IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-6, Issue-1 E-ISSN: 2347-2693

Techniques of Sentiment Classification, Emotion Detection, Feature Extraction and Sentiment Analysis: A Comprehensive Review

Mamatha M[1*], Thriveni J[2], Venugopal K R[3]

[1*]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore, India
[2]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore, India
[3]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore, India

Corresponding Author: mmtha.s@gmail.com, Mobile: +91-9844527012

Available online at: www.ijcseonline.org

Received: 06/Jan/2018, Revised: 10/Jan/2018, Accepted: 25/Jan/2018, Published: 31/Jan/2018

Abstract—Sentiment Analysis(SA) is a way of analyzing the sentiment/emotion in the sentence. Entity, sentence or document-level sentences have been carried out to discover the polarity of text. Sentiment is measured using various approaches such as lexical-based and supervised machine learning methods. This survey focuses on various methods for Feature Extraction and Emotion Detection. Before applying any algorithm for sentiment polarity identification, preprocessing is to be carried over. Machine learning(ML) techniques such as Naive Bayes, SVM, and Max Entropy have been applied to identify the polarity score that are classified as positive, negative or neutral. Feature Selection method identifies a subset of most functional features from the entire set of features. The major challenge in Opinion mining lies in identifying the emotion expressed in the text. This survey provides an insight into the efficient techniques, methods and future scope in opinion mining investigation.

Keywords- Sentiment Analysis; Support Vector Machine; Emotion Detection; Feature Selection; Machine Learning

I. INTRODUCTION

In recent past, an opinion of the customer is an important source for growing business services in any area that is either in the form of text or in the form of review ratings. Meanwhile, when we consider social media, the mobile platform is the most popularly used to view the ratings or comments. The purchase decision of a product is done on basis of the review comments. These reviews and comments reflect on the future business of any product in the market. Consequently, Sentiment Analysis and Opinion Mining have a significant impact on reviews and comments as they perform automatic analysis based on the user request [1]. Analysis of sentiment/ opinion can be applied to audio, visual and image (multimodal SA).

Sentiment Analysis(SA) is an area of Data Mining [2] [3] [4] [5] that deals with opinion oriented natural language data processing. SA is applicable in wide areas such as social media network, movie reviews, stock market review etc. A vital role of SA in all these areas is to perform classification and find the polarity in the text. Due to the growth in text data, text mining is applied to locate hidden knowledge in the text. In domains such as business sectors and movie reviews effort has been made to identify the sentiments and opinions of the customers towards company's products and their services through the expression in the text. However, finding sentiments in large text data is an extremely difficult task. In general, SA has been scrutinized at three different levels:

1) Document level: It classifies the document as containing positive or negative sentiment. For example, given a product, the system finds if the review of the product that is expressed is positive or negative. This is a method of classification and it is done on one entity. This cannot be applied to documents for multiple entities evaluation.

2) Sentence level: Here, it classifies sentences and determines the expression as positive, negative, or neutral sentiment. This analysis relates to classification based on subjectivity that distinguishes sentences based on factual information and one that expresses subjective views and opinions.

3) *Entity/Aspect level:* This feature level performs fine-grained analysis. The sentiment is not expressed clearly in the document and sentence level, hence, aspect level analysis is performed that looks directly at the opinion itself [6].

ML techniques and deep neural network methods are used to find the semantic orientation of text: *positive, negative or neutral*. Most of the common techniques applied to perform sentiment classification(SC) are supervised learning and unsupervised learning techniques.

Contribution: The main contributions of this survey are:

- This paper presents a review of the literature in the various research field of sentiment analysis including Emotion Detection, Feature Selection.
- Large number of articles are categorized based on the techniques used. This helps the researchers to choose appropriate technique for their application.
- Various techniques and algorithms are categorized in brief detail to solve the issues in SA.

Organization: The rest of this paper is organized as shown: Section I contains Introduction to Sentiment Analysis, Section II gives an insight on Opinion Classification of text. Section III describes the Feature Selection method. Section IV introduces Emotion Detection. Section V contains the Sentiment Analysis methods and Section VI concludes the research work with Conclusions and Future Work.

II. OPINION CLASSIFICATION

Text classification is a process of extracting textual data from reviews and classifying them into positive and negative polarity. Classification is performed by applying machine learning methods or Heuristic-based methods [7]. The ML methods involve Support Vector Machine(SVM), Naive Bayes(NB) and Maximum Entropy. K-Nearest neighbourhood, ID3, C5, centroid classifier, winnow classifier, and the N-gram model are the most popular machine learning methods. There are several issues to be overcome in the sentiment classification. Intensive research is required on the following issues:

1. Ambiguity in the text data where multiple sentiments are related to two or more issues.
2. Identification of the most effective polarity in a given document that contains both positive and negative sentiment is a major issue.
3. Achieving successful results is tedious due to lot of noisy statements in the dataset.

To perform machine learning classification two sets of documents are needed, namely, training and test set. To identify different characteristics in documents, an automatic classifier is used along with a training set. Test data set is to evaluate the performance of automatic classifier [8].

A. SVM based Sentiment Classification

SVM is said to be an effective traditional text classification technique. It is a simple linear classification algorithm that tries to find a hyperplane which separates the data into two classes as many points as possible. Classification performed using SVM gives best results when compared to Naive Bayes and max entropy methods.

Tang et al., [9] proposed sentence level segmentation which is known as a joint framework approach and the polarity is classified using segmentation results. This framework uses three models: (i) Candidate segmentation of a sentence is given by the candidate generation model, (ii) a segmentation ranking model to obtain the result of each segment in a given sentence, and (iii) a classification model which predicts the polarity of each of the segments. The joint framework method performs better than the pipelined method. The drawback of the pipelined method is that it suffers from fallacy propagation, because mistakes from linguistic-driven and Bag-of-Words segmentations is difficult to be corrected by the sentiment classification model.

Jindal et al., [10] presented SVM and Decision-tree based methods to prevent electrical theft in power system. This technique is capable of detecting as well as locating real- time electricity theft at every level of power transmission and distribution (T&D). Merit is, it identifies thefts at both dispatch and customer levels and it identifies fraudulent customers and the false positive rate is less. High time complexity in SVM's and lack of memory in large scale to perform quadratic programming are drawbacks.

Michael Gamon et al., [11] used SVM for sentiment classification. By applying huge initial feature vectors and feature reduction based on log-likelihood ration classification can be performed on noisy domains of customer feedback. Based on the linguistic analytic features classification accuracy in sentiment is obtained.

Zhou et al., [12] presented novel calculations for extracting target opinions and opinion outline in microblogs in CMiner for Chinese microblogs. CMiner is easy to learn the views embedded in microblog messages. The disadvantage is that only one target opinion can be found for each sentence through the label propagation algorithm.

Ye et al., [13] incorporated the classification of sentiment from travel blogs in mining platform and compared supervised ML algorithms Bayesian classifier, support vector machines and N-gram model. In terms of accuracy, SVM, and N- gram model achieves performance better than the Bayesian classifier. These strategies can lead to an automatic analysis of the opinion mentality on travel goals. Due to increase in the training dataset size, the distinction is not all that noteworthy and since consumers change their perspectives as often as possible it is good to do a longitudinal review to find the difference between various time periods.

Laboreiro et al., [14] present an approach to text classification that tokenizes Twitter data. The SVM classifier is used for isolating tokens at certain discontinuity characters. With the help of classification-Fbased approach F-measures can be achieved successively, better than the baseline rule-based tokenizer.

Saleh et al., [15] applied SVM for testing datasets of different domains and utilizing several weighting schemes and also explored new research areas. The goal is to check how classification is affected when ML algorithms are applied with different features. Different weighting schemes and N- grams approach is included along with SVM. The performance increases based on corpus size and corpus domain but the investigation of external knowledge that is integrated like SentiWordNet is not addressed.

Moraes et al., [16] use BOG model where document is represented as a vector, Naive Bayes learning method, Support Vector Machine (SVM). The advantages are that the experiments indicate that ANN creates better outcome than SVM's, on the benchmark dataset of movie ratings. Results indicate that an imbalance of data occurs, SVM is not very much hit by noisy terms than ANN and time of preparation of a neural system is significantly much larger than SVM. The disadvantages are that a review between SVM and ANN by including highlights like part-of-speech tags and joint topic- opinion estimations is not addressed.

Mohammad et al., [17] describe two cutting-edge SVM classifiers: Tweets and SMS (message-level task) that recognizes the sentiments in the messages and another recognize the feeling of a term inside the data (term-level task). In both these tasks, the performance is good. A variety of surface-form, semantic, and opinion highlights with opinion-word hashtags, and twitter data with emoticons have been implemented. Initially, a gain of 5 F-score points is provided by the dictionary- based features over all others.

Liu et al., [18] suggested a semi-supervised topic-adaptive sentiment classification (TASC) model on a set of familiar features and labelled information from various topics. Hinge loss is reduced by adapting to unlabeled data topic-related opinion words characteristics and sentiment connections. Topic- specific features dependent on the collaborative selection of unlabeled data are updated by TASC learning algorithm. It helps in selecting more reliable tweets to increase the accuracy. A model is also designed for dynamic tweets along with a timeline(TASC-t). This model is effective for dynamic tweets. The drawback is the difficulty of adapting to topics not predicted and labeled data, and extremely sparse text.

Smailovic et al., [19] applied the active learning approach to analyzing the tweet streams in the stock market. This strategy was adjusted by a stream-based setting by utilizing incremental dynamic learning technique. This gives the ability to select new training information from a data stream for hand-labeling. The disadvantages are that this technique is sophisticated and adaptable with time and tweet concept drift detection and recognition of irony and sarcasm methods are not incorporated.

Canuto et al., [20] solve the problem of classifying the opinion in short text msg by utilizing the information derived from meta-level features that is specially designed to analyze sentiments in these messages such as (i) data got from the sentiment circulation among the *k* nearest neighbors through a small test is recorded as x, (ii) their neighbors have been distributed with distances of x and (iii) the neighbors obtain the document polarity by unsupervised lexical-based strategies. The preferred standpoint is to keep the standard formulation of the computerized classification issue and to adequately and efficiently capture critical data from exceedingly noisy outside domains. The disadvantage is that feature selection technique are not applied to different levels of sentiment analysis.

Wu et al., [21] applied sentiment ontology to perform context-sensitive SA on opinions of online posts in stock markets. This integrates sentiment analysis with ML methods based on SVM and generalized autoregressive conditional heteroskedasticity modeling. The advantage is investors and stock companies are facilitated by investment decision making and risk perception respectively. The accuracy of the sentiment calculation can be improved.

Sharma and Dey [22] suggested sentiment classification approach that uses back-propagation artificial neural network (BPANN) using information Gain, and three popular sentiment lexicons. It is the best approach to deal with quality of BPANN in terms of accuracy with intrinsic subjectivity knowledge that is accessible in the sentiment lexicons. BPANN is not adaptable for classification in applications like Blogs, Twitter and FB.

Wang et al., [23] describe that hashtag-level sentiment classification generates the polarity score automatically for the hashtag given in a certain time period, where it differs in terms of sentence and document-level analysis. The first (i.e. subjectivity) classifier decides if a tweet is impartial or subjective while the next one allocates a subjective tweet with positive or negative extremity. A graph model is used to boost the results that improve the hashtag- level classification which successfully consolidates the tweets data and hashtags co-event relationship. Fig. 1 represents the various sentiment classification techniques that are discussed in detail.

B. Naive Bayes classifier for Sentiment Classification

In machine learning, NB classifier is the best approach for textual categorization and are highly scalable. In the super- vised learning setting, Naive Bayes classifier can be trained efficiently. It uses a method that identifies the probability of different classes based on various attributes.

Trupthi et al., [24] proposed a ML classification method (naive bayes) as in Fig. 1 with different feature se- lection schemes (chi-square algorithm) that is used to obtain a sentiment analysis model for feature extraction, sentiment classification, and sentiment evaluation. The feature extraction result is identified as positive and negative opinions. High Information Features results in higher accuracy than Bigram Collocation during classification. The drawback is that it works for low dimensionality.

Rui Yao and Jianhua Chen [25] describes a simple version of the sentiment-aware autoregressive model. This model holds sentiment analysis along with ML methods that help in knowing the relation between online reviews and performance of box office revenue in movie. To predict box office revenue a classification model (NB Classifier) along with sentiment data is applied. The Autoregressive model can be used to identify the current day's information based on previous day's data. This model is effective for predicting the performance of sales at box office using online data of reviews. The analysis is performed only to limited datasets.

Bhoir et al., [26] used lexicon-based method that holds both positive and negative words and proposed sentiWordNet model to identify the orientation of opinions extracted. The subjectivity result represents that the performance of NB methods is higher than SWN approach. The (i) Customers can easily understand the product and its service (ii) Product quality or services can be improved by the manufacturers. The drawback is that it is applied only on Hindi movies reviews.

Ekawijana et al., [27] say that to solve the sentiment method problem, a model is used in which NB classification assumes that every word is independent in a sentence. Both Naive Bayes Classification and the author's prototype is applied to obtain the accuracy of sentiment score. This is a far-reaching technique that ascertains all letters and images in a sentence and also identifies the semantic meaning. It could become a weakness if they use a poor lexicon wherein a large number of sentences cannot be figured and words in the sentence cannot be calculated.

Jha et al., [28] proposed Hindi Opinion Mining System (HOMS) for movie review. The opinions are identified at the document level and further classified. It uses two approaches: ML method and Part-Of-Speech (POS) tagging to perform this task. To perform POS tagging they applied the Naive Bayes Classifier, and furthermore, the sentiment words are considered as adjectives. This method handles negative sentences efficiently and it relies on the nature of extracted documents from the web page. It can be extended to handle discourse relation. Vociferous content/textual information can be reduced by applying text mining strategies.

Jose et al., [29] say that the sentiment in tweets can be classified by using a combination of both ML classifiers with dictionary-based classifier. This system holds six different modules: i) Acquiring data ii) Pre-processing iii) Classification using SentiWordNet iv) Sentiment Classification using NB v) Sentiment Classification using HMM vi) Classification using an ensemble. Accuracy can be increased in terms of classification by applying ensemble approach rather than the individual classifiers.

C. Hybrid Approach for Sentiment Classification

A hybrid approach is a combination of different classification techniques, such as k-means, max entropy, Naive Bayes and SVM can be combined to obtain better accuracy. This novel approach is applied to classify words, sentences or lexicons into positive, negative and neutral.

Raja et al., [30] present sentiment analyzer correlation with other existing classifiers (NB, Centroid-based and k-NN). The advantages are the performance of existing classifiers at sentence level is better than Bag-Of-Words(BOW) approach from the experimental output on the digital camera review datasets. BOW method can be extended to compute individual reviews weight by identifying the design of descriptive word or adverb and noun phrases combination. The learning method is generally sophisticated to learn the review database and characterize new occurrences precisely.

Andriotis et al., [31] identified the distinct communities over the network and gives approaches and metrics to re- late the internal relationship among the members associated with the network by applying social network analysis (SNA). The grouping is conducted through SVM, nomial naive Bayes(MNB) and maximum entropy. It identifies the interaction among the entities in the friend's list and phone contacts in networking applications. It is fast, efficient, automated and the credentials of the persons under investigation are not considered to generate social mapping of their phones. The aggregation/accumulation of data is not done from multiple social eco-systems.

Wang et al., [32] proposed SentiView, an interactive visualization system that combines the uncertainty modeling and model-driven adjustment by looking for and correlating the frequently occurring textual words. The astrolabe along with the sentiment helix helps to focus on time-varying evolution trend and attributes comparison between individuals. The relationship map on sentiment visualize, the interest and relation between the various participants. The enhancement of river- like diagram visualize and verifies the simulated results based on real information, that is helpful for managing and predicting the opinions effectively of topics on web.

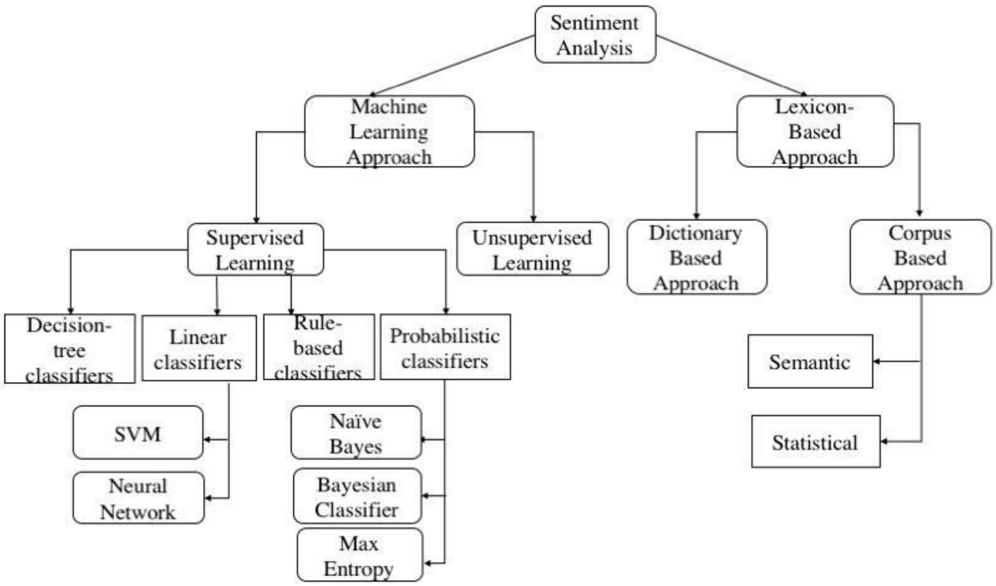

Fig 1: *Sentiment Classification Techniques*

Yang and Dorbin Ng [33] suggested distance-based K- means clustering technique that identifies the themes in a discussion in Web social networks and also the interactions of active participants. It allows the users of the internet across the world to interact among others and exhibit their own views/opinions. The SDC performs better than DBSCAN with both microaccuracy and macroaccuracy. Adaptive techniques have not been used here to balance on configuring density- based clustering.

Troussas et al., [34] chose four classification algorithms to determine the differences and similarities on the student characteristics for comparison, namely the k-means algorithm, k-Nearest Neighbors algorithm, SVM and the NB classifier. The information may be used in order to craft a more appealing curriculum, predict student performance and even draw conclusions regarding social phenomena that may be linked to student performance and attendance. The disadvantage is that data clustering is not used, which improves the efficiency and performance.

Chen et al., [35] proposed a workflow to integrate qualitative research methodology and large-scale data mining techniques using SVM, M3L methods, and Naive Bayes Multi- Label Classifier. It is useful to researchers who are learning analytics, education mining, and learning technologies. It provides an idea for analyzing the media data for the purpose of education. The disadvantages are requirement of human effort to perform analysis and interpretation of data and to identify the prominent themes with many number of tweets in the data.

Haddi et al., [36] describe that ML techniques such as NB, maximum entropy and support vector machine (SVM) are used for sentiment classification to achieve accuracies. Preprocessing technique is used to minimize the noise that is found in text. The classifier performance increases through filtering and unnecessary features are removed using chi-squared method. The disadvantages are that translation of investor opinion into a signal for buying/selling is not represented and it does not explain how investor's sentiments are associated with stock prices fluctuation.

Liu et al., [37] proposed a hybrid approach(NB-SVM) that integrates features of deep learning and shallow learning. It can also realize bilingual text sentiment classification. It introduces the expression of semantic information, and it enhances the capacity of learning and understanding the importance of con- tent. With the expansion of measurements of vector features, the exactness diminishes and the models cost substantially more.

Da et al., [38] presented a method to consequently order the feelings in tweets by utilizing classifier troupes and dictionaries. It is valuable for customers who use analysis of sentiment to look for items. Organizations use this to monitor their brands public sentiment. The advantages are improved classification accuracy and classifier groups built by expanded elements are good to examine sentiment tweets. The study on neural tweets, in which the datasets are improved with simple domain datasets are not considered.

Khan et al., [39] present twitter opinion mining and polarity classification algorithm for twitter feeds classification and performs sentiment analysis based on a hybrid approach along with the steps for pre-processing on the text that is fed to the classifier. The challenges of the past methods are classification accuracy, data sparsity, and sarcasm, incorrect classification of tweets and hence it is classified as neutral. The advantages are it has good accuracy as pre-processing occurs before the text is fed for classification.

Shi et al., [40] proposed sentiment classification algorithm(K-means) that works on Probability Word- List to recognize various opinions of dissimilar threads on the same topic from Chinese Web forums. Higher precision is obtained when compared to SVM and Naive Bayes.

Tang et al., [41] present a technique for learning word embedding for characterization of Twitter information. It addresses by learning sentiment particular embedding of words(SSWE). It encodes data having sentiment in it in a ceaseless representation of words. Three neural networks to viably fuse the supervision from sentiment extremity of text (e.g. sentences or tweets) in their loss functions. SSWE feature performs on par with hand-made features.

Yang et al., [42] suggest PIWI, a novel graph visual analytics model that perform tasks such as community-related graph analysis, identifying the connection among communities, identifying the relationship between attribute-structure, and selecting vertices that has structural features and attributes. PIWI facilitated great perceptions and interactions based on graph community structure better than Node-Link Diagrams (NLDs). The advantage is that a large number of vertices are selected based on structural features and vertex attributes.

Xia et al., [43] give a relative investigation of the viability of ensemble method for classification of sentiments. With a target of productively coordinating feature sets and to incorporate a more accurate grouping strategy ensemble framework is used. The ensemble framework is used for coordinating diverse capabilities but incurs high computational cost.

Jiang et al., [44] focus on target-dependent Twitter sentiment; where, if a query is given, the sentiments of the tweet data are classified into pos, neg or neut. Twitter feeling characterization can be enhanced by 1) incorporating target characteristics, and 2) mulling over related tweets. The problems are investigating the relations between an objective and any of its extended targets. Exploring relations between Twitter accounts for classifying the sentiments of the tweets published by them has not been addressed.

Sokolova et al., [45] present a methodical investigation of twenty four performance measures used as a part of the entire range of ML characterization errands, i.e., binary, multi-class, multi-labelled, and hierarchical. It describes the changes in a confusion matrix to particular qualities of information for every classification. The examination focuses on changes that happen on confusion matrix without change in the measure.

Vo et al., [46] propose estimation driven and standard embeddings, and a rich arrangement of neural pooling capacities to part a tweet into a left setting and a correct setting as per the given target. Here, a three-way sentiment classification is performed. It is independent of external syntactic analyzers and gives better performance compared to the earlier methods that use syntax. It also solves the potential impediment of linguistic structure based method by avoiding the influence of noise by automatic syntactic analyzer. Precision can be accomplished by multiple embeddings, numerous pooling functions, sentiment lexicons and by offering rich sources of feature information which is a disadvantage in this method.

Speriosu et al., [47] explain that there is appeal for computerized instruments that assign polarity to microblog content, for example, tweets (Twitter posts). A label propagation algorithm is connected and this opponents a model directed with in-space clarified tweets. It outperforms the noisy classifier as well as a dictionary-based polarity proportion classifier. Just by viewing the page to be linked, it is easy to identify the polarity by using this method. This is incorporated in label propagation by including nodes for those pages plus edges between them and all tweets that reference them. The drawback is that they do not discover overall gains from using the follower chart as executed here.

Hu and Li [48] proposed two sentence-level sentiment portrayal models, namely positive and negative Topical Term Description Models to perform narrative-level sentiment grouping. It is easy to understand positive and negative context data effectively in a supervised way. It gives a path to develop more effective TTDMs for classifying sentiment. The disadvantage is that the subjectivity of a sentence is to be decided for classification.

Zhang et al., [49] depict that Recurrent Neural Networks (RNN),

TABLE I; COMPARISON OF SENTIMENT CLASSIFICATION METHODS

Author	Algorithm	Advantages	Disadvantages
Jindal et al., [10]	SVM and Decision- tree based	Identify fraudulent consumers and thefts at both transmission and consumer level	High complexity in SVMs´ and lack of memory to perform quadratic programming.
Canuto et al., [20]	Comprehensive Attention Recurrent Neural Networks (CA-RNN)	Captures important information from noisy domains efficiently	Does not perform feature extraction
Troussas et al., [34]	Classification algorithms- k-means, the K-Nearest Neighbors, SVM and the Naive Bayes Classifier	Information may be used to craft a more appealing curriculum, predict and draw conclusions regarding student performance and attendance	Data clustering is not implemented
Da et al., [38]	Classification approach using lexicons and classifier ensembles	Accuracy and classifier ensembles are improved by diversified components that are promising for tweet SA	Datasets that are enriched with analogue domain which holds neutral tweets are not considered
Zhou et al., [12]	CMiner	Embedded opinions in the microblog messages can be easily understood	For each sentence only one target opinion is found

can be used to manage arrangement issue of variable- length sentences. The standard RNN gets the sentence to past setting. The Comprehensive Attention Recurrent Neural Networks (CA-RNN) architecture stores past, succeeding and local contexts of any position in a group. To access past and future information while a convolutional layer is utilized to capture nearby data, the two-way recurrent neural net- works (BRNN) is used. Standard RNN is supplemented by Long Short-Term Memory (LSTM) and Rated Recurrent Unit (GRU), to improve the effectiveness of the new architecture.

Taboada et al., [50] present a lexicon based technique to ex- tract opinion from text. Using the lexicon of words commented on with their semantic introduction (polarity and quality) and consolidating the strengths and negations is done by Semantic Orientation CALculator (SO-CAL). Its performance is reliable crosswise over spaces/domains and on invisible data. For a vocabulary-based approach, a physically constructed dictionary is the solid foundation and it is the one needed to obtain complete benefit from a framework like SO-CAL. SO-CAL performs well across various sorts of reviews, good cross-domain performance; dictionary-based strategies for sentiment examination are robust, and can be effectively changed using numerous knowledge sources.

Davidov et al., [51] proposed supervised sentiment classification approach that depends upon information from Twitter. Automated identification of diverse sentiment types is more advantageous for most of the NLP systems such as review summarization framework, dialogue systems and open media investigation framework. The need for labor intensive manual explanation is avoided by this framework and it permits identification and grouping different sorts of sentiment in short text messages.

Celikyilmaz et al., [52] investigated the effectiveness of the word class-based features on NLU tasks. The representation techniques used for Natural Language Understanding are: (i) User Intent Detection (ii) Semantic Tagging *via* Slot Filling. The advantages of word class-based features are no overhead amid translating and it is task independent. The disadvantages are intent detection and slot filling tasks convey additional overhead either during training or decoding.

Maas et al., [53] present a combination of unsupervised and supervised procedures to understand word vectors capturing semantic term document data as well as rich sentiment information. This method is superior to LDA, which models inert points directly. The unsupervised model is expanded further into sentiment data and demonstrates how it can control plenitude of opinion-labeled texts on the web to yield word portrayals that catch both sentiment and semantic relations. These errands involve relatively simple sentiment information and the model is exceedingly adaptable. It can be applied in areas where opinion analysis and recovery is performed and is also used to portray various annotations.

Dragut et al., [54] describe that sentiment dictionaries have a lot of inaccuracies in them. Other than the same word having different polarity in various lexicons, other cases of inconsistencies are reduced. The authors have proposed the use of 2 fast SAT solvers to discover irregularity in the lexicon. The advantage is that the inconsistency across dictionaries and within individual are identified. The issues that are not solved are the reason behind the inconsistencies is not found. Finding the correlation rate in a dictionary after repairing the inconsistencies is not done and the inconsistencies are not found based on subjective and objective word senses. A comparison of various recent Sentiment Classification techniques is shown above in Table I.

III. FEATURE SELECTION

Feature Selection (Content-free features) is divided into 4 different categories:

A. Syntactic Feature

It uses any one of the tags among parts-of- speech(POS)/word tagging, N-grams, phrase patterns or punctuation.

B. Semantic Feature

Its focus is on identifying the relation between words, phrases, signs and symbols. Score-based technique is applied together with semantic features. Using such approaches message sentiment are classified on the basis of title sum of comprised positive and negative semantic features.

C. Link-base Feature

Link or citation analysis is to look for sentiment in Web artifacts and documents. Using the relations and link that lie among them, link base samples are classified.

D. Stylistic Feature

This is used by artist's in trying to pass a message to them. Use of symbolism: Here is where the artist uses symbols to describe, represent or characterize a person, thing or place.

Content-specific features and without content are extracted from the text data collection. To extract the features, it needs to perform POS tagging on collected information. Sentiment score can be calculated after performing POS tagging, where every word is obtained by looking up in a sentiment-based lexicon.

Chen et al., [55] proposed a RESOLVE (Ranking Emotional Synonyms for language Learner's Vocabulary Expansion) method for second language learners who are not be able to express the emotion of the sentence. The performance of RESOLVE is evaluated in comparison to commonly-adopted mutual information and machine learning algorithms for classification. The students of low and intermediate proficiency are benefited more from this method than the highly-proficient students. The drawback is that word sense disambiguation is not performed which is to be done before pattern extraction.

Some of the issues in Feature Selection process are noted below:

1. Large feature space problem: This causes downgrade in performance due to computational problems and hence, effective feature selection techniques are necessary.
2. Redundancy: N-grams are highly redundant. Thus, problems are caused in univariate and multivariate methods.
3. Domain Dependency: Clustering based feature extraction technique performance is domain dependent, and thus creates cross domain and generalization problems.

TABLE II- COMPARISON OF FEATURE EXTRACTION METHODS

Author	Algorithm	Advantages	Disadvantages
Chen et al., [55]	RESOLVE (Ranking Emotional Synonyms for language Learner's Vocabulary Expansion) method	Students of low and intermediate proficiency find this method effective than highly-proficient students	Word sense disambiguation not performed
Tang et al., [56]	Learning sentiment-specific word embeddings	Sentiment embeddings are verified on word level, sentence level and lexical level	Applicable only for words present in the dictionary and not new words
Celikyilmaz et al., [61]	Model for the effects of the word class-based features	Feature sets are automatically generated, task independent and consistently improves the performance of the tasks	Class based features examination for Deep Neural Networks and long short term memory networks not implemented
Poria et al., [62]	Multimodal sentiment analysis method	Performance analysis of sentiment on text is good and visual features outperform the state-of-the-art	Gaze- and smile-based features is not considered and it is not culture- and language-independent
Agarwal et al., [63]	Information gain (IG) and Minimum redundancy maximum relevancy (mRMR) methods	Extracts basic, composite and Prominent features effectively, accuracy is good	Highly computationally expensive for feature selection

4. POS tagging problem: Accuracy of heuristic based feature selection method depends on POS tagging accuracy. Hence, designing an efficient POS tagger specially for languages other than english is a major issue.
5. Limited work on Lexico: structural features are carried out unlike syntactic and semantic features.

Tang et al., [56] suggested learning sentiment-specific word embeddings dubbed sentiment embeddings to overcome the problem of existing word embedding learning algorithms using only word context yet they disregard the feelings in the text context. Hence, the words with opposite sentiment polarity, for example, great and awful, are drawn to neighboring word vectors. Advantage is that sentiment embeddings are verified and analyzed on 3 stages: (i) On word level SA (ii) On sentence level classification, and (iii) On lexicon level errand, i.e., constructing sentiment dictionary. It deals only with words which are there in the dictionary.

Schumaker et al., [57] proposed the Arizona Financial Text (AZFinText) framework which is a news article prediction for characterizing the sentiments as obstinate/subjective or verifiable/objective. It can automatically recognize the terms used in the financial news article and emotional cues. A price prediction model is built for evaluating news articles in future. The advantage is that the subjectivity of the articles may affect the trading behavior and for investors reacting strongly to negative articles, the outcomes are attributable. The disadvantage is that the role of verbs and adverbs as a text representation strategy is not investigated for better predictivity and the negation in the Opinion Finder tool presents a little amount of noise into the results that impacts price direction.

Ren et al., [58] proposed a framework known as Social and Topical Context incorporated matrix factorization (ScTcMF) system. This structure performs well for best in class collective filtering strategies, and demonstrates that social context and topical context are viable in enhancing the user-topic sentiment prediction execution.

Yi et al., [59] suggested Sentiment Analyzer(SA) to obtain feelings about a subject from online text documents. SA determines the emotion in each reference utilizing the strategies of NLP by identifying all references to the given subject, SA sentiment lexicon and pattern database for the analysis. SA reliably exhibits looks for subject related characteristic terms from review articles in web, allowing finer granularity. The drawback is that human expertise could be unnecessary in the formulation to deal with the semantics correctly.

Asghar et al., [60] discusses about the vast methods that already exists in SA to perform feature extraction. The impediments are feature space reduction, redundancy removal and evaluating performance of hybrid methods of feature selection. Various approaches of feature selection used are NLP based, ML or cluster based, Statistical and Hybrid.

Celikyilmaz et al., [61] recognizes the impacts of word class based components for the group of models concentrating on NLU undertakings. This system exhibits that the change in execution can be ascribed to the regularization impact of the class-construct features on the underlying model. The task performance of feature sets is improved consistently across several modeling methods investigated. The feature sets are automatically generated and there is no overhead during decoding and it is task independent. The examination of language independence to accomplish similar gains for classification and tagging tasks in other languages and the examination of class based features for Deep Neural Networks, particularly recurrent neural networks and long short term memory networks needs to be implemented.

Poria et al., [62] suggested a technique to perform multi- modal SA, that identifies emotions in Web videos by exhibiting a model that uses audio, visual and textual modalities as sources of information. To integrate the data collected from many models, both feature and decision-level fusion approaches are used. The textual sentiment analysis performance is good and the visual features better than the new methods. Analysis on gaze and smile-based features are not performed and is not applicable for culture and language-independent multi-modal sentiment classification framework.

Agarwal et al., [63] describes sentiment classification using

Unigrams, bi-grams and dependency features from textual data. The bi-tagged features using various feature selection procedures, i.e., information gain (IG) and minimum redundancy maximum relevancy (mRMR) eliminates clamorous and inappropriate characteristics from the vector. These methods extract basic, composite and prominent features effectively and the accuracy obtained is good. The IG and GA(Genetic Algorithm) methods are computationally expensive for feature selection.

Pasarate et al., [64] suggests feature extraction techniques to Approach extracts features of candidates using syntactic rules. Domain specific and not overly generic opinion features are identified *via* the intercorpus statistics IEDR criterion.

Duric et al., [65] suggested strategies for feature selection that entities reviewed from subjective expressions that portray those elements in polarity terms. Generalized Iterative Scaling (GIS) and Improved Iterative Scaling (IIS) algorithms are applied. Statistical approaches are fully automated but do not distinguish features that carry opinions. The problems are classification of review documents based on a scale is not addressed as the focus is on the binary classification task.

Li et al., [66] suggested domain adaptation method for sentiment and topic-lexicon co-extraction without using labeled data. The system is twofold. Initially, some high-certainity opinion and topic seeds in the target domain are generated and next, a Relational Adaptive bootstrapping (RAP) algorithm approach is used to enlarge the seeds in the target area by exploiting labeled source domain data and the relationships between topic and opinion words. Topic lexicons and precise sentiments are extracted from the target domain. Further, the sentiment lexicons that is extracted can be effectively applied to perform classification. Homogeneous relationships across topic and emotion words and polarity identification of the extracted sentiment lexicons have not been addressed.

Uuguz and Harun [67] develops text categorization by reducing the dimensionality of the feature space to improve performance. The feature extraction uses: (i) information gain(IG) method ranking every term in the document is ranked. Next, to the terms classified in descending order, (ii) genetic algorithm (GA) and (iii) principal component analysis (PCA). Feature selection and extraction is applied on highly important terms resulting in low computational time. Since it is two- stage, a high classification performance is acquired with less features in text categorization.

Elawady et al., [68] resolves the difficulty in high dimensionality of feature vector and sentiment classification using Machine learning(ML) methods viz., SVM and NB classifiers. mRMR removes redundant features which are exceptionally associated among each other, and retains pertinent features having least correlation. A large portion of inapplicable and noisy features are reduced by Rough Set Attribute Reduction(RSAR). An integrated approach removes the unnecessary features and obtains minimal characteristics set with low time complexity for classification.

Abbasi et al., [69] built Entropy Weighted Genetic Algorithm

overcome the importance of the word as it can be distinctive in various circumstances. The techniques are: i) Total Weighted Score Computing Method used for predicting semantic orientation of reviews. ii) Neutral/Polar/Irrelevant Classification Model where preprocessing and classification is done on the tweets as positive and negative. iii) Weighing and Aggregation Scheme which is applicable on document level and entity level classification using SentiWordNet. iv) Intrinsic and Extrinsic Domain Relevance

(EWGA) for effective feature selection. This was applied to expand precision and locate the key components in each of the sentiment class. EWGA significantly out performs the *no feature selection baseline* and GA on all test beds. The method of utilizing elaborate and syntactic features together with EWGA feature selection method achieves a high level of accuracy. The effectiveness of different forms of GA hybridization, such as using SVM weights are not addressed.

Agarwal et al., [70] focuses on identifying good feature sets for sentiment analysis. Concept extraction algorithm is based on a novel concept parser scheme to extract semantic features that exploit semantic relationships between words in a natural language text. Minimum Redundancy and Maximum Relevance feature selection approach was utilized. This approach outperforms other existing sentiment analysis approaches. A supervised fuzzy learning algorithm could be used to enhance the efficiency of the system.

Matsuyama et al., [71] proposed an automatic expressive opinion sentence generation mechanism for pleasant conversational frameworks. From the number of reviews on the internet, opinions are extracted and it is positioned in terms of contextual relevance, length of sentences, and measure of data spoken to by the recurrence of descriptive words. It is the best mechanism for finding the dynamic topics in written documents.

Shelke et al., [72] depicts that given a gathering of review data on text, idea is to recognize individual product viewpoint remarks by analysts and identify if they are positive or negative. The unsupervised learning technique is used to identify feature specific opinions. It is domain independent, wherein it works with same efficiency in any application.

Wang et al., [73] suggested feature-based vector method and weighting algorithm to perform SA of Chinese item reviews. This model classifies reviews into positive and negative, and represents strength of the sentiment by adverb of degree and it also modifies the relationships between words and punctuations in review texts. The drawback is that this model is domain dependent.

Xianghua et al., [74] presented an unsupervised approach, Multi-aspect Sentiment Analysis method based on LDA for Chinese Online Social Reviews (MSA-COSRs) to identify aspects in Chinese social audit, and it finds the opinions indicated in them. This strategy distinguishes the part of the nearby point by feature extraction. This approach gives better performance in topic identification as well as sentiment analysis. Picking an appropriate topic number to prepare the LDA demonstrate is hard. This could have been resolved using hierarchical Dirichlet process.

Yu et al., [75] describe that to extract the frequent patterns of human interaction the flow of human collaboration in a discussion is in the form of a tree. Hence, tree-based interaction mining calculations are intended to investigate tree structure and extract interaction stream patterns. Tree-based interaction mining approaches are intended to break down the tree structure and extract interaction flow patterns. The mining results obtained would be helpful for examination for comparison of meeting information, summarization, and indexing. The burden is that the present meetings are task oriented. Embedded tree mining for hidden interaction pattern discovery needs to be explored.

Rong et al., [76] suggest Auto-Encoder based Bagging Prediction Architecture (AEBPA), a bootstrap process, an auto-encoder-based prediction model, an aggregating process and embedding-boosted auto-encoder to perform SA on reduced dimensionality. The auto-encoder-based model reduces dimensionality and high request characteristics from raw representations is extracted in an unsupervised manner. It is beneficial to utilize extra information to train a better embedding by utilizing more experimental datasets. It is better to use other optimization techniques, for example, Newton and LBFGS.

Alshari et al., [77] investigate the difficulty in predicting rating based on user's comments by applying information retrieval and the Vector Space Model. An investigation is done on the impact of integrating SA for the rating forecast. Merits are that the ratings and comments can be used for making purchasing decisions. It is difficult to predict the rating of products based on comments and does not include advanced NLP techniques.

Giachanou and Crestani [78] explored conventional time series analysis strategies and their appropriateness on topic and sentiment trend investigation. A feature based method is used to predict the sentiment changes and Time Series Decomposition technique is applied to get better understanding of data. The opinions have a tendency to take up a contradictory conduct implying that when positive sentiment pattern is expanding, the negative is diminishing and the other way around and peaks in subjects' popularity and in the feelings. No methods are implemented for automatic detection of sentiment change and the reason behind it.

Krouska et al., [79] describe the required data to obtain preprocessed evaluation in order to identify the sentiment and analyze whether it is positive or negative. The role of text preprocessing in opinion and an examination of sentiment polarity categorization techniques for Twitter content are ad- dressed. Classification performance is affected positively by feature selection and representation.

Liu et al., [80] suggests method to identify product features using Latent Semantic Analysis (LSA). This model is designed and developed in the mobile environment for movie-rating and review-summarization. The accuracy of classification and response time of the system is considered to build the system. This method uses a factual approach to identify sentiment phrases; the rating and review-summarization framework can be reached out to other product-review domains easily. It utilizes a LSA-based filtering mechanism that enables the clients to select features in which they are intrigued and the outline size can be reduced efficiently using this mechanism.

Qiu et al., [81] proposed double propagation method based on bootstrapping which propagates data between opinions and targets. It performs the task in which many syntactic relations exist that connect opinion words and targets. An initial lexicon is used to begin the process of bootstrapping and works on semi supervised approach. It solves the problem of sentiment lexicon extension and sentiment target extraction. Accuracy of feelings word extraction can be enhanced by addressing issues on sentiment word pruning techniques and noisy target pruning.

Wang et al., [82] proposed Fisher's discriminant ratio, an effective feature determination strategy for subjectivity content opinion classification along with Support Vector Machine classifier. As with rapid rise of internet and with people expressing their opinions, attitudes and feelings, this method is being useful. So the subjectivity texts must be classified into various sentiment orientation categories automatically. The advantage is it quickly classifies and organizes on-line reviews automatically. The disadvantage is Sentiment Analysis is not performed on vocabulary, syntactic and semantic words.

A comparison of recent Feature Extraction methods is shown in Table II.

IV. EMOTION DETECTION

Emotion Detection is a classification task. It is an inclination that portrays the state of mind. For example: happiness, outrage, love etc. To separate feelings, physiological characters such as voice, outward appearances, hand signals, body movements, pulse, circulatory strain and text data are used [83]. Techniques of Emotion Detection is partitioned into dictionary based methodologies and machine learning methods. There are four types of emotion recognition methods in text:

A. Keyword Spotting Method

It deals with the identification of keywords to be searched in text. The input is a text document and the output is generated as an emotion class.

B. Lexical Affinity Method

This approach assigns arbitrary words a probabilistic 'affinity' for a particular emotion. It is more sophisticated than keyword spotting. The probabilities assigned by this method are part of linguistic corpora.

C. Learning Based Approach

It tries to recognize the emotions based on previously trained classifier/results, which map with various machine learning classifiers such as SVM.

D. Hybrid Based Approach

It is a combination of machine learning and keyword spotting methods collected from training datasets.

The basic emotion types are: Satisfaction, Trust, Fear, Surprise, Sadness, Disgust, outrage and Anticipate. These act as a base to find other advance emotions.

E.g.: Basic Emotion + Basic Emotion = Advance Emotion

Joy + Trust = Love

There are several issues that occur in the Emotion Detection technique. Few issues are as listed below:

1. Sentiment is Domain Specific: The word meaning changes based on the context they are used in.
2. Negation Handling: Handling negation in SA can be tricky.
3. Spam Sentiment Analysis: Refers to fake reviews that mislead potential readers by giving inappropriate opinions to the object.
4. Multiple Opinions in a Sentence: Single sentence may contain multiple opinions along with subjective and factual portions.

Brown et al., [84] demonstrate that predicting a user's task is done accurately and also requires some user personality attributes by applying ML approach to deal with interaction data. Here, an approach is followed where the members perform a visual inquiry undertaking and to the three encodings of the client's collaboration information well-known machine learning algorithms are applied. This can be extended for evaluating additional personal traits.

Liu et al., [85] propose emotional tendency investigation approach based on the context view to tackle the problem of limited customary emotional examination techniques on micro-blog and ignoring the context information. Feelings word reference of micro-blogs are established and favored expressions are added in the lexicon and classified as emotions section, text group and the context related to a tweet. The advantage is that along with finding the sentimental tendency of micro-blog texts, context information is also taken into consideration.

OF́laherty et al., [86] traces the improvement of a cloud- based self-benefit Predictive Decision Support System (PDSS) for customer retention and endeavors to position the results of this implementation in the context of decision support, information administration and work practice evolution. The advantages are robustness and appropriateness for advancement of making decision from unstructured to organized (i.e. stock control administration).

Poria et al., [87] web opinion examination includes a clear understanding of natural language text by machines. To this end, concept-level SA extends from word-level analysis of text and gives way for sentiment mining. SA empowers a more productive passage from (unstructured) text data to (structured) machine-processable information. The disadvantage is that unstructured information is produced for human consumption which is not directly machine processable. The limitation in concept-level SA and the ELM classifier's accuracy is less than the one obtained by means of sentic patterns.

Munezero et al., [88] presented opinion mining to differentiate between subjective terms and how they relate to each other. To solve such inconsistencies, the author tries to clarify the differences between the five subjective terms: affect, feelings, emotions, sentiments and opinions. Opinion Mining(OM) method is used to find the subjective terms. The difference between the terms- affect, feelings, sentiments, emotions and opinions and understanding of how they relate to one another is made clear. The analysis is not performed on attitude, mood, sensation, and temperament subjective terms.

Sintsova et al., [89] proposed a distant supervision(Dystemo) method Balanced Weighted Voting(BWV) to produce sentiment classifiers automatically from twitter data which are labeled using easy-to-produce emotion lexicons. It improves the micro F1-score and introduces heuristics to automatically find neutral tweets. The drawback is, it identifies only emotions and not the sentiment in the text.

Su et al., [90] presents a method capable of deriving recommendations for improving the product quality and in- crease the products market competition by using Social Media- based Product Improvement Framework (SM-PIF). The recommendations obtained are more clear and less biased than classical methods due to the nature of big data. Statistics for product improvement is gathered in an innovative way through gathering, sifting and inspecting applicable data. The opinionated resources are also gathered to take decisions easily. The disadvantage is that the battery of the SM-PIF needs to be improved.

Quan et al., [91] says that a Latent Discriminative Model (LDM) can efficiently perform the task of extracting the feelings of readers through online documents. By considering emotional dependency

TABLE III- COMPARISON OF EMOTION DETECTION METHODS

Author	*Algorithm*	*Advantages*	*Disadvantages*
OFlaherty et al., [86]	Cloud-based self-service predictive decision support system (PDSS)	Improved robustness and appropri- ateness and decision-making evolu- tion from unstructured to structured	Specific to a single knowl- edge domain
Sintsova et al., [89]	(Dystemo) Distant Super- vision method	Improves the micro F1-score and introduces heuristics to automati- cally find neutral tweets	Identifies only emotions and not the sentiment in the text
Munezero et al., [88]	Opinion Mining(OM)	Inconsistency between the terms are solved	Analysis is performed only on the specified subjective terms
Su et al., [90]	Social Media-based Prod- uct Improvement Frame- work (SM-PIF)	Gathering of information to im- prove the product through collect- ing and filtering and Relevant in- formation is examined easily	Battery of the SM-PIF is to be improved
Quan et al., [91]	Latent Discriminative Model (LDM)	New Emotional Dependency-based LDM (eLDM) gives good results than LDM	Imbalance problem(i.e.., the variation of votes count received for each emotion)

with LDM, a new Emotional Dependency-based LDM (eLDM) is obtained. Limitation here is that imbalance problem (i.e.., the variation of votes counts received for each emotion) is not considered.

Kouloumpis et al., [92] investigates how sentiment features are detected using linguistic features on Twitter messages. A supervised approach was used, but existing hashtags in the Twitter data for building training data is also grasped. The drawback is that by including microblogging features, the benefit of emoticon training data is reduced.

Perez et al., [93] suggested an approach for multimodal classification of sentiments by introducing a multimodal dataset, that identifies the sentiment in language proficiency visual data streams. The benefit here is that multimodal sentiment analysis can be performed in sentiment annotated utterances extracted from video reviews and audio-Visual Emotion Analysis. It works only on multimodal sentiment classification and not on multimodal fusion methods which is integration of the visual, acoustic, and linguistic modalities.

Bollen et al., [94] says that text content analysis of twitter feeds is done in view of two mood tracking tools- Sentiment identifier that measures positive versus negative mood and Google-Profile of Mood States (GPOMS) that computes mood in 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). From the content of large-scale Twitter feeds, the changes in the public mood can be tracked by a simple text processing approach. This improves the performance of the most basic models to anticipate Dow Jones Industrial Average(DJIA) closing values. Decisions on whether positive or negative and happiness or satisfaction are relatively expensive and time- consuming.

Solakidis et al., [95] considers two emotionally rich features: emoticons and lists of emotionally intense keywords and evaluates them by applying a semi-supervised approach which has prompted two new procedures for collecting training data automatically without human intervention. Feature vectors with particular pos-labels are used as adjectives and adverbs. Specific punctuation imprints, special characters and combination of emoticon and keywords are not evaluated.

Wang et al., [96] proposes a Joint factor graph model to address the problem of correlation between different emotions to understand monolingual and bilingual information. The challenge here is to explore bilingual information of each post and connect different emotions in a single post. A framework is proposed to predict emotions in code-switching text that considers both bilingual and emotion data. A comparison of various recent Emotion Detection techniques is presented in Table III.

V. SENTIMENT ANALYSIS

Sentiment Analysis focuses on knowledge discovery from text neural networks are applied to solve the problems in Opinion few of the major issues in SA. Some of the major issues are:

1. Grammatically Incorrect Word: Implementation to identify grammatical errors is not done. Hence, SA results can be improved by mapping such errors to correct words.
2. Incremental Approach: It permits updating the existing results by using only individual data instances, without re-processing the past instances. This can save the reanalysis time taken.
3. Ironic and sarcastic statements: Identifying irony and sarcasm from expressions in text is a difficult task.
4. Improving the precision of algorithm: Opinion detection output depends on accuracy of the algorithm. An algorithm that reduces human effort and increases accuracy is essential.

Lu et al., [98] proposed a visual analytics approach for event prompting, utilizing media information. A time series intervention model was used to test when discourse develops over time. It checks whether the level of framing is distinctive earlier or after a given date. This intervention method can be applied to any temporal variable and abnormality detection strategies. The intervention models frequently have an extensive false positive rate. The disadvantage is that anomaly models together with intervention models is not used. A blend of SA, frames and clustering for defining geo-political regions that offer basic surrounding systems could have been addressed.

Asur et al., [99] proposed a model to show how web-based social networking content is utilized to anticipate real-world outcomes. It demonstrates the way in which extracted opinions from Twitter can be used to enhance the forecasting power of social media. To figure box-office incomes in the film industry, the chatter in Twitter.com is used. It is illustrated to show how the estimating power of media content results are based on only twitter dataset which does not seem to be efficient.

Clavel et al., [100] proposed the integration of SA in direct human-agent interactions. Semantic rules and ML approaches are combined using (i) the multimodal nature of sentiment- related phenomena, (ii) the variation in temporal and decision frames, (iii) varying levels of complexity required by the timing constraint of the inter-action, and (iv) the heterogeneity of the contextual information. SA methods are investigated as input and output of both small and big-term strategies. To deal with Automatic Speech Recognition(ASR) outputs is difficult.

Kang et al., [101] propose senti-lexicon and enhanced Naive Bayes algorithm for SA of restaurant reviews. When the algorithm uses unigrams and bi-grams as features, the difference in performance between positive and negative is small in the

and information retrieval. This is the final stage of identifying sentiment in text data by analyzing the given information. The data is collected from reviews, online e-learning systems, discussion groups, forums and blogs. Sentiment Analysis performs Classification, Feature Selection and Emotion Detection. Emoticons are extracted if present in the data and then it is processed. Domain dependency is a major issue is SA. Working on domain specific data gives accurate results rather than domain-independent data. Deep learning models such as deep neural networks and convolutional

Mining[97]. Word ambiguity, Sarcasm detection and handling are

experimental results. Features are extracted by applying the part-of-speech tagger in every language. Review analysis can be performed in any language but review analysis in multiple languages is a difficult task.

Psomakelis et al., [102] develop the BOW, n-grams and n- gram graphs techniques. For each method, the authors evaluate the accuracy of a dictionary-based and 7 learning-based classification algorithms (namely SVM, Naive Bayesian Networks, Logistic Regression, Multilayer Perceptrons, Best-First Trees, Functional Trees and C4.5) as well as their combinations, using a set of 4451 manually annotated tweets. Superiority of learning based techniques are demonstrated through results particularly with n-gram graph approaches to predict opinion in tweets. This represents that combinatory method shows good effects on n-grams.

Nguyen et al., [103] proposed SVM with linear kernel to foresee stock value movement using the sentiment from online networking. The drawback is the number of topics and sentiment specified beforehand for the LDA and JST-based method. A non-parametric topic model is used to overcome this weakness that can derive the quantity of themes and opinions naturally. The limitation is that only the historical prices and opinions derived from web-based social networking are considered.

Krening et al., [104] examines a software agent that learned to play the Mario Bros. game using explanations. The object focused advice and object focused-Q learning is used to classify the explanation into clauses. The negative sentences (false negative) are split into clauses and reclassified and hence the sentiment filter accuracy increases. The drawback is that free- form explanations can be processed by sentiment filter but a number of sentences are not significant and cannot be directly utilized as advice.

Li et al., [105] proposed Supervised User-Item based Topic(SUIT) incorporating the user and item information for SA. This model acts as an integration between supervised LDA (sLDA) and Probabilistic Matrix Factorization (PMF) and achieves efficient outcome and shows the importance of both content information and user-item information. This model can be used in applications, such as recommender system, computational ads. It is not an efficient method for large-scale data sets.

Li et al., [106] uses SA and text mining methods to study the online forums hotspot detection and forecast. Results show that SVM forecasting achieves highly consistent results with K-means clustering. Results of top 4 hotspot forums of the year are the same for both support vector machine forecasting and K-means. Hotspot

prediction may not be accurate on past information. This is because many of the hotspots are emergent events that has no correlation with past hotspot history. A practical system, which is of a form of a web site portal needs to be designed and developed.

Lima et al., [107] proposes a framework for Twitter messages

that combines approaches and automatic contextual module. This framework is known as polarity analysis framework. Various classifiers such as, NB; SVM; Decision Trees(J48); and Nearest Neighbors(KNN) have been considered. It provides a good method to perform polarity analysis of twitter messages automatically with good performance levels. The disadvantages are that the assignment of weights towards emoticons is not done automatically in order to establish classification priorities. Semi-supervised learning algorithms might be better for the classification module.

Chen et al., [108] set an objective to develop a prototype that extract and analyze the sentiments of customer's product reviews. In positive and negative analysis, the precision of the system gives best accuracy. It can be improved by taking into consideration more grammar orientation of a sentence and the output of the system can be made into more interactive to users.

Yan et al., [109] propose a new feature-level Sentiment Analysis and extended PageRank algorithm to extract the features of products and construct expandable, context-dependent sentiment dictionary for online product reviews. Consumers would discover the feature-oriented and context-dependent review SA easy to understand others opinions on particular elements as the opinion terms have distinctive polarities in various contexts. The dataset is very limited and does not consider the computational complexity. In all the three methods,
the performance is evaluated based on precision of sentiment analysis. Fig 2 represents the various stages in Sentiment Analysis.

Li et al., [110] introduces the clustering-based SA approach. Stable clustering results can be obtained by applying a Term Frequency-Inverse Document Frequency(TF-IDF) weighting method, a voting mechanism and importing term scores. The methodology has competitive advantages over symbolic techniques and supervised learning methods. It is a well- performed, productive and non-human taking part strategy to deal with the difficulties in SA. With respect to accuracy, human participation and efficiency
the performance of clustering- based SA is the most balanced. The disadvantages are results are not stable and the report measure that is set may impact the result in terms of clustering- based opinion mining technique.

Ghag et al., [111] discusses about the different techniques in Sentiment Analysis. For polarity identification, SA could be carried out by techniques that might use a lexicon. Some opinion analyzers could be language dependent or language independent, likewise called as multilingual feeling analyzers.

Smallovic et al., [112] developed an active learning approach which was applied to tweet streams in the stock market domain to perform SA. The results are improved using SVM classifier. The study shows that examination of mind-set of the public that is derived from Twitter feeds can be used to eventually forecast movements of individual stock prices. The SVM neutral zone gives the capacity to order tweets into the neutral class. Sophistication and versatility along the timeline should be improved.

Wei et al., [113] proposed the text semantic orientation analysis method that works on Hidden Markov Model(HMM). HMM is developed on the basis of markov chain. This method is applied to improve accuracy of network public opinion orientation analysis. The HMM model approach has better recall rate and precision rate. The algorithm analysis speed is to be improved to ensure the classification efficiency.

Wang et al., [114] presented SentiView, an interactive visualization system that analyzes the sentiment of the public posted text on media such as forums. It predicts the short-term trend of the sentiments about events being talked about. It examines open opinions for popular subjects on the web and is adaptable for various networking platforms; For example, Twitter, blog and forum. It has not considered performing the simulation of incomplete or future data using the sentiment prediction model.

Calais et al., [115] suggests that real time SA is performed by a topic-based real time sentiment analysis by using transfer learning approach. Consistency in human bias enables robustness and real time analysis of opinions with unlabeled information on topics in which polarization among user opinions occur is demonstrated. Strategy to contexts with lower degrees of polarization is not evaluated. Theoretical guarantees regarding performance based on the level of polarization among sentiment holders are not derived.

Trilla et al., [116] investigate expressive speech synthesis where sentiment analysis procedure is applied as an input feature. To identify the correct adaptation procedure, distinctive combinations of text features and classifiers are evaluated. The merit is that considering the most relevant unigrams gives good classification adequacy compared to using additional features, for ex, bigrams, POS tags, stems, synonyms, emotional dimensions and negations. The training data size is very small and evaluation of results is to be done using TTS system, with different languages. It considers a temporal examination for the development of a conversation.

Gonccalves et al., [117] describe that messages express opinions about events, items, and services, political perspective or the

TABLE IV- COMPARISON OF SENTIMENT ANALYSIS METHODS

Author	Algorithm	Advantages	Disadvantages
Lu et al., [98]	Visual analytics framework for event cueing using media data	Can be applied to any temporal variables, anomaly detection methods, intervention models and others	Combination of anomaly models and intervention models is not used
Clavel et al., [100]	comparative state of the art and sentiment detection methods	A psycholinguistic model is used to characterize human-agent affective dialogs	Dealing with ASR outputs is difficult
Psomakelis et al., [102]	BOW, n-grams, n-gram graphs and 7 learning classification algorithms	Demonstrates the superiority of learning-based methods	Outcome quality can be enhanced using new techniques
Nguyen et al., [103]	SVM with the linear kernel	Both topic and sentiment are identified	Only sentiments and historical prices derived from social media are considered
Krening et al., [104]	object focused advice and object focused-Q	False negative sentences are reclassified to obtain positive polarity	Sentiment filter processes only free-form explanations

emotional state and mood of author. A web service called iFeel provides an open API to access and compare the results over various strategies of sentiment for a given content. The combination of these methods relies on the following: PANAS-t, Emoticons, SentiStrength, SentiWordNet, Sentic- Net, SASA and Happiness Index. Eight different methods are used to obtain the result. The results are accurate and the comparison of results is easy through the API iFeel. The limitation is that input text strings are used which can be extended to use input files.

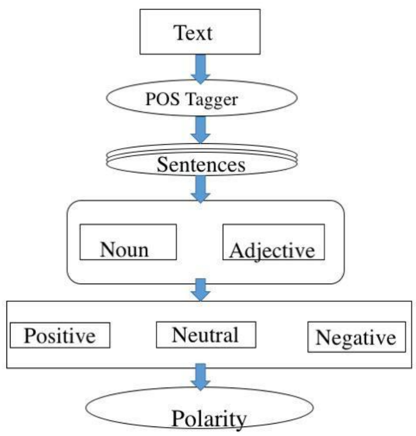

Fig. 2. Flow Diagram of Sentiment Analysis

Yu et al., [118] proposed a sales prediction model, known as ARSA(Auto Regressive Sentiment-Aware) model. This is used for solving mining review problems from movie domain for predicting product sales performance. The accuracy of prediction is further improved by presenting ARSQA model, to use the sentiments and quality to predict the performance of product sales. The outcome of ARSA prompts noteworthy learning that can readily be utilized by decision makers. It is very common and can be connected to different situations other than the Sentiment- Probabilistic Latent Semantic Analysis(S-PLSA) model. The role in clustering and grouping of surveys in view of their opinions is not available.

Hiroshi et al., [119] proposed another model for performing interpretation from text documents to a set of sentiment units. For machine translation the deep language analysis methods are applied to the task of text mining. A high-precision SA system is developed that incurs low costs. This uses an existing transfer-based machine translation engine. The precision of the sentiment polarity is much higher than for the conventional methods. The sentiment units created by this system were less redundant and more informative than when using naive predicate-argument structures.

Dermouche et al., [120] applies a topic model for topic-specific opinion modeling from text to solve the issue of joint topic-sentiment modeling. Gibbs sampling process is derived based on the inference algorithm. The experiments performed on two review datasets show that this model outperforms the widely used ML-based methods in terms of predicting topic- specific sentiments. The overall topic-specific distribution is done over sentiment polarities.

Claypo et al., [121] describes automatic keyword selection to select subsets of keywords for SA utilizing the data of class dependency and dissimilarity. The shared data and distance among the classes for choosing the subsets of catchphrases is done easily. Since the dataset includes 3,073 keywords, it is difficult for classifiers to classify them.

Jongeling et al., [122] examined whether the investigation devices concur along with sentiments identified by human evaluators and among each other. While conducting a study on software engineering, the effect of selecting SA tool is learned. This leads to contradictory conclusions as there is a necessity for sentiment analysis tools whose target is towards the software engineering domain.

Fang et al., [123] proposed a multimodal Aspect-Opinion Model (mmAOM) that solves the multimodal mining problems for entities. This captures the relationship between aspects and views while relation between text and visual modalities are also captured. The drawback is that noisy and insignificant data in the information is overlooked. Diverse types of social media data is not validated to find the viability of multimodal aspect-opinion modeling for different knowledge mining tasks.

Lee et al., [124] proposed Electronic Word-of-Mouth (eWOM) that is used to collect reviews in the movie industry. To calculate and compare the star ratings given by reviewers, SA makes use of Semantic Orientation Calculator (SO-CAL). Results of this study would be pertinent in forecasting box-office sales. This study illustrates how a forecasting model can benefit from Word of Mouth information.

Cambria and Erik [125] provides a technique for mining of sentiments and SA that enables an efficient way to convert from unstructured text data to organized machine-processable information, in any domain. It is popular due to its accessibility and economy keyword spotting. The limitation of semantic approaches is that; knowledge representation is usually specified. As the inference of semantic and affective features related with ideas is limited by fixed, level portrayal it does not enable diverse idea nuances to take care.

Vural et al., [126] designed a sentiment-focused Web crawling framework to discover and fetch the sentiment in web contents quickly. This method grabs web pages with huge content and identifies the sentiment, but the performance is the same when compared to small pages of web with sentiment content. It has not considered the metric of discovering of Web pages that contain diverse sentiments (both positive and negative).

Li et al., [127] executed a non-specific stock price pre- diction framework. The Harvard psychological dictionary and Loughran-McDonald financial sentiment dictionary were used to develop an opinion space. They evaluated the model's pre- diction precision and observationally looked at their execution at various market classification levels. Here, the prediction ac- curacy is improved through SA. Focus on positive and negative measurements and hence does not give good predictions. The methods used to identify opinion polarity does not perform well in every test. Table IV presents a summary of various recent Sentiment Analysis techniques.

VI. CONCLUSION AND FUTURE WORK

Sentiment Analysis is a method of identifying the emotion behind the words, attain an understanding of attitudes, views and emotions indicated within an online mention. Nowadays, business organizations and academic institutions are applying efforts to identify the best system for SA. This paper reviews the SA of data that is performed in four dimensions viz. Sentiment Classification, Emotion Detection, Feature Extraction and Sentiment Analysis. After analyzing the articles in this paper, there is scope for further research in the fields of ED, FS and SC. These are reviewed based on the features selected, approaches and techniques employed, the pros and cons in each paper. The most familiar techniques used are

SVM, Naive Bayes and Max Entropy. Data from social media, movie reviews and forums, are used in Opinion Mining. More research is necessary on multimodal analysis and identification of emotions in music. There are several challenges that exist in this area such as word contradiction, domain dependency and inconsistent data and these can be resolved using different approaches.

The future research directions in the context of Opinion Classification, Emotion Detection, Feature Selection and Sentiment Analysis are presented. In terms of Feature Extraction, minimal work is being performed on identification of implicit features as they are comparatively hard to identify than explicit features. Many existing Feature Extraction techniques are not able to address the pertinent features in redundant feature space. Thus, dimensionality reduction of large attribute spaces and performance evaluation of hybrid methods of feature selection are the areas for future research. In Sentiment Analysis, handling of bi-polar sentiments, improving the sentiment word identification algorithm and spam detection SA are the future areas of research. The future directions in Emotion Detection is to study how emotional expression changes over time or between genders or between ethnic groups. Search technique based on emotions is to be improved.

REFERENCES

[1] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, “Sentence Compression for Aspect-Based Sentiment Analysis,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2111–2124, 2015.

[2] D. Sejal, V. Rashmi, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, “Image Recommendation based on Keyword Relevance using Absorbing Markov Chain and Image Features,” *International Journal of Multimedia Information Retrieval*, vol. 5, no. 3, pp. 185–199, 2016.

[3] K. C. Srikantaiah, N. K. Kumar, K. R. Venugopal, and L. M. Patnaik, “Web Caching and Prefetching with Cyclic Model Analysis of Web Object Sequences,” *International Journal of Knowledge and Web Intelligence 2*, vol. 5, no. 1, pp. 76–103, 2014.

[4] R. H. Vishwanath, S. V. Leena, K. C. Srikantaiah, S. Kumar, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, “APST: Approximation and Prediction of Stock Time-Series Data using Pattern Sequence,” *International Conference on Data Mining Workshops*, 2013.

[5] G. U. Vasanthakumar, B. Prajakta, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, “PIB: Profiling Influential Blogger in Online Social Networks, A Knowledge Driven Data Mining Approach,” *Procedia Computer Science*, vol. 54, pp. 362–370, 2015.

[6] B. Liu, “Sentiment Analysis and Opinion Mining,” *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[7] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, “Sentiment Classification: The Contribution of Ensemble Learning,” *Decision support systems*, vol. 57, pp. 77–93, 2014.

[8] G. Vinodhini and R. M. Chandrasekaran, “Sentiment Analysis and Opinion Mining: A Survey,” *International Journal of Advanced Re- search in Computer Science and Software Engineering*, vol. 2, no. 6, pp. 282–292, 2012.

[9] D. Tang, B. Qin, F. Wei, L. Dong, T. Liu, and M. Zhou, “A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1750–1761, 2015.

[10] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, “Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005–1016, 2016.

[11] M. Gamon, “Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors and the Role of Linguistic Analysis,” *Proceedings of the 20th international conference on Computational Linguistics*, pp. 841–847, 2004.

[12] X. Zhou, X. Wan, and J. Xiao, “CMiner: Opinion Extraction and Summarization for Chinese Microblogs,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 7, pp. 1650–1663, 2016.

[13] Q. Ye, Z. Zhang, and R. Law, “Sentiment Classification of Online Reviews to travel Destinations by Supervised Machine Learning Approaches,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 6527– 6535, 2009.

[14] G. Laboreiro, L. Sarmento, J. Teixeira, and E. Oliveira, “Tokenizing Micro-Blogging Messages using a Text Classification Approach,” *Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data*, pp. 81–88, 2010.

[15] M. R. Saleh, M. T. Mart ́ın-Valdivia, A. Montejo-Ra ́ez, and L. Uren ̃a- Lo ́pez, “Experiments with SVM to Classify Opinions in Different Domains,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 14 799–14 804, 2011.

[16] R. Moraes, J. F. Valiati, and W. P. G. Neto, “Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN,” *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013.

[17] S. M. Mohammad, S. Kiritchenko, and X. Zhu, “NRC-Canada: Building the state-of-the-art in Sentiment Analysis of Tweets,” *Seventh International Workshop on Semantic Evaluation Exercises (SemEval- 2013)*, 2013.

[18] S. Liu, X. Cheng, F. Li, and F. Li, “TASC: Topic-Adaptive Sentiment Classification on Dynamic Tweets,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1696–1709, 2015.

[19] J. Smailovic ́, M. Grcˇar, N. Lavracˇ, and M. Zˇnidarsˇicˇ, “Stream-Based Active Learning for Sentiment Analysis in the Financial Domain,” *International Journal of Information Sciences*, vol. 285, pp. 181–203, 2014.

[20] S. Canuto, M. A. Goncalves, and F. Benevenuto, “Exploiting New Sentiment-Based Meta-Level Features for Effective Sentiment Analysis,” *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 53–62, 2016.

[21] D. D. Wu, L. Zheng, and D. L. Olson, “A Decision Support Approach for Online Stock Forum Sentiment Analysis,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 8, pp. 1077–1087, 2014.

[22] A. Sharma and S. Dey, “A Document-Level Sentiment Analysis Approach using Artificial Neural Network and Sentiment Lexicons,” *ACM SIGAPP Applied Computing Review*, vol. 12, no. 4, pp. 67–75, 2012.

[23] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, “Topic Sentiment Analysis in Twitter: A Graph-Based Hashtag Sentiment Classification Approach,” *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pp. 1031–1040, 2011.

[24] M. Trupthi, S. Pabboju, and G. Narasimha, "Improved Feature Ex- traction and Classification Sentiment Analysis," *International Conference on Advances in Human Machine Interaction (HMI)*, pp. 1–6, 2016.
[25] R. Yao and J. Chen, "Predicting Movie Sales Revenue using Online Re- views," *IEEE International Conference on Granular Computing (GrC)*, pp. 396–401, 2013.
[26] P. Bhoir and S. Kolte, "Sentiment Analysis of Movie Reviews using Lexicon Approach," *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–6, 2015.
[27] A. Ekawijana and H. Heryono, "Composite Naive Bayes Classification and Semantic method to enhance Sentiment Accuracy Score," *4th International Conference on Cyber and IT Service Management*, pp. 1–4, 2016.
[28] V. Jha, N. Manjunath, P. D. Shenoy, K. Venugopal, and L. Patnaik, "HOMS: Hindi Opinion Mining System," *IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, pp. 366– 371, 2015.
[29] R. Jose and V. S. Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach," *International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp. 64–67, 2016.
[30] M. A. M. Raja, S. G. Winster, and S. Swamynathan, "Review Analyzer: Analyzing Consumer Product Reviews from Review Collections," *International Conference on Recent Advances in Computing and Software Systems (RACSS)*, pp. 287–292, 2012.
[31] P. Andriotis, G. Oikonomou, T. Tryfonas, and S. Li, "Highlighting Relationships of a Smartphone's Social Ecosystem in Potentially Large Investigations," *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 1974–1985, 2016.
[32] C. Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang, "SentiView: Sentiment Analysis and Visualization for Internet Populars," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 620–630, 2013.
[33] C. C. Yang and T. D. Ng, "Analyzing and Visualizing Web Opinion Development and Social Interactions with Density-Based Clustering," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1144–1155, 2011.
[34] C. Troussas, M. Virvou, and S. Mesaretzidis, "Comparative Analysis of Algorithms for Student Characteristics Classification using a Methodological Framework," *6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–5, 2015.
[35] X. Chen, M. Vorvoreanu, and K. Madhavan, "Mining Social Media Data for understanding Students Learning Experiences," *IEEE Trans actions on Learning Technologies*, vol. 7, no. 3, pp. 246–259, 2014.
[36] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-Processing in Sentiment Analysis," *Procedia Computer Science*, vol. 17, pp. 26–32, 2013.
[37] G. Liu, X. Xu, B. Deng, S. Chen, and L. Li, "A Hybrid method for Bilingual Text Sentiment Classification based on Deep Learning," *17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/ Distributed Computing (SNPD)*, pp. 93–98, 2016.
[38] N. F. Da Silva, E. R. Hruschka, and E. R. Hruschka, "Tweet Sentiment Analysis with Classifier Ensembles," *International Journal of Decision Support Systems*, vol. 66, pp. 170–179, 2014.
[39] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter Opinion Mining Framework using Hybrid Classification Scheme," *International Journal of Decision Support Systems*, vol. 57, pp. 245–257, 2014.
[40] L. Shi, B. Sun, L. Kong, and Y. Zhang, "Web Forum Sentiment Analysis based on Topics," *Ninth IEEE International Conference on Computer and Information Technology, 2009. CIT'09*, vol. 2, pp. 148– 153, 2009.
[41] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification," *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,* pp. 1555–1565, 2014.
[42] J. Yang, Y. Liu, X. Zhang, X. Yuan, Y. Zhao, S. Barlowe, and S. Liu, "Piwi: Visually Exploring Graphs based on their Community Structure," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 6, pp. 1034–1047, 2013.
[43] R. Xia, C. Zong, and S. Li, "Ensemble of Feature Sets and Classification Algorithms for Sentiment Classification," *International Journal of Information Sciences*, vol. 181, issue. 6, pp. 1138–1152, 2011.
[44] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-Dependent Twitter Sentiment Classification," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pp. 151–160, 2011.
[45] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *International Journal* of *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
[46] D. T. Vo and Y. Zhang, "Target-Dependent Twitter Sentiment Classification with Rich Automatic Features," *Proceedings of the Twenty- Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pp. 1347–1353, 2015.
[47] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, "Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph," *Proceedings of the First workshop on Unsupervised Learning in NLP*, pp. 53–63, 2011.
[48] Y. Hu and W. Li, "Document Sentiment Classification by Exploring Description Model of Topical Terms," *International Journal of Computer Speech & Language*, vol. 25, issue. 2, pp. 386–403, 2011.
[49] Y. Zhang, M. J. Er, R. Venkatesan, N. Wang, and M. Pratama, "Sentiment Classification using Comprehensive Attention Recurrent models," *International Joint Conference on Neural Networks (IJCNN)*, pp. 1562–1569, 2016.
[50] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon- Based Methods for Sentiment Analysis," *Computational linguistics*, vol. 37, issue. 2, pp. 267–307, 2011.
[51] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced Sentiment Learning using Twitter Hashtags and Smileys," *Proceedings of the 23rd International Conference on Computational Linguistics: posters*, pp. 241–249, 2010.
[52] A. Celikyilmaz, R. Sarikaya, M. Jeong, and A. Deoras, "An Empirical Investigation of Word Class-Based Features for Natural Language Understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 6,

pp. 994–1005, 2016.

[53] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol 1*, pp. 142–150, 2011.

[54] E. C. Dracut, H. Wang, P. Sistla, C. Yu, and W. Meng, "Polarity Consistency Checking for Domain Independent Sentiment Dictionaries," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 3, pp. 838–851, 2015.

[55] W.-F. Chen, M.-H. Chen, M.-L. Chen, and L.-W. Ku, "A Computer- Assistance Learning System for Emotional Wording," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 1093– 1104, 2016.

[56] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, "Sentiment Embeddings with Applications to Sentiment Analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 496–509, 2016.

[57] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, "Evaluating Sentiment in Financial News Articles," *International Journal of Decision Support Systems*, vol. 53, issue. 3, pp. 458–464, 2012.

[58] F. Ren and Y. Wu, "Predicting User-Topic Opinions in Twitter with Social and Topical Context," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 412–424, 2013.

[59] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment Analyzer: Extracting Sentiments about a given Topic using Natural Language Processing Techniques," *Third IEEE International Conference on Data Mining*, pp. 427–434, 2003.

[60] M. Z. Asghar, A. Khan, S. Ahmad, and F. M. Kundi, "A Review of Feature Extraction in Sentiment Analysis," *Journal of Basic and Applied Scientific Research*, vol. 4, no. 3, pp. 181–186, 2014.

[61] A. Celikyilmaz, R. Sarikaya, M. Jeong, and A. Deoras, "An Empirical Investigation of Word Class-Based Features for Natural Language Understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 6, pp. 994–1005, 2016.

[62] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, "Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content," *Neurocomputing*, vol. 174, pp. 50–59, 2016.

[63] B. Agarwal and N. Mittal, "Prominent Feature Extraction for Review Analysis: An Empirical Study," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 3, pp. 485–498, 2016.

[64] S. Pasarate and R. Shedge, "Comparative Study of Feature Extraction Techniques used in Sentiment Analysis," *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS- INBUSH)*, pp. 182–186, 2016.

[65] A. Duric and F. Song, "Feature Selection for Sentiment Analysis based on Content and Syntax Models," *International Journal of Decision Support Systems*, vol. 53, no. 4, pp. 704–711, 2012.

[66] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-Domain co-Extraction of Sentiment and Topic Lexicons," *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Paper, vol 1*, pp. 410–419, 2012.

[67] H. Uguz, "A Two-Stage Feature Selection Method for Text Categorization by using Information Gain, Principal Component Analysis and Genetic Algorithm," *International Journal of Knowledge-Based Systems*, vol. 24, issue. 7, pp. 1024–1032, 2011.

[68] R. M. Elawady, S. Barakat, and N. M. Elrashidy, "Different Feature Selection for Sentiment Classification," *International Journal of Information Science and Intelligent System*, vol. 3, no. 1, pp. 137–150, 2014.

[69] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Transactions on Information Systems (TOIS)*, vol. 26, no. 3, pp. 1–34, 2008.

[70] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, "Concept- Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach," *International Journal of Cognitive Computation*, vol. 7, no. 4, pp. 487–499, 2015.

[71] Y. Matsuyama, A. Saito, S. Fujie, and T. Kobayashi, "Automatic Expressive Opinion Sentence Generation for Enjoyable Conversational Systems," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 2, pp. 313–326, 2015.

[72] N. Shelke, S. Deshpande, and V. Thakare, "Domain Independent Approach for Aspect Oriented Sentiment Analysis for Product Reviews," *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, pp. 651–659, 2017.

[73] H. Wang, L. Liu, W. Song, and J. Lu, "Feature-Based Sentiment Analysis Approach for Product Reviews," *Journal of Software*, vol. 9, no. 2, pp. 274–279, 2014.

[74] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Multi-Aspect Sentiment Analysis for Chinese Online Social Reviews based on Topic Modeling and HowNet Lexicon," *International Jouranl of Knowledge-Based Systems*, vol. 37, pp. 186–195, 2013.

[75] Z. Yu, Z. Yu, X. Zhou, C. Becker, and Y. Nakamura, "Tree-Based Mining for Discovering Patterns of Human Interaction in Meetings," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 4, pp. 759–768, 2012.

[76] W. Rong, Y. Nie, Y. Ouyang, B. Peng, and Z. Xiong, "Auto-Encoder Based Bagging Architecture for Sentiment Analysis," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 840–849, 2014.

[77] E. M. Alshari, A. Azman, N. Mustapha, S. C. Doraisamy, and M. Alksher, "Prediction of Rating from Comments Based on Information Retrieval and Sentiment Analysis," *2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP)*, pp. 32–36, 2016.

[78] A. Giachanou and F. Crestani, "Tracking Sentiment by Time Series Analysis," *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1037– 1040, 2016.

[79] A. Krouska, C. Troussas, and M. Virvou, "The Effect of Preprocessing Techniques on Twitter Sentiment Analysis," *7th International Conference on Information, Intelligence, Systems Applications (IISA)*, pp. 1–5, 2016.

[80] C.-L. Liu, W.-H. Hsaio, C.-H. Lee, G.-C. Lu, and E. Jou, "Movie Rating and Review Summarization in Mobile Environment," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 3, pp. 397–407, 2012.

[81] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation,"

Computational Linguistics, vol. 37, no. 1, pp. 9–27, 2011.

[82] S. Wang, D. Li, X. Song, Y. Wei, and H. Li, "A Feature Selection Method based on Improved Fishes Discriminant Ratio for Text Sentiment Classification," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8696–8702, 2011.

[83] N. Shelke,"Approaches of Emotion Detection from Text, "*International Journal of Computer Science and Information Technology Research*, vol. 2, no. 2, pp. 123–128, 2014.

[84] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang, "Finding Waldo: Learning about Users from their Interactions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1663–1672, 2014.

[85] C. Liu, L. Xu, Z. Wang, and N. Wang, "Sentiment Analysis method based on Context View," *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC- FSKD)*, pp. 1577–1581, 2016.

[86] B. OF ́laherty and C. Heavin, "Positioning Predictive Analytics for Customer Retention," *Journal of Decision Systems*, vol. 24, no. 1, pp. 3–18, 2015.

[87] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic Patterns: Dependency-Based rules for Concept-Level Sentiment Analysis," *International Journal of Knowledge-Based Systems*, vol. 69, pp. 45–63, 2014.

[88] M. D. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text," *IEEE Transactions on Affective Computing*, vol. 5, no. 2, pp. 101–111, 2014.

[89] V. Sintsova and P. Pu, "Dystemo: Distant Supervision Method for Multi-Category Emotion Recognition in Tweets," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, p. 13, 2016.

[90] C.-J. Su and Y.-A. Chen, "Social Media Analytics Based Product Improvement Framework," *International Symposium on Computer, Consumer and Control (IS3C)*, pp. 393–396, 2016.

[91] X. Quan, Q. Wang, Y. Zhang, L. Si, and L. Wenyin, "Latent Discriminative Models for Social Emotion Detection with Emotional Dependency," *ACM Transactions on Information Systems (TOIS)*, vol. 34, no. 1, pp. 2–19, 2015.

[92] E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter Sentiment Analysis: The Good the Bad and the Omg!" *Proceedings of the Fifth International Conference on Weblogs and Social Media*, vol. 11, no. 538- 541, p. 164, 2011.

[93] V. Perez-Rosas, R. Mihalcea, and L.-P. Morency, "Utterance-Level Multimodal Sentiment Analysis," *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (1)*, vol. 1, pp. 973–982, 2013.

[94] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[95] G. S. Solakidis, K. N. Vavliakis, and P. A. Mitkas, "Multilingual Sentiment Analysis using Emoticons and Keywords," *IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 2, pp. 102–109, 2014.

[96] Z. Wang, S. Y. M. Lee, S. Li, and G. Zhou, "Emotion Analysis in Code-Switching Text with Joint Factor Graph Model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 3, pp. 469–480, 2017.

[97] Pushpa C N, Thriveni J, Venugopal K R and L M Patnaik, "Radial Basis Neural Network Technique based Web Page Recommendation System", *International Journal of Computer Sciences and Engineering (IJCSE*), vol. 2, no. 9, pp. 1-7, September 2014.

[98] Y. Lu, M. Steptoe, S. Burke, H. Wang, J.-Y. Tsai, H. Davulcu, D. Montgomery, S. R. Corman, and R. Maciejewski, "Exploring Evolving Media Discourse through Event Cueing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 220–229, 2016.

[99] S. Asur and B. A. Huberman, "Predicting the Future with Social Media," *IEEE/WIC/ACM International Conference on Web Intelligence*, vol. 1, pp. 492–499, 2010.

[100] C. Clavel and Z. Callejas, "Sentiment Analysis: From Opinion Mining to Human-Agent Interaction," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 74–93, 2016.

[101] H. Kang, S. J. Yoo, and D. Han, "Senti-Lexicon and improved Naive Bayes algorithms for Sentiment Analysis of Restaurant reviews," *International Journal of Expert Systems with Applications*, vol. 39, no. 5, pp. 6000–6010, 2012.

[102] E. Psomakelis, K. Tserpes, D. Anagnostopoulos, and T. Varvarigou, "Comparing Methods for Twitter Sentiment Analysis," *6th Conference on Knowledge Discovery and Information Retrieval*, 2015.

[103] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment Analysis on Social Media for Stock Movement Prediction," *International Journal of Expert Systems with Applications*, vol. 42, no. 24, pp. 9603–9611, 2015.

[104] S. Krening, B. Harrison, K. M. Feigh, C. L. Isbell, M. Riedl, and A. Thomaz, "Learning from Explanations using Sentiment and Advice in RL," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 1, pp. 44–55, 2017.

[105] F. Li, S. Wang, S. Liu, and M. Zhang, "SUIT: A Supervised User- Item Based Topic Model for Sentiment Analysis." *Proceedings of the Twenty-Eighth Conference on Artificial Intelligence*, vol. 14, pp. 1636–1642, 2014.

[106] N. Li and D. D. Wu, "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast," *International Journal of Decision support systems*, vol. 48, no. 2, pp. 354–368, 2010.

[107] A. C. E. Lima, L. N. de Castro, and J. M. Corchado, "A Polarity Analysis Framework for Twitter Messages," *International Journal of Applied Mathematics and Computation*, vol. 270, pp. 756–767, 2015.

[108] Y. Y. Chen and K. V. Lee, "User-Centered Sentiment Analysis on Customer Product Review," *International Journal of World Applied Sciences*, vol. 12, pp. 32–38, 2011.

[109] Z. Yan, M. Xing, D. Zhang, B. Ma, and T. Wang, "A Context- Dependent Sentiment Analysis of Online Product Reviews based on Dependency Relationships," *Association for Information Systems*, 2014.

[110] G. Li and F. Liu, "Application of a Clustering Method on Sentiment Analysis," *Journal of Information Science*, vol. 38, no. 2, pp. 127–139, 2012.

[111] K. Ghag and K. Shah, "Comparative Analysis of the Techniques for Sentiment Analysis," *International Conference on Advances in Technology and Engineering (ICATE)*, pp. 1–7, 2013.

[112] J. Smailovic, M. Grcar, N. Lavrac, and M. Znidarsic, "Stream-based Active Learning for Sentiment Analysis in the Financial Domain," *International Journal of Information Sciences*, vol. 285, pp. 181–203, 2014.

[113] W. Wei and T. Yongxin, "Network Public Sentiment Orientation Analysis based on HMM Model," *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2269–2273, 2016.
[114] C. Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang, "SentiView: Sentiment Analysis and Visualization for Internet Popular Topics," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 620–630, 2013.
[115] P. H. Calais Guerra, A. Veloso, W. Meira Jr, and V. Almeida, "From Bias to Opinion: A Transfer-Learning Approach to Real-Time Sentiment Analysis," *Proceedings of the 17th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 150–158, 2011.
[116] A. Trilla and F. Alias, "Sentence-Based Sentiment Analysis for Expressive Text-to-Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 223–233, 2013.
[117] P. Goncalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and Combining Sentiment Analysis Methods," *Proceedings of the first ACM Conference on Online Social Networks*, pp. 27–38, 2013.
[118] X. Yu, Y. Liu, X. Huang, and A. An, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 4, pp. 720–734, 2012.
[119] K. Hiroshi, N. Tetsuya, and W. Hideo, "Deeper Sentiment Analysis using Machine Translation Technology," *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 494–500, 2004.
[120] M. Dermouche, L. Kouas, J. Velcin, and S. Loudcher, "A Joint Model for Topic-Sentiment Modeling from Text," *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pp. 819–824, 2015.
[121] N. Claypo and S. Jaiyen, "Automatic Keyword Selection for Sentiment Analysis using Class Dependency and Dissimilarity," *13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1–6, 2016.
[122] R. Jongeling, S. Datta, and A. Serebrenik, "Choosing your Weapons: On Sentiment Analysis Tools for Software Engineering Research," *IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pp. 531–535, 2015.
[123] Q. Fang, C. Xu, J. Sang, M. S. Hossain, and G. Muhammad, "Word- of-Mouth Understanding: Entity-Centric Multimodal Aspect-Opinion Mining in Social Media," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2281–2296, 2015.
[124] H. Lee, H. J. Lee, R. A. Cruz, and J. L. Enriquez, "Electronic Word of Mouth (eWOM) in the Movie Industry," *Proceedings of the 18th Annual International Conference on Electronic Commerce: e-Commerce in Smart Connected World*, pp. 1–7, 2016.
[125] E. Cambria, "An Introduction to Concept-Level Sentiment Analysis," *Mexican International Conference on Artificial Intelligence*, pp. 478-483, 2013.
[126] A. G. Vural, B. B. Cambazoglu, and P. Karagoz, "Sentiment-Focused Web Crawling," *Proceedings of the 21st ACM International Conference on Information and Knowledge Management,* vol. 8, no. 4, pp. 1–21, 2014.
[127] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News Impact on Stock Price Return via Sentiment Analysis," *Int. Journal of Knowledge-Based Systems*, vol. 69, pp. 14–23, 2014.

Author's Profile

Mrs. Mamatha M pursued Bachelor of Engineering from University Visveswaraya College of Engineering, Bangalore University, India in 2008 and Master of Technology from Dr. Ambedkar Institute of Technology, VTU in year 2011. She is currently pursuing Ph.D. Her main research work focuses on Sentiment Analysis, Data Mining and Multimodal Analysis. She has 4 years of teaching experience.

Mrs. Thriveni J pursued Bachelor of Engineering from Dayanand Sagar College of Engineering and Master of Engineering from University Visveswaraya College of Engineering, Bangalore University, India. She obtained Ph.D. in the year 2010 in the area of Mobile AdHoc Networks. She has 4 years of Industrial Experience, 22 years of Teaching Experience and 12 years of Reseach Experience. She is currently working as Associate Professor in Department of Computer Science and Engineering, University Visveswaraya College of Engineering, Bangalore University. She has published more than 75 reseach papers in International Conferences and Referred Journals and a book on AdHoc Networks.

Mr K R Venugopal pursued Bachelor of Engineering from University Visveswaraya College of Engineering, Bangalore University, India and Master of Engineering from Indian Institute of Science, Bangalore. He pursued Ph.D. in Computer Science from Indian Institute of Technology, Madras and Ph.D. in Economics from Bangalore University, Bangalore. He is currently working as Principal, University Visveswaraya College of Engineering, Bangalore since June 2007 till date. He is a member of IEEE, Member of AICTE,Bangalore. He has published more than 600 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. He has filed 101 patents and authored and edited 62 books. He has received IEEE Fellow for contributions in Computer Science and Electrical Engineering Education. His main research work focuses on Data Mining, Optical Networks, Ad Hoc Networks, Sensor Networks, Digital Signal Processing and IoT. He has 36 years of teaching experience and Research Experience.

