3D Visualization of Sentiment Measures and Sentiment Classification using Combined Classifier for Customer Product Reviews

Siddhaling Urologin

Department of Computer Science
Birla Institute of Technology and Science, Pilani-Dubai,
Dubai, U.A.E.

Sunil Thomas

Department of Electrical and Electronics Engg.,
Birla Institute of Technology and Science, Pilani-Dubai,
Dubai, U.A.E.

Abstract—**The Internet has wide reachability making many users to buy the products online using e-commerce websites. Usually, users provide their opinions, comments, and reviews about the products in social media, e-commerce websites, blogs, etc. The product review comments provided by the customers have rich information about the usage of the products they bought and their sentiments towards those products. In this research, we have collected reviews from Amazon.com and performed sentiment analysis to collect sentiment information. We have proposed 3D visualizations to represent sentiment information, such as sentiment scores and statistics about words used in the reviews. The 3D visualizations are useful to represent large sentiment related information and to have an in-depth understanding of sentiments of users. We have developed a combined classifier using Logistic Regression, Decision Tree and Support Vector Machine. From the reviews, we formed N-gram features using a bag of words and performed sentiment classification using combined classifier. On 10 fold cross-validation, a maximum classification rate for combined classifier of 90.22% is obtained for sentiment classification.**

Keywords—Sentiment analysis; 3D visualization; sentiment classification; natural language processing; product reviews

I. INTRODUCTION

Many users are buying products online through e-commerce sites as they are widespread reachable by internet [1], [2]. The internet also provides an opportunity to the user to give their opinions in various forums such as social media, blogs, online e-commerce sites, etc. [2]. Many users provide reviews and opinions in natural language for the product they have come across. These reviews have wealth of knowledge [1] about the products and feeling of users towards the products. Sentiment analysis involves in mining the naturally expressed text to understand the feeling of people towards the entity of interest. Sentiment mining and analysis has found many application in areas of healthcare [3], [4], tourism [5], fraud detection [6], finance [7], politics [8], business [9], few more applications are listed in [10]. In [11] informatics, theoretic approach is used for classification of sentiments. A lexicon for sentiment analysis and concept level sentiment analysis is presented in [12]. The opinion mining of Amazon data is carried out in [13] using Support Vector Machines to summarize the unstructured data. As-LDA model is used in [14] for sentiment classification. Multimodal Naïve Bayes and

Decision tree classifiers are discussed in [15] for tweeter data sentiment analysis. A cloud integrated system for Support Vector Machines, Naïve Bayes and Neural Networks is presented in [16] for blog data sentiment classification. Several lexicon dictionaries such as [17], [18], have been used while determining sentiment analysis. In [18], Valence Aware Dictionary for Sentiment Reasoning (VADER), which is a rule-based model for sentiment analysis, is presented. VADER uses a lexicon list with sentiment measures to compute sentiment score for textual data. Sentiment analyzer examines the textual input to mine the feeling of user, which is present in the text and provides sentiment information indicative of feelings expressed by the users. Customer reviews for products are collected through e-commerce sites, blogs, social media etc., have rich knowledge about the products and their usability. Sentiment analyzer can be used to undermine sentiment information from the customer reviews. A large number of reviews being collected every day, better schemes are highly desirable to analyze sentiments of reviews and to provide visualization of sentiment information. In [19], authors have presented SentiView, which an interactive visualization system used to analyze people sentiments towards selected topic. It mines the data from online posts and uses uncertainty modeling to depict the changes in sentiments. Megan K et al. [20] have collected Twitter data during Gulf Oil Spill 2010 and they analyzed emotion of the broadcasted information in twitter. The emotion classification and analysis results are given with various visualizations. In [21], author has discussed the design, implementation of tools to extract, analyze and explore public messages. Then authors have performed sentiment analysis with a web application.

Motivated by these facts, we propose novel 3D visualization schemes to represent sentiment information obtained by sentiment analyzer. We utilize VADER [18] to mine the sentiments of electronic products reviews gathered from Amazon.com. The VADER provides the sentiment information such as compound score, positive score, negative score and neutral score, etc. The sentiment information has been used to construct 3D visualizations such as 3D surface plot, 3D column charts, 3D scatter plots, etc. in this research. The 3D visualization provides better schemes to undermine sentiment information especially for large review sets. Furthermore, we developed a new combined classifier for the sentiment classification. The combined classifier is built from

three base classifiers such as Logistic Regression, Decision Tree and Support Vector Machine. The voting technique is used to determine the resultant of combined classifier. The feature extraction is carried out in this research by taking N-gram features using a bag of words similar to method in [22]. This research paper is organized as follows. Review data preprocessing and filtering has been described in Section II. In Section III, the lexicon list and their measurements of VADER [18] are presented. Section IV elaborates the architecture on sentiment visualization and combined sentiment classifier. The experimental results are given in Section V and the conclusion is covered in Section VI.

II. DATA PREPARATION AND FILTERING

The customer review about the products have been gathered originally from Amazon.com and we have downloaded from [23], [24]. The customer review gathered from [23], [24] are in JSON format and an example for customer review has been shown in Table I. The format has various fields such as Title, Author, ReviewID, Overall, Content, and Date. Here Overall Rating is a five-star rating where the user can assign zero (lowest) to five (highest) stars. The content is textual description given by the user as their review comments for a product. Each customer review thus obtained will undergo preprocessing and filtering. All the text parts are retained and converted to lower case letter. Stop words are identified using the list of [18] and these stop words are removed from the reviews to retain only essential words. Then, we utilized VADER lexicon and sentiment analyzer [18] to compute the score of words and determine the sentiment score of the customer reviews. Then overall sentiment polarity of the review is found to be positive, negative or neutral based on the total sum of sentiment score of a review.

TABLE I. JSON FORMAT OF CUSTOMER REVIEW FROM AMAZON.COM

JSON Format	

{
 Title: …….
 Author: ………..
 ReviewID:…….
 Overall:……
 Content:………
 Date:……
 }
	{"Reviews": [{"Title": "Overall a nice laptop for around $1100", "Author": null, "ReviewID": "R2Y4WQSYMYCE24", "Overall": "4.0",
	
	 "Content": "... did not really review Model 2681CU1 as shown on this website. This models comes with Pentium 4 -Mobile 1.8 Ghz,… ","Date": "October 30, 2003"},

III. LEXICON FOR SENTIMENT SCORE

In this research, we have used VADER lexicon and sentiment analyzer [18] to determine sentiment scores of a review. The VADER lexicon is a well-trained list of words with the polarity values especially for micro blogs of social media. This lexicon focuses on sentiment emphasis by different social media web sites. The list also includes many of generally used emotions such as ":)", ":(", acronyms such as "LOL" and slang "nah", etc. The lexicon contains the word with their polarity value between -4 to +4. In Table II few examples for VADER lexicon words with polarity values are presented.

TABLE II. EXAMPLE FROM VADERLEXICON

Word	Polarity
Okay	+0.9
Good	1.9
Great	3.1
Horrible	-2.5
:(	-2.2

IV. SENTIMENT VISUALIZATION AND CLASSIFICATION

After collecting customer reviews in the JSON format, the content field is extracted, which is the commentary description given by a customer. The sentiment analysis of the review comments are carried out as shown in Fig. 1 using VADER sentiment analyzer of [18]. A list of lexicons related to micro blog along with their sentiment measures is used in VADER sentiment analyzer. This analyzer uses five heuristic rules, based on grammatical and syntactical conventions used in natural language while expressing human emotions. Using VADER sentiment analyzer, we obtain sentiment information such as compound sentiment score, positive score, negative score and neutral score for a review comment. Motivated by current research [19]-[21], we are proposing novel schemes for representing sentiment information of customer reviews into 3D visualization charts. The 3D representations of sentiment information are most useful to gain further insight of the customer reviews.
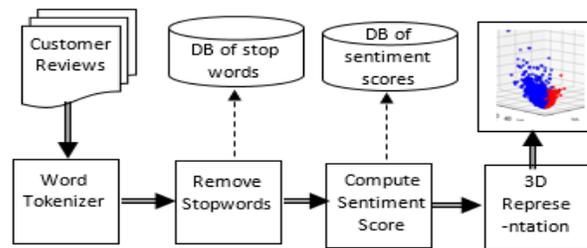

Fig. 1. Customer reviews to 3D representation.

We propose a new combined classifier of Logistic Regression, Decision Tree and Support Vector Machine classifier using voting technique to perform sentiment classification. The combined classifier using voting technique is depicted in Fig. 2. The customer reviews undergo preprocessing in which sentence tokenization and stop words removal are carried out. Thereafter N-gram feature vector is constructed for each review using bag of words similar to [22]. The sentiment classification of the reviews is performed using a combined classifier, which consists of three base classifiers such as Logistic Regression, Decision Tree and Support Vector Machine. The resultant of combined classifier is found by taking majority in voting of base classifiers.
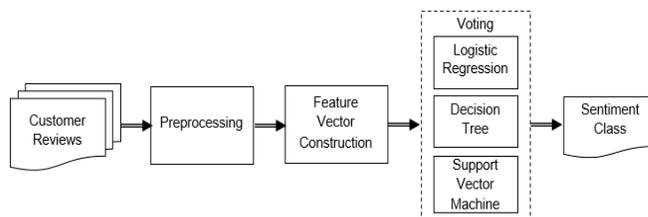

Fig. 2. Combined classifier using voting technique.

TABLE III.   STATISTICS FOR THE SENTIMENT OF REVIEWS [SENTIWORDS IS SENTIMENT WORDS]

Review	Negative Sentiment Score	Neutral Sentiment Score	Positive Sentiment Score	Compound Sentiment Score	Num. Of Negative SentiWords	Num. Of Neutral SentiWords	Num. Of Positive SentiWords
looking external dvd player kids could watch dvds asus eee netbook , exactly got ….	0.126	0.679	0.195	0.8174	6	84	9
dont buy unit . dvd-rom needed dvd writer . read dvds read cd . absolute waste money .	0.177	0.823	0	-0.4215	1	13	0
needed good thing works , sometimes take cd input , guess dumb moments	0.217	0.592	0.191	-0.1027	1	9	1
worth every penny ! shippment speedy ! performed better expected . would recommend…	0	0.565	0.435	0.816	0	11	3
small asus netbook decided needed external dvd drive . one works fine . far 've imported cds...	0.115	0.809	0.077	-0.2263	1	19	1
mine shipped instruction manual documentation whatsoever . plugged , whirred though sensed...	0	1	0	0	0	26	0
dvd/cd drive laptop ended failing six months ago , leaving stranded laptop needed reformatted ….	0.043	0.912	0.046	0.0772	2	100	2
great buy inexpensive price.there n't much instructions came dvd/cd drive . think one small paragraph..	0.072	0.797	0.131	0.4295	1	25	1
received sooner expected ! part christmas gift twins go netbooks . fact came early , bonus , worrying	0.114	0.569	0.317	0.6476	1	12	2
p.o.s dead arrival . cheap chinese junk chinglish manual . power supply included . purchased…	0.203	0.797	0	-0.68	1	18	0

V.   EXPERIMENTAL RESULTS

We have gathered customer reviews for electronic products from [23], [24], which are originally collected fromAmazon.com. The electronic review data set contain 1,689,188 user comments. In our experiments, the sentiment visualization is presented on four electronic products and the sentiment classification is carried out on 12,500 review comments. We have carried out preprocessing by removing URL links, tags, stop words and each comment is converted to lower case letters. Each review is subjected to sentiment analysis using VADER analyzer [18]. In Table III, we have tabulated sentiment information obtained from VADER for few reviews on the electron products. In column 1, user review comments given (partial comments are shown). Next four columns show negative, neutral, positive and compound sentiment scores of a review. The number of negative, neutral and positive sentiment words found in that review is given 6, 7 and 8th columns respectively.

VI SENTIMENT VISUALIZATION

Further analysis and visualization are carried out on four electronic products reviews to reveal customer opinions in depth. We have obtained sentiment information using [18] on reviews of products External USB DVDCD, GE 72887 Superadio III Portable AMFM Radio, NETGEAR Prosafe FS105NA, Panasonic On-Ear Stereo Headphones. There are 199 comments on External USB DVDCD with 11,630 words, for GE 72,887 Superadio III Portable AMFM Radio 265 reviews with 33,973 words, for NETGEAR Prosafe FS105NA423 comments with 25310 words and for Panasonic On-Ear Stereo Headphones 1692 comments with 1,06,284 words. The sentiment of a review is considered as positive when its compound score is greater than zero, it is neutral when the compound score is equal to zero and when the compound score is less than zero its sentiment is negative. Each word in a review contribute in compound sentiment score. It is interesting to observe the variation of compound sentiment score against number of positive and negative words.

In Fig. 3, a representation of compound score is shown as 3D surface against number of positive sentiment words and negative sentiment words. In this figure, compound score is accumulated for all the reviews of a product with respect to number of positive words and negative words. In Fig. 3(a), accumulated compound sentiment score is plotted as a 3D surface for reviews of External USB DVDCD. Similarly in Fig. 3(b), 3(c) and 3(d) accumulated compound score is plotted against number of negative and positive words.
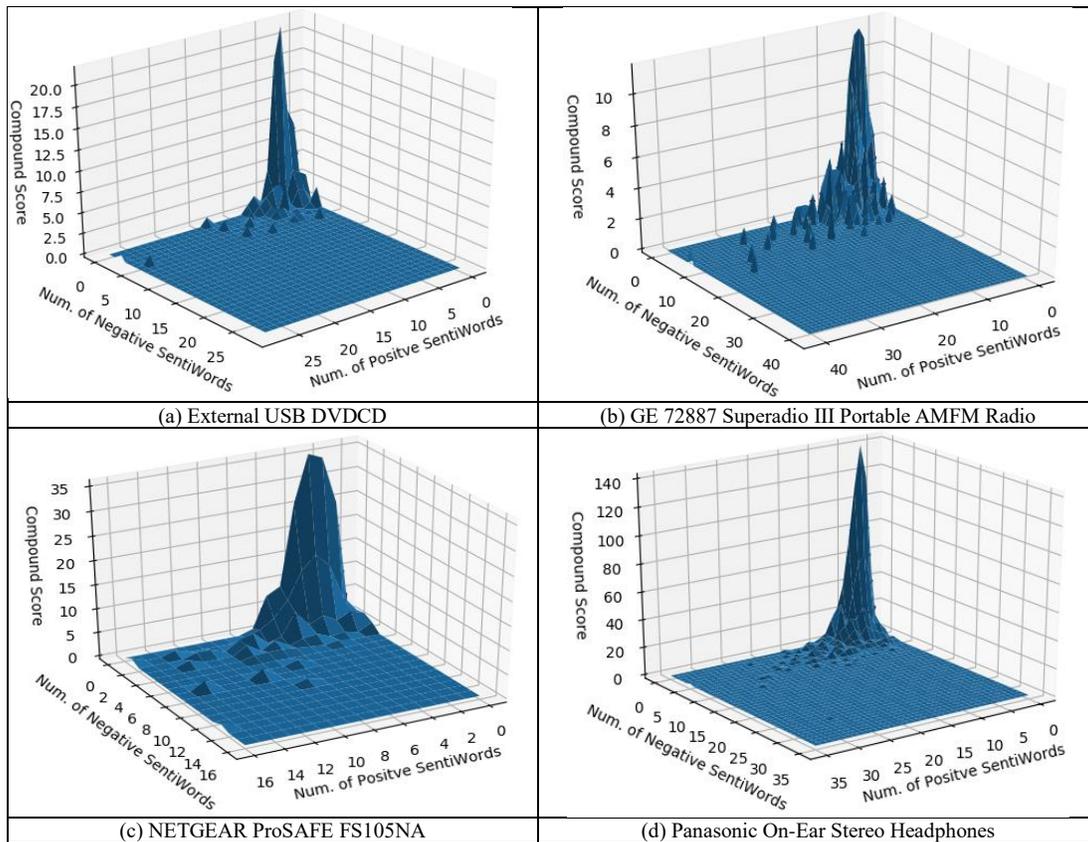
| (a) External USB DVDCD | (b) GE 72887 Superadio III Portable AMFM Radio |
| (c) NETGEAR ProSAFE FS105NA | (d) Panasonic On-Ear Stereo Headphones |

Fig. 3.　(a) External USB DVDCD; (b) GE 72887 Superadio III Portable AMFM Radio; (c) NETGEAR Prosafe FS105NA; (d) Panasonic On-Ear Stereo Headphones.
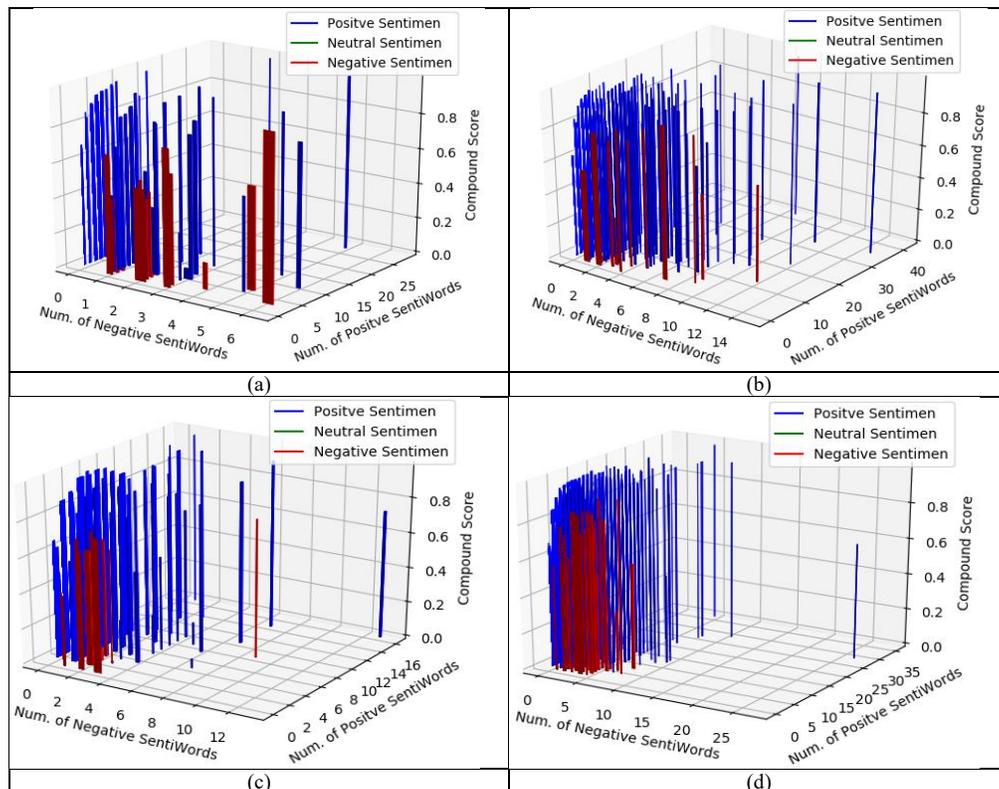

| (a) | (b) |
| (c) | (d) |

Fig. 4.　(a) External USB DVDCD; (b) GE 72887 Superadio III Portable AMFM Radio; (c) NETGEAR Prosafe FS105NA; (d) Panasonic On-Ear Stereo Headphones.
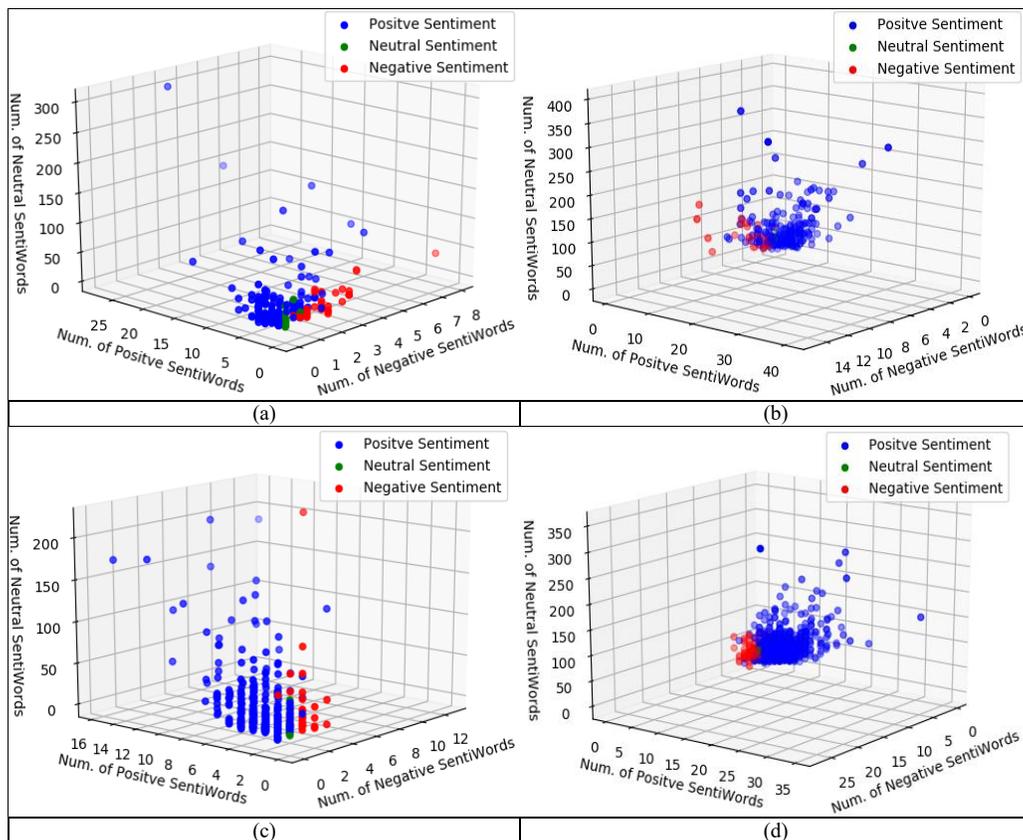
Fig. 5.   (a) External USB DVDCD; (b) GE 72887 Superadio III Portable AMFM Radio; (c) NETGEAR Prosafe FS105NA; (d) Panasonic On-Ear Stereo Headphones.

These representations reveal contribution of number of positive and negative words for compound sentiment score. As depicted in Fig. 3(a), the highest score in the accumulated compound score observed as 21.64 with number of negative words as zero and positive words as three. More positive comments have been observed by customers than negative and majority of customers used around 10 positive words. There are few negative comments observed with fewer than seven words. More variation in the accumulated compound sentiment score is observed for reviews of GE 72887 Superadio III Portable AMFM Radio as shown in Fig. 3(b).

Fig. 4 shows 3D column chart of compound scores for reviews of a product. The compound review score is depicted against number of positive and negative words. In Fig. 4(a) compound scores for 199 reviews of External USB DVDCD is represented against number positive and negative words. As shown in Fig. 4(a), more positive sentiments can be observed compared to negative. Also few number of words are used to express sentiment negatively, while more number of words are used express positive opinions. The 3D surface views as in Fig. 3 and 3D column view as shown in Fig. 4 of compound sentiment score highlight usability of a number of words to express sentiment, which is important to understand the characteristics of a product reviews.

In Fig. 5, a review is represented as a point with coordinates as a number of positive sentiment words along the x-axis, negative sentiment words along they-axis and neutral sentiment words along the z-axis. Fig. 5(a) is a 3D scatter plot for all 199 reviews indicating the distribution of sentiments type with respect to the number of words present in reviews. This figure indicates that more positive reviews than negative or neutral are given by customer on External USB DVDCD. The 3D scatter plot reveals the distribution of reviews observed on sentiment type with respect the number of words.

We have collected 10 words having maximum sentiment score and minimum sentiment score for each product and represented using donut charts in Fig. 6 and 7. The top 10 words having maximum sentiment score represent positive sentiment are depicted in Fig. 6. For each word, its sentiment score along with the percentage contribution to top 10 maximum sentiment score words has been shown. Fig. 7 shows top 10 words with minimum sentiment scores representing negative sentiment. The compound sentiment score and the percentage contribution to top 10 negative sentiment words have been depicted in Fig. 7.
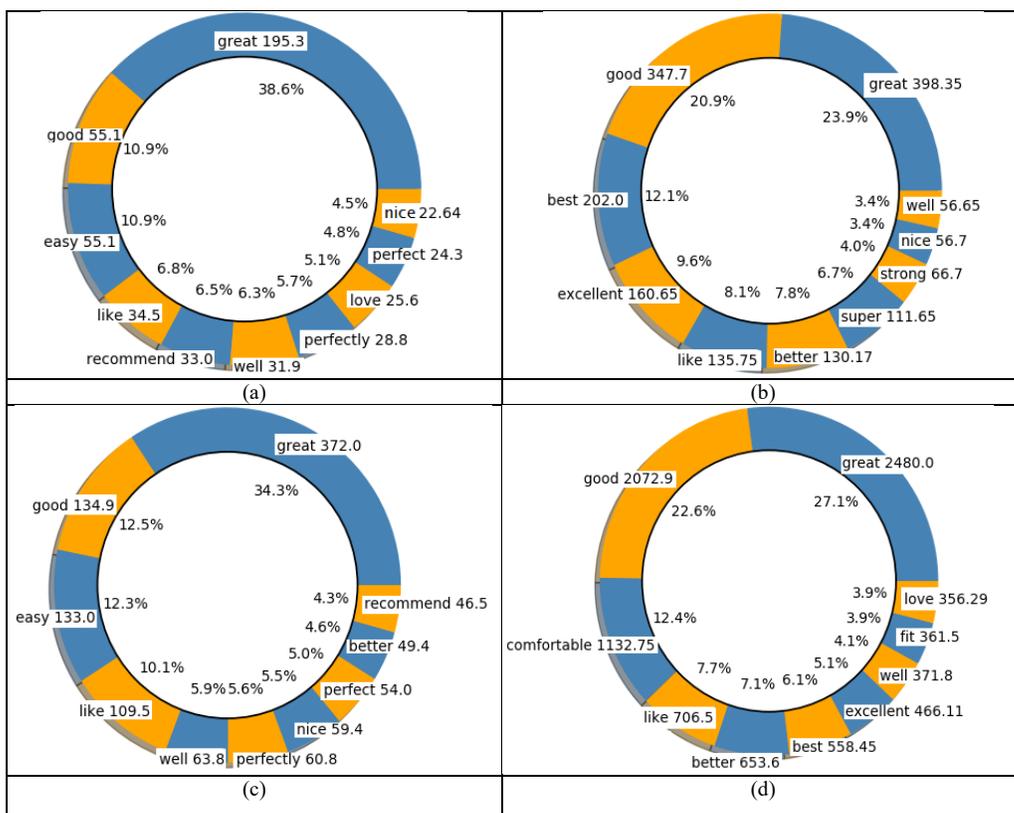
Fig. 6.   (a) External USB DVDCD; (b) GE 72887 Superadio III Portable AMFM Radio; (c) NETGEAR Prosafe FS105NA; (d) Panasonic On-Ear Stereo Headphones.
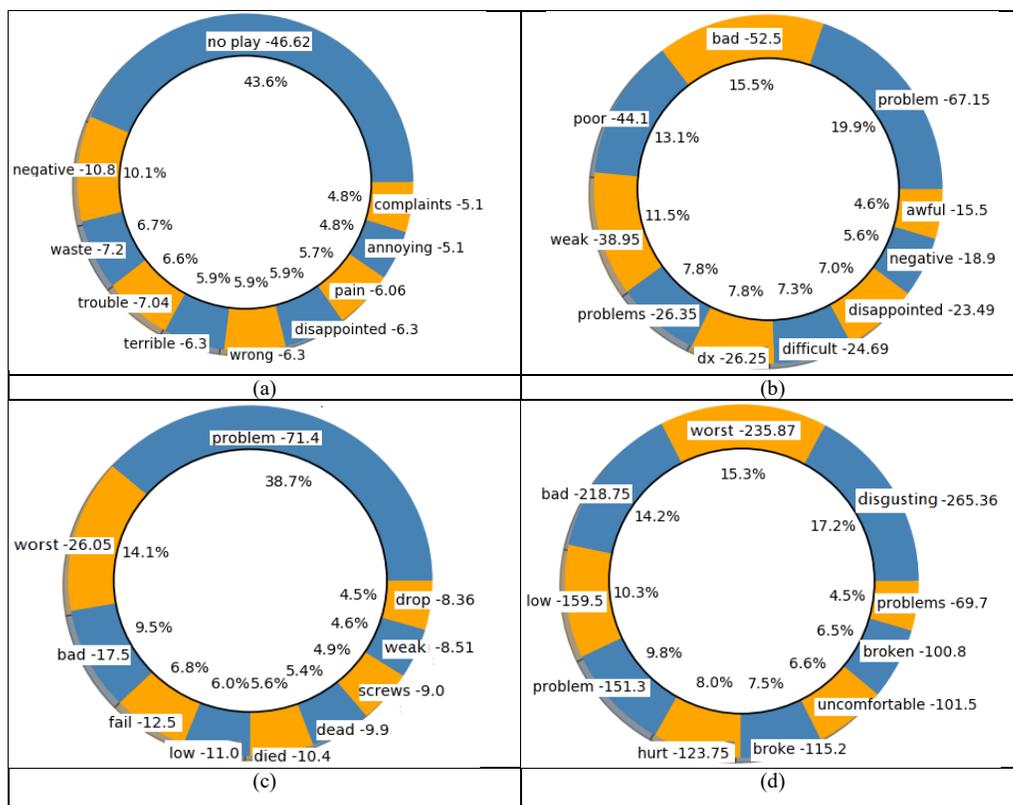

Fig. 7.   (a) External USB DVDCD; (b) GE 72887 Superadio III Portable AMFM Radio; (c) NETGEAR Prosafe FS105NA; (d) Panasonic On-Ear Stereo Headphones.

VI.   SENTIMENT CLASSIFICATION

We have considered 12500 review comments on electronic products from Amazon data set. Each review comment is subjected to preprocessing where in each sentence is tokenized and stop words are eliminated. Then we utilized sentiment analyzer of VADER [18] and determined the sentiment information for user comments. The sentiment analyzer VADER provides the scores such as positive score, negative score, neutral score and compound score. When compound score of a review is greater than zero then review is considered as positive and when compound score is less than zero it is taken as negative sentiment. Thereby we constructed a dataset of 12500 review comments along with their sentiment type as positive or negative and carried out sentiment classification. For a review comment, a bag of word features vector is constructed by taking N-gram size feature matrix similar to the method in [22]. To select the words from the collection of 12500 reviews, we have computed frequency of occurrence of each word in the dataset. Then the bag of words is created by using selected 'n' number of most frequent words. Ten most frequent important words along with their frequency are depicted in Fig. 8. In the experiments we have chosen 'n' as 2000, 3000, 4000, 5000 and 6000.
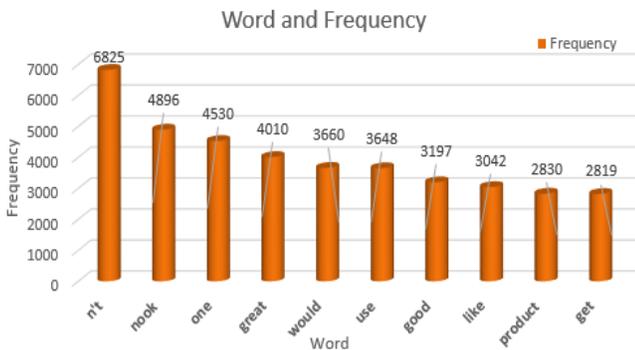

Fig. 8.   Ten most frequent word in the dataset of user reviews.

TABLE IV.   NAÏVE BAYES CLASSIFIER PERFORMANCE.

Sl. Num.	n	Classif. Rate in %	Misclassif. Rate in %
1	2000	82.94	17.06
2	3000	83.43	16.57
3	4000	84.26	15.74
4	5000	84.90	15.10
5	6000	85.14	14.86

Further, we have conducted experiments to predict sentiment of a review as either positive or negative sentiment using Naïve Bayes Classifier of NLTK [25]. The 10-fold cross validation is carried out on Naïve Bayes Classifier to perform classification of sentiments with 'n' as 2000, 3000, 4000, 5000 and 6000. The cross-validation accuracy in terms of classification rate and misclassification rate are shown in Table IV.

Next, the sentiment classification is carried out using three classifiers such as Logistic Regression, Decision Tree and Support Vector Machine independently. The 10-fold cross validation for all three classifiers with varying feature vector size n=2000, 3000, 4000, 5000 and 6000 is conducted and results are shown in Table V. In columns 3, 5 and 7 is shown the classification rate for Logistic Regression Classifier, Decision Tree Classifier and Support Vector Machine Classifier, respectively. In columns 4, 6 and 8 the misclassification rates are given. Out of three classifiers, Logistic Regression Classifier gives better classification rates ranging from 88.52% to 88.93%. The Decision Tree Classifier provides the least classification rate between 82.34% and 83.09% as tabulated in Table V.

A novel combined classifier using voting technique is developed for sentiment classification. Three classifiers such as Logistic Regression, Decision Tree and Support Vector Machine are used in the combined classifier and resultant is determined by majority voting technique as shown in Fig. 2.

TABLE V.   LOGISTIC REGRESSION, DECISION TREE AND SUPPORT VECTOR MACHINE PERFORMANCE INDEPENDENTLY

Sl.Num.	n	Logistic Regression Classifier		Decision Tree Classifier		Support Vector Machine	
		Classif. Rate in %	Misclassif. Rate in %	Classif. Rate in %	Misclassif. Rate in %	Classif. Rate in %	Misclassif. Rate in %
1	2000	88.62	11.38	82.34	17.66	86.42	13.58
2	3000	88.52	11.48	83.03	16.97	86.42	13.58
3	4000	88.72	11.28	83.03	16.97	86.42	13.58
4	5000	88.63	11.37	82.99	17.01	86.42	13.58
5	6000	88.93	11.07	83.09	16.91	86.42	13.58

TABLE VI.   COMBINED CLASSIFIER PERFORMANCE

Sl.Num	n	Combined Classifier in %	
		Classif. Rate	Misclassif. Rate
1	2000	89.12	10.88
2	3000	89.10	10.90
3	4000	90.22	9.78
4	5000	89.55	10.45
5	6000	89.15	10.85

The performance of the combined classifier is measured on 10-fold cross validation. The results of cross validation such as classification rate and misclassification rate are given in Table VI. Maximum classification rate of 90.22% with feature vector size n=4000 and minimum classification rate of 89.10% with feature vector size n=3000 are observed. For most of feature vector size n=2000, 3000, 4000, 5000 and 6000 the performance of the combined classifier is better than base classifiers such as Logistic Regression, Decision Tree and Support Vector Machine.

VII. Conclusion and Future Work

In recent years e-commerce websites are gaining popularity. Users find it easier and convenient to buy products online through various e-commerce websites. Moreover, with the increase in reachability of internet, there is a substantial growth in customer provided information in terms of reviews comments for various products. Due to which, there is interest in mining useful information from customer review and comments. In this research, we performed sentiment analysis on reviews gathered from Amazon.com for electronic products. We utilized the VADER, which is a rule based sentiment analyzer and uses lexicon with sentiment measures to compute sentiment scores. VADER provides sentiment information such as compound, positive, negative, neutral score, etc. for the reviews. In a large set of data, it is important to visualize the sentiment information and related statistics for a better insight of reviews. In this research, 3D visualization such as 3D surface, 3D column and 3D scatter plot charts are presented. The 3D scatter plot, provide insight into the distribution of sentiment type with respect to a number of words in the reviews. The 3D surface and 3D column charts show variations in sentiment score with respect to a number of positive and negative words. Furthermore, we have utilized donut charts to represent most 10 positive and negative sentiment words in a set of reviews. In addition, we developed a new combined classifier for sentiment classification. The combined classifier has three base classifiers such as Logistic Regression, Decision Tree and Support Vector Machine. The combined classifier utilizes voting technique to determine resultant class. A bag words for reviews is created to collect N-gram features from reviews and performance of the classifiers is evaluated on 10-fold cross validation. For the combined classifier a maximum classification rate of 90.22% for 10-fold cross validation was obtained. The future research work will utilize the 3D visualizations such as, 3D surface, 3D column and 3D scatter plots and determine crucial features to perform sentiment classification.

References

[1] Z. Hu, J. Hu, W. Ding and X. Zheng, "Review Sentiment Analysis Based on Deep Learning," 2015 IEEE 12th International Conference on e-Business Engineering, Beijing, 2015, pp. 87-94.doi: 10.1109/ICEBE.2015.24.

[2] Youcef Baghdadi, "A framework for social commerce design, In Information Systems, Volume 60, 2016, pp. 95-113, ISSN 0306-4379, https://doi.org/10.1016/j.is.2016.03.007.

[3] Vincenza Carchiolo, Alessandro Longheu, and Michele Malgeri, "Using Twitter Data and Sentiment Analysis to Study Diseases Dynamics", 2015 Proceedings of the 6th International Conference on Information Technology in Bio- and Medical Informatics - Volume 9267,pp. 16-24. doi:10.1007/978-3-319-22741-2_2

[4] Denecke K, "Sentiment Analysis from Medical Texts. In: Health Web Science. Health Information Science".2015 Springer, Cham. doi:https://doi.org/10.1007/978-3-319-20582-3_10

[5] D Gräbner, M Zanker, G Fliedl, M Fuchs, "Classification of Customer Reviews based on Sentiment Analysis",2012 Information and Communication Technologies in Tourism pp 460-470.

[6] Gann W-JK, Day J, Zhou S, " Twitter analytics for insider trading fraud detection system" 2014. Proceedings of the sencond ASE international conference on Big Data. ASE. May 27 - May 31, 2014, Stanford, CA, USA, 94305

[7] Siddhaling Urolagin, "Text Mining of Tweet for Sentiment Classification and Association with Stock Prices," 2017 International Conference on Computer and Applications (ICCA), Doha, 2017, pp. 384-388.doi: 10.1109/COMAPP.2017.8079788

[8] Kartik Singhal, Basant Agrawal, Namita Mittal, "Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data", Advances in Intelligent Systems and Computing, vol 339. Springer, New Delhi, pp 469-477.

[9] F. Xing and J. Zhan, "Sentiment Analysis Using Product Review Data," Journal of Big Data, vol. 2:5,2015.https://doi.org/10.1186/s40537-015-0015-2

[10] M.Walaa, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," Ain Shams Engineering Journal,vol.5, no. 4,pp. 1093—1113, 2014

[11] Lin Y, Zhang J, Wang X, Zhou A," An information theoretic approach to sentiment polarity classification". 2012,Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality, ACM, New York, NY, USA.pp 35–40.

[12] A. Mudinas, D. Zhang and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis,"2012, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, Beijing, China ugust 12 , 2012, Article No. 5, doi:10.1145/2346676.2346681

[13] T. H. A. Soliman, M. A. Elmasry, A. R. Hedar and M. M. Doss, "Utilizing support vector machines in mining online customer reviews," 2012 22nd International Conference on Computer Theory and Applications (ICCTA), Alexandria, 2012, pp. 192-197. doi: 10.1109/ICCTA.2012.6523568

[14] J. Liang, P. Liu, J. Tan, and S. Bai, "Sentiment Classification Based on AS-LDA Model," Procedia Computer Science, vol. 31, pp. 511—516, 2014.

[15] A. P. Jain and P. Dandannavar, "Application of Machine Learning Techniques to Sentiment Analysis,"2nd International Conference on Applied and Theoretical Computing and Communication Technology, Bangalore, 2016, pp. 628--632.

[16] R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques,"Cluster Computing, https://doi.org/10.1007/s10586-017-1200-1

[17] Y. Wang, Y. Zhang, and B. Liu,"Sentiment Lexicon Expansion Based on Neural PU Learning, Double Dictionary Lookup, and Polarity Association,"Conference on Empirical Methods in Natural Language Processing, Copenhagen, 2017, pp. 7–11.

[18] C.J. Hutto and Eric Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,"Eighth International AAAI Conference on Weblogs and Social Media, 2014. [available online: 08 Nov 2017].

[19] Changbo Wang, Zhao Xiao, Yuhua Liu, Yanru Xu, Aoying Zhou, Kang Zhang, "SentiView: Sentiment Analysis and Visualization for Internet Popular Topics", in IEEE Transactions on Human-Machine Systems, Volume: 43, Issue: 6, Nov. 2013.

[20] Megan K. Torkildson, Kate Starbird, Cecilia Aragon, "Analysis and Visualization of Sentiment and Emotion on Crisis Tweets", in International Conference on Cooperative Design, Visualization and Engineering CDVE 2014: Cooperative Design, Visualization, and Engineering pp 64-67,2014.

[21] Farina J., Mazuran M., Quintarelli E., "Extraction, Sentiment Analysis and Visualization of Massive Public Messages",. In: Catania B. et al.

(eds) New Trends in Databases and Information Systems. Advances in Intelligent Systems and Computing, vol 241. Springer, Cham, 2014.

[22] A. Deshwal and S.K. Sharma, "Twitter sentiment analysis using various classification algorithms," 5th International Conference on Reliability, Infocom Technologies and Optimization, Noida, 2016, pp. 251-257

[23] J. McAuley and A.Yang "Addressing Complex and Subjective Product-Related Queries with Customer Reviews," Proceedings of the 25th International Conference on World Wide Web,Montreal, 2016, pp. 625—635

[24] L. Jure and S.Rok, "SNAP: A General-Purpose Network Analysis and Graph-Mining Library", ACM Transactions on Intelligent Systems and Technology, vol.8, no. 1, pp. 1—20, 2016.

[25] Edward Loper, Steven Bird, "NLTK: the Natural Language Toolkit", ETMTNLP '02 Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1, Pages 63-70 , 2002.