# AN APPROACH FOR THE STUDY OF OPINION MINING IN THE FIELD OF CLOUD BASED SERVICES

**B. Anil Kumar, A. Ravi Kumar**

*Abstract*— **The rapid revolution of Social Network Sites (SNS) around the globe is presenting wide range of data that can be used in studies of sentiment analysis about certain products, brands, services ... etc. In addition, cloud computing fields had been one of the most interesting fields in research studies. In this paper, we used the sentiment analysis of top leading cloud service providers namely; Amazon and Microsoft Azure to analyze their customers' opinions and reviews. To do that, two datasets are extracted which are consisting of tweets that had either organizations' names or cloud names. We study, and analyze the way customers think about them. In this regard, many organizations tend to find out what do customers think or tweet about their products in order to effectively plan marketing campaigns and try to gain the positive impact of Word-of-Mouth. Results are analyzed and explained in details in term of polarity and emotions classifications to show the impact of sentiment analysis to support organizations decisions. We can note from the emotions classification results that, "joy" category is better for Microsoft Azure comparing to Amazon, The "sadness" percentage is larger for Amazon comparing to Microsoft Azure. Furthermore, we can note from the polarity classification that Microsoft Azure has 65% positive tweets compared 45% for Amazon. In addition, the results show that Amazon has 50% negative polarity compared 25% for Microsoft Azure.**

*Index Terms*— **Social Network Analysis (SNA); Cloud Computing; Amazon Web Services (AWS); Microsoft Azure; Sentiment Analysis**

## I. INTRODUCTION

It had been mentioned that, in the re cent years, Social Networking (SN) websites had been spreading widely, and their users are increasing rapidly [1]. For example, in 2003, users of Friendster had reached millions, and in the beginning of 2008, MySpace users were increased to be over 300 million, while Facebook users had exceeded over 1 billion usersi, 85% of which were U.S Students. As weil as Cy world, the most famous Social Network in Korea, had 18 million users and had been expanding their website to china since 2001, and after that to USA in 2006 [2]. Online Social Network Services (SNS) enable their users to get connected and have the chance to communicate and share activities, Interests, and Ideas within their network. SNS can be considered as web-based services that allow users to create a profile of their own which can be accessed by set of chosen users giving them the chance to communicate deeply with each other.

SNS can be defined as web-based services that are based on certain valuable relationships such as friendship, Interests and activities; enabling users to communicate for several and different purposes such as networking, sharing information and Ideas [1]. Moreover, SNS are representing a new stage in the internet and web services' development which depends on the user who can articulate and specity what he/she exactly wants to share, converting the static web pages into interactive ones characterized by the user within contents [3]. As a result, these SNS had created huge data about customers which organizations have interests to analyze and understand as a first step. Whereas the second step to be taken is to be able to predict the opinion of customers. This is all about forming a chance for predictive analytics purposes. [4]

Moreover, data filtering which is provided by SNS via knowledge based systems gives user's connections about his or her choices, products and service providers they have dealt with [5]. In other words, those SNS help in the process of marketing as weil as advertising for chosen products or services to eventually making service providers pay more attention to their type and quality of products and services they provide to customers in order to build highly reputed and recommended services. With millions of users who are daily registering in SNS and the ease of communication they provide, service providers are struggling to get involved with this wide range of people. As a matter of fact, organizations cannot ignore the revolution of Social Networking through the internet, which enables them to cultivate strong relationships with their publics [6].

Many researchers had studied the relationship between organizations and SNS to find that there is a deep relationship between both. As [7] had listed many processes and procedures which being executed using SNS from social relationship between management and employees to marketing, staffing and recruiting, etc., this will give an indication about the future of those SNS and their role in the overall business functions. For [7], the future of marketing and sales will be personal; it means that not only organizations are going to send custom messages to individual customers each by his own, but also in the means of releasing unprotected media contents over SNS or Social Blogs which will be spread all over the internet and seen by unlimited number of people just by one elick, which also means that messages are going to be sent from friends to other friends.This might be considered as a recommendation for the organization as weil as its product or service.

***Manuscript received Jan, 2018***.

**B. Anil Kumar** *P.G Student-Department of CSE,* ***SSJ Engineering College, Hyderabad****, Telengana, India. Phone/ Mobile No :9440852004*

***A. Ravi Kumar, Associate Professor, Department of Computer Science and Engineering SSJ Engineering College, Hyderabad****, Telengana, India Mobile No.,9440898357.*

In order to enable service providers to understand their current state on SNS and analyze opinions of their customers, several researchers had used text mining along with sentiment analysis. This paper discusses the application of sentiment analysis on two of the most recognized c10ud service providers namely; Microsoft and Amazon. Datasets of extracted tweets about their c10ud products are to be analyzed for the purpose of understanding the customers' opinions about them.

The remainder of paper is organized as folIows: section [[includes literature review about c10ud computing, AWS and Azure, and Sentiment Analysis. Section [[[includes the proposed sentiment analysis methodology. Section IV introduces discussion of results generated. Section V lists conclusions and future work.

## II. LITERATURE REVIEW

Surfing the web for a c10ud service provider, you can find plenty of them with different services, prices and quality of service (QoS). Following subsections explain the mechanisms of two leading c10ud computing providers, Amazon as the top leader followed by Microsoft Azure.

### *A. Amazon cloud*

Amazon is the top leader of cloud computing market with its AWS (Amazon Web Services). A WS was launched in 2006, which means they are about to close 10 years of expertise in this field soon. Also, Amazon was the first cloud provider to offer Infrastructure as a service (IaaS), allowing individuals and organizations to rent virtual computers.

Amazon Elastic Compute Cloud (EC2) is the basic part of A WS that allows users to allocate required computers in order to run applications they need on them. It allows creating virtual mach in es, "instances" as Amazon calls them, for attaining scalable deployment of applications the user needs. These instances can be created, launched and terminated by a user as per demand. Payment for active instance is computed per hour which represents meaning of elasticity term. [8], [9], [10]

### *B. Microsoft Azure Cloud*

Microsoft is the second leading eloud provider after Amazon. Their c10ud platform and infrastructure called Azure was first released in February 2010. It provides both PaaS (Platform as a Service) and [aaS (Infrastructure as a Service) delivery models, and supports many frameworks and various programming languages. By using Azure, a user can build, deploy and manage both applications and services through a global network ofMicrosoft-managed datacenters.

Azure provides the user with the ability to use various types of instances at hourly rate fees just like Amazon. Nevertheless, they compute the cost of used recourses per minute. Which means if a user allocated a resource for one hour and a half, then payment is computed for the exact period of time without estimating the ceiling forthe period oftime. However, Microsoft has dealt with pricing models they are using with secrecy that affects their transparency with users. Hence, affects the number of users convinced to use it. [10], [11]

### *C. Related Work*

Sentiment analysis can be defined as studying peoples' opinions, attitudes and feelings towards an event, product or an organization computationally. [15], [[6] and [17]. With the massive reviews people post online about their personal decisions about various items they use, food they eat or anything they do, huge amount of data is available to conduct sentiment analysis. Usually this is conducted depending on a single word extracted from a sentence or post while even sentence structure should be involved as well as implicitly expressed opinions.

Text mining is the process of extracting knowledge from available text online whether on SNS, public forums or internet generally regarding certain topic. Text mining is used in information extraction, information retrieval and natural language processing fields. Many researchers had used text mining in their studies regarding various topics. The authors in [12] had conducted text mining on digital library's documents in order to extract metadata of them. Then, they could tag more appropriate items mentioned in the text. And they concluded that this process can enrich documents of library as well as user experience. Another study conducted by [13] had used text mining to classify E-learning resources and documents by identifying similarities among different topics. Another study by [14] had targeted posts by video streaming students to be analyzed in order to discover learning patterns and technology issues discussed.

Furthermore, [15] had applied text mining on online forums hotspot detection and forecast. They had used datasets from Sina sports forums, with a range of 3 [ different topic forums and 220,053 posts. Then, created an algorithm to analyze emotion polarity for extracted text with values given to text. Afterwards, in order to have unsupervised text mining approach, they combined their algorithm with Support Vector Machine (SVM) and K-Means clustering techniques. Having clustered the forums into groups, the center of each group was considered as hotspot forum among others. Conclusions had shown that both SVM and K-Means had same results for top 4 hotspot forums, while results differ for top 10 hotspot forums as SVM forecasting resembles 80% of K-means clustering results.

In [16], the authors applied 4 of text mining techniques on systematic reviews, namely; Automatic Term Recognition (ATR), document clustering, classification and summarization.

They conducted this research to prove that text mining application supports the identification of pertinent studies in systematic reviews efficiency. Their conclusion was that applying text mining would be of a positive impact and support the reviewing process at various stages. Nevertheless, text mining is not common in the field of systematic reviews and some required evaluation then methods development are essential before assessing text mining application.

In [17] the authors stated that availability of internet and SNS made it possible for customers to express their opinion about products, issues, topics and much more events than ever before. These online opinions provides potential homogenous data source for evaluating sentiment. It can be defmed as the process of extracting feeling whether verbal or non-verbal communication. Sentiment analysis has three main types as [18] stated, namely; sentiment classification, feature based sentiment analysis and comparative sentences and relation analysis. While [16] had found that sentiment analysis is a vital part of text mining. Which makes it possible to know customer's opinion through analyzing content and structure of text he had written.

Many research studies were conducted to examine the sentiment of users regarding various topics. In [17], the authors presented a logical approach for extracting sentiment from widely spread SNS. This approach is combination of categorical grammar, lexicon acquisition and annotation.

Then, semantic networks are used to analyze sentiments extracted from text. This research tried to solve some issues researcher may face when dealing with Machine Learning Algorithms such as labeled training data and unexpected results. Results had shown that presented logical approach found to provide extremely precise results than machine learning ones.

In [19], the authors used twitter to extract tweets about companies' stock, conduct sentiment analysis and assign a sentiment value for each company. After wards, comparison was conducted with improvement of companies' stock price in real time environment. Used techniques were n-gram and "word2vec" along with random forest classification algorithm to find the sentiment of tweets. Resulted values are evaluated then compared with actual companies' stock prices. Their conclusions shown that consumer facing companies actual stock prices and sentiment were positively correlated, while opposite is applicable on the rest companies. Companies such as Microsoft and Walmart showed strong positive correlation. On the contrary, companies like GoldmanSachs and Cisco Systems showed strong negative correlation.

Another related work is a study conducted by [20] for the sake of Urbn transportation in Milan. Dataset was extracted from twitter about this company in order to be analyzed for evaluating quality of services. Text mining technique was designed to work with Italian language and used to specify tweets discussing different events of Urban such as accidents, traffic, driving ... etc. while sentiment was conducted via (SVM) technique. This study had supported Urban with their commuters' opinion about experiences they had with them, level of their quality of services and the chance to plan for required improvements to achieve commuters' needs.

Moreover, [21] had presented an Aspect and Sentiment Unified Model that can specify the exact category of text especially when the online text is plain text without any emotional symbols. Their model can be applied on any online reviews.

Online merchants usually ask their customers to post reviews about items they bought. Whenever an interested customer is willing to buy an item, these reviews can help in decision of buying it. In [22], the authors presented a novel machine learning system for mining customers' opinions about products available online. This system identify opinion expression, orientation and classify them into negative and positive opinion about every recognized item.

## III. METHODOLGY

In this paper we have chosen two of the top leading cloud service providers namely; Amazon Web Services and the other is Microsoft Azure, by taking their pages on twitter in order to conduct sentiment analysis. Two datasets were extracted from twitter. The first one is for A WS by using the query: "A WS OR aws cloud OR ec2cloud or Amazon Web Services". The second data set is for Microsoft Azure by the query: "Azure OR Microsoft Azure". Afterwards, sentiment analysis using the Naive Bays Classifier is conducted on extracted tweets. We choose Naive Bayes (NB) to do the sentiment analysis because it is a simple, easy to implement and combines between the efficiency with acceptable accuracy. Furthermore, two types of sentiment will be investigated: the first one based on polarity lexicon, and the second one is based on emoticons lexicon.

Figure 1 describes full steps of followed methodology.

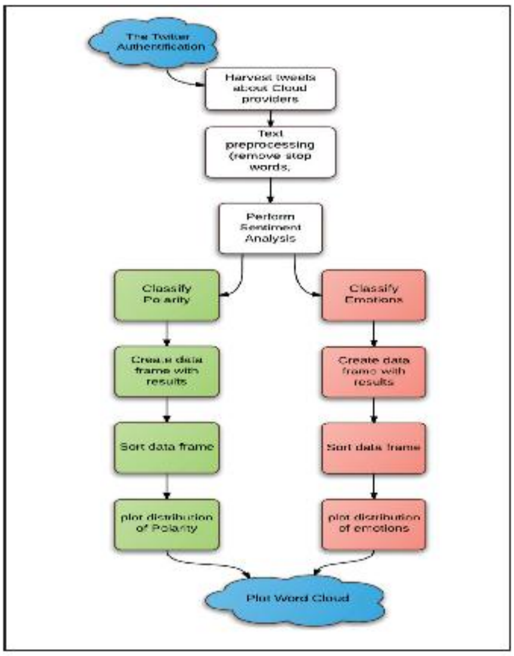
The Twitter Authentification
Harvest tweets about Cloud providers
Text preprocessing (remove stop words,
Perform Sentiment Analysis
Classify Polarity
Classify Emotions
Create data frame with results
Create data frame with results
Sort data frame
Sort data frame
plot distribution of Polarity
plot distribution of emotions
Plot Word Cloud

Fig. 1 Methodolgy flow chart

*A. Naive Bayes Classifier*

Naive Bayes (NB) is a one of the most popular classifiers. [They been widely used because of its simple probabilistic model that assurne all the data attributes are independent. Also, it showed reasonable performances in various tasks [27]. The probabilistic model uses the Bayes theorem to solve the classification problems such as the maximum posterior probability of the class label given the attributes set is calculated. Bays theorem is given by the Equation 1 [25], [26] and [28].

$$P(C|T) = \frac{P(T|C)P(C)}{P(T)} \qquad (1)$$

Where C is a class, and T is a tweet, which is represented by a vector of words T = {tl, t2, ... , tm}, while P(C) and P(TIc) are the prior probability of a given class and the conditional probability of the words given the class, respectively. P(TIc) is computed based on the product of probabilities using the following Equation:

$$P(T|C) - P(t1, t2, \ldots, t_m|C) - \prod_{i=1}^{m} P(x_i|C) \qquad (2)$$

Finally, the class label of T is predicted as the class C which has the highest P(qT).

*B. Dataset Collection and Preprocessing*

On twitter, a search had been conducted for official pages of AWS and Azure and the following information were collected about both cloud computing providers: number of followers was 269,427 and 354,391, number of tweets 5,544 and 16,882, favorites 76 and 1,382 respectively. Afterwards, dataset of each cloud computing provider was extracted with 1500 tweets from both pages and users, to build a good experiment. Then, we have started to prepare the extracted datasets by cleaning them from any unnecessary characters such as retweet and usernames symbols, hashtags, numbers, punctuations, stop words, whitespaces and html links.

*C. Polarity Classification*

The most basic method is to label words from single dimension based of semantic variability called "semantic orientation". Usually used lexicons are available freely online and has certain amount of words labelled as "positive" or "negative". [29]

The first used NB classifier was trained on training data set and makes use of a polarity lexicon based on the Janyce Wiebe's subjectivity lexicon [23]. The training data set is annotated to three classes: positive, neutral and negative tweets. Neutral tweets are taken into account to generalize the sentiment analysis model. The NB polarity classifier uses polarity lexicon based on the matching criteria between the tweet words and lexicon words.

### *D. Emotion Classification*

Emotions represents a vital factor in sentiment analysis as it describes briefly and relatively the response to evaluation of an event, product, organization ... etc. emotions detection can help organizations to determine satisfied and unsatisfied customers. This leads to determine reasons of unsatistying these customers and working on solutions to such reasons. [29]

The second used NB classifier is trained on training data set and makes use of emotions lexicon based on the Carlo Strapparava and Alessandro Valitutti's emotions lexicon [24]. The training data set is annotated to seven classes: anger, disgust, fear, joy, sadness, surprise, and unknown tweets. Like the polarity classification, the matching criteria between the tweet words and emotions lexicon words.

## IV. RESULTS

### *A. Environment*

We ran the experiments on the PC containing 6GB of RAM, 4 Intel cores, i7 (2.0GHz each). For our experiments, we used RStudio to implement the proposed methodology.

### *B. Results*

Both used datasets consist of 1500 records representing tweets on twitter accounts by both customers and official page of cloud computing provider. After applying the two sentiment analysis models explained in the previous section. Visualization of emotions and polarity classification for both A WS and Azure are shown in figure 2, and 3 respectively. Accuracy of model was calculated based on whether words in tweets are existed in the both lexicons.

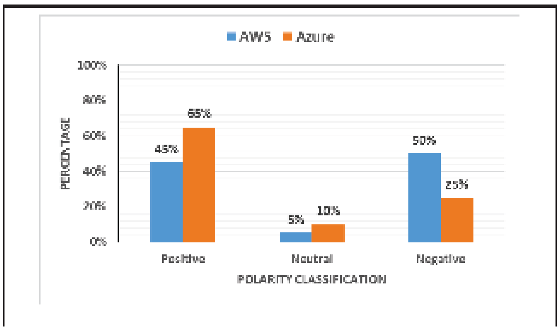

Fig. 2 Polarity Classification for A WS, and Azure polarity classifications.

Polarity classifications for AWS are 45% "Negative", 50% "Positive" and 5% "Neutral". While Azure's Polarity classifications are 25% "Negative", 65% "Positive" and 10% "Neutral".

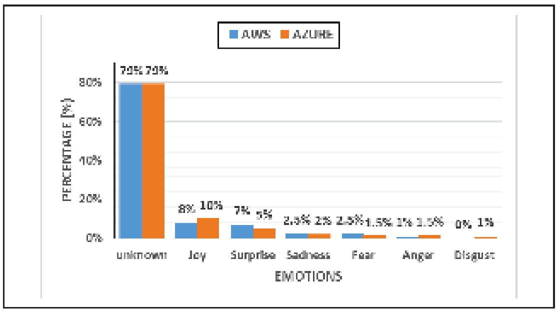

Fig. 3 Polarity Classification for AWS, and Azure

Figure 3 showed Azure emotions categories which were 79% labeled as "unknown", 10% "J oy" , 5% "Surprise", 2% "Sadness", 1.5% "Fear", 1.5% "Anger" and 1 % for "Disgust". AWS emotions categories with 79% labeled as "Unknown", 8% "Joy" , 7% "Surprise", 2.5% "Sadness", 2.5% "Fear", 1% "Anger" and 0% for "Disgust".

Word cloud representation was used to identify the most frequent words in each emotions category. The word cloud results for Amazon and Azure are shown in Figure 4, and Figure 5, respectively. The words with biggest font size was most frequent ones and font gets smaller when frequency decrease. Furthermore, each type of emotions has different color to distinguish between their words. For Aamzon cloud, the word cloud results show that that most frequent word was "turn" followed by "help" when discarding the amazon word. On the other hand, Azure word cloud showed that most frequent words were "Louder" and "Mic", followed by "high".

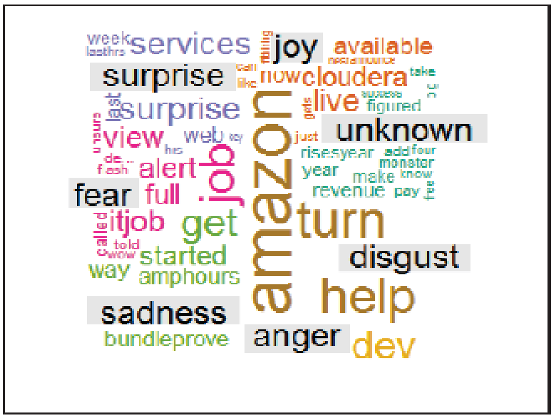

Fig. 4 Word cloud of AWS emotions classifications

## V. CONCLUSION

Sentiment Analysis is being one of the most attractive fields of study and implementation for various types of organizations and service providers. Having reviewed many of its applications as mentioned in the related work section, we had an idea of connecting this interesting field to cloud computing providers. This is because the attention has been focused towards cloud computing in recent years of research in addition, this study was conducted to apply sentiment analysis on two of top leading cloud computing providers to identify the opinion of customers around each one of them, take useful information that helps in marketing and compare their results.

As results showed that Azure's positive polarity was higher while AWS negative polarity was higher. This means that for example an announcement in Azure page would circulate to higher number of satisfied people than in A WS page. As for emotion classification A WS "unknown" category was exactly same as Azure's one with 79%. While "joy" category was 8% for A WS and 10% for Azure which represents slight difference. "Surprise" category, was 7% for A WS but 5% for Azure, which is also a slight difference. This indicates that customers of both providers are alm ost at the same level of satisfaction. Other categories differs in slight percentages, which can be considered insignificant.

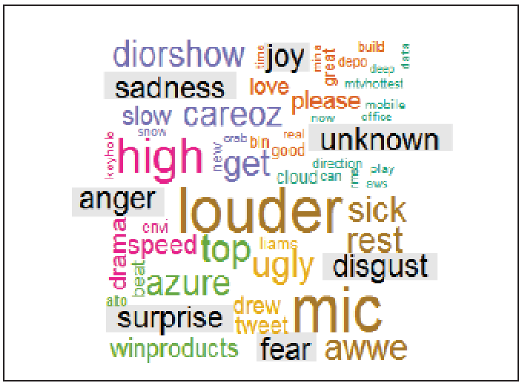

Fig. 5 Word cloud of Azure emotions classifications

Finally, it can be concluded that A WS is launching a page on twitter for marketing and professionalism purposes but higher attention would make it more effective for them. While, Azure can benefit from their twitter page by increasing advertising offers to their customers to gain more customer satisfaction and loyalty. They also, can take benefit from the huge amount of customers connecting to their page by increasing their posts, tutorials, stories of success of some current customers, etc. Such methods can attract new customers and increase loyalty of current customers.

## REFERENCES

[1] Kwon, 0., & Wen, Y. "An empirical study ofthe factors affecting social network service use. Computers in human behavior", 26(2), 254-263, 2010.

[2] Cyworld statistical report. Retrieved April 19, 2015, fromhttp://www.cyworld.com

[3] SmithaW., Kidderb D. "You've been Tagged! (Then again maybe not): Employers afId facebook. Business Horizons", 5(53), pp. 491-499. Available through: Elsevier database. 2010.

[4] Campbell, W. M., Dagli, C K., & Weinstein, C 1. "Social Network Analysis with Content and Graphs". Lincoln Laboratory Journal, 20(1), ,2010.

[5] Ya-li C, Wen-dong W., Xiaflg-Yaflg G., Yu-hong L., CafI-feng C, Jiafl M., "Mobile Ecommerce Model Based on social Network Analysis". The Journal of China Universities of Posts afId Telecommunications [ejournal] (15), pp. 79-83,97,2008.

[6] Watersa R., Burnettb E., Lammb A., Lucasb J., "Engaging Stakeholders through Social Networking: How nonprofit Organizations are using facebook. Public Relations Review", 2(35), pp. 102-106,2009.

[7] Langheinrich M., Ka~jothb G., "Social Networking and risk to compaflies afId institutions. Information Security Technical Report". 2010.

[8] LaMonica, M. "Amazon web services adds resiliency to EC2 compute service", 2010.

[9] Macias, M., & Guitart, 1. "A genetic model for pricing in c10ud computing markets". In Proceedings of the 2011 ACM Symposium on Applied Computing (pp. 113-118). ACM, 2011.

[10] Cloudyn Ebook: https://www.c1oudyn.comlwpcontentluploads/2014/08/Who-moved-my-c1oud-ebook. pdf

[l1] Auction Pricing System: https://www.purecommerce.com/dictionary/ecommerce/Auction_Pricing_System.cfm

[12] Witten, I. H., Don, K . .I., Dewsnip, M., & Tablan, V. 'Text mining in a digital library". International Journal on Digital Libraries, 4(1), 56-59,2004.

[13] Tane, 1., Schmitz, C., & Stumme, G. "Semantic resource management for the web: an e-Iearning application". In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters(pp. 1-10). ACM, 2004.

[14] M'hammed Abdous, He, W., & Yen, C. 1. "Using data mining for predicting relationships between online question theme and final grade". Journal of Educational Technology & Society, 15(3), 77-88, 2012.

[15] Li, N., & Wu, D. D. "Using text mining and sentiment analysis for online forums hotspot detection and forecast". Decision Support Systems, 48(2), 354-368,2010.

[16] Thomas, 1., McNaught, 1., & Ananiadou, S. "Applications oftext mining within systematic reviews". Research Synthesis Methods, 2(1), 1-14, 2011.

[17] Kasture, N. R., & Bhilare, P. B. "An Approach for Sentiment Analysis on Social Networking Sites". In Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on (pp. 390-395). IEEE, 2015.

[18] Liu, B .. "Web data mining: exploring hyperlinks, contents, and usage data". Springer Science & Business Media, 2007.

[19] Dickinson, 8., & Hu, W. "Sentiment Analysis of Investor Opinions on Twitter". Social Networking, 4(03), 62, 2015

[20] Candelieri, A., & Archetti, F. "Detecting events and sentiment on Twitter for improving Urban Mobility. Emotion and Sentiment in Social and Expressive Media".

[21] .10, Y., & Oh, A. H. "Aspect and sentiment unification model for online review analysis". In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 815-824). ACM, 2011.

[22] .Iin, W., Ho, H. H., & Srihari, R. K. "OpinionMiner: a novel machine learning system for web opinion mining and extraction". In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1195-1204). ACM, 2009.

[23] Wilson, T., Wiebe, .1., Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: Procs. of HL T -EMNLP, 2005.

[24] Strapparava and A. Valitutti. "WordNet-Affect: anaffective extension of WordNet". In Proc. of 4 the Interna-tional Conference on Language Resources and Evalua-tion (LREC 2004) , pages 1083 - 1086, Lisbon, May,2004.

[25] Liu, X., & Pedrycz, W. "The development of fuzzy decision trees in the framework ofaxiomatic fuzzy set logic". Applied Soft Computing, 7(1), 325-342,2007.

[26] Kantardzic, M. "Data mining: concepts, models, methods, and algorithms" . .lohn Wiley & Sons.

[27] Kim, S. B., Han, K. S., Rim, H. C., & Myaeng, S. H. "Some effective techniques for naive bayes text c1assification. Knowledge and Data Engineering", IEEE Transactions on, 18(11), 1457-1466,2006.

[28] Mozina, M.; Demsar, 1.; Kattan, M.; Zupan, B. "Nomograms for Visualization ofNaive Bayesian Classifier". Proc. PKDD-2004. pp. 337-348,2004.

[29] Martin, .l.H. and .Iurafsky, D., 2015. Speech and language processing. international Edition.

**B Anil Kumar** has obtained his M.C.A Degree in the year 2005 from the prestigious Osmania University, Hyderabad. He is pursuing his M.Tech Degree from SSJ Engineering College. He is currently working as corporate trainer MNC's

**Aakuthota Ravi Kumar** has obtained his B.Tech Degree in the Year 2001 from the prestigious Karnataka University and M.Tech Degree in the year 2005 from Bharath University. Currently he is pursuing his PhD in the field of Software Engineering. Presently he is working as a Associate Professor and Head of CSE department at SSJ Engineering College. He has published almost 12 Papers in the reputed International Journals.