzed to extract useful information. People find it very easy to write their opinion on various social media sites, with this easiness the participation of user has been increased drastically. Most of the internet user now a days post their opinion and view related to a product. They share their experience regarding that product. Organizations also involved in finding the Customer interest in their product, the freedom of expression of the consumer opinion has given a new way to understand the true feedback of product. These review for opinions are not only beneficial for Customer but they are beneficial for organization, so it is very important to know true and unbiased feedback of a product. Internet has become one such medium where consumer of that product can write freely about various positive and negative feature of that product. As the data is growing day by day there is need for developing new technology new methods to analyze this data. Decision maker, manager or various data analyst have been successfully analyzing structured data by using tools available like SPSS, which can give good data analysis on this data. On the other hand unstructured data which is available in large volume on web requires different approach in order to explore data for benefit of organization. Unlike structured data the mining of unstructured data especially textual information needs involvement of NLP and text mining techniques. Web data mining is one of the upcoming fields to mine enormous information available on web in various forms in this research work we have focused on techniques which are used to mine unstructured data i.e mostly text data in the form of consumer reviews which is available in the form of product reviews. Companies are now a days exploring this text data mostly online product reviews to gain insight for their product. Every organization is competing with other to remain ahead in this market this required a detailed analysis of their product one of the best way to know about the product is to explore the end user review on that product. Thousands of product review are available online in which consumer has shared their experience and feedback about the product .In order to acquire the maximum number of customer it is necessary to provide them quality product with best features and low cost. this possible only when companies are able to understand consumer's likes or dislikes about their product. Organizations can improve on these features which are most usable by the customer. Thus by analyzing consumer attitude or sentiment about the product companies can improve the quality of their product.The technologies for generating and collecting data have been advancing rapidly. Large volumes of business transactions are recorded in the enterprise-scale data warehouse every day. The enormous increase in unstructured data in the form of online product reviews gives us motivation to extract the useful hidden information for better insight about product. This will further gives knowledge to enhance the product quality since we can focus on area of improvement. By understanding the customer needs better we can enhance the business opportunities by making timely and better decisions.

Data Mining Techniques: There are number of data mining techniques which can be classified broadly into supervised learning and unsupervised learning. In supervised learning techniques like decision tree, neural network or machine learning techniques, we know that what is outcome class. Data can be divided into training and testing and a model can be constructed for prediction. These techniques extract the hidden information which is not known previously and can be very useful. source. This data which is collected from various online or offline sources can be structured or un structured in nature. The example of structure data is data stored in data marts or data warehouse, while the unstructured data may be like a textual or image data available on web in different forms. As of late, information mining innovation has pulled in a lot of consideration in the business world, both in the private segment and in government associations, because of the immense measure of information that has been changed over into helpful data. The way toward extricating standards and examples from concealed information is called information mining. The tenets and examples are created are translated as diagrams and reporting instrument. In the learning presentation, the perception of mined information is connected. Figure 1.1 diagram showing overall process of data mining, it shows that data is collected from various sources may have different format of data. Data pre-processing is one of the very important task in data mining process. ETL (extraction transform loading) is the sequence of data pre-processing steps, there are several data pre-processing techniques like, data cleaning, data reduction and data transformation.

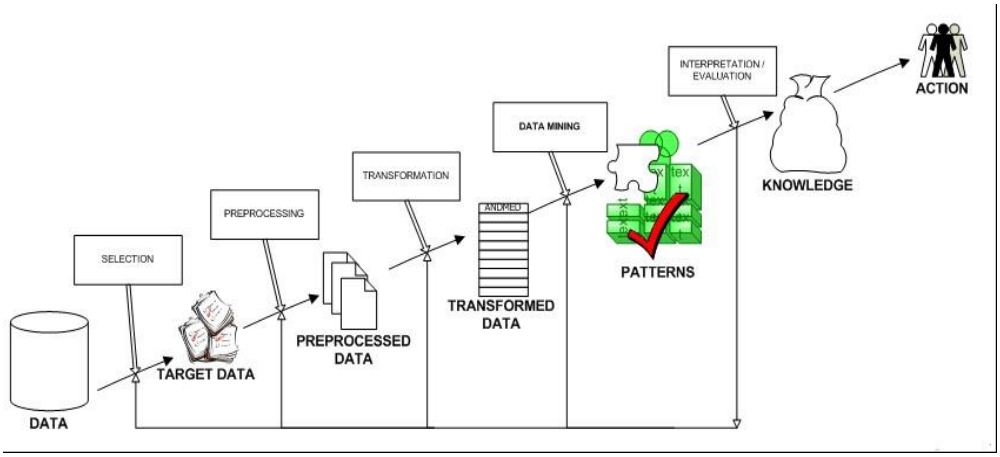

Figure1.1: Knowledge discovery process using DM techniques, source: www.kMining.com

Data cleaning is done to handle missing values in data this process also removes noisy data by various data cleaning techniques like binning and discretization. Data reductions techniques are used in order to reduce the size of data, dimensionality reduction attribute subset selection, numerosity reduction are some popular data reduction techniques. Data transformation is done to convert the source data from one format to another format. Data mining techniques can be used to extract information from both structure and unstructured type of data. Text mining and web mining are also the part of data mining; text mining mainly focuses on the various methods for mining the text data. This includes the involvement of NLP (natural language processing). Both supervised and un-supervised learning methods can be used to mine the text data. Figure 1.2 shows that how data mining as process is divided in various domains. Data required or traditional data mining is different from data required for text mining. Large volume of documents and text in the form of forums, blogs and reviews are used for extracting useful information hidden in these documents.

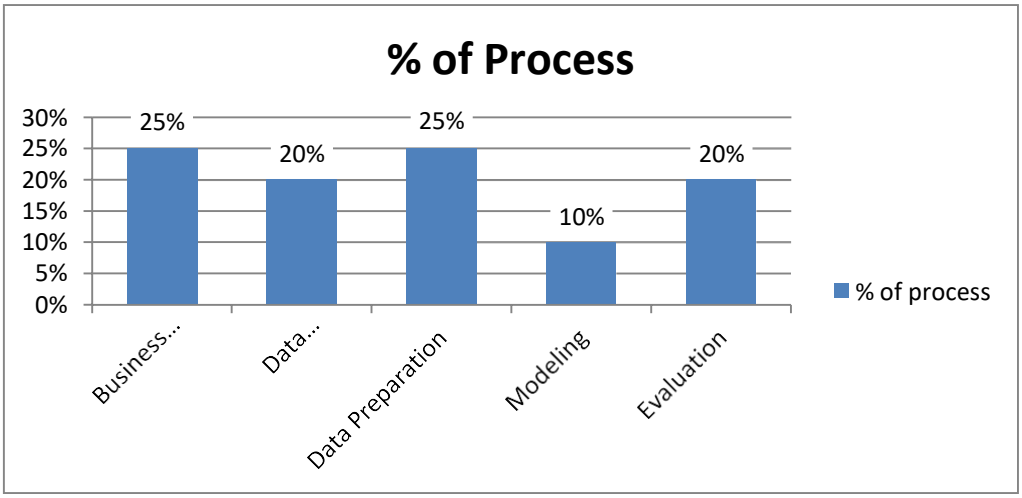

Figure 1.2: Data mining as process

Supervised learning techniques: Supervised learning is very useful for real- world applications. These techniques are used in several domains which includes, healthcare, business, research and development. Supervised learning is also known as predictive mining or classification or inductive learning in machine learning. These techniques are useful in deriving business intelligence from past data. In supervised learning techniques the outcome class label is known to us. For example if a company wants to know the potential customers to whom loan can be given, they can explore their previous data. The main aim of a supervised learning technique is to develop a classifier or model from a set of data. These techniques can used various types of data most of the data is structured in nature, but unstructured data like text data or image data can also be used in supervised learning techniques.

Step 1: Data is collected from various sources.

Step 2: Split the data into test data and training data.

Step 3: Develop a classifier using classification technique.

Step 4: Repeat the execution on training data for best classifier accuracy.

Step 5: Apply the test data on classifier to generate unseen results.

Figure 1.3 shows the various steps to build the classifier.

The training data collected can be gathered from various sources. Data pre-processing is done before splitting the data into training and test data. The basic learning process is shown in figure 1.4.

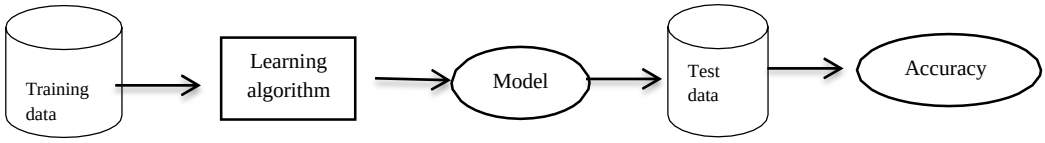

Figure: 1.4: The basic learning process training and testing

There are various data pre-processing steps like, data cleaning, data integration, data transformation etc. In this technique we use past data to develop a classifier which can further predict useful information from new test data Some of the supervised learning techniques are listed below:

(i) Decision Tree Induction

(ii) Neural Network and Back Propagation

(iii) Support vector machine

(iv) Classification based on association rules

(v) Naïve Bayesian Classification

Unsupervised learning techniques for Text Mining: Clustering techniques are the part of unsupervised learning techniques, the main difference between supervised techniques and Un-supervised technique is that outcome class label is not known. These techniques can also be applied for document clustering. Using clustering techniques on text data is different from using data on numeric or categorical data. Clustering methods applied on text documents can group the documents into different cluster. The data within the cluster is same while these cluster differs from each other. There are several clustering methods which can be applied on text data. The clustering techniques can be divided according to the nature of data. Most popular clustering methods are partition based or density based. There are number of domains and application where these clustering methods are applied some of the applications are listed below:

(i) Organizing Documents & Browsing: Organizing document sequentially is one of the important clustering applications. By doing proper cluster analysis of various documents we can arrange them systematically, which can be very helpful from an organization point of view. Since lots of textual information can be found on web related to various documents, the unsupervised learning techniques are very useful.

(ii) Summarizing Corpus: Unsupervised techniques are helpful in corpus summarization, the unstructured text can be processed with various text mining methods and the corpus can be summarize to form cluster digests (D. Lewis., 1992) or clusters of words (L. Cai, T. Hofmann, 2002, S. Chakrabarti, S. Roy, M. Soundalgekar, 2003), which are utilized so as to offer summary information into the whole content of the cluster corpus. Various types of such systems for example, sentence grouping would also supportive to archive abridge. Those cases for claiming grouping is likewise personally attached to that about dimensionality decrease and subject sentence demonstrating.

(iii) Classification of documents: Clustering method is essentially a unsupervised learning method, even now these methods can be utilized to increase the quality in supervised version. Specifically, statement groups or cluster of words (L. Cai, t. Hofmann, 2002,S. Chakrabarti, S. Roy, M. Soundalgekar, 2003) . Unsupervised learning techniques can be used to enhance the performance of classification techniques word clustering and parallel training methods can be used to identify the documents which may belong to different groups. There are several clustering methods like partition based, hierarchical based and density based under these methods there are several popular clustering techniques like k-means, DBSCAN, AGNES, EM- algorithm. All these algorithms can be used to handle large volume of text data, which can be represent in numeric form also. The different clustering methods have different effectiveness in classification of documents. (F. Sebastiani, 2002) have shown in the experiment the comparative analysis of various clustering techniques.

Overview of Web data Mining & Text Classification: Now a days lots of data is available on web in different data formats. This data can be utilized effectively for enhancing productivity and performance in an organization. Web mining is one of the applications of data mining techniques which can be used to predict useful information from the web data. The overall process of Web mining can be divided into, web content mining, web structure mining and web usage mining. Web content huge amount of data which is mostly unstructured in nature, this data is available in the form of blogs, reviews, comments etc. Many people share their opinions about a product that they have purchased in the form of comments or reviews. These reviews are large in numbers and are treated as text documents. Text document available online are such source of un-structured data. These text documents are available in various ways such as product reviews, blogs, twitter data etc. There is lot of hidden useful information is available in these documents or reviews. Mining text data from web is tedious task , there are number of text mining techniques available , we discuss some important text mining method in this section..

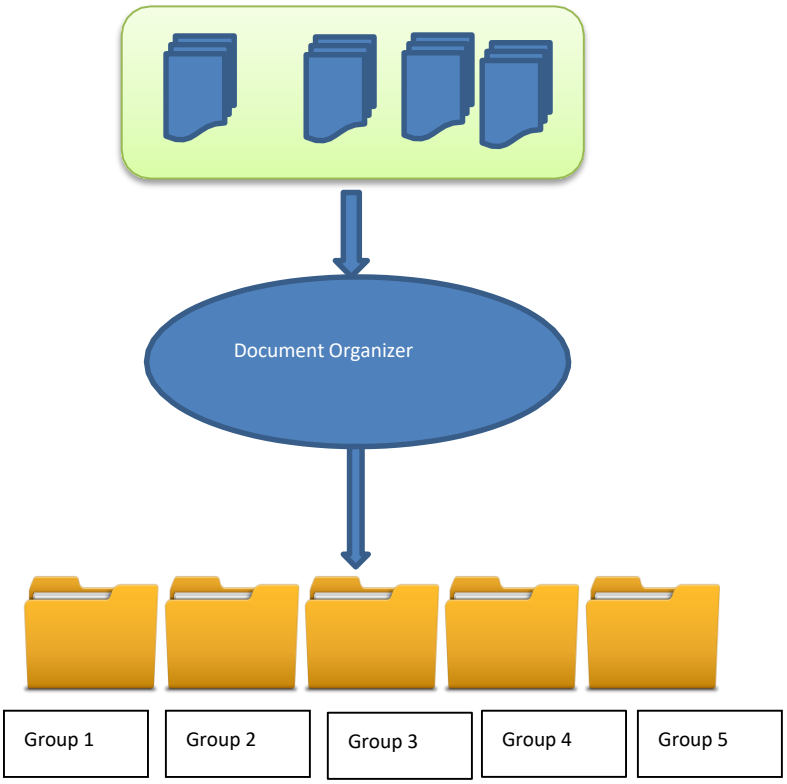

Figure 1.5 Document classification using text mining

Figure 1.5 shows one such example of document classification using text mining. The data mining techniques are little bit different from text mining techniques because the type of data which is mostly used with data mining techniques is structured in nature. On the other hand the data used for text mining is textual unstructured data, which is usually in document pattern. Now, text radically differs from numbers. We will analyse ways in this study akin to those used for data-mining. These methods have turned out to be quite effective sans the need to understand the particular text properties like grammatical concepts/word meanings. In order to mine the text document there can be the words which are less in frequency. The main concept behind application of machine learning methods to text data is that textual information can be converted into numerical forms. Therefore despite of the fact that the presentation is different of textual data, we are able to apply data mining methods because of available encoding schemes for converting textual information into numeric. The text mining methods are not much different from data mining methods once the unstructured textual data is converted. There are several ways through which we can transform the text data into numerical form, using spreadsheet method is one such approach. The occurrence or non-occurrence of word can be computed from spreadsheet cell in each document. Classification techniques are used with data warehouse , database and other information retrieval forums. These techniques are very useful in order to classify the data in almost every domain like business, science, social media etc. The implementation of these techniques can be done using various data mining tools like R, Weka, & rapid miner. In order to construct a classifier, the data is split into two parts, first is training data and second is test data. The classifier model is built using training data using any appropriate classification technique like decision tree, neural network or SVM. Text classification methods are used to classify textual data into positive or negative. One of the useful applications of text classification is to classify the product reviews into positive or negative review. Semantic orientation is computed using various text mining methods like point wise mutual induction, score function etc. There are number of text mining methods which can be used to explore the text data which is available in the form of reviews, blogs, comments etc. Frequent words can identified from the large text corpus, one can also find the various features from the product reviews.

Literature Review

Data Mining techniques have been used in past to explore and analyze the data in order to find better business ways in an organization. The huge amount of unexplored data is freely available on web and data mining (DM) techniques have been applied to extract some hidden useful information which may be useful to enhance the business of an organization. There is literature available that supports this fact that DM techniques have been used in past to develop new business opportunities. There are various applications of DM techniques, sentiment analysis and opinion mining is one

of them which can be applied on un-structured data. Sentiment analysis or Opinion Mining is a deterministic technique for classifying and evaluating other people's opinions. Now a day's people builds their perception and make decisions by analyzing the facts and reviews of other people either manually or computationally. Since everything is online now a day's , hence internet has become an integrated part of human lives and is thus used for exchanging all aspects of human life viz. sentiments, emotions, affection, support, opinions, trade, business etc. With the onset of social media there has been numerous platform such as *blogs, discussion forums, reviews and social networks* where an individual can post his or her reviews, feedbacks and list their likes and dislikes for a product's attributes or features or comparison of different products (same or different feature).
These reviews are gathered and are analyzed to evaluate the overall orientation of the collected reviews. This chapter focuses the past work done related to sentiment analysis and opinion mining. We have presented the outcome of research papers which have shown the application of machine learning techniques on online reviews. This chapter also discusses the research papers in which methods and techniques used for gathering and analyzing the reviews, extracting the phrases based on the Subjectivity (Esuli, A. and F. Sebastiani, 2006) and thereafter some work is also discussed for calculating the semantic orientation of the collected reviews. Sentiment analysis is the part of Subjectivity analysis (Akkaya, C., J. Wiebe, and R. Mihalcea, 2009) which is also very popular by the name Opinion mining. Opinion Mining is mainly concern with analysis of linguistic natural expression of individual's opinion about certain product or any other area where public opinion or review matter a most. Subjectivity analysis aims at determining the attitude of the writer or author of opinion with respect to some topic or product or services or the overall contextual polarity or tonality of a document or review (Hassan, A. and D. Radev, 2010). The attitude may involve the user's experience, evaluation, judgment, the emotional state or intended emotional effect. It is a Natural Language Processing (Indurkhya, N. and F. Damerau, 2010) and Information Extraction task that identifies the writer's feelings and experiences expressed in positive and negative comments, questions and requests, by analyzing monstrous amount of information available over the web. The major force behind the emergence of Opinion Mining today, is the exponential increase in Internet usage and exchange or share of public views and opinions (Dellarocas, C., X.M. Zhang, and N.F. Awad, 2010). It was observed by (Jijkoun, V., M.d. Rijke, and W. Weerkamp ,2010) that the some of the opinions can be topic based where documents are classified into predefined topic classes, e.g , science, sports, entertainment , politics etc. Topic related words are important in topic based classification.

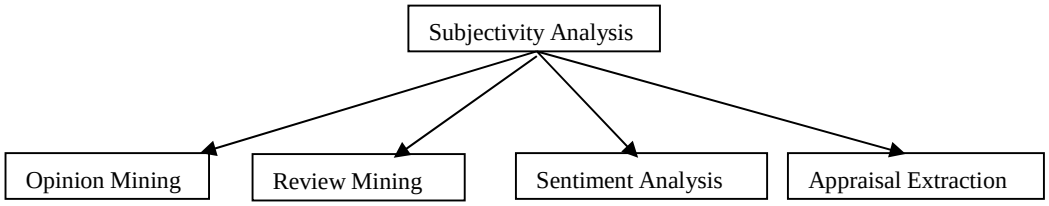

Figure 2.1: Names often interchangeably used for Sentiment analysis

However, in Sentiment classification (Li, S., S.Y.M. Lee, Y. Chen, C.R. Huang, and G. Zhou, 2010) they are least considered. Here, the classification is at document- level, where whole document is classified based on its polarity i.e words that indicate negative or positive opinions (sometimes neutral) are important e.g, great, poor, excellent, bad, disgusting etc. This classification can also be extended to sentence level, comparative sentence (Jindal, N. and B. Liu, 2006) i.e to classify each sentence as expressing a positive, negative or neutral opinion (Sugato C; Sabyasachi B; Karishma B, 2013). The client created online reviews are reprimanded by a few scientists to be seen as having lower believability and trust than traditional word of mouth because of the nonattendance of source signs on the virtual world like the Internet (Smith et al., 2005).

A few scientists have likewise reported in writing that conventional word of mouth frequently depends on meaningful gestures (e.g., social connection between word of mouth communicators) that can upgrade word of mouth convincingness (Zhang et al., 2010). Be that as it may, in circumstances like online reviews, these relevant signs may not be accessible (Gupta and Harris, 2010). The absence of meaningful gestures in online reviews powers customers to assess their influence exclusively in view of accessible constrained substance. In any case, significant discoveries of a past examination likewise uncovered that buyers assessed the online reviews as more reliable and valuable while seeing an assertion between the reviews and their own sentiments (Xia and Bechwati, 2008).To enable the above visualization, identify product's review's phrases in which customers have presented their views These opinions consist of user's viewpoint, fancy, attitude, sensibility, etc. The reviews can be of product's feature, its attributes or it could contain the comparison of different products of same realm (Li, S., C. Lin, Y. Song, and Z. Li., 2010). Completely dissecting and arranging conclusions includes undertakings that identify with some genuinely profound semantic and syntactic investigation of the content. These incorporate perceiving that the content is subjective, as well as figuring out what the conclusion is about, and which of numerous conceivable positions the holder of the opinions communicates with respect to that subject. Next we present the some of the research paper summary, it present the work which has been done in sentiment analysis.

Classification and Feature Extraction from Reviews

Overview of Sentiment Classification: The analysis of sentiments or emotions selectively classify positive or negative. The classifier takes input from the database. The property of selective attributes filter out the relevant attributes. Feature selection method can be used to select important

features in order to have better result. The validating cross- validation step enabled and estimated the accurate performance of the predictive model. The online reviews of mobile phones extracted from various product based websites are raw. The explosion in the mobile online market has created tremendous competition edge. In today's chance there is gigantic rivalry in this field as bunches of new organizations are entering in this business sector. In light of the high aggressive business sector the cost of the cellular telephones are falling quickly. There is an intense change in the mobile phone business sector on the off chance that we analyze the old cost and new cost. Shopper is additionally confounded before purchasing any mobile phone whether online or offline. Thousands of online product reviews are available for mobile phones, in our work we are characterizing such mobile phone reviews on the different perspectives like there features, over all polarity, frequently search features etc.

Decisions are the fundamental to all human exercises since they are key influencers of market practices. In this present reality, companies want to know the customer purchasing interest. Individual buyers likewise also wants know the post purchase experience of other person before purchasing any new item. The market trend presently, depends upon the sentiments or reviews by various clients. If one need to purchase an item or product, one is no more limited to approaching one's known ones for product feedback. In fact there are numerous consumer surveys and reviews are available out in the open gatherings on the web about the product. Textual information in the form of reviews is available online in huge amount and this is useful for an organization as it need not to depend on various tedious data collection methods. It might never again be important to conduct survey or to take opinion polls as huge amount of data is already available freely on web every webpage commonly contains an immense volume of customers' judgements or opinions that is not generally effectively summarized in long online reviews. The investigation of opinions might be subject based (Vijay Kumar and Kumar, 2010) where documents are ordered into predefined theme classes, e.g., science, sports, stimulation, legislative issues and so on. There are two types of classification topic based and document based. The keywords play an important role in the topic based classification. However in sentiment based classification (Li et al, 2010a) they are least bothered. Here, the classification is at document level, where entire report is grouped taking into account its polarity i.e., assessment words that show positive or negative opinions or neutral opinions. This classification can likewise be stretched out to sentence level, comparative sentence (Jindal and Liu, 2006) i.e., to characterize every sentence as communicating a positive, negative or unbiased opinion.
This research work presents utilization of text mining, machine learning and natural language processing techniques for sentiment analysis and review analysis for evaluation orientation from each mobile phone review. In this chapter we are discussing various classification techniques which are used to classify the each review based on class recommended.

Experiment and Results

The experimental results on the mobile reviews of the proposed work are presented in this chapter. We have collected the mobile reviews from Amazon.com. There are total 2000 reviews out of which 1000 are positive and 1000 are negative. As there is not directly mentioned on the website which is negative and which is positive review because there is a rating given to each comment from 1 star to 5 star. We assume that all 5 star and 4 star rated comments are positive comments and all one and two star rating comments are negative comments. We are not considering 3 star rated comments because these reviews are treated as neutral reviews and in this paper we consider only negative and positive comments. There is another reason for not considering 3 star rated comments because these comments do not have capability to provide any sentiment. We have collected these reviews using web crawler developed in python from Amazon.com so we can kept each comment in a separate text file. Now we have to create a dataset from these separate comments, so we created a dataset. The sample dataset with few reviews has been shown figure 4.1. The initial textual preprocessing was done on these reviews and then these reviews were divided into positive and negative opinions. We have used support vector machine classifier to develop a classification model which computes the orientation of the mobile review. The performance of the SVM classifier is evaluated and it is presented in term of overall accuracy, recall and precision. The performance of classifier on negative and positive reviews is shown in Figure 4.2 and Figure 4.3. In this work we have trained the support vector machine SVM. We have also selected some other default parameters; the feature selection was done using Information gain.

Sample structure of of Mobile Reviews:

'Phone seems to be good. But the battery was horrible, just 20 days up and it\'s not functioning. Cannot charge. The irritating thing is we can not return it as return time period is up to 10 days only. Now have to take this to bangalore. I purchased this phone for a comfort but it became an headache now.',NEG

'Never Buy redmi prime 2... They have the worst customer service.my mobile started showing digital lines in the screen. Eventhough the screen is not at all damaged (not even a scratch) the service people said that the problem is due to my mishandling. They are just not even ready to listen. Don\'t buy it...',NEG

'I received a defective piece.. Microphone is not working properly, person on the other end receives very low voice even if the call volume is full. Display flickers sometimes.. Wanted a replacement for this piece. Please help me out to replace this one. I have not used this phone more than a month.',NEG

'Dear Team,,,i have first time buy the cell phone apart from Sony...this is my worst experience...during second month..its side key not working...cell getting hanged...i dont know why manufacturer put so much features if cell dont have capacity..i dont want to buy or recommend this product to anyone..',NEG

'My recentky bought, MI2 prime is getting so heated like its going to blast the very next second. To get fully charged it takes 5hours, and its takes only 40 to 50 minutes to get law better warning, even You don\'t use it for a second.\r\nCan somebody help with company\'s email id or contact details?',NEG

'Mobile is not charging! Not even two months, I will have to visit service centre now',NEG

'Same hanging problem like samsung phones.Camera can say worst. Overall exp is worst.',NEG

Figure 4.1: Sample data set of online mobile Reviews

The number of features which we have selected is in the range which is varied from small selection like 50 features to large selection like 2000 features. We have divided the whole dataset into ten different equal sized sets. These sets contain the data for training and testing, we use nine data sets for training the SVM, and remaining data sets are used for testing. We have repeatedly done this experiment in by varying the size of training and testing data sets in order to get mean accuracy. The best accuracy that we find was around 81-82%. We have successfully classified the reviews into positive and negative class and classified the reviews based on sentiment orientation the overall accuracy of the model using SVM was 81 percent. These reviews were also tested on other machine learning techniques and the comparative result of those techniques is presented in terms of confusion matrix.

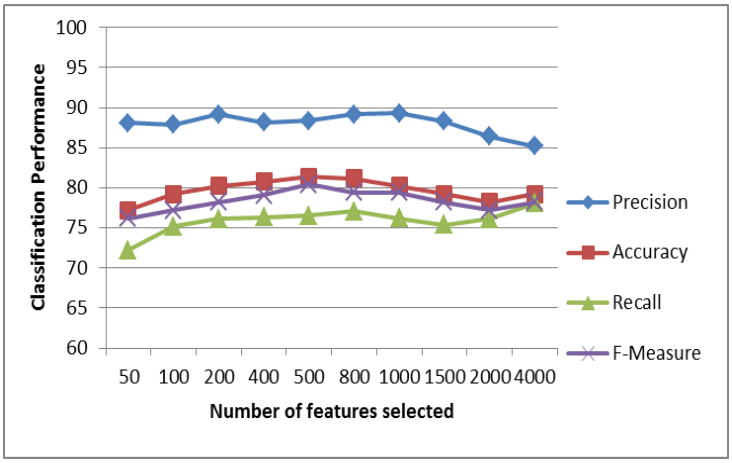

Figure 4.2: Sentiment Classification Performance for Negative Online Reviews

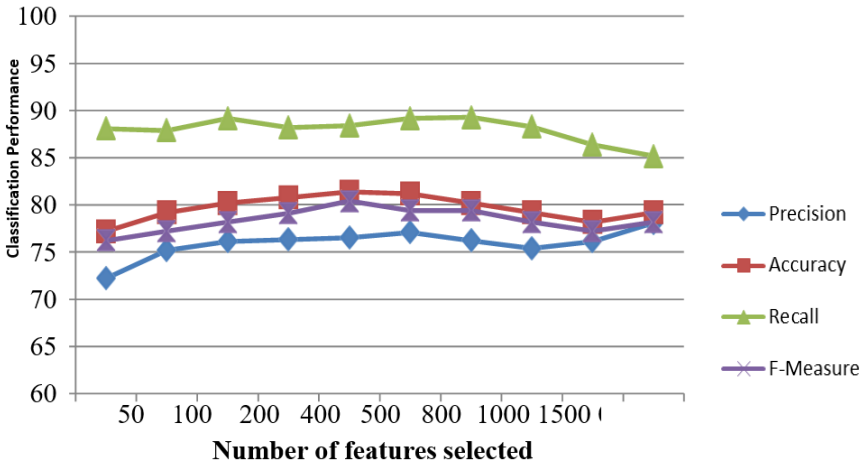

Figure 4.3: Sentient Classification performance for positive online reviews

Comparing accuracy of classifiers: We have found that overall accuracy of SVM was found 81%. The accuracy of classifier was measured in terms of confusion matrix where we have stored true negative and true positive instance along with predicted negative and predicted positive The similar process was repeated with naïve Bayes and decision tree classifier and the mean accuracy was obtained was
78% and 75% respectively. We haves tested after training on 1000 reviews data set and then dividing the test data set into two parts of 500 reviews each. Below mention table 4.1 and table 4.2 shows the result for test1 data and test 2 data using naïve Bayes containing 500 reviews and table 4.3 & table 4.4 shows the result for test1 data and test 2 data using decision tree containing 500 reviews.

Table 4.1: Confusion matrix test1 data for naïve Bayes

	True Negative	**True Positive**
Predicted Negative	**198**	**58**
Predicted Positive	**52**	**192**

Table 4.2: Confusion matrix test2 data for naïve Bayes

	True Negative	**True Positive**
Predicted Negative	**200**	**45**
Predicted Positive	**50**	**195**

Table 4.3: Confusion matrix test1 data for decision tree

	True Negative	**True Positive**
Predicted Negative	**188**	**66**
Predicted Positive	**62**	**184**

Table 4.4: Confusion matrix test2 data for decision tree

	True Negative	**True Positive**
Predicted Negative	**188**	**66**
Predicted Positive	**62**	**184**

User Interface for entering the reviews: We have also developed an interface to compute the frequency of mostly looked features in a phone. This research work has also identified the most frequent positive words in a positive review and most frequent negative word in negative review. The user interface is developed using

PHP, the interface has provision to compare products and to have a detailed information regarding product. We can perform document level mining by entering a whole document.

Conclusion and Future Scope

In this research work we have identified the various reasons for benefit of customer and organizations from online mobile reviews. The experiments showed that how an organization can also improve the mobile phone quality by focusing on improving by features of mobile phone. The overall conclusion of this work is explained further. This research has demonstrated some methods for knowing the opinions of the consumers through online reviews on mobile phones. This research work aimed at incorporating sentiment analysis in order to derive marketing intelligence from the analysis of mobile phone reviews. Another reason as to why this mobile phone review system may help the consumers and the mobile phone companies is that a lot of important results related to the behavior of the customers may be derived by analyzing the mobile phone reviews. Henceforth, the conclusions derived from this study have contributed in a massive way in knowing the consumer's attitude and thinking and services related to mobile phones as well as their other aspects. According to the recent studies the current generation and broad use of honest reviews available online is an opportunity to derive useful information for and mobile phone users and mobile phone industries. It is also stated via practical findings that considering online mobile phone reviews would profit potential buyers as well mobile companies. Currently, the mobile phone buyers consider online reviews on the internet for useful information of the desired cell phone before thinking about purchasing them. This implies that the opinions of other consumers strongly affect the decision making of buying mobile phones. The research provides ways to conduct automatic investigation of the reviews of the buyers which can be beneficial to potential visitors as well as mobile phone proprietors to make the right decision by looking into these reviews. This research is centered on mobile phone reviews but is also true for various other products and services. Analyzing opinions and views from online reviews and from other sections of products and services have always been an interest to various expert and analyst of an organization and gives better and deeper insight about a product using NLP and text mining techniques. Online reviews being a well-recognized source for user information have benefitted marketing managers and practitioners in the research department. We have experimented various mobile reviews using machine learning techniques to classify the reviews and for feature extraction and summarization. This can also be further extended for any other product as the reviews can be collected for other products also. But the applicability of sentiment analysis is yet to be determined in other domains and need further investigations. Also, opinions of the consumers are temporary, thus it would be interesting to include longitudinal analysis to track reviews periodically. We have also explain detailed methods to mine text data sentiment analysis and opinion mining is also explained in detail we have shown the various types of sentiment analysis at various level for example document level sentence level and aspect level the concept of opinion mining also explain in detail how are one of the objective was to find the polarity of an online review we have presented several machine learning techniques in chapter 1 we have also discuss various types of classifier which has been used in this research work to develop a classifier this research work has also presented a way by which we can improve decision making process in an organization In this research work we have demonstrated the use of data mining techniques for the benefit of organizations. This research work has demonstrated that how effectively we can mine the data which is available in huge amount on web. This research work has explained that how the enhanced use of social media and information sharing can be useful from both individual and organizational point of view. This research work has shown that how we can utilize the unstructured text data, which is freely available in the form of consumer opinion, can be utilized effectively for enhancing the business of an organization. This work has shown some techniques of sentiment analysis and opinion mining for automatically extracting consumer opinions from online reviews on mobiles phones.

Promoting Intelligence from Overall Consumer Sentiment

The summarized purchaser notion and sentiment on mobile reviews can be utilized for determining advertising knowledge as item or feature level attitude, mindfulness and affiliation. Mobile reviews analyzed at different level of granularity, like document level, sentence level and feature level. The detailed analysis gave deep insight about reviews in terms of classification and frequent feature selection. The classification of reviews at different level gives easiness to user to instantly classify the reviews as positive or negative. At feature level the identification of different features in the reviews can be further used to enhance their quality in mobile phones. For example if feature ‘camera’ is found as a frequent feature , its quality can be improved. By improving features and incorporating new one the sale of particular brand phone can be enhanced. This may be used as marketing strategies by different mobile companies.

REFRENCES

1. Dubey, G., Rana, A. and Ranjan, J. ‘Fine-grained opinion mining of product review using sentiment and semantic orientation’, Int. J. Business Information Systems, *INDERSCIENCE,* Vol. 25, No. 1, pp.1–17, 2017.
2. Dubey, G., Rana, A., Ranjan, J. A Research Study of Sentiment Analysis and various techniques of Sentiment Classification, Int. J. of Data Analysis Techniques and Strategies, *INDERSCIENCE,* Vol. 8, No.2 pp. 122 – 142, 2016.
3. Dubey, G., Rana, A., Shukla, N.K .User Reviews Data Analysis using Opinion Mining on Web, *IEEE* Conference Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015 Pages: 603 - 612, DOI:

10.1109/ABLAZE.2015.7154934.

4. Dubey, G., Rana, A., Ranjan, J. Summarizing and computing Orientation of user reviews using Bayesian Classification, International Journal of Emerging Technologies in Computational and Applied Sciences, 5(3), June-August, 2013, pp. 299-305.
5. Dubey, G., Rana, A., Ranjan, J. Opinion Mining From User Reviews using Sentiment Classification, presented in International Conference on Facets of Business Excellence (FOBE 2012), Nov 29-30.
6. Adnan D., Song F., Feature selection for sentiment analysis based on content and syntax models. Decis Support Syst 2012;53:704–11.
7. Akkaya, C.,Wiebe J., Mihalcea R., Subjectivity word sense disambiguation. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP- 2009), 2009.
8. Andrew L. M., Raymond E., Daly, Peter T., Oham D., Huang A. Y., 'Learning word vectors for Sentiment Analysis' .Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 142–150, Portland, Oregon, June 19-24, 2011.
9. Antonio, R., Paolo R., Making objective decisions from subjective data: detecting irony in customer reviews. Decision Support System 2012;53:754–60.
10. Argamon-Engelson S., Koppel M., Avneri, G., 'Style-based text categorization: What newspaper am I reading?', Proc. of the AAAI Workshop on Text Categorization, pp.1–4., 1998.
11. Agrawal, S., Chaudhari, S. and Narasayya, V., 'Automated Selection of Materialized Views and Indexes in SQL databases', In 26th International Conference on Very Large Data Bases (VLDB 2000), Cairo, Egypt, pp. 495-505, 2000.
12. Aouiche, K.,Darmont, J., 'Data mining-based materialized view and index selection in data warehouse', In Journal of Intelligent Information Systems, Pages 65 – 93, 2009.
13. Asad B. Sayeed, Hieu C. Nguyen, Timothy J. Meyer, and Amy Weinberg. Expresses-an- opinion-about: using corpus statistics in an information extraction approach to opinion mining. COLING (Posters), pp. 1095-1103, 2010.
14. Algur, Siddu P., Amit P. Patil, P. S. Hiremath, and S. Shivashankar. "Conceptual level
15. similarity measure based review spam detection." In Signal and Image Processing (ICSIP), 2010 International Conference on, pp. 416-423. IEEE, 2010.
16. Bai X., Predicting consumer sentiments from online text, Decision Support System 2011;50:73242.
17. Balahur, A., Hermida, J. M., Montoyo, A. 'Detecting implicit expressions of emotion in text: a comparative analysis'. Decision Support System 2012;53:742–53.
18. Berger A., Della S., Pietra Della V., 'A maximum-entropy approach to natural language processing', Computational Linguistics, March, Vol. 22, No. 1, 1996.
19. Bollen, J. and Mao, H. 'Twitter mood as a stock market predictor', IEEE Computer, October, Vol. 44, No. 10, pp.91–94,2011.
20. Boldrini E., Balahur A., Martı´nez-Barco P., Andre´ s M., Using EmotiBlog to annotate and analyse subjectivity in the new textual genres. Data Min Knowl Discov 2012;25:603–34.
21. Baralis, E., Paraboschi, S. and Teniente, E. 'Materialized View Selection in a Multidimansional Database', In 23rd International Conference on Very Large Data Bases (VLDB 1997), Athens, Greece, pp. 156-165, 1997.
22. Brin, S., Motwani, R., Ullman, J.D., Tsur, S. "Dynamic Itemset Counting and Implication Rules for Market Basket Data", SIGMOD Record, Volume 6, Number 2: New York, June 1997, pp. 255 - 264.
23. Cai, L., Hofmann, T., Text categorization by boosting automatically extracted concepts. ACM SIGIR Conference, 2003.
24. Carvalho, V.R., Cohen, W., On the collective classification of email "speech acts", ACM SIGIR Conference, 2005.