Opinion Mining and Client Feedback Mining for Quality Improvement

J. Vijaya Chandra, G. Ranjith, B. Bhagya Lakshmi, Bhattu Murali Manohar Raju, Mohammad Azaruddin.

Abstract: As E-Commerce growing rapidly, now a day's customers reviews about product has become most crucial for development of the business. The opinion mining is most crucial in ecommerce websites, furthermore advantageous with client's feedback. The current research work mainly focuses on the client engagement review analysis. In this paper we concentrated on mining reviews from clients who are seeking consulting services from the IT industries, which enables the companies to understand the data patterns and identify some Key Performance Indicators and develop accordingly. The Client feedback mining allows the user to allow Clients to convey what they think and feel about the services. This opinion or review about the product will help purchaser to get an idea regarding the product, and the seller regarding required improvements or updates, hence opinion mining plays a major role in ecommerce. We also concentrated on the sentimental analysis that means a reviewer can write about the product positively or negatively, it depends on many parameters such as emotions they want to express, opinion of the product they purchased, received time, package and condition they received the product. It clearly gives an idea to the seller and purchaser about demand, supply and quality improvement. Key Performance Indicator is measured based on metrics that business organizations track in order to measure their progress towards goal within marketing networks. We designed a mathematical model and executed for results. We discussed different algorithms based on opinion mining and we implemented cloud based practical implementation of a simulated model for understanding of results and given graphical analysis along with result analysis.

Keywords: opinion mining, marketing, ecommerce, feedback mining, data mining.

J. Vijaya Chandra, HOD and Assistant Professor, Department of CSE, Warangal Institute of Technology and Science. Warangal, Telangana, India. Email: vijayachandra.phd@gmail.com

Dr. Ranjith Gyaderla, Vice Principal and Associate Professor, Department of CSE, Warangal Institute of Technology and Science, Warangal, India. Email: gyaderlaranjith@gmail.com

B. Bhagya Lakshmi, Assistant Professor, Department of CSE, Warangal Institute of Technology and Science, Warangal, India. Email: bhagya712@gmail.com

Bhattu Murali Manohar Raju*, M.Tech Scholar, Department of CSE, Warangal Institute of Technology and Science. Warangal, Telangana, India. Email: bmmraju83@gmail.com

Mohammad Azaruddin, M.Tech Scholar, Department of CSE, Warangal Institute of Technology and Science. Warangal, Telangana, India. Email: Azaruddin.testing@gmail.com

I. INTRODUCTION

The Opinion Mining or sentiment analysis comes under the type of Natural Language Processing (NLP), Text-analysis and Computational-Linguistics which are major concepts to identify key words using word alignment models and extract subjective information such as opinion words in source material. Data mining is an analytical exposure for words for opinion targets with dynamic data analysis, in which opinion mining deals with the associated services such as feedbacks, reviews, comments and rating. Data mining retrieves the hidden text even when the information is in the form of patterns or emojis such as smiley, angry, thumb-up or down. The research problem is identifying and extracting the opinion targets, levels of sentiment analysis such as word level, sentence level and finally the document level, even though major words and opinion target focuses on alignment of words and opinion words. The different levels in opinion mining are retrieval, classification and summarization, based on the keywords regarding the opinion the data will be retrieved, where some data mining and classification algorithms are implemented on the retrieved data, finally the opinion mining analysis is done by summarizing the opinion. The review or feedback given by the ecommerce customer will be either positive or negative, based on the frequency of assurances the opinion summary is created. For regular feed back analysis and ranking procedure is supported by the datamining algorithms such as Navies Bayes, logistic regression and Senti-WordNet. The steps involved in the opinion mining and sentiment analysis. text extraction, source code extractor, list of products, display review list, stop word dictionary, algorithm selection, calculate performance, display the classification result and finally positive and negative opinion dictionary. When opinion mining is done related to datasets using different machine learning algorithms for retrieving the results and classification of data in the dataset, where the datasets contains the opinions of authors, general citations, comments, summarized data points, reviewers opinion, contents of review reports, journal citations in computational perspective [1].

Opinion mining provides knowledge to the seller on different aspects such as "market prediction", "product pricing", "competitive intelligence", recently election forecasting also added to the opinion mining which benefits the political leaders, in the area of the banking also implementing of customer and bank relationship, predicting risk management and loan approvals . The opinion mining is also used in various fields for the quality improvement process. Now a day's online review sites are playing major roles in collecting the opinion mining data and sentiment analysis.

In modern e-commerce world the opinion mining is used to enhance their business, by knowing the rating for further improvements in customer services. Research on sentiment analysis is done by identifying affective words from the statements that are responsible to formulate an opinion by extracting individual statements or keywords such as good, needs improvement, excellent, poor, bad, average and ordinary on particular subject, sentiment analysis is a vast field that is inter-related with the Artificial Intelligence, Natural Language Processing, Data Mining and Machine Learning. The opinion targets are the objects on which a customer express his expressions or opinion to give view about the sentiment about the object [2].

II. RELATED WORK

E-Commerce and online shopping majorly depends on the online reviews that is opinions of the persons who already used or purchased the product, where these reviews increase the confidence towards the reliability of the product by the purchaser but most of the online reviews now a days are fake or spam's. The major task is fine grained opinion mining, the key component of which involves detecting opinion relations among words, knowledge-based system with co-ranking algorithms are associated for extracting opinion words and for the ability to determine the legitimate reviews and spam or malicious reviews. With the rapid development in E-commerce, E-marketing and E-Banking, an enormous number of product reviews that obtain the opinion of customer can get product information that is qualities, characteristics, reliability of the product. It also helps the manufactures to obtain immediate feedback on the product so that they can do changes or updates the product as per the comments or reviews. The fake or malicious reviews on internet are misguiding the customer and producer, which is becoming a challenging problem to be solved by the researchers. we compare using syntax-based method by extracting opinion targets where these are nouns or noun phrases. Mining opinions is the major task for the improvement of the ecommerce, it became a research issues and many researchers are attracted towards this problem.

The objective of opinion mining is defined as the object about which users express their opinions, where a customer and producer should get a clear picture regarding the product. The most adaptive technique is implementing nearest neighbor rules that is nearest adjective or verb to a noun or noun phrase and syntactic patterns is that the opinion relations among the words. With the rapid increase in information flow through different mediums such as the social networking, websites, blogs, you tube videos and many other communication mediums through internet, made classification of reviews became most important [3]. Online Customer reviews determines the business over internet, most of the customers read reviews before taking decision regarding purchasing an item either online or offline, now reviews became key factors and they are essential for e-commerce. Data mining is a mechanism that is used for extraction of meaningful and useful form of knowledge information from a given set of data. The knowledge discovery deploys most of the principles and methods for retrieval of required data. There is a great need of mining and summarizing data as the product sold in different sites and opinions will be at different sites, as the opinions are needed not only for the customer but it is also

required for the manufacturer, as the manufacturer does not produce only one kind of product and he will not sell it on only at one site. We concentrated on the syntactic patterns, partially supervised word alignment model, nearest neighbor rules and error propagation. To change the opinions generated by the customers, the rival companies may use unethical methods by observing and collecting information and resources of desired websites, using social engineering methods and psychological tricks the attacker tries to change the public opinion. [4]. The role of Sentiment analysis is identifying or extracting the subjective information, majorly concentrates on the classifying opinions in the text of reviews which may be either positive or negative and it may be even neutral. It is a survey analytics based on the reviews on web. Mathematical Evaluation is done for the mechanism for co-extracting opinion targets and opinion targets and opinion words by using a word alignment model. The Graphical analysis is given to get opinion relation graph and final analysis is given along with results on beneficiaries for extracting opinion targets and opinion words [5].

A. Types of Opinion Mining

Based on the Natural Language Processing (NLP) the text information is generally classified as facts and opinion; facts are objective expressions and about the product whereas the opinion is the subjective information about the object. The NLP problems are identified as the classification problems which means that two sub problems must be resolved. They are subjectivity classification and polarity classification. The subjectivity classification classifies a sentence as a subjective or objective, where as the polarity classification is a sentence expressing positive, negative or neutral [6]. The other types of opinions are Directive or comparative opinion, when the object is only one to describe then it is called directive opinion, if the objective is two or more and they are compared then the opinion is called as comparative opinion.

B. Opinion Mining Use Cases, Models and Patterns

i. Nearest – Neighbor Rules: It is mainly used in the field of pattern or text recognition for classification and regression. These are broadly used in the opinion retrieving and opinion classification, but only considering opinion relations are not sufficient as the text recognition while classification may not give the accurate results, for example *LCD* and *LED* are both different types of Televisions, computer screens and other electronic devices, in such cases the opinion ranking is in-sufficient.

ii. Syntactic Patterns: For word alignment and opinion relations in sentences the syntactic patterns are used, where the opinion mining is classified based on alignment and syntax. Syntactic Patterns are used for the capturing opinion relations in sentences and even estimates the association between words and word pairs.

iii. Word Alignment Model: This model captures the opinion relations in sentences. Next, opinion targets were retrieved in a regular random procedure framework where nouns and noun phrases are considered. Bi-lingual word alignment is possible based on potential opinion targets.

The word-based alignment model has capability of performing only mono-lingual word alignment which has been widely used in many tasks such as association extraction and tag suggestion.

iv. **Partially supervised word Alignment Model:** This is a system used to incorporate partial alignment links into the alignment process.

v. **High – Precision Syntactic Patterns**: It provides a high-level effectiveness of this model in usage of capturing opinion relations in the sentences. The syntactic patterns are used to targets or words to get the final co-ranking process for output.

C. Opinion Mining Algorithms

The linear regression is used as predictive analysis and used for verifying the opinions of the customers either they are of positive or negative.

The K- Nearest Neighbor algorithm is used in pattern recognition and machine learning techniques. It comes under the categories of classification where generally used for image classification, prediction of disease, win-loss prediction of horse races or games, prediction of natural calamity like earthquakes, floods etc., the other algorithms that too support classification are Random forest, Decision trees and Navies Bayes algorithm. These classification algorithms can be used for opinion mining [7].

D. Opinion Mining Metrics and Evaluation

Opinion of the customers on the reviews is basically either positive or negative or even moderate. Knowledge based Text selects a subset of the original text and identifies the keywords and decides the category which it belongs.

Spam reviews contain malicious negative opinion at large scale on good product from competitive rivals. Ason web any one can write their opinion that is anything they want, so there is no proper control on the internet user reviews, this made the spam classification as the major research issue. For spam and legitimate classification, knowledge-based machine learning classifiers are used. Logistic regression is the probability based statistical model which uses binary values 0 or 1 with respect to the Boolean expression majorly concentration either true or false that is a straight-forward interpretation [8].

81 % of internet users will do the online research to get a conclusion or an opinion regarding a product before purchase.

73% to 85% of customers depends on the reviews before purchases, reviews have a significant

5-star rated, and 4-star rated items are majorly purchased in online shopping's.

The single star rated products sales are found less in ecommerce sites and apps.

III. DESIGN OF OPINION CLASSIFIER

Researchers proposed many opinion mining schemes and mechanisms to protect different Products from fake reviews and provides a clear idea to the customer. These mechanisms are used to detect the user requirements and reliable products that a user shows interest even alerts can be provided by the system [9].

The Online review statements or the feedback obtained from customers or reviewers on a particular product is stored in the database. Based on the keywords and the strings related to the

product and ranks, the classifier retrieves the strings and classifies the opinion and produces the information and submits it to the ranking algorithm, which outputs the ranking on the opinions [10].
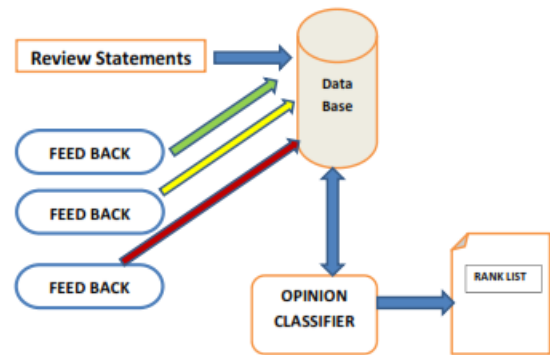

FIGURE 1: BLOCK DIAGRAM OF OPINION CLASSIFIER

Classification plays a major role in the opinion mining where the classification of legitimate reviews and fraud reviews is the major task, were a keyword-based classification is done, so based on the feedback the keywords should be updated in the database. Most spam reviews are the bulk reviews which are posted with an automated program where there will not be any involvement of the human [11]. Spam reviews are the un-solicited junk produced by the companies at a high level to sell their products or exploit the opponent products, these reviews are even used for the promotion of the company products which as not as good as explained in the reviews, classifiers identifies the probability of spam reviews based on the characteristics of the feedbacks in the database and finally produces the output as a rank list [12].

IV. MATHEMATICAL ANALYSIS

An Opinion is a Quintuple (O_β, $f_{\beta\gamma}$, $SO_{\alpha\beta\gamma\delta}$, h_α, t_α) where O_β is a target object, $f_{\beta\gamma}$ is a feature of the targeted object O_β, $SO_{\alpha\beta\gamma\delta}$ is the sentiment value of the opinion of the opinion holder h_α on feature $f_{\beta\gamma}$ of object O_β at time. $SO_{\alpha\beta\gamma\delta}$ is positive, negative or neutral or more granular rating. h_α is an opinion holder and t_α is the time when the opinion is expressed. t_α an important tuple since most of the opinions are expressed at non-working hours, where as the spams will be expressed mostly at working hours who are hired by companies but the spammers are mostly part time writers mostly they write reviews at week-ends, so even though t_α is an important tuple, it is difficult to come to a conclusion based on that particular tuple. Duplicate opinions or bulk opinions are posted by web-spam so identifying genuine opinions also a major task. h_α is the person who holds opinion that is who expresses the opinion. As the opinions are not comments on whole objects they are only comments on features of the object that is the tuple $f_{\beta\gamma}$, where in $SO_{\alpha\beta\gamma\delta}$ the subscripts α represent time variant, β represent Object, γ represent opinion holder, δ represents the features, hence Sentiment value can be calculated. Given a online review statement with n strings, where the statement is represented by S and Strings are represented by

$$S = \{S_1, S_2, S_3, S_4, ..., S_n\}$$

$$\hat{S_A} = \arg \max P(S_A \,|\, S)$$

Where (X, Sax) means that a noun phrase or noun at position x is aligned with an adjective at position Sax.

The opinion relations obtain with respect to adjectives or noun phrases must be aligned with the following results of word transformation model as a sample we collected three models for modeling purpose where

$$P_1(S_A|S) \alpha \prod_{j=1}^{n} C(S_j|S_{aj})$$

Where C is the co-occurrence of information and other samples as

$$P_2(S_A|S) \alpha \prod_{j=1}^{n} C(S_j|S_{aj}) d(j|S_{aj}, n)$$

Where $d(j|S_{aj}, n)$ models the string positions information a probability of string position.

$$P_3(S_A|S) \alpha \prod_{i=1}^{n} n(\phi_i|S_i) \prod_{j=1}^{n} C(S_j|S_{aj}) d(j|S_{aj}, n)$$

Where $n(\phi_i|S_i)$ models the fertility of strings where alignment is given with the relation of one-to many. As there are chances of capturing opinion mining as the estimation of probability is defined as follows

$$P(S_N|S_A) = \frac{COUNT(S_N|S_A)}{COUNT(S_A)}$$

Where $P(S_N|S_A)$ is the translation probability between nouns or noun phases and adjectives.

The candidate's importance is calculated with reference to the score of

$$CI = \frac{tf - xdf(c)}{\sum tf - idf(c)}$$

Where to obtained CI scores should be normalized by the sum of scores of all the candidates. Where tf is the term frequency, df is computed by using ranking algorithm. Finally, the candidates' importance is calculated for getting the result as rank list [13].

V. EXPERIMENTAL ANALYSIS

As we know that opinion mining is the responsible feature for the online marketing and e-business, we did experiment using the java eclipse programming language, where the posting opinion is done with input date and time which is collected automatically where the input of the product name is taken along with the rating with five options such as Excellent, very-good, good, not bad, poor and also a text box for memo is given for the information of the opinion. After giving opinion on the product the user will click on the option button, that is Add my opinion, where the opinions are stored in databases [14].

The tool weka is used to apply opinion mining, where weka uses the most popular algorithms which are used in artificial intelligence and machine learning such as logistic regression, decision tree, fuzzy logic, neural network and supports vector machine. The weka tool is compatible with the Java Eclipse, as weka jar file can be incorporated by building path through Add External Archives. Here you will find the filters such as values, instances and many other based on the attributes the classification is done, and final output of the opinion results will be given as output. Where the metrices to calculate the output are the accuracy, recall, precision and F-Measure. Where the recall is the ratio of correctly classified by the correctly classified and misclassified [15].

It analyzes data not only by pre-processing, it also implements clustering, classification and association rules. Weka classifier visualizes the graphical representation of the result for the better and easy understanding of the output. It provides the option of clusters for grouping the similar feature instances [16].

Opinion Mining effecting a lot the E-Commerce revenues, Sentiment Analysis provides convenient data and information Analysis systems, due to targeted cyber-attacks on E-Commerce Websites, Industrialists and Researchers all over the world concentrated on the opinion mining, as cloud provides convenient data storage system for industry, government, researchers and academic users with cost-effective access to distributed services through internet, the main challenge is the data privacy and information security. Identifying the spam opinion or false review is done using filtering methods, were tested and examined by customers, Entrepreneurs, E-commerce merchants and researchers and categorized into three groups that is manual inspection, Dynamic Auto Filtering mechanisms and content-based filtering. The content-based filtering is further categorized or divided as Network filters, Rule-based filters, Machine Learning filters. The Classifiers are generally used along with the regression, Natural Language Processing, Transfer Learning, Clustering, Word embeddings. The general algorithms that are used in this process are Navies Bayes Classifiers, Random forest Classifiers, K-Nearest Neighbor algorithm, support vector machine and decision tree under the data science and machine learning research [17].

The experimental purpose we use the jupyter notebook, the open-source application for machine learning and data science project, and the python for scientific computing supported by Scipy and numpy, we also use the Anaconda the python distribution and other supporting packages such as matplotlib and pandas. We took the data of Amazon reviews and rating the dataset is called as Amazon Product Reviews, The dataset contains 8,28,30,000 Ratings, users of 2,09,80,000 and items of 93,50,000 items, in the timespan of May 1996 to July 2014, this large crawl of product reviews from amazon. Metadata has the following data items, reviews and rating, item-to-item relationships, timestamps, price, category and ranks [18].

Books	5-core (8,898,041 reviews)	ratings only (22,507,155 ratings)
Electronics	5-core (1,689,188 reviews)	ratings only (7,824,482 ratings)
Movies and TV	5-core (1,697,533 reviews)	ratings only (4,607,047 ratings)
CDs and Vinyl	5-core (1,097,592 reviews)	ratings only (3,749,004 ratings)
Clothing, Shoes and Jewelry	5-core (278,677 reviews)	ratings only (5,748,920 ratings)
Home and Kitchen	5-core (551,682 reviews)	ratings only (4,253,926 ratings)
Kindle Store	5-core (982,619 reviews)	ratings only (3,205,467 ratings)
Sports and Outdoors	5-core (296,337 reviews)	ratings only (3,268,695 ratings)
Cell Phones and Accessories	5-core (194,439 reviews)	ratings only (3,447,249 ratings)
Health and Personal Care	5-core (346,355 reviews)	ratings only (2,982,326 ratings)
Toys and Games	5-core (167,597 reviews)	ratings only (2,252,771 ratings)
Video Games	5-core (231,780 reviews)	ratings only (1,324,753 ratings)
Tools and Home Improvement	5-core (134,476 reviews)	ratings only (1,926,047 ratings)
Beauty	5-core (198,502 reviews)	ratings only (2,023,070 ratings)
Apps for Android	5-core (752,937 reviews)	ratings only (2,638,172 ratings)
Office Products	5-core (53,258 reviews)	ratings only (1,243,186 ratings)
Pet Supplies	5-core (157,836 reviews)	ratings only (1,235,316 ratings)
Automotive	5-core (20,473 reviews)	ratings only (1,373,768 ratings)
Grocery and Gourmet Food	5-core (151,254 reviews)	ratings only (1,297,156 ratings)
Patio, Lawn and Garden	5-core (13,272 reviews)	ratings only (993,490 ratings)
Baby	5-core (160,792 reviews)	ratings only (915,446 ratings)
Digital Music	5-core (64,706 reviews)	ratings only (836,006 ratings)
Musical Instruments	5-core (10,261 reviews)	ratings only (500,176 ratings)
Amazon Instant Video	5-core (37,126 reviews)	ratings only (583,933 ratings)

Table 1: Amazon Data Set Reviews and Ratings.

The complete review data is divided into the raw review data, user review data and finally the product review data [19].

VI. GRAPHICAL ANALYSIS

The Dataset consists of a large quantity of reviews, our datasets contains nine attributes they are as follows

- *reviewerid* is unique identity of the person giving review it is given to avoid un-authorized reviews or to identify the authorized reviewer.
- *asin* is amazon standard identification number whereas it is a unique identification given to every product in amazon database.
- *reviewName* name of the person giving the review.
- *Helpful* thumbs-up or down, it is a symbol indication for positive or negative voting regarding the product.
- *reviewText* review of the product.
- *overall* is a class label rating from 1 to 5.
- *summary* the complete description about the product and review
- *unixReviewTime* time taken for reviewing based on operating system time that is Unix.
- *reviewTime* area the time constraints that is the time taken for reviewing.

We computed correlation between each and other attributes, the attributes are not highly corelated, we choose the review text to generate features, the overall class label is 1 to 5.
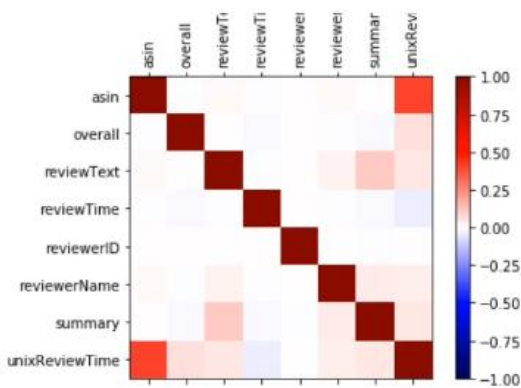

Fig. 1.Co-relation matrix between Attribuutes.

VII. PERFORMANCE - EVALUATION

The performance measures of machine learning algorithm is identified by implementing it on a dataset, as we taken amazon opinion mining dataset, the next step here is testing different types of classifiers and compare the result based on the constraints such as precision, recall and accuracy [20].

The model accuracy is identified by the total number of correct classifications divided by the total number of classifications done. The percentage of misclassifications are calculated by the Error Rate. Precision and recall are the major performance measures of a supervised learning model. The precision gives the proportion of positive predictions which are truly positive, recall gives the proportion of true positive cases over the positive cases [21].

The other performance measures generally used are the specificity and sensitivity. Specificity is a model which measures the proportion of negative examples which have been correctly classified. Sensitivity of a model measures the proportion of true positive examples or positive cases which were correctly classified [22]. F-measure is another measure

of model performance which combines the precision and recall. It takes the harmonic mean of precision and recall.

Methodology	Precision	Recall	Accuracy
Navies Bayes Classification	0.712	0.752	70.2%
Neural Networks	0.674	0.71	65.2%
Support Vector Machines (SVM)	0.483	0.521	47.6%
Logistic Regression	0.761	0.80	75.2%
Decision Tree Classifier	0.75	0.79	74.8%
Random Forest	0.724	0.76	71.2%

Table 1: Evaluation of algorithms on Amazon data set

A. Designed Methodology

- To Improve the performance of the Machine Learning Methodology, we used ensemble learners where multiple models are combined to solve a problem and for the better results and accuracy. It gives better predictive performance then existing algorithms [23].
- Bootstrapping Aggregating is an ensemble technique that is used to enhance the performance of the algorithms by combining related algorithms and makes a hybrid algorithm which enhances the performance and accuracy.
- We combined the random forest algorithm with random tree algorithm to create a hybrid algorithm for better results.

B. Results

We implemented the Bootstrapping Aggregating methodology on two machine learning algorithms that is Random forest algorithm and Decision Tree Classifier algorithm on the Amazon Data Set, the performance evaluation is compared with existing algorithm with designed algorithm.

Methodology	Precision	Recall	Accuracy
Decision Tree Classifier	0.75	0.79	74.8%
Random Forest	0.724	0.76	71.2%
Bootstrapping	**0.80**	**0.84**	**79.3%**

VIII. CONCLUSION

Opinion Mining and Client Feedback mining are the major tasks evaluated on the Amazon Data Set successfully, where we also concentrated on the false opinions or fake opinions, so the classification of legitimate and spam opinions are done by using different algorithms, finally we designed a new algorithm using ensemble learning techniques and implementing, bootstrapping method on random forest and decision trees algorithms, we received better performance constraints [24].

REFERENCES

1. F. Bodendorf and C. Kaiser, "Mining Customer Opinions on the Internet - A Case Study in the Automotive Industry," *2010 Third International Conference on Knowledge Discovery and Data Mining*, Phuket, 2010, pp. 24-27.

2. P. Akre, H. Patil, A. Khandare and M. Atique, "Mining topical relations between opinion word and opinion target," *2017 2nd International Conference for Convergence in Technology (I2CT)*, Mumbai, 2017, pp. 389-392.

3. Kang Liu, Liheng Xu and Jun Zhao, *"Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model",* IEEE Transactions on Knowledge and Data, Volume: 27, Issue: 3, Pp. 636-650, 2015.

4. J.VijayaChandra, Narasimham Challa, Sai Kiran Pasupuleti, "Intelligence based Defense System to Protect from Advanced Persistent Threat by means of Social Engineering on Social Cloud Platform", , Indian Journal of Science and Technology, Vol: 8,Issue: 28, October 2015.

5. S. Anupkant, P. V. M. S. Kumar, N. Sateesh and D. B. Mahesh, "Opinion mining on author's citation characteristics of scientific publications," *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, 2017, pp. 348-351.

6. I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*, Beijing, 2017, pp. 542-546.

7. J. Chen, Y. Liu, G. Zhang, Y. Cai, T. Wang and H. Min, "Sentiment Analysis for Cantonese Opinion Mining," *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, Xi'an, 2013, pp. 496-500.

8. J. V. Chandra, N. Challa and S. K. Pasupuleti, "A practical approach to E-mail spam filters to protect data from advanced persistent threat", *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, Nagercoil, pp. 1-5, 2016.

9. V. B. Raut and D. D. Londhe, "Opinion Mining and Summarization of Hotel Reviews," *2014 International Conference on Computational Intelligence and Communication Networks*, Bhopal, 2014, pp. 556-559.

10. A. Angelpreethi and S. B. R. Kumar, "An Enhanced Architecture for Feature Based Opinion Mining from Product Reviews," *2017 World Congress on Computing and Communication Technologies (WCCCT)*, Tiruchirappalli, 2017, pp. 89-92.

11. J. Vijaya Chandra, NarasimhamChalla and M. Ali Hussain, "Data and Information Storage Security from Advanced Persistent Attack in Cloud Computing", International Journal of Applied Engineering Research, 2014.

12. C. C. Hnin, N. Naw and A. Win, "Aspect Level Opinion Mining for Hotel Reviews in Myanmar Language," *2018 IEEE International Conference on Agents (ICA)*, Singapore, 2018, pp. 132-135.

13. J. Vijaya Chandra, Narasimham Challa, Sai Kiran Pasupuleti, Thirupathi RK, Krishna RV. "Numerical Formulation and Simulation of Social Networks using Graph Theory on Social Cloud Platform". Global Journal of Pure and Applied Mathematics, 11(2):1253–64, 2015.

14. J. VijayaChandra, Narasimham Challa and Sai Kiran Pasupuleti, "Advanced Persistent Threat defense system using self-destructive mechanism for Cloud Security," *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, Coimbatore, 2016, pp. 7-11.

15. G. Holmes, A. Donkin and I. H. Witten, "WEKA: a machine learning workbench," *Proceedings of ANZIIS '94 - Australian New Zealnd Intelligent Information Systems Conference*, Brisbane, Queensland, Australia, 1994, pp. 357-361.

16. T. Ahmad and M. N. Doja, "Opinion Mining Using Frequent Pattern Growth Method from Unstructured Text," *2013 International Symposium on Computational and Business Intelligence*, New Delhi, 2013, pp. 92-95.

17. R. Rastogi, "Machine Learning @ Amazon," *2017 IEEE 24th International Conference on High Performance Computing (HiPC)*, Jaipur, 2017, pp. 182-182.

18. A. Ejaz, Z. Turabee, M. Rahim and S. Khoja, "Opinion mining approaches on Amazon product reviews: A comparative study," *2017 International Conference on Information and Communication Technologies (ICICT)*, Karachi, 2017, pp. 173-179.

19. T. Karthikayini and N. K. Srinath, "Comparative Polarity Analysis on Amazon Product Reviews Using Existing Machine Learning Algorithms," *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Bangalore, 2017, pp. 1-6.

20. A. Onan and S. Korukoğlu, "Ensemble methods for opinion mining," *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, Malatya, 2015, pp. 212-215.

21. K. L. S. Kumar, J. Desai and J. Majumdar, "Opinion mining and sentiment analysis on online customer review," *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Chennai, 2016, pp. 1-4.

22. Y. A. Solangi, Z. A. Solangi, S. Aarain, A. Abro, G. A. Mallah and A. Shah, "Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis," *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, Bangkok, Thailand, 2018, pp. 1-4.

23. P. Tian, Y. Liu, M. Liu and S. Zhu, "Research of Product Ranking Technology Based on Opinion Mining," *2009 Second International Conference on Intelligent Computation Technology and Automation*, Changsha, Hunan, 2009, pp. 239-243.

24. P. K. Verma, S. Agarwal and M. A. Khan, "Opinion mining considering roman words using Jaccard similarity algorithm based on clustering," *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Delhi, 2017, pp. 1-4.

AUTHORS PROFILE

J.VijayaChandra is Head of the Department and Assistant Professor, in Department of Computer Science and Engineering at Warangal Institute of Technology and Science, Warangal, Telangana, India; Interested Research areas are Cloud Security, Network Security, Machine Learning, Intelligence Security and Data Security. Published 15 Research Papers for International Journals. He is Oracle Certified Associate and Machine Learning Certified. Member of IEEE and ACM.
Email id: vijayachandra.phd@gmail.com

Dr.Ranjith Gyaderla, Vice Principal and Associate Professor, Department of Computer Science and Engineering, Warangal Institute of Technology and Science, Warangal. His Research Areas are Ontology, Novice, Theory of Computation and Machine Learning. Published 10 Research Papers for International Journals.
Email id: gyaderlaranjith@gmail.com

B. Bhagya Laxmi, Assistant Professor, Department of Computer Science and Engineering, Warangal Institute of Technology and Science, Warangal, Telangana, India. Hers Research Areas are E-Commerce, Data Mining and Data Warehousing and Computer Graphics. She Published 5 Research Papers for International Journals.
Email id: bhagya712@gmail.com

Bhattu Murali Manohar Raju is M. Tech Scholar, Department of Computer Science at Warangal Institute of Technology and Science, Warangal, Telangana, India. His Interested areas are Cloud Computing, Machine Learning, Network Security and Artificial Intelligence.
Email id: bmmraju83@gmail.com

Mohammad Azaruddin is M. Tech Scholar, Department of Computer Science at Warangal Institute of Technology and Science, Warangal, Telangana, India. His Interested areas are Testing Tools, Software Testing and Quality Improvement, Machine Learning, Network Security and Artificial Intelligence.
Email id: Azaruddin.testing@gmail.com