iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VII **Month of publication:** July 2018

DOI: http://doi.org/10.22214/ijraset.2018.7056

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com

Current Scenario of Opinion Mining & Sentimental Analysis

Kirti Kushwah[1], Rajendra Kumar Gupta[2]
[1]*Department of Computer Science & Information Technology*
[2]*Madhav Institute of Technology & Science, Gwalior (M.P.) 474005, India*

Abstract: ***Nowadays, social networking and online shopping sites are being used by the public to express their opinion about a particular entity. Entity may be any product, event or people etc. Corporate houses, government organizations and other rating agencies are always eager to analyze these opinions/views, which help them in taking strategic decisions. The analysis of opinions for some useful, hidden and previously unknown information is called Opinion Mining. Sentiment Analysis is the process of detecting the emotional reaction from the collection of opinion/ review. The goal of Opinion Mining and Sentiment Analysis is to make computer more intelligent to recognize and express emotions of user's. Currently some opinion mining and sentiment analysis techniques are available in the literature that processes the review, which are gathered from different user's view. However, these techniques are unable to remove noise from the data (review), discovering of sentiment and their polarity in complex sentences and extraction of multiple opinions from the same document etc. This paper describes a present scenario of available techniques and algorithms that are being used currently in opinion mining and sentiment analysis. Authors have also tried to put some light on some challenges and other related issues relevant to these techniques and algorithms.***
Keywords: Opinion Mining, Sentimental Analysis, Sentiment classification, Machine Learning

I. INTRODUCTION

Before the Internet era, people would seek opinions on products and services from the traditional sources such as friends, relatives or consumer reports etc., which was difficult, complicated as well as time consuming. Now due to easy accessibility of internet to common men it has become much easier to collect opinions from different users around the globe. Presently users visit opinion/review sites such as: - CNET, Epinions.com, Trip Advisor, Rotten Tomatoes and Yelp etc. before purchasing any product or hiring service. The numbers of buyer who buy products and services online are increasing expeditiously. As per the data available in literature [31], the growth rate of buyer globally in 2017 was 45%. Flipkart, One of the India's largest e-commerce marketplace, has expanded its buyer base by 33 per cent in the past six months to 100-million registered users. It had announced to breaching the 75-million user milestone in 2018. These data reveals that future of online shopping is bright. Opinion Mining and Sentiment Analysis will play a brighter role in online shopping sites. Natural Language Processing (NLP) is a field of computer science and artificial intelligence that mainly deals with human-computer language interaction. Opinion Mining is a type of NLP for tracking the mood of the public or user's about a particular entity. Sentiment Analysis is a kind of text classification based on the sentiment of opinions they contain, it thus plays an important part of NLP [1]. The main aim of opinion mining and sentiment analyses is to find out what others think. Nowadays opinion and sentiment of other people's highly influence the human behavior and it becomes central to almost all decision making activities. The fast growing online shopping website has significantly contributed to this user generated data including reviews, comments, opinions, services and events etc. This data is useful for the customer as well as the seller/manufacturer. The manufacturers can get feedback about their product strength and weaknesses based on the opinion and sentiment of the customers. The objective of the opinion mining is to fetch reviews of a product of various companies and select the best product for the consumer by analyzing user reviews [2]. Many algorithms have been developed, for searching the best product, on the basis of the reviews and feedback available on the various online shopping sites.

A. Types of Opinion

There are two types of opinion-

1) *Direct Opinion-* It is expressed sentiment expression directly on an entity or one of its aspects.
e.g. - "The picture quality of this camera is great."
2) *Indirect Opinion-* It is expressed on an entity based on its effect on some other entities. It shows relation that expressing similarities and differences of compare to more than one entity. e.g. - "Car x is cheaper then car y."

©IJRASET: All Rights are Reserved

B. Component of Opinion Mining

To express opinion, there are three components:-

1) *Opinion Holder:* The person or organization that holds a particular views or opinion about an object or entity.
2) *Object:* An object is that Item about which the opinion holder is showing his/her opinion.
3) *Opinion:* Opinion is used for determine whether the opinion holder's sentiment about any targeted object is positive, neutral or negative.

Sentiment analysis is the process of detecting the emotional reaction or contextual polarity of the opinion (text). It determines whether given text is positive, negative or neutral. It is otherwise called as opinion mining too, since it derives the opinion or attitude of the speaker. For this analysis, the opinions are collected from the users, which can be employed for further [4].The social networks act as a medium where the users can post or express many opinions a day and these blogs are used for classification.

The analysis of Sentiment and Opinion Mining are almost the same, but there is a small difference between them that is an extract and analyzes the people's opinions about an entity is known as opinion mining, while the analysis of feeling, seeks the feeling of words / expressions in a text and then analyze it is known as sentiment analysis.

As this paper highlight some point of current scenario and relevant challenges of opinion mining and sentiment analysis.

This paper is organized as follows: Section 2 presents literature survey that include the related works that have been done before in opinion mining and sentimental analysis sector and also discuss about the several proposed methods like – Machine Learning Method, Lexicon Based method that is used in sentimental analysis. Section 3, while presenting the comparison and consolidation of the various algorithms that is used in this sector. Section 4, while presenting the list of several tools that is used to extract the useful information from the collection of data and mine the opinion and sentiment regarding the entity. Section 5, while presenting various application areas where we used sentimental analysis. Section 6 presents conclusion and future works.

II. LITERATURE SURVEY

In today's world Internet or Online Shopping Sites have attracted the attention of everyone, which attempts to understand and analyze the thinking regarding any product by expressing their opinions and feelings. So now-a-days opinion mining has become an important field in research area due to its importance in business community. After critical analysis of currently available literature, authors put some research idea in this sector as discussed in following:-

1) Balahur et al., [5] presented a comparison on the techniques which may be utilized for mining opinions from quotations in news articles. The challenges in the task possibility of various targets as well as a huge set of affect phenomena which quotes comprise. The proposed methods evaluated were evaluated utilizing annotated quotations taken from news given by the EMM news collecting engine. A general OM system needs usage of both huge lexicons as well as specialized training as well as testing data [6].
2) Ding et al. [7] proposed and described a new type of tree known as a tree of opinions. A flexible system of opinions was formulated on the basis of the tree of opinions and therefore the extraction of medium and fine-grained opinions was conducted in a unified flexible model.
3) Anwer, et al., [9] who read the reviews word by word to finally summarize the results in terms of frequency and likelihood of opinions. The Bayesian probability was useful for obtaining accurate results and real forecasts. Because the frequency results were graphical, its use by new customers could lead to the decision to buy the product shown. Frequency based results were understood by the consumer and the results of the frequency were verified by the results of the Bayesian frequency.
4) Ji et al., proposed a sentiment mining and retrieval system which mines useful knowledge from product reviews. Furthermore, the sentiment orientation and comparison between positive and negative evaluation were presented visually in the system. Outcomes of experiments on a real-world dataset have shown the system is both feasible and effective.
5) Conditional Random Fields (CRFs) model was adopted by Qi and Chen [10] to perform OM tasks. Many elements in the model were integrated to optimize training and decoding. The new method was compared to the OM-based Lexicalized Hidden Markov Model (L-HMMs) method which showed that it was very accurate method in many respects.
6) A novel collaborative fuzzy set theory based filtering model which integrated subjective as well as objective data was suggested by Cheng and Wang [11]. The technique offered comprehensive results as well as solved issues of traditional Collaborative Filtering (CF) systems.
7) Huang et al. They proposed the processing of product attribute extraction as a sequential labeling activity and used discriminatory learning models that use conditional random fields (CRFs) to address it.

©IJRASET: All Rights are Reserved

8) A grouping of product characteristics for OM proposed by Zhai, et al. [11], looked at aggregators and e-commerce sites that were examples of companies that rely on OM to produce product-quality feature summaries. This model identifies product features and collects their positive and negative opinions to produce good and bad point's summary.

9) New Avenues in OM and Sentiment Analysis, which have valuable, vast, and unstructured information about public opinion was proposed by Cambria, et al., [12] where history, current use, and future of OM and sentiment analysis were discussed, with techniques and tools.

10) Steinberger et al. [13] suggested a semi-automatic method for creating dictionaries of opinion in different languages and launched high-level opinion dictionaries for two languages and automatic translation into a third language. The words discovered in the word list of the target language are normally used similarly to the meaning of the word in the two source languages.

11) Veeraselvi and Saranya [14] presented the opinion and organization survey subsystem, which had already been integrated into the proposed broader question answer system. This survey subsystem act as a subjectivity classification system that used the learning technique based on the genetic method (GBML) which considered subjectivity as a semantic problem.

A. Related Work

Now a day's many user on Social Networking site uses Internet to express their opinion and view (Feelings) regarding to a particular object. So Social Networks have attracted the attention of users, and here they understand and analyze other user's opinion and view regarding the object. This paper shows a survey of various techniques that is used for opinion mining. Opinion Mining can be classified into three categories-

1) *Sentence Level Mining:* The task of classifying a sentence as subjective or objective is often called classification of subjectivity in the existing literature. The resulting subjective sentences are also classified as an expression of positive, negative or neutral opinions, which is called the classification of sentiment at the level of prayer. Sentence-level mining has two subtasks:

a) Subjectivity Classification: The main purpose of Subjectivity Classification is to find out the sentence is Subjective or objective.

i. Subjective: Subjective sentence often has a basis in reality, but reflects the perspective through with the user's views reality. It shows the feeling of users.

e.g. - It is such a nice phone.

ii. *Objective:* Objective sentence is completely unbiased. It is not linked by the user's previous experiences. It does not have personal feelings.

e.g. - I bought a Samsung Galaxy a few days ago.

- *Sentiment classification:* The main purpose of Sentiment Classification is to classify the sentences or clauses, and find out that the sentence is positive or Negative.

2) *Document Level Classification:* They classified document according to sentiment instead of the topic. In this approaches whole document is considers as a single entity and the analysis approaches which is applied on the whole document. The result generated in document level sometimes not appropriate.

3) *Feature Based Level Mining:* Identify the sentiment of the opinion holder based on the features of a particular object. Because a person may like some features and some other may not like it. It gives better and in depth analysis of the product than the sentence level and document level sentiment classification [1].

Feature-based level mining contains step-

a) Identifying object features

b) Determining opinion orientations

c) Grouping synonyms

B. Techniques of Opinion Mining and Sentiment Analysis

Opinion Mining and Sentiment Classification techniques are based on the following approaches [15]-

1) Machine Learning Based Approach

2) Lexicon Based Approach

3) Hybrid Approach

©IJRASET: All Rights are Reserved

The Machine Learning Approach (ML) applies the various famous machine learning algorithms (Supervised, Unsupervised, Semi-Supervised) and uses linguistic features. The Lexicon-based Approach depends on a sentiment lexicon, a sentiment lexicon is a collection of known sentiment terms [16]. It is classified into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The various approaches and the most popular algorithms of opinion Mining and sentiment classification are illustrated in Fig.1.

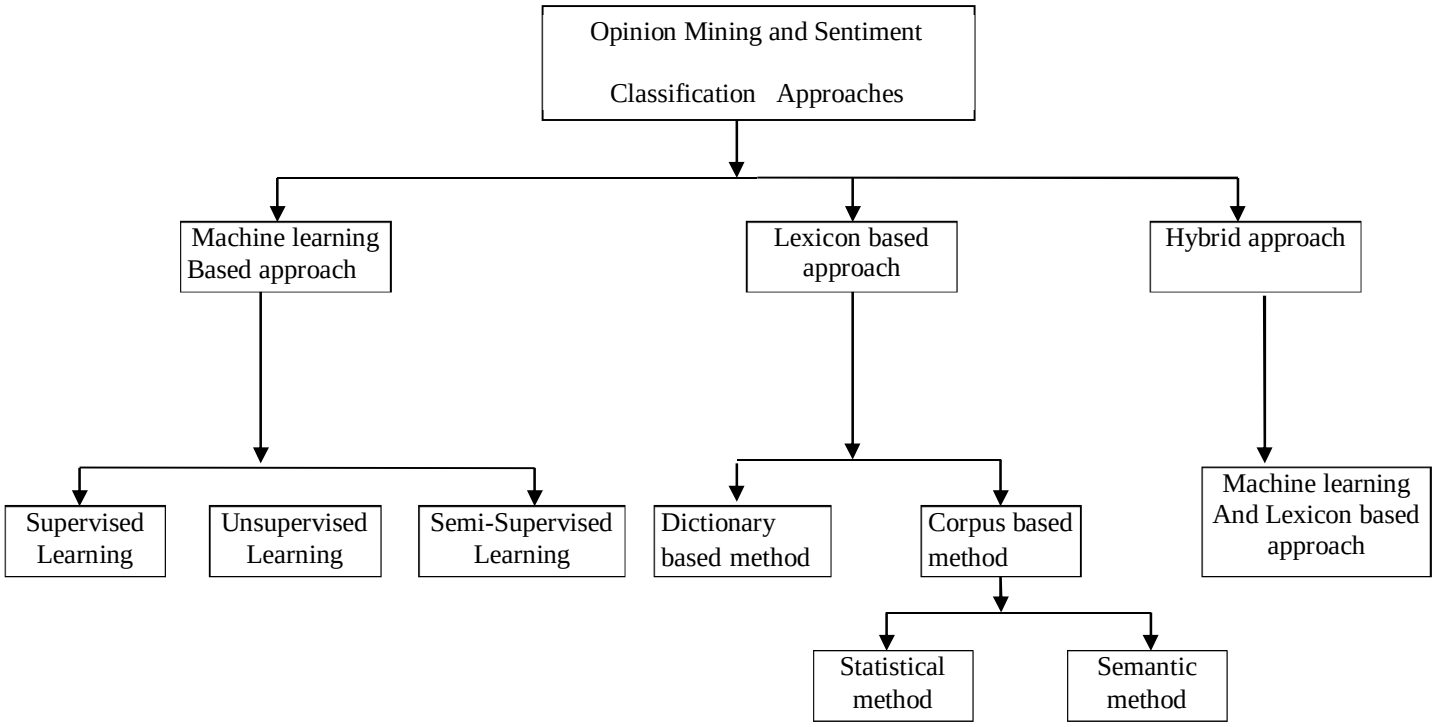


Fig. 1: Techniques of Opinion Mining & Sentiment Analysis

1) Machine Learning Based Approach

Machine Learning-Based Approach work by training an algorithm with a training data set before applying it to the actual data set. Machine learning techniques first trains the algorithm with some particular inputs with known outputs so that later it can work with new unknown data [15]. Machine Learning-Based Methods can be classified into three types-

a) Supervised Learning: This method contains two set of documents-Training Set and Test Set. Training Set is used by the classifier to learn about the document and, Test Set is used by the classifier to validate the document [17]. The supervised learning methods depend on the existence of labeled training documents [18]. There are many techniques for opinion classification -

i) Decision tree classifier: The main aim of Decision tree classifier is to decomposition of the training data space in which we divide the data by applying the condition on the attribute value. The condition is the presence or absence of one or more words. The division of the data space is performed recursively until the leaf nodes contain a minimum number of records used for classification. Yes, then the rules are also generated by the decision tree. The implementations of the decision tree are used in the classification of the text [3]. It shows small variations on standard packages such as ID3 and C4.5. The C5 algorithm which is a advance version of the C4.5 algorithm. The concept of this algorithm Depend on the tree. An approach has been proposed to extract the content structures of the terms of the argument in the context at the sentence level using the MASS (Maximum Spanning Tree) structure to discover the links between the term of the " t " argument and the relative words of context. Topical terms are those terms that specified entities or terms or certain aspects of entities in a particular domain. They extract some topical terms from text based on their domain term-hood. Then, these topical terms are used to differentiate document topics. .This approach is different from the machine learning based approach but it has the ability to learn the positive and negative sentiment term [6].

©IJRASET: All Rights are Reserved

ii) Linear Classifier

1. *Support vector machine (SVM):* Support Vector Machines are one of the type of linear classifier in supervised learning methods which is used for classification. In sentimental classification Support Vector Machines (SVM) classifier is used very frequently. Basically the goal of SVM classifier is to find a linear decision boundary that separates the collection of review data in such a way that the margin is maximized. SVM perform this by taking a series of points and separating those points using mathematical formulas.
2. *Neural network (NN):* Neural Network consists of many neurons where the neuron is its basic unit. The inputs to the neurons are denoted by the vector over line Xi. There are a set of weights A which are associated with each neuron used in order to compute a function of its inputs. Based on inputs and weights output is generated. Multilayer neural networks are used for non-linear boundaries. As Neural Network uses single layer, but Multilayer Neural Networks are used multiple layers to originate multiple piecewise linear boundaries, which are used to approximate enclosed regions that belongs to a particular class. The outputs of the neurons in the previous layers feed into the neurons that present in the later layers. This training process is more complex because the errors need to be back-propagated over different layers.

iii) Rule based classifier: This is supervised learning in which the system is trained to learn if any new input is given. In rule based classifiers, the data space is modeled with a set of rules. The left hand side represents a condition on the feature set while the right hand side is the class label. The conditions are on the term presence. The absence of terms is rarely used because it is not informative in the scattered data. There are numbers of criteria in order to generate rules, the training phase and construct all the rules depending on these criteria. The most two common criteria are support and confidence.

1. *Support:* There are several rules are generated, so support is the absolute number of instances in the training data set relevant to the rule.
2. *Confidence*: As there are many rules are generated, so the Confidence refers to the conditional probability that present on the right hand side of the rule is satisfied, if the left-hand side is satisfied [113].

b) Probabilistic classifier

Probabilistic Classifiers is one of the important classifier of Supervised Learning method. For classification purpose the probabilistic classifiers use mixture models. Basically the mixture model assumes that each class is a component of the mixture. Probabilistic classification is the study of approximating a joint distribution with a product distribution. Bayes rule is used to estimate the conditional probability of a class label, and then assumptions are made on the model, to decompose this probability into a product of conditional probabilities. Three of the most famous probabilistic classifiers are:-

i) Naive Bayes (NB): The Naive Bayes classifier is the simplest and most commonly used classifier. Naive Bayes classifier computes the posterior probability of a class, based on the distribution of the words in the document. So it is one of the probabilistic classifier and this classifier is mainly used when the training set is small. Probabilistic classifier is based on Bayes theorem. By Bayes theorem determined the conditional probability that an event A occurs given the evidence B is-

P (A/B) = P(A) P(B/A) / P(B)

So for finding the sentiment the equation is -

P(Sentiment/Sentence)=P(Sentiment)P(Sentence/ Sentiment)/P(Sentence)

1. *Bayesian Network (BN):* Naïve Bayes (NB) classifier is the independence of the features. But in Bayesian Network all the features are fully dependent. Basically Bayesian Network (BN) model is a directed acyclic graph whose nodes represent random variable, and edge represents conditional dependencies. But the computation complexity of Bayesian Network (BN) is very expensive that is why it is not frequently used.
2. *Maximum Entropy Classifier (ME)*The maximum entropy classifier is also known as conditional exponential classifier that converts feature sets labeled into vectors. This encoded vector is used to calculate the weights of each characteristic that can then be combined to determine the most likely label for a set of characteristics. This classifier is parameterized by a set of weights, which is used to combine the characteristics of the joint generated by a set of characteristics by means of an encoding. Basically, this encoding associates each pair (feature set, label) with a vector. The ME classifier was used to detect parallel sentences between any pair of languages with small amounts of training data. The other tools developed to automatically extract parallel, non-parallel corpora data using specific linguistic techniques. Their results have shown that ME classifiers can produce useful results for any linguistic pair.

©IJRASET: All Rights are Reserved

ii) *Unsupervised Learning Method:* The training data set is not required for the classification of unsupervised machine learning. In contrast of supervised learning, unsupervised learning has no clear targeted output that is associated with input. Class label for any instance is unknown so unsupervised learning is about to learn by observation instead of learn by example. Clustering is a technique used in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters. The uses of different grouping algorithms, such as the K-Mean grouping, Hierarchical grouping to classify data into classes. The neural network can be used to define the threshold values of words and then classify them as defined the values.

iii) *Semi-Supervised Learning Method:* Semi-Supervised learning (SSL) method is a new machine learning method that SSL method is tries to overcome the drawback of both supervised and unsupervised methods. The major drawback of supervised learning which is learning only from labeled data. Hence, using SSL with unlabelled Data specially when there is fixed amount of labeled data, can help to achieve improvement over supervised learning [14]. Also, the constraints of unsupervised learning approaches are not present with SSL if we include some form of prior knowledge to unsupervised models.

2) *Lexicon Based Approach:* Lexicon Based Method uses sentiment dictionary with opinion words and then match these opinion words with the data to determine the polarity. So that we can assigns the sentiment to the opinion words .Basically there are three types of sentiment- Positive, Negative and Neutral. Positive opinion words show some desired states, while negative opinion words shows some undesired states [21]. It consists of two methods-

a) *Dictionary-based Method:* In this approach manually we collect a small set of opinion words from the database then this set of opinion words are searching in the corpora WordNet and matches for their synonyms and antonyms. WordNet is an example of dictionary based method i.e. is used to develop a lexicon called SentiWordNet. If the word is not found in dictionary, it means this is the new word and this newly found words are added to the dictionary, so that whenever these words are repeated they will find in dictionary. This searching process is continuing until no new words are found. The major disadvantage of dictionary based approach is that it is not able to find opinion words with context specific orientations [22].

b) *Corpus-based Method:* As the dictionary based approaches are unable to find opinion words with context specific orientations, so this disadvantage overcome in the Corpus-based approach. The Corpus-based approach helps to find opinion words with context specific orientation. It begins with a seed list of opinion words and then searches for other opinion words in a large corpus. This approach is depending on the syntactic patterns of the opinion words in a large corpus[11]. One of the method were represented to identify additional adjective opinion words and their orientations, they started to use a list of seed opinion adjectives, and used them along with a set of linguistic constraints

For connectives the constraints are like – AND, BUT, OR, EITHER-OR, etc. There are also task, so The corpus-based approach is performed using statistical approach or semantic approach-

i) *Statistical Approach:* This approach used the co-occurrence of adjectives in a corpus. The polarity of a opinion word can be find by studying the occurrence frequency of the word in a large corpus of texts. If the word occurs more frequently with positive texts, then the words polarity is positive. If it occurs more frequently with negative texts, then the words polarity is negative. If the word has equal frequencies, then it is a neutral word. The similar opinion words frequently appear together in a corpus. This is the main observation that the state of the art methods are based on. Therefore, if two words appear together frequently within the same context, they are likely to have the same polarity [21]. This approach overcomes the problem of the unavailability of some words in the text, if the used corpus is not large enough.

ii) *Semantic Approach:* The Semantic approach provides sentiment values directly and relies on different principles for computing the similarity between words. This principle gives similar sentiment values to semantically close words. WordNet is the example that provides different kinds of semantic relationships between words, that is used to calculate sentiment polarities [23]. The Semantic approach is used in many applications to build a lexicon model for the verbs, nouns and adjectives to be used in Sentimental Analysis. One model shows the detailed subjectivity relations among the actors in a sentence expressing separate attitudes for each actor. These subjectivity relations are labeled with information by identify of the emotions (positive vs. negative) of the attitude [25]. They could find the weaknesses of the product, as it was probably the most unsatisfied aspect in customers' reviews, or the aspect which is more unsatisfied when compared with their competitor's product reviews [26].

3) *Hybrid Approach:* The hybrid Approach combines both Machine Learning Approach and Lexical Based approach.

©IJRASET: All Rights are Reserved

III. COMPARISISON AND CONSOLIDATION

The comparison between of the two main approaches i.e. Machine Learning Based Approach and Lexicon Based Approach used in sentiment analysis is shown in Table Sentiment Analysis performed by many different-different approaches that will produce different-different results because every approach has its own way of processing, so that every approach has its own advantage and disadvantage. By considering the key factors like performance, efficiency, and accuracy, the machine learning approach yields the best result and most of the work has been done in this approach.

Table 1: Comparison of Machine Learning and Lexicon Based Approaches

Approaches	Classification	Advantages	Disadvantages
Machine Learning Based approach	Supervised and Unsupervised Learning	• Good for Document level • Roboust • No need of Dictionary • Substantiate the high accuracy of classification.	• Classifier trained on the texts in one domain in most cases does not work with other domains.
Lexicon Based Approach	Unsupervised learning	• Labelled data and the procedure of learning is not required.	• Requires powerful linguistic resources which are not always available.

Table 2: Comparison of different Classifier

Methods	Advantages	Disadvantages	Title, Author, Publication	Accuracy
Super Vector Machine (SVM)	• Good prediction in a variety of situations • Easy to interpret result • High dimensional input space • Work well even if data is not linearly separable in the base of feature space • High accuracy • Few irrelevant features. • Document vectors are sparse	• Speed and size requirement both in training and testing is more • High complexity and extensive memory requirements for classification in many cases • Data collection is tedious	Title: Sentiment Analysis: Measuring Opinions Author: Chetashri Bhadane, Hardi Dalal, Heenal Doshi Publication: Science Direct, 2015 [4]	78%
Neural Network (NN)	• Good prediction generally • Some tolerances to correlated inputs • Incorporating the predictive power of different combinations	• Not roboust to outliers • Difficult in dealing with big data with complex model • Susceptible to irrelevant feature • Computationally expensive	Title: Neural networks for sentiment analysis on Twitter [19] Author: Brett Duncan, Yanqing Zhang Publication: IEEE, 2015	NA

©IJRASET: All Rights are Reserved

	of inputs • Fast to train the data			
Naive Bayes (NB)	• Simple to implement • It predicts accurate results for most of the classification and prediction problem • Great computational efficiency and classification rate	• Mainly used when the size of the training set is less. • For obtaining good results it requires a very large number of records	Title: Applying Supervised Opinion Mining Techniques On Online User Reviews [20] Author: Ion Smeureanu, Cristian Bucur Publication: Informatica Economică, 2012	80% (NB, n-gram)
Maximum Entropy(ME) Classifier	• Requires only small number of connections in neural network • This method does not assume the independent features like NB method • Can handle large amount of data.	• Learns only static input-output mapping • Simplicity is hard	Title: Twitter Sentiment Mining Framework Based Learners Emotional State Classification And Visualization For E-Learning System [21] Author: M.Ravichandran, G.Kulanthaivel Publication: Journal of Theoretical and Applied Information Technology, 2014	95% (SVM, ME)
Decision Tree Classifier	• It is able to handle both numerical and categorical data • Performs well with large dataset • Can handle missing values • A single tree is highly interpretable	• Relatively less predictive in many situation • Cannot work on combination of features • Decision tree can lead to overfitting	Title: Analysis And Indentifying Variation In Human Emotion Through Data Mining [27] Author: Jasakaran Kaur, Sheveta Vashisht Publication: International Journal of Computer Technology and Applications, 2012	NA

IV. TOOLS USED IN OPINION MINING

Free of charge and easy availability are two key driving forces behind open-source tools. The objective of saving money dictates many to focus exclusively on open source test tools and some licensed tools available commercially. But selecting open source tools only helps set aside the initial license cost, not other costs associated. There are so many open source text analysis tools that are used for processing natural language, such as information extraction and classification, which can also be applied to the analysis of feelings. The tools are listed below-

1) *WEKA:* The purpose of this tool is used for Data pre-processing, Classification, Regression, Clustering, Visualization, Association Rules and Machine learning algorithm for Data Mining.
2) *NLTK:* The purpose of this tool is Provides lexical resources such as WordNet, Tagging, Parsing, Semantic reasoning, Classification, Tokenization and Stemming.
3) *STANFORD CORENLP:* The purpose of this tool is used for POS tagging, Named entity recognizer, Parsing, Co reference resolution system and Sentiment analysis.

©IJRASET: All Rights are Reserved

4) *GATE:* The purpose of this tool is used for Sentence Splitter, POS tagging, Tokenizer, Gazetteer, Named entities transducer and Coreference tagger.
5) *Ling Pipe:* The purpose of this tool is used for Entity extraction, POS tagging, Clustering and Classification.
6) *Robust Accurate Statistical Parsing:* This tool is used for Statistical Parser, Tokenization, Tagging, Lemmatization and Parsing.
7) *Pattern:* This tool is used to enables users to find products based on features.

This paper describe about the many solution providers are present in the market of opinion mining tools. Most of these applications consider customer comments on products and services and analyze the feelings felt by positive / negative feelings when interpreting natural language.

V. APPLICATION AREAS

The most useable applications of this field are the following:

A. Market Analysis

For a long period of time marketers were highly dependent on the traditional marketing like:-TV commercials, newspaper advertisement, posters and even radio commercials. These methods have served for marketing (launching) a product well in the past. Some of these methods are time consuming. So now days as businesses expanded into new marketing categories and new technologies were adopted. During the past years online shopping has grown. Now days it is being used extensively by many businesses to promote their products and services through online advertisements. Companies are using online shopping site such as Flipkart, Paytm, and Amazon etc. to provide various services and interact with customers.

The current trends of market can also be analyzed by using opinion mining and sentimental analysis as one of the techniques. With the help of these techniques, businessmen or marketer can find out the list of products which are liked and disliked by the end users. Marketers not only sell the products but also determine the people's feeling, reaction, sentiments and suggestions regarding their new product. These determinations are based on the techniques of opinion mining and sentimental analysis. It can help the users to decide what is recommended and what is not recommended. Market Intelligence (MI) is designed to meet four needs of business managers such as:-

1) Opportunities and threat determination
2) Having competitors
3) Contributing to anticipating competitors' actions
4) Helping to make effective marketing decisions

In this regard, Yung-Ming Li, Tsung-Ying Li [28] framed a framework of market intelligence with the help of various tasks such as:-

a) *Trendy Topics Detection:* This is the first task of MI, mainly the aim of this task to detect the trendy (hot) topic that is currently in very growth (famous). This task was accomplished by assigning a thematic scoring trend with the help of:- The term frequency, Inverse frequency of the document and Models Appearing in a term .This can be done by using some techniques of opinion mining and sentimental analysis.
b) *Classification Review:* This is the second task that was performed in two phases (a) Subjectivity classification based on self developed lexicon (b) Sensation of classification according to SVM with the RBF (Radial Basic Function) and NB (Naive Bayes) kernel with emotions, unigram, bigram, unigram + bigram and a series of subjective words, as a representation of the characteristics**.** With the help of these two phases given review can be classify.
c) *Credibility Assessment:* This is the third task, the term Credibility Assessment refers to any attempt to ascertain truthfulness. Other terms that have been used to refer to the assessment of credibility include the identifying deception and lie detection. The goal of assessing credibility is usually to determination of the truth of the varity of a statement or series of statements. The third task was the calculation of credibility based on the follower-followee relationship.
d) *Numerical Synthesis:* Finally, In the last task the semantic score and the credibility score were added in numerical summarization task.

For the experiment, follower and followed relationship and tweets on three brands viz. Google, Microsoft, Sony, for three products namely iPhone, iPad, and Mac book have been collected from Twitter. For the training, 11,929 tweets have been collected with two emoticons ":)" and ":(", SVM and NB have been trained on 9,165 preprocessed tweets. So, improvement can be made by considering meronym patterns, PageRank consideration for the credibility assessment etc.

©IJRASET: All Rights are Reserved

Opinion mining and Sentimental Analysis is very beneficial to manufacturers. By using this, they can get feedback (both, positive as well as negative) from the customer's side regarding their products or services. Based on this feedback, they can make the necessary changes that helps improve the lacking areas and increase their business.

B. Buying a Commodity or Service

Opinion mining helps the people to interpret and understand other people's experience or review about any product or service. It helps them in taking the right decision from a variety of options. From a large amount of data available on the internet, Opinion mining takes into consideration user reviews and opinions, interprets it and displays it to the users in a presentable and easily understandable form.

C. Spam Identification

With the rapid increase in the number of internet users, the amount of content being uploaded is increasing at a very fast pace. This leads to the chances that spam content being posted on the web. Sometimes some users may intentionally upload spam content just to confuse other users. Opinion mining can be used to differentiate between spam content and authentic content

D. Policy Development

policy makers can use the extraction of opinions to receive feedback and comments from a particular entity and decide whether or not to make changes based on the approval or disapproval of the people.

E. Detection of "flame"

Opinion mining can be used to find out negative or arrogant words by analysing various social media, forums and blogs.

F. Business Intelligence

Opinion mining is well-suited for Business intelligence (BI) which is utmost importance now days. For example- a car manufacturer, who is worried about the unexpected low sales, tries to answer the following question: "Why aren't consumers buying our cars?" Although important specifications such as the car's weight or the price of the competitor's model are quite relevant, the manufacturer needs to focus more on consumer reviews of such objective as well as subjective characteristics to answer this question.

G. Voting Advise Applications

It helps voters to compare different political parties. It helps to understand which political parties have closer positions to theirs. For example, Smart Vote, asks the voter to declare its degree of agreement with a number of policy statements. Then it matches its position with the political parties.

In this paper, a wide range of applications are covered by opinion mining and sentiment analysis. In almost all spheres, it can be used. It is applied in query response systems, search engines, recommendation systems, etc. Interactions with the human computer can be improved and valuable information for further analysis is provided by Opinion mining.

VI. CHALLENGES AND ISSUES

A. The major challenge in the process of semantic analysis is the handling of ambiguity in Natural Language Process (NLP) and evaluation of non-standardized data. Every user has its own distinct style of expressing their opinions. The user may or may not use correct grammar, which may result in ambiguity. Also, now-a-days the use of acronyms and short forms is very common. This serves as a major hurdle when processing natural language and interpreting sentiments [29].

B. Also there are certain linguistic issues in opinion mining. The language used may not always be English. It is tough to interpret any data which is not in English without translating it. Also, data available in English might be tricky to interpret especially when the data contains jargons (special words or expressions that are used by a particular profession or group) or words from local language.

C. Another challenge faced is the cost of opinion mining software/tools that can only be afforded by the government and other large organizations.

D. Another challenge that is faced in opinion mining is the domain dependent nature of sentiment words. A set of features may perform well in one domain, but at the same time, it may not perform well in some other domain.

©IJRASET: All Rights are Reserved

E. Natural language processing (NLP) overheads like ambiguity, co-reference, inference, implicitness etc create obstacles in opinion mining.

F. The software's that are used for opinion mining are highly expensive and it can only be afforded by the large organizations. It is beyond a common man's expectation. These software's should be available to common man also so that every individual can benefit from it.

VII. CONCLUSION AND FUTURE WORKS

This survey paper presented an overview of the current scenario of Opinion Mining and Sentimental Analysis that contain various techniques and applications in various fields. It helps in classifying and summarizing reviews in real time applications. Mainly this paper focuses on sentiment analysis techniques, machine learning, that available and challenges in sentiment analysis. Various sentiment analysis methods and its different levels of analyzing sentiments have been studied in this paper. Machine learning methods like SVM, NB, Maximum Entropy methods were discussed here in brief, along with some other interesting methods that can improve the analysis process in one or the other way In the world of Internet majority of people depend on social networking sites to get their valued information, analysing the reviews from these blogs will yield a better understanding and help in their decision-making. There are still some open challenges in this area, such as the discovery of feeling and its polarity in complex sentences, the implicit identification of aspects, the extraction of multiple opinions from the same document. It is a vast research area with several challenges. However, just like all other things, it has certain challenges that need to be overcome in order to get the best possible results. The one of the challenge is now days online shopping sites are used by users to express their opinion regarding to products, events, entity etc. But there is chances that reviews available on these sites may contain noise so, it is necessary to make data noise-free so that it can be used for opinion extraction.

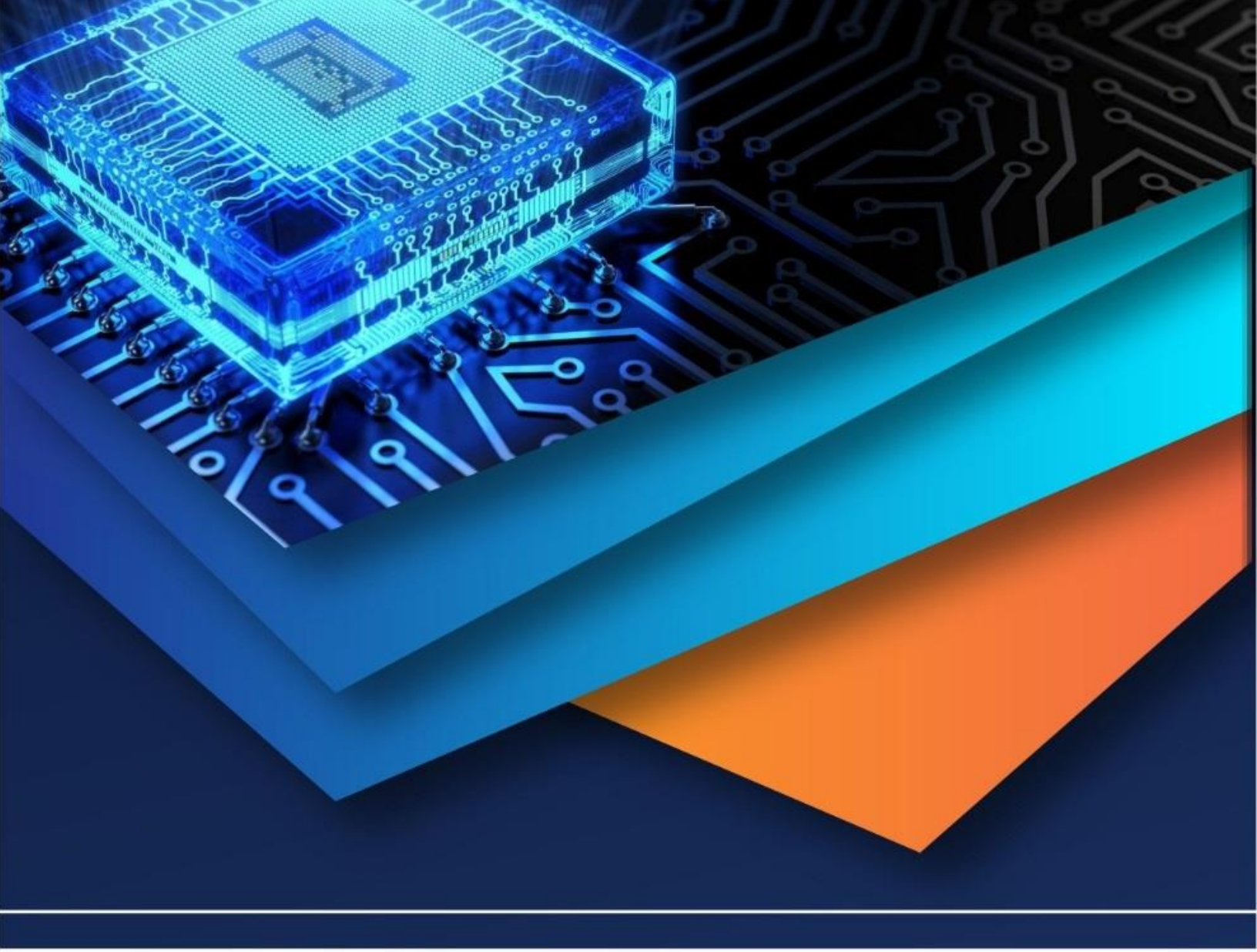
REFERENCES

[1] Solanki Yogesh Ganeshbhai, Bhumika K. Shah, "Feature Based Opinion Mining: A Survey", IEEE International Advance Computing Conference (IACC), 2015.
[2] Ananta Arora1 , Chinmay Patil2 , Stevina Correia3, 11, November 2015, "Opinion Mining: An Overview", International Journal of Advanced Research in Computer and Communication Engineering ,Volume- 4
[3] Weiyuan Li, Hua Xu, 2013, "Text-based emotion classification using emotion cause extraction", ELSEVIER .
[4] Chetashri Bhadane,Hardi Dalal, Heenal Doshi, 2015, "Sentiment Analysis: Measuring Opinions", Science Direct, Volume-45 .
[5] Balahur A., and Montoyo A., October 2008, "A feature dependent method for opinion mining and classification. In Natural Language Processing and Knowledge Engineering", International Conference on IEEE.
[6] Balahur A., Steinberger R., Goot E. V. D., Pouliquen B. and Kabadjov M. , September 2009, "Opinion mining on newspaper quotations In Web Intelligence and Intelligent Agent Technologies", IEEE/WIC/ACM International Joint Conferences IET.
[7] Ding J., Le Z., Zhou P., Wang G. and Shu W, November 2009, "An Opinion-Tree Based Flexible Opinion Mining Model. In Web Information Systems and Mining", International Conference on IEEE.
[8] Jiang P., Zhang C., Fu H., Niu Z. and Yang Q., December 2010, "An approach based on tree kernels for opinion mining of online product reviews. InData Mining" , IEEE 10th International Conference on IEEE.
[9] Anwer N., Rashid A. and Hassan S., August 2010, "Feature based opinion mining of online free format customer reviews using frequency distribution and Bayesian statistics. In Networked Computing and Advanced Information Management (NCM)", Sixth International Conference on ,IEEE
[10] Qi L. and Chen L., August 2011,. Comparison of model-based learning methods for feature-level opinion mining", IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society.
[11] Zhai Z., Liu B., Wang J., Xu H. and Jia P. 2012, "Product Feature Grouping for Opinion Mining". IEEE Intelligent Systems
[12] Cambria E., Schuller B., Xia Y. and Havasi C, 2013, " New avenues in opinion mining and sentiment analysis".
[13] Josef Steinberger, Polina Lenkova, Mohamed Ebrahim, Maud Ehrmann, Ali Hurriyetoglu, Mijail Kabadjov, Ralf Steinberger, Hristo Tanev and Vanni Zavarella, 24 June 2011, "Creating Sentiment Dictionaries via Triangulation", EC Joint Research Centre 21027, Ispra (VA), Italy.
[14] S. J. Veeraselvi and C. Saranya, March 2014, "Semantic orientation approach for sentiment classification", International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE).
[15] Marouane Birjalia, Abderrahim Beni-Hssanea, Mohammed Erritalib, 2017, "Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks", The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)Volume-113.
[16] Nilesh M. Shelke, Shriniwas Deshpande, Vilas Thakre, November 2012. "Survey of Techniques for Opinion Mining", International Journal of Computer Applications Volume- 57– No.13.
[17] Gautami Tripathi and Naganna S, 2, June 2015, "Feature Selection And Classification Approach For Sentiment Analysis", Machine Learning and Applications: An International Journal (MLAIJ) Volume- 2.
[18] Vidisha M. Pradhan, Jay Vala, Prem Balani, January 2016. "A Survey on Sentiment Analysis Algorithms for Opinion Mining", International Journal of Computer Applications, Volume- 133 – No.9
[19] Brett Duncan, Yanqing Zhang, 14 September 2015, "Neural networks for sentiment analysis on Twitter", IEEE 14th International Conference
[20] Ion Smeureanu, Cristian Bucur, February 2012, "Applying Supervised Opinion Mining Techniques on Online User Reviews", Informatica Economic, Volume-16.

©IJRASET: All Rights are Reserved

[21] M. Ravichandran, G.Kulanthaivel, 10 November 2014, "Twitter Sentiment Mining Framework Based Learners Emotional State Classification And Visualization For E-Learning System", Journal of Theoretical and Applied Information Technology, Volume-69
[22] Nawaf A. Abdulla , Nizar A. Ahmed, Mohammed A. Shehab and Mahmoud Al-Ayyoub, 2013, "Arabic Sentiment Analysis: Lexicon-based and Corpus-based", IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT).
[23] Samir Tartir , Ibrahim Abdul-Nabi, 2017, "Semantic Sentiment Analysis in Arabic Social Media", Journal of King Saud University – Computer and Information Sciences, Science Direct
[24] Deepa Anand, Deepan Naorem, 2016 , "Semi-supervised Aspect Based Sentiment Analysis for Movies using Review Filtering", 7th International conference on Intelligent Human Computer Interaction, Science Direct, Volume-84.
[25] Weiyuan Li, Hua Xu, "Text-based emotion classification using emotion cause extraction", ELSEVIER, 2013.
[26] Kai Gao, Hua Xu, Jiushuo Wanga, 2015, "A Rule-Based Approach To Emotion Cause Detection For Chinese Micro-Blogs", ELSEVIER.
[27] Jasakaran Kaur, Sheveta Vashisht, 2012, "Analysis And Indentifying Variation In Human Emotion Through Data Mining", International Journal of Computer Technology and Applications, Volume-3(6)
[28] Yung-Ming Li, Tsung-Ying Li, "Deriving market intelligence from microblogs", Decision Support Systems 55 (Science Direct), (2013).
[29] Doaa Mohey El-Din Mohamed Hussein, 18 April 2016, "A survey on sentiment analysis challenges", Science Direct
[30] Shikha parashar , R.K. Gupta, 7 –JULY 2017," A Survey on "Citizen Opinion Mining Initiatives In Government Decision Making", IJSART, Volume -3.
[31] Dr Kishore Kumar Das, Affreen Ara, 4, July 2015, "Growth of E-Commerce in India", International Journal Of Core Engineering & Management (IJCEM), Volume 2.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)