IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-5, Issue-4 E-ISSN: 2347-2693

A Systematic Literature Review of Sentiment Analysis Techniques

J. Kaur[1*], S.S. Sehra[2], S.K. Sehra[3]

[1*] Dept. of CSE, Chandigarh University, Chandigarh, India
[2]Dept. of CSE, Guru Nanak Dev Engineering College (I.K. Gujral PTU), Ludhiana, India
[3]Dept. of CSE, Guru Nanak Dev Engineering College (I.K. Gujral PTU), Ludhiana, India

[*]Corresponding Author: jasleenbhullar66@gmail.com, Tel.: +91-9592526960

www.ijcseonline.org

Received: 19/Mar/2017, Revised: 02/Apr/2017, Accepted: 20/Apr/2017, Published: 30/Apr/2017

Abstract— Development of Web 2.0 has resulted in enormous increase in the vast source of opinionated user generated data. Sentiment Analysis includes extracting, grasping, arranging and presenting the feelings or suppositions communicated in the information gathered from the clients. This paper exhibits an efficient writing survey of different strategies of sentiment analysis. A model for sentiment analysis of twitter data using existing techniques is constructed for comparative analysis of various approaches. Dataset is pre-processed for noise removal and unigrams as well as bigrams are used for feature extraction with term frequency as weighting criteria. Maximum accuracy is achieved by using a combination of SVM and Naïve Bayes at 78.60% employing unigrams and 81.40% employing bigrams as features.

Keywords— Sentiment Analysis, Crowdsourced data, Twitter, Machine Learning Techniques.

I. INTRODUCTION

The technology stack of Web 2.0 tools, encouraged individuals to contribute their idea or persuasion on web that enhanced the extent of data which is being generated by the users. Ubiquity of this stack has made intuitiveness on the web a great deal and worthy [1]. The client produced content made through these has been a believable hotspot for investigation of both subjective and factual data. The significance of content mining tools is developing equal to the rampant development of electronic content [2]. This data is present in various structures as blog entries; microblog posts etc. and demonstrated the simplicity of observing assessments in regards to any brand, one can see miniaturized scale blogging as a promising wellspring of competitive intelligence [3]. Microblogging sites, portrayed by a free organization of messages and also simple openness, tend to move Internet clients to Microblogging sites, for example, Twitter from customary stages, for example, blogs.

Due to dynamic and unpredictable nature of Twitter, the sentiment areas of interest cannot be anticipated in advance and become evident only as any real-world event unfolds. Sentiment Analysis involves elicitation, comprehension and classification of the feelings or conclusions communicated by the clients or users. Thus, Sentiment examination framework, takes any archive as information, investigates and produces a definite answer for the feelings elaborated in information report. The machine based identification of sentiments are vital for many application domain in today's era. It gives organizations offering items with intend to examine published reviews and gauge the degree of item acknowledgment. For customers, the superfluity of opinions from distinct sources helps them to use wisdom of crew, and in settlement of more educated and better choices [4].

Generous increment in portable and tablet applications in the current years has engaged clients to remark and examine on social networking platforms whenever they want through distributed computing, mobiles, Internet [5]. However the real test in exploring client created information lies in the mind boggling extensiveness of covered point. The short length of microblogs infers any assumption covered by them is minimal and in this way introduces many difficulties in its extraction. However, for identifying answer to particular question, this classification of the sentiments of content as positive or negative based on its positive or negative assumptions with respect to that inquiry is strenuous. Classifier trained in a particular area may or may not perform well when connected to some other domain inferable from befuddle between the domain particular words. Different difficulties are forced by amusing expressions or mockery. Amusing/wry statements should be identified for deducing the genuine importance or genuine focus of the creator. This in turn increases the accuracy [6]. A negative slant or explanation can be communicated in an inconspicuous way without containing a solitary word that appears to be clearly negative. Likewise, if nullification happens in the content, the extremity is evaluated to the inverse of the extremity of content. But in some cases, certain expressions that join refutation words strengthen as opposed to evolving extremity.

Section II of this paper puts forward the objectives of carrying out this review of sentiment analysis techniques. Section III explains the current status of contribution in sentimental analysis for answering our examination objectives. Section IV and section V explain in detail the methodology used for comparative analysis of techniques and results achieved respectively. At last, conclusion of this review has been presented in section VI.

II. OBJECTIVES OF THIS REVIEW

Our target of summarizing the exploration that has been executed in the past in sentiment analysis is separated into three research questions:

Q1- What are the different element extraction and weighting strategies that have been utilized as a part of the past?

Q2- What supervised and unsupervised methods used for sentiment analysis?

Q3- How do these methods handle different difficulties e.g. area independency or mockery in slant investigation?

III. SENTIMENT ANALYSIS METHODS

The following section explains the current status of contribution in sentimental analysis to answer our examination objectives.

Q1 - Extraction of features

Features regarding opinion mining comprise the terminology or language which heartily passes on the assessment of the content either as positive or as negative. It can be deduced that components affect the course of given content when contrasted with different words which have been actually displayed in that content.

Classification a word in a corpus according to its linguistic component, on start of its definition or setting is insinuated as POS labeling. Bermingham and Smeaton [7] noticed syntactic illustrations which are spoken to by n-gram highlights and Part of Speech contain more genuine specifications than unigrams. They concluded that Support Vector Machine beat Multinomial Naïve Bayes in thought gathering of online diaries however the opposite was observed in microblogs which is opposite to the revelations of Go et al [8]. Pak and Paroubek [9] used the appropriation in the assignment of POS names between two sets for perceiving messages as subjective or objective. MNB classifier using POS-marks with N-gram achieved the highest accuracy. Highlights consolidating earlier extremity of words or expressions with their particular POS labels were observed to be most vital for characterization assignments [10]. The precision using POS names as parts upgraded for Naïve Bayes, decreased in SVM and remained same with MaxEnt [11].

N gram is an adjacent connection of n terms from any corpus of content or address. Diverse estimations of N affect the precision of analysis or classification to different extents. Consolidating an N-gram classifier with an already existing learned feeling marker enhanced execution significantly [12]. Akaichi et al [13] attempted different mixes of N-grams as components and the subsequent accuracies were looked at. They established that Support Vector Machine and Multinomial Naïve Bayes achieved their most astounding exactness by consolidating unigrams and bigrams. Further, less accuracy was achieved by mix of these, which is not in line with the findings of Zhai et al [14]. Zhai et al [14] concluded that bigrams tend to outperform other N-gram features. Abbasi et al [15] presented Feature Relation Network, an administered technique for content element choice that could choose highlights in a more effective way and beat univariate, multivariate and hybrid element determination strategies. It considered syntactic connections and semantic measurements among different N-gram highlights.

TF-IDF is a basic measure for mirroring the significance or noteworthiness of a specified term in a gathering of archives or corpus. This causes in allocating weights to different terms in the corpus and hence characterizing better. Pang et al [11] concluded that taking into account occurrence of a feature provides unrivalled outcomes in opinion order than recurrence. Ghag and Shah [16] concentrated on the normal recurrence tally circulation that depended on corresponding recurrence and relative nearness number dispersion differentiating it from conventional methodologies like delta TF-IDF that depended on both recurrence tally of the term and relative nearness check dissemination. Accuracy improved to 71.3% utilizing this approach which indicated change from 66.5% that was achieved by utilizing Delta TF-IDF. Deng et al [17] proposed a managed term weighting plan so as to enhance execution. It was subject to significance of the term in that report and the term's quality to express notion. Their exploratory outcomes demonstrated that their approach beat the past unsupervised methodologies and created the best precision.

Q2 - Existing Supervised and Unsupervised approaches

Opinion mining systems can be for the most part grouped into two distinct groups in particular: Lexicon based methodologies and Machine learning approaches.

Machine Learning methods utilize two different datasets: training dataset and testing dataset for characterization. Other machine learning algorithms such as support vector machine, k-nearest neighbours, artificial neural networks (ANN), naïve bayes, fuzzy rationale, Maximum Entropy Neural systems are some of broadly utilized classifiers. Support vector machine and Naïve Bayes are mainly utilized [7-9, 11, 13, 18, 19]. Jiang et al [20] proposed an improved kNN calculation for arrangement of content by joining compelled one pass grouping calculation with content classification utilizing kNN which ordered test records on premise of bunch vectors as opposed to unique content specimens. They discovered that this technique performed best on high speed, imbalanced, multi-dimensional content when contrasted with SVM, NB. Cambria et al [21] proposed a classifier that depended on consolidated utilization of scaling in numerous measurements and ANN. Coordination of a model propelled from natural handling with standard segment examination gripped the non-linearity of resultant vector space and in this way the instinctive sound judgement and also thinking capacities of the framework under thought were increased. Tekchandani and Dhole [22] dealt with the issue of mis-spellings in the client surveys on web based business destinations by utilizing fuzzy string seeking and Levenshtein distance.

Unsupervised systems are broadly utilized at spots where no or less training information is accessible. Bollen et al [23] completed productive opinion examination by means of a syntactic, term-based approach requiring no learning as opposed to machine learning opinion investigation systems that yield precise grouping only when provided with expansive preparing and testing sets. Bagheri et al [24] presented an unsupervised aspect discovery approach which is independent of domain for conclusion mining of client audits. This model utilized unsupervised training set and investigated the dataset utilizing recurrence based data to discover aspects. The model outflanked the current methods yet could not categorize sarcastic statements.

The technique to map words to their aforementioned extremity expands the exactness of classification [10]. Barbosa and Feng [25] put forward two components sets: meta data about terms and trademark conduct in composing tweets. The term was classified accordingly using its earlier subjectivity and extremity. They presumed that components which were surmised from syntactic elements of tweets alongside Meta data of expressions/words were better than unigrams in assigning subjectivity and extremity. Huang et al [26] proposed a technique for self-regulating administration of field-particular or domain-subjected sentiment dictionary based on restricted label generation. Investigation of chunk dependent processed data based in prior generic lexicon was performed and opinion terms for constructing domain-particular sentiment lexicon were extracted.

The particle swarm optimization (PSO) is an advancement procedure established using development and knowledge of swarms. It is a populace based transformative calculation which depends on conduct of flocks. It is being mostly utilized for solution of optimization problems and feature extraction. Numerous previous researches have optimized SVM and NB with PSO for increasing accuracy. Lin et al [27] utilized this technique for deciding parameters and selecting elements of SVM. Huang and Dun [28] proposed a hybrid SVM-PSO model to enhance the exactness of order with a little component subset. Basari et al [29] utilized PSO to enhance the choice of optimal parameters for explaining the dual optimization issue in SVM.

Q3 - Handling challenges and domain dependency

The significant test of investigating client produced information lies in the mind boggling broadness of theme. Content being created is generally boisterous. Classifier that is prepared in a particular space may not perform well when connected to some other area attributable to crisscross between the spaces of particular words. Different difficulties are forced by unexpected expressions or mockery. Incongruity, mockery, refutation and analogies are critical parts of document that can prompt for augmenting in the extreme. Carvalho et al [30] discussed incongruity in words or expressions and were to analyse that expressions portraying giggling, recurrence of accentuation imprints and citations etc. are most helpful examples. Riloff et al [31] displayed a bootstrapping algorithm that naturally learned arrangements of good and awful circumstance phrases from snide tweets and thus enhanced review for mockery acknowledgment. An analogy is an examination between two not at all like things. Qadir et al [32] introduced an administered arrangement structure for perceiving emotional extremity in analogies utilizing lexical, semantic, and estimation includes for training. Wilson et al [33] concentrated the impact of invalidation on the extremity of the content and found that expansion in features utilized prompted to an expansion in correctness and accuracy.

Sentiment analysis manages finding the feelings of any individual, whereas subject based grouping characterizes distinctive records into comparative/diverse classifications in light of their included content. Pang et al [11] contrasted the consequences of opinion mining and theme based grouping for movie characterization and discovered Feature recurrence performed better for subject based and presence for opinion mining. Ku et al [34] utilized best quality level to assess the sentiment extraction execution at word, sentence and in addition record level and found that the accuracy of machine learning strategies expanded in the wake of applying subject recognition algorithms. However these techniques were not reasonable for sentiment analysis at unit level. Cai et al [35] discussed, procedures to recognize the themes related with positive or negative suppositions for helping examiners to get outline of opinion degree alongside the drivers.

Algorithms or procedures connected in one area may or may not conceivably perform well in different spaces. Subsequently area particular and space free systems have been considered independently. Jambhulkar and Nirkhi [36] conducted a study on Cross-Domain Sentiment Analysis. They concentrated on these methods: spectral feature alignment, sentiment sensitive thesaurus and structural correspondence learning and discovered them diverse in the way of enhancing the element vector, assessing relationships between words and afterward in the technique utilized. Bisio et al [37] handled two primary parts of the notion characterization: multifaceted nature of the given structure and its capacity to work in broadened business spaces by utilizing semantic systems, relevant valence shifters and a predictive model based on distance.

O'Hare et al [38] performed area particular notion investigation of monetary sites. They utilized DiffPost calculation for examination that used the way that the undesirable substance or commotion would be reproduced over various articles whiles the significant and in this manner helpful content would be particular to that specific article. The best outcomes were accomplished by utilizing content extraction approaches in light of words and binary weighting plan. Integrating topic-based text extraction techniques showed effectiveness over usual technique of utilizing the whole document for training as well as testing. Schumaker et al [39] actualized a machine learning technique that was empowered to learn value developments in the past utilizing particular terms and passionate words as components and subsequently assembled a model for value forecast of to be published news articles. They discovered subjectivity as an essential angle as financial specialists responded all the more emphatically to negative articles.

IV. METHODOLOGY

This segment displays the procedure utilized as a part of the present work for Sentiment Analysis of Twitter information on the premise of existing strategies.

A. Data Collection

Twitter is an eminent micro-blogging website. This work utilizes Twitter to gather tweets communicating conclusions or estimations with respect to particular themes. Twitter gives an API (Application Programming Interface) that aide in getting to tweets in an automatic way by utilizing a word as question. Twitter utilizes OAuth for giving a secured and approved access to its API. OAuth is a convention for confirmation which permits distinctive clients to concur for operation of utilization for their sake without uncovering password. We have confined the Twitter API to force 1000 tweets at any given moment.

B. Pre-processing

The following pre-processing steps have been incorporated in our work in order to improve accuracy:

- **Punctuation Erasure:** It allows erasing punctuation marks like the period, exclamation point, comma, apostrophe, question mark, quotation mark and hyphen.
- **Number Filter:** It filters all those terms that consist only of numbers.
- **N chars filter:** It filters all those terms that consist of words with less than the pre-specified number of 3 characters.
- **Case Converter:** It converts all the terms present in the text to lower case.
- **Stemmer:** It stems the terms present in the text by the stemmer algorithms which are provided by the Snowball stemmer library. Porter snowball stemmer is used in our project.
- **Filtering stop words:** It filters all the terms that represent stopwords like the, is, on, at etc.

C. Feature creation

Features employed in the present work are:

- **Creation of Bag of words:** A Bag of Words consist of two columns, one containing documents and the other containing the terms occurring in the corresponding document.
- **N grams:** N grams of words are extracted from the documents and their frequencies are calculated.
- **Term Frequency:** The relative term frequency of each term is computed according to each

document and a column containing the term frequency value is added to the data table. This value is calculated by dividing the absolute frequency of each term according to a document by sum of all terms.

- **Row Filter:** It is used to keep only those terms that occur in minimum 2% of the total documents in the corpus.

D. Classification

The present work employs different algorithms like Support Vector Machine (SVM), Naïve Bayes (NB), Decision trees, Neural Networks, Ensemble learning using bagging and PMML with Support Vector Machine and Naïve Bayes at different sizes of the corpus. Training set is given to the corpus for learning and then the accuracy of different classifiers on the test set are compared to each other.

V. RESULTS AND DISCUSSION

Three data indexes comprising 1000, 2000 and 10000 tweets were contemplated. Expanding the measure of preparing dataset from 1000 to 2000 extensively enhanced the exactness of the considerable number of classifiers. The accuracy utilizing unigrams and bigrams as components with different approaches for a set of 2000 tweets is illustrated in Figure 1. The outcomes utilizing bagging strategy for various classifiers are presented in Figure 2. Figure 3 shows the accuracy utilizing PMML with SVM and NB.

SVM, Neural Network and Decision Tree played out their finest with training sets of 2000 tweets while NB achieved its most astounding exactness utilizing 10000 tweets. Subsequently an expansion in the training set has its most noteworthy effect on the execution of Naïve Bayes. Additionally, the correctness of classifiers for a similar text was diverse for bigrams and unigrams. Given a training set containing 2000 tweets, the accuracy by utilizing unigrams acquired by NB, SVM, Neural Network and Decision Tree added up to 69.70%, 73.70%, 69.10% and 67.60% and by utilizing bigrams to 68.80%, 74.70%, 69% and 71.50% respectively. Utilization of Bagging (Ensemble learning procedure) affected exactness of all calculations with unigrams and in addition bigrams. Bagging enhanced the accuracy for SVM (unigrams-73.90%, bigrams-76.40%) and Decision Tree (unigrams-74.60%, bigrams-71.50%) barely yet prompted to a decline in the exactness of NB (unigrams-69.30%, bigrams-68.60%).

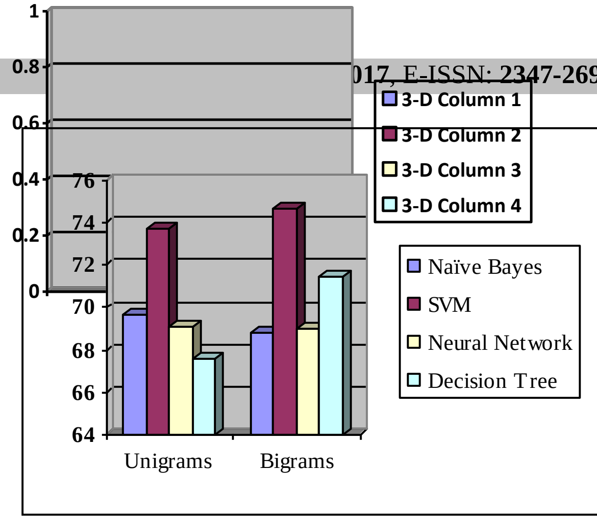

Figure 1. Accuracy for different classifiers for unigrams and bigrams

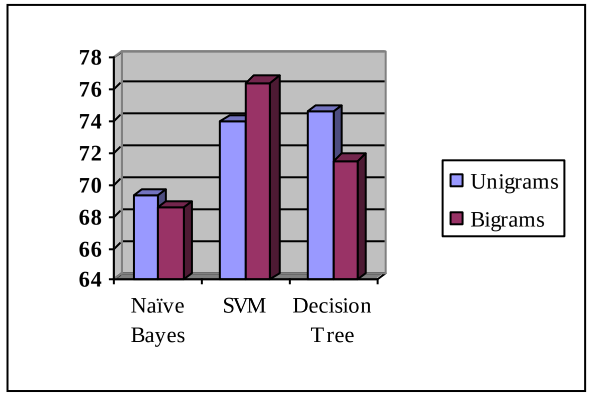

Figure 2. Accuracy for different classifiers using bagging

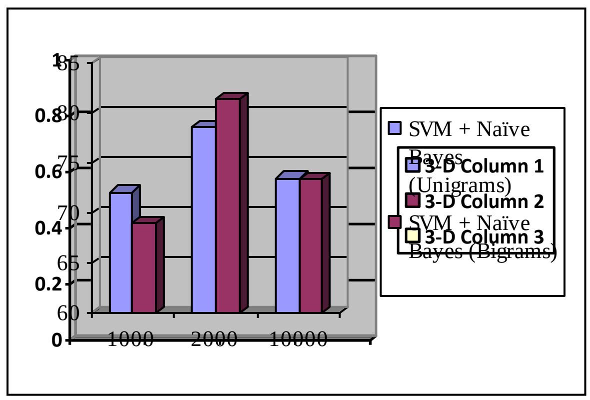

Figure 3. Accuracy using hybridization of SVM and NB

Nonetheless, the most noteworthy accuracy is achieved by utilizing both SVM and NB at 78.60% with unigrams as features and 81.40% with bigrams. With huge dataset i.e. corpus comprising of 10000 tweets, no distinction appears utilizing unigrams or bigrams.

VI. CONCLUSION

Given the tremendous and quickly developing means of generating information by the sharing-period of Web 2.0, examining retrieval of information like mining clients' goals and sentiments about various terms at various levels of granularity needs to be clearly studied [40]. Sentiment mining is essentially a utilization of NLP (Natural Language Processing) utilizing computational etymology and content mining with a specific end goal to foresee conclusion of content notion, mostly as positive or negative [41]. Dictionary based and machine learning approaches have been widely utilized as a part of this field in the past. This paper shows an efficient writing audit of slant examination methods and their correlation utilizing a model. Correctness of various classifiers approaches are processed as well as introduced. For choosing feature, Bigrams turned out to be more successful when compared to Unigrams. Utilizing Bagging builds the exactness of Support Vector Machine and Decision Tree however diminishes the same in Naïve Bayes. Highest accuracy of 81.40% is acquired by utilizing hybridization technique for SVM and NB using bigrams.

REFERENCES

[1] H Lee, P Ferguson, N OHare, C Gurrin, AF Smeaton, "*Integrating interactivity into visualising sentiment analysis of blogs"*, In the Proceedings of the first international workshop on Intelligent visual interfaces for text analysis, USA , pp 17-20, 2010.

[2] S Agarwal, S Godbole, D Punjani, S Roy, "*How much noise is too much: A study in automatic text classification*", In the proceeding of Seventh IEEE International Conference on Data Mining, USA, pp. 3-12, 2007.

[3] BJ Jansen, M Zhang, K Sobel, A Chowdury, "*Micro-blogging as online word of mouth branding",* In: CHI'09 Extended Abstracts on Human Factors in Computing Systems, USA, pp. 3859-3864, 2009.

[4] V Sindhwani, P Melville, "*Document-word co-regularization for semi-supervised sentiment analysis*", In the proceeding of Eighth IEEE International Conference on Data Mining, Itly, pp. 1025-1030, 2008.

[5] L Zhang, K Hua, H Wang, G Qian, L Zhang, "*Sentiment Analysis on Reviews of Mobile Users*", Procedia Computer Science , Vol:34, Issue.3, pp.458-465, 2010.

[6] K Buschmeier, P Cimiano, R Klinger, "*An impact analysis of features in a classification approach to irony detection in product reviews*", In the Proceedings of the 5th Workshop on Computational Approaches to Subjectivity Sentiment and Social Media Analysis, Maryland, pp. 42-49,2014.

[7] A Bermingham, A F Smeaton, "*Classifying sentiment in microblogs: is brevity an advantage?"* In the Proceedings of the 19th ACM international conference on Information and knowledge management, NY, pp.1833-1836,2010.

[8] A Go, R Bhayani, L Huang, "*Twitter sentiment classification using distant supervision*", CS224N Project Report of Stanford, Vol.1, Issue.12, pp.1-23, 2009.

[9] A Pak, P Paroubek, " *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*", LREC, Vol.10, Issue.4, pp.1320-1326, 2010.

[10] C. Nanda, M. Dua, "*A Survey on Sentiment Analysis*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.2, pp.67-70, 2017.

[11] D. Chafale, A. Pimpalkar, "*Review on Developing Corpora for Sentiment Analysis Using Plutchik's Wheel of Emotions with Fuzzy Logic*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.10, pp.14-18, 2014.

[12] M. Fernandes , "*Data Mining: A Comparative Study of its Various Techniques and its Process*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.1, pp.19-23, 2017.

[13] J Akaichi, Z Dhouioui, MJ Lopez, "*Text mining facebook status updates for sentiment classification*", In 17th International Conference on System Theory Control and Computing (ICSTCC), USA, pp 640-645,2013.

[14] Z Zhai, H Xu, B Kang, P Jia, "*Exploiting effective features for chinese sentiment classification*",, Expert Systems with Applications, Vol.38, Issue.8, pp.9139-9146, 2011.

[15] A Abbasi, S France, Z Zhang, H Chen, "*Selecting Attributes for Sentiment Classification Using Feature Relation Networks*", IEEE Transactions on Knowledge and Data Engineering , Vol.23, Issue.3, pp. 447-462, 2011.

[16] KV Ghag, K Shah, "*ARTFSC-Average Relative Term Frequency Sentiment Classification*", INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY, Vol.12, Issue.6, pp.3591-3601, 2014.

[17] ZH Deng, KH Luo, HL Yu, "*A study of supervised term weighting scheme for sentiment analysis*", Expert Systems with Applications, Vol.41, Issue.7, pp. 3506-3513, 2014.

[18] A Amolik, N Jivane, M Bhandari, M Venkatesan, "*Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques*", International Journal of Engineering and Technology Vol.7, Issue.6, pp. 2038 - 2044, 2015.

[19] V. Kapoor, "A New Cryptography Algorithm with an Integrated Scheme to Improve Data Security", International Journal of Scientific Research in Network Security and Communication, Vol.1, Issue.2, pp.39-46, 2013.

[20] S Jiang, G Pang, M Wu, L Kuang, "*An improved K-nearest-neighbor algorithm for text categorization",* Expert Systems with Applications, Vol.39, Issue.1, pp. 1503-1509, 2012.

[21] E Cambria, T Mazzocco, A Hussain, "*Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining*", Biologically Inspired Cognitive Architectures, Vol.4, Issue.3, pp. 41-53, 2013.

[22] MP Tekchandani, M A Dhole, "*Opinion mining using fuzzy string searching",* IJARCCE, Vol.4, Issue.4, pp.526-530, 2015.

[23] J Bollen, H Mao, A Pepe, "*Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena*", ICWSM, Vol.11,, Issue.2, pp. 450-453,2011.

[24] A Bagheri, M Saraee, F de Jong, "*Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews",* Knowledge-Based Systems. Vol.52, Issue.2, pp. 201-213, 2013.

[25] L Barbosa, J Feng, "*Robust sentiment detection on twitter from biased and noisy data",* In the Proceedings of the 23rd International Conference on Computational Linguistics: Posters, China, pp 36-44,2010.

[26] S Huang, Z Niu, C Shi, "*Automatic construction of domain-specific sentiment lexicon based on constrained label propagation*", Knowledge-Based Systems, Vol.56, Issue.9, pp.191-200, 2014.

[27] SW Lin, KC Ying, SC Chen, ZJ Lee, " *Particle swarm optimization for parameter determination and feature selection of support vector machines",* Expert systems with applications, Vol.35, Issue.4, pp.1817-1824, 2008.

[28] CL Huang, JF Dun, "*A distributed pso-svm hybrid system with feature selection and parameter optimization",* Applied Soft Computing, Vol.8, Issue.4, pp.1381-1391,2008.

[29] ASH Basari, B Hussin, IGP Ananta, J Zeniarja, "*Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization*", Procedia Engineering Vol.53, Issue.2, pp.453-462, 2013.

[30] P Carvalho , L Sarmento, MJ Silva, E De Oliveira, "*Clues for detecting irony in user-generated contents: oh...!! its so easy;-)",* In the Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, China, pp 53-56, 2009.

[31] E Riloff, A Qadir, P Surve, L De Silva, N Gilbert, R Huang, "*Sarcasm as Contrast between a Positive Sentiment and Negative Situation*", EMNLP, Washigtan, pp.704-714, 2013.

[32] A Qadir, E Riloff, MA Walker, "*Learning to Recognize Affective Polarity in Similes",* In: EMNLP, Washigtan, pp.23-34, 2015.

[33] T Wilson, J Wiebe, P Hoffmann, "*Recognizing contextual polarity in phrase-level sentiment analysis",* In the Proceedings of the conference on human language technology and empirical methods in natural language processing, China, pp.347-354,2005.

[34] LW Ku, YT Liang, HH Chen, " *Opinion Extraction, Summarization and Tracking in News and Blog Corpora",* In: AAAI spring symposium on Computational approaches to analyzing weblogs, China, pp.100-107, 2006.

[35] K Cai, S Spangle, Y Chen, L Zhang, " *Leveraging sentiment analysis for topic detection*", International Journal of Web Intelligence and Intelligent Agent Technology, Vol.8, Issue.3, pp.265-271, 2008.

[36] P Jambhulkar, S Nirkhi, "*A survey paper on cross-domain sentiment analysis*",, Int J Adv Res Comput Commun Eng, Vol.3, Issue.1, pp.5241-5245, 2014.

[37] F Bisio, P Gastaldo, C Peretti, R Zunino, E Cambria, " *Data intensive review mining for sentiment classification across heterogeneous domains*", In: International Conference on Advances in Social Networks Analysis and Mining, Canada, pp 1061-1067, 2013.

[38] N O'Hare, M Davy, A Bermingham, P Ferguson, P Sheridan, C Gurrin, AF Smeaton, "*Topic-dependent sentiment analysis of financial blogs",* In: Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, China, pp 9-16, 2009.

[39] RP Schumaker, Y Zhang, CN Huang, H Chen, "*Evaluating sentiment in financial news articles*", Decision Support Systems, Vol.53, Issue.3, pp.458-464, 2012.

[40] L Robaldo, L Di Caro, "*OpinionMining-ML*", Computer Standards & Interfaces, Vol.35, Issue.5, pp.454-469, 2013.

[41] MM Mostafa, "*More than words: Social networks' text mining for consumer brand sentiments",,* Expert Systems with Applications, Vol:40, Issue.10, pp. 4241-4251, 2013.

Authors Profile

Ms. Jasleen Kaur pursued Bachelor of Technology in computer science and engineering from Guru Nanak Dev Engineering College, I.K. Gujral PTU, India in 2006 and Master of Technology in computer science and engineering from Guru Nanak Dev Engineering College, I.K. Gujral PTU, India in 2014. She is currently working as Assistant Professor in Department of Computer Science and engineeering in Chandigarh University, Chandigarh, India since 2016. Her main research work focuses on Big Data Analytics, Data Mining and Sentiment analysis.

Mr Sukhjit Singh Sehra pursed Bachelor of Technology from Guru Nanak Dev Engineering College, Ludhiana in 2002 and Master of Technology from Punjab Agricultural University, Ludhiana in year 2005. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science & Engineering Guru Nanak Dev Engineering College, Ludhiana since 2006. He is a member of various professional bodies at National and International level. He has published more than 40 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. His main research work focuses on Spatial data analysis, Big dta analytics and, Data Mining.

Mr Sumeet Kaur Sehra pursed Bachelor of Technology from Guru Nanak Dev Engineering College, Ludhiana in 2002 and Master of Technology from Punjab Agricultural University, Ludhiana in year 2005. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science & Engineering Guru Nanak Dev Engineering College, Ludhiana since 2006. He is a member of various professional bodies at National and International level. He has published more than 48 research papers in reputed international journals. Her research area focuses on Software Engineerin, Project Planning.

