Volume 9, Issue 1
February 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Novel Method for feature Specific Sentiment Analysis of Product Reviews

[1]Nilesh Shelke and [2]Shrinivas Deshpande

[1]Department of Computer Science and Engineering, PIGCE, Nagpur, India

[2]P. G. Department of Computer Science and Technology, Amravati, India

E-mail: [1]nileshshelke08@gmail.com, [2]shrinivasdeshpande68@gmail.com

ABSTRACT

Social media platforms like facebook, flikr etc. are flooding with large number of product reviews. These platforms are motivating their users to record opinions, describe events and beautiful moments of person's life. Millions of users are posting such tweets everyday. If machines can make to understand and analyze these reviews, it can be beneficial to several applications. Sentiment analysis is extraction of the opinion about the features of a product from the given review. It includes NLP tasks, computational linguistics and classifying the opinion i.e. positive, negative or neutral. Feature engineering is very much primary and mandatory task for Sentiment Analysis. This research attempts to test the effectiveness of proposed approach for extraction of features for feature specific Sentiment Analysis of Product Reviews. Feature extraction from reviews is critical for knowing the exact sentiments of the customers. SentiWordNet has been used for calculation sentiment values of sentiment bearing phrases. Continuous positive reviews or continuous negative reviews about any politician, movie, product have direct impact of the persons' personality or on the sale of the products. And the proper analysis of such reviews can increase the sale also.

Keywords: SentiWorNdet, sentiment score, synset, hypernym

1. INTRODUCTION

Sentiment Analysis is a way recognizing opinions of other people regarding any entity. Recently research in Sentiment analysis has seen a spike in its popularity among researchers [1]. Prominent reason behind this is that people are expressing their opinions largely on social media. Millions of people are expressing their views regarding products, events, books, restaurants etc. daily through social media platforms. It has been very challenging to understand the opinions of others as opinion-rich resources such as online review sites and personal blogs are rapidly increasing. Major companies are agreed with the fact that the opinion expressed on internet influences the opinions of readers and ultimately influences brand and purchasing options. Mining of these reviews leverages many valuable applications [2].

Different studies have examined the recognition of reviewer's view from several ways like text, voice and image. First step in the Sentiment analysis is preprocessing of reviews. Feature extraction is the second step. Many researchers are using machine learning approaches along with additional steps for harvesting features from the reviews. Next step include extracting the sense of the reviewer regarding each feature of the product. It is also critical to know with what intensity reviewer liked a feature of a product. Last step includes the overall polarity about a review and its overall intensity. It is also useful to know which the prominent features of a product users are most discussing. Fig. 1 is representing the major steps in Sentiment Analysis.

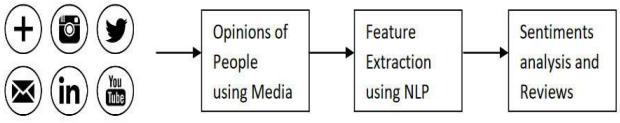

Fig. 1. Process of Sentiment Analysis

Basic requirement of the research problems like sentiment analysis is proper dataset. Researchers have increased their trend in this domain as many social networking sites are providing their APIs to download the posts on their web-sites in standard formats.

Feature engineering is very much primary and mandatory task for Sentiment Analysis. Mechanism of transformation of natural text to a feature vector is the fundamental step for sentiment analysis. Consider the following example reviews from Table 1.1.

For illustration, restaurants, computer and camera reviews are considered. In Table 1.1, the bold words (picture quality, appetizers, restaurant, lunch, delivery service, prices, screen size, netbook, product, battery life) are the products features, or issues about which the opinions are expressed. Underlined words (highly recommended, ok, slow, casual, perfect, fast, better quality) are semantic features that express opinions.

Table 1.1 Samples from data-sets

#	**Review**
1.	A few of my work constituents owned the g2 and highly recommended the canon for **picture quality**.
2.	The **appetizers** are ok, but the service is slow.
3.	There **restaurant** is very casual, but perfect for **lunch**, and their **delivery service** is always very fast. Chinatown definitely has better quality with cheaper **prices.**
4.	This computer has an excellent **screen size** and quality for a **netbook**.

5.	This wonderful **product** was easy to **unbox** and **install** !!
6.	It is a *fantastic* **camera** and well worth the **price**.
7.	I charge it at night and skip taking the **cord** with me because of the good battery life.

2. RELATED WORK

Different approaches regarding how to extract features and how to find sentiments regarding that features have been studied in [3]. Major work of research in Sentiment Analysis is done in machine learning [4]. Many of researchers have preferred deep learning for sentiment analysis. Study of [5] has used Subjectivity and Polarity Detection in Online Courses. Many researchers have also worked upon sentiment analysis for different languages. Research work in [6] revealed different demerits in Arabic language. There is problem of limitations of words in sentiment dictionaries. Study of [7] has designed to examine the effects on sentiment analysis with extended sentiment analysis dictionary. This research work was done in Chinese language.

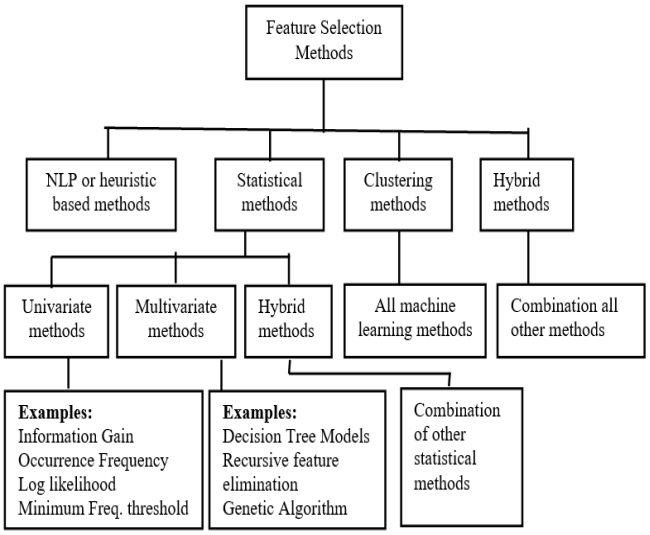

Fig. 2 Summary of Features Selection Techniques

Fig 2 is the summary of feature selection methods researchers worked upon. Statistical feature selection methods include univariate, bivariate, multivariate and hybrid methods. Univariate methods are simplest type of features selection techniques where only one variable exist. Example of variable includes "salary" and "age" of a person. Univariate methods do not look on salary and age at the same time. It also not try to see any relationship between age and salary. Univariate methods focuses on statistical parameters like standard deviation, quartiles, mean and mode etc. Bivariate methods are used to know the relationship between two variables. Bivariate methods tries two to find the relationship between two variables like food intake and weight of a person [8]. Multivariate methods include three or more variable. Example of multivariate methods are cluster analysis, regression etc. In text mining domain like opinion mining, there are two important fields namely text classification and text clustering. Text classification subfield employs information gain, chi-square, and entropy as feature extraction methods while text clustering uses term count, term contribution as feature extraction methods. Genetic Algorithm also has been used for feature selection [9]. Multivariate methods has a serious drawback of being computationally expensive. Researchers are also preferring embedded methods for feature selection. Embedded methods. They blends the merits of all the methods previously described. FRMT algorithm is one of the examples of such types of methods [10].

Feature selection methods using text clustering employs PMI. Frequency determines the head word of the cluster. PMI Similarity is calculated by comparing this head word with other word. And if the distance between these two words crosses some threshold value, major features are declared as candidate features. But, such techniques has serious drawback that they can collect prominent features only [10]. Machine learning algorithms are heavily used in feature selections [11].

3. DATASET

Experiments are carried on customer reviews. It includes cameras, restaurants, antivirus software, computer etc. These are available on Amazon.com. Products on this site have a large number of reviews. These reviews are available in the standard text files.

Sentiment values are calculated from SentiWordNet dictionary. SentiWordNet is an advancement to WordNet.

Table 1: Sample Scores in SentiWordNet 3.0.

Synsets	Positive	Negative	Objective
lucky#1	0.75	0	0.25
roofless#2	0	0.875	0.125
frozen#1	0	0.375	0.625
unconstrained#1	0.5	0	0.5
unfriendly#1	0	0.875	0.125
social#1	0.125	0.25	0.625
ungrudging#1	0.375	0.375	0.25
ungenerous#1	0.375	0.25	0.375
chilly#3	0.125	0.625	0.25
intimate#2	0.125	0	0.875

Every synset is related with negative, positive or objective connotation. SentiWordNet has been made by its developers for the purpose of research. Each synset has positive, negative and objective score. Values of all these score lie between 0.0 and 1.0, sum of which will always be 1.0. Therefore every synset has positive, negative and objective sentiment score as per the context in which those are used. Table 1 represents sample scores available in SentiWordNet 3.0.

4. PROPOSED SYSTEM

Prominent steps for feature extraction for opinion mining has been described in section I and fig. 1. Reviews contains numerical values like camera with 128 MB RAM or the price, pixel capacity of the camera. Reviews also contains dates sometimes. We can either remove numbers/dates or convert

Volume 9, Issue 1
February 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

them into their textual representations. Regular expressions are used to remove the numbers/dates.

Table 2: Shortcuts and spelling errors in the reviews

Ill-Formed Word	Correct Word
Abt	About
b'day/bday	birthday
M	Am
lov/luv/lv	Love
C	See
l8r	later
l8	Late
h8	Hate
b4	Before
eve/evng	evening
b'coz/cos/bcoz/coz	because
R	Are

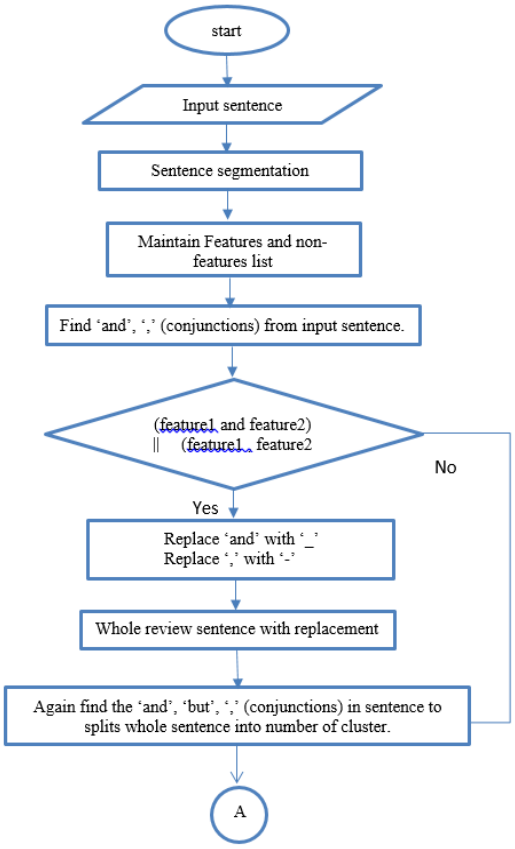

Fig. 3 (a) Proposed method for exaction of product reviews.

First step in pre-processing is lingo corrections. Many commercial word level error detection and correction tools are available which uses the dictionary like Table 2.

All the words that contain less than three characters are also discarded as such words do not contain any feature.

Identification of product features is very critical for boosting the performance of the classification. Fig. 3 (a and b) exhibits the proposed system for extraction of features. For illustrating the proposed method algorithm has been provided with detailed example.

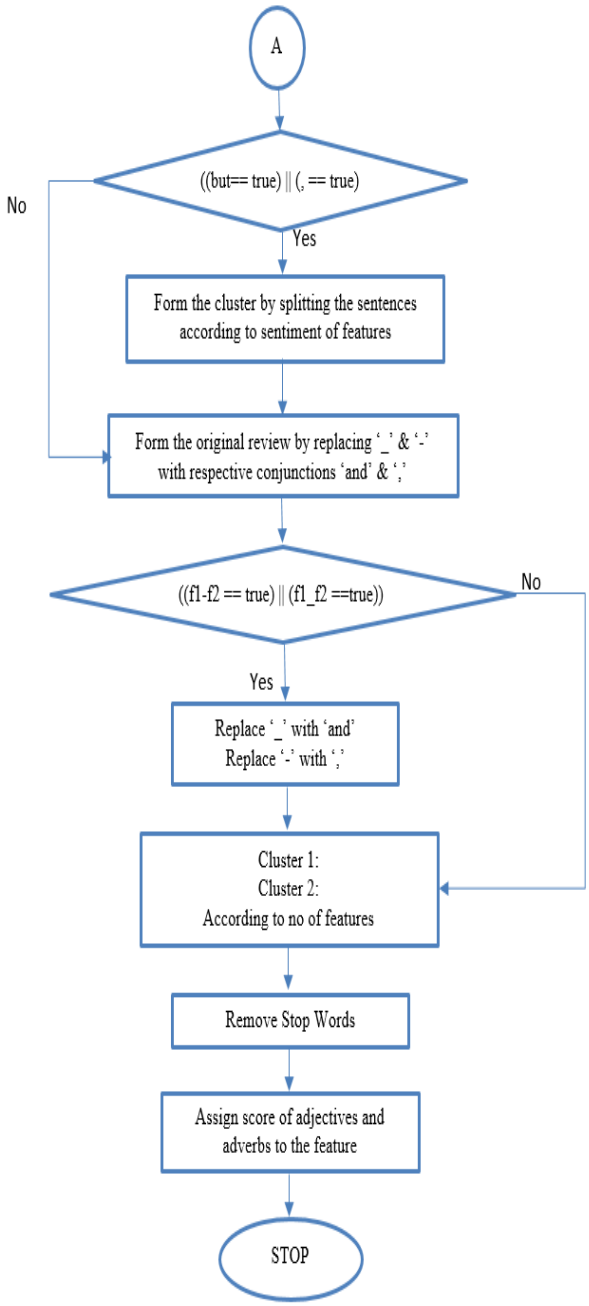

Fig. 3 (b) Proposed method for exaction of product reviews.

Clusters obtained from the implementation of this algorithm thus contains features as well as sentiment bearing phrases. These clusters are provided as an input to the polarity classification module. Block diagram of this module is shown in Fig 4. Generally adjectives appear with nouns and verbs appear with adverbs. Semantic scores of these phrases are calculated from SentiWordNet dictionary. In many cases it happens that we do not get sentiment bearing phrases in SentiWordNet mentioned in the reviews, in such cases we look for synonyms of these phrases.

Language uses a single word in many senses. Table 3 shows the different senses of the same word (Trouble) in SentiWordNet. If multiple senses occur then average value of sentiment score is used for nullifying the effect of choosing any wrong sense.

Volume 9, Issue 1
February 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Step 1:- **Start**

Step 2:- **Input Sentence** inputString;
"The food is good but waiter, parking and services are bad, music is awesome and sushi was best."

Step 3:- **Sentence segmentation-Divide the Sentence into Words and assign index value.**
1-The
2-food
3-is
4-good…….etc.

Step 4:- **Feature List Formation-Feature list formed from feature class according to Domain and maintain Feature in Array.**
Feature List - food, waiter, parking, services, music, sushi.

Step 5:- **Finding feature index from original Indexing**
FeatureIndex-2(food),6(waiter) 7(parking), 9(services), 12(music), 16(sushi).

Step 6:- **Finding Non-Feature index by deleting feature index from array**
Non-FeatureIndex-1(The), 3(is), 4(good), 5(but), 8(and), 10(are), 11 (bad), 13(is), 14(awesome), 15(and), 17(was), 18(best).

Step 7:- **Find the conjunctions from sentence such as ',', 'and'.**

Step 8:- **Check for the conjunction conditions to replace them with particular '_' and '-'.**
If(feature1 and feature2)
// 'and' surrounded with the features
{
Replace 'and' with '_'.
}
"The food is good but waiter, parking_services are bad, music is awesome and sushi was best."
If(feature1 , feature2) // Features separated with ','
{
Replace the ',' with '-'.
}
"The food is good but waiter-parking_services are bad, music is awesome _and sushi was best."

Step 9:- **Again Check for the conjuctions (but, ',', and) in sentence to split the sentence in number of sentences to form a cluster of features.**
If(but==true || ,==true || and==true)
// presence of 'but' keyword in sentence
// presence of ',' keyword in sentence
// presence of 'and' keyword in sentence
{
Split the sentence from 'but' or ',' or 'and' keyword.
}
Cluster 1- "The food is good."
"waiter-parking_services are bad, music is awesome and sushi was best."
Cluster 2- "waiter-parking_services are bad."
"music is awesome and sushi was best."
Cluster 3-"music is awesome".
Cluster 4-"sushi was best".

Step 10:- **Form the original review by replacing the '_', '-' with respective keywords and punctuation 'and', ','**
If(feature1_feature2==true){
Replace '_' with 'and'
}
Cluster 2- "waiter-parking and services are bad."
If(feature1-feature2==true){
Replace '-' with ','
}
Cluster 2- "waiter, parking and services are bad."

Step 11:- **Final Clusters.**
Cluster 1- "The food is good."
Cluster 2- "waiter, parking and services are bad."
Cluster 3-"music is awesome".
Cluster 4-"sushi was best".

Step 12:- Remove Stop Words.

Step 13 Assign score of adjectives and adverbs to the feature.
Assign score of good from SentiWordNet to food.
Assign score of bad from SentiWordNet to waiter and parking.
Assign score of awesome from SentiWordNet to music.
Assign score of best to sushi.

Step 14 Stop

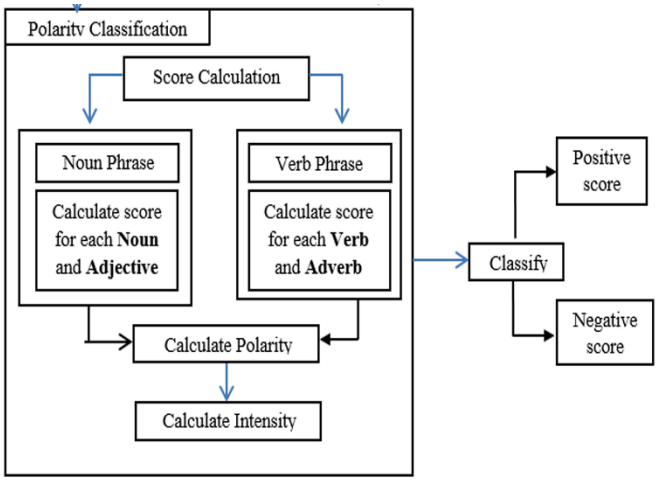

Fig. 4: Block diagram of polarity classification module

Table 3: Fragment of SentiWordNet with different senses of the same word

Sr. No.	Word	Sense	Positive score	Negative score
1	Trouble	trouble#1 a "This book troubled me"; "A troubling thought"	0.25	0.375
2		trouble#2 put_out#1 bother#3 to	0	0.875

Volume 9, Issue 1
February 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

		"Sorry to trouble you, but..."		
3		trouble#3 perturb#1 cark#1; "She was rather perturbed by the news that her father was seriously ill"	0	0.125
4		trouble#5 inconvenience "She have not troubled to call his mother on her birthday"; "Don't bother, please";	0	0.5

5. RESULT ANALYSIS AND DISCUSSION

Evaluation parameters of 1000 restaurants reviews are shown in Table 4 and Fig. 5 and Table 5 and Fig 6. are result analysis of camera reviews.

Table 4: restaurant reviews

	Precision	Recall	FM	Accuracy
Proposed System	0.917	0.943	0.929	0.9

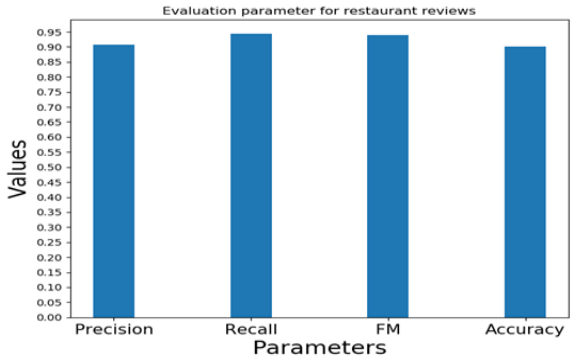

Fig. 5 restaurant reviews

Table 5: camera reviews

	Precision	Recall	FM	Accuracy
Proposed System	0.933	0.9300	0.93	0.841

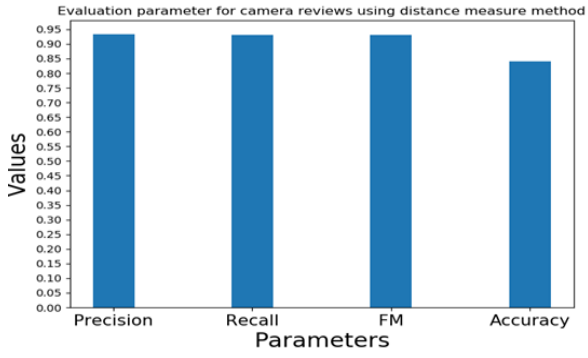

Fig. 6 Camera reviews

For knowing overall much discussed prominent features and customer feedback towards those features, proposed system is trained so that if total reviews are inputted to the system at once, it provides opinion analysis of customer towards the various features of the product. Fig. 8 and 9 illustrates this fact. This type of sentiment analysis is beneficial for corporate groups, who can evaluate trends among all properties in a single snapshot. This way they can understand as a group or brand where improvements need to be made.

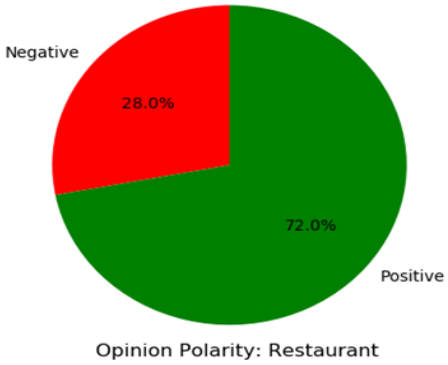

Fig. 7 Overall opinion mining

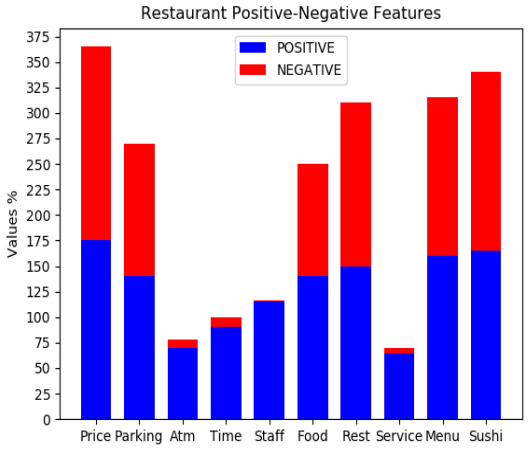

Fig. 8 Prominent Features along with polarity from 1000 restaurant reviews

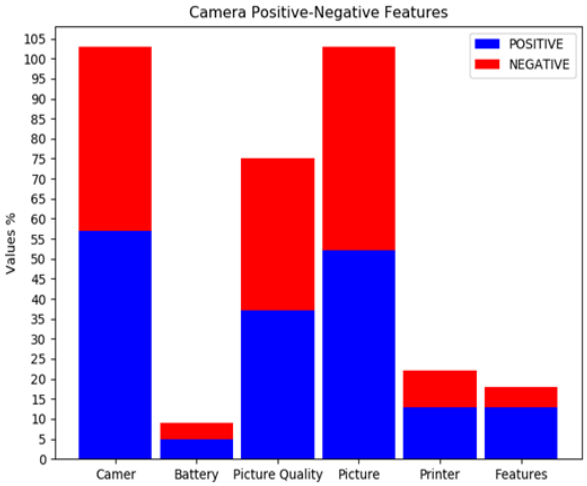

Fig. 9 Prominent Features in camera reviews along with polarity

6. CONCLUSION

Currently, many approaches have been introduced which works on distance measures and heavily use machine learning/ deep learning algorithms for extraction features from reviews. Contribution of this study is to introduce feature selection method from product reviews. It is found that implemented method is performing well. For polarity classification, SentiWordNet has been found very handy.

There are seemingly large applications and advantages. Work presented here can be pursued further in several directions. Customer satisfaction, work performance, appraisal, quality enrichment, reputation management, market search, forecasting are some of the potential applications.

All the internet users are not conversant in English. There is strong growing need for construction of sentiment analysis tools for languages other than English. Other languages have lack of resources. This research work also can be extended for fake or bogus opinion refer opinion spam detection sentiment analysis. Implementation and evaluation of sentiment analysis methods for the effectiveness of close monitoring of social customer relationship management is a good research avenue and big data technologies adoption can healthy platforms for it.

REFERENCES

[1] Devendra Kamalapurkar, Ninad Bagwe, R. Harikrishnan, "Sentiment Analysis of Product Reviews", International Journal of Engineering Sciences & Research Technology, Vol. 3 Issue 2, pp. 456-460, 2017.

[2] Ghowsalya Balu, "Forecast rating based on public comments by sentiment analysis" (2019, March 19) available on www.adlibris.com

[3] Mahesh G., Satish Kumar T., Shreya S., Sushmitha N., Sripad T.,"Exploring Classification Techniques for Sentiment Analysis",International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-6, March 2020, pp. 720-726.

[4] A NithyaKalyani, S Ushasukhanya, TYJ Nagamalleswari and S Girija, "Rating prediction using textual reviews", National Conference on Mathematical Techniques and its Applications (NCMTA 18), pp. 202-207, 2018.

[5] S. JayaKumar, Priyank Patel, Rajat Kumar Singh, Shivraj, "Election Result Prediction Using Deep Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-4, pp. 74-77, April 2019

[6] Gerardus Blokdyk, "Sentiment Analysis A Complete Guide" (2018, May 12), available on walmart.com.

[7] Mark Keenan, "Advanced Positioning, Flow and Sentiment Analysis in Commodity Markets", (2nd ed.), 2019.

[8] Arindam Chaudhuri, "Visual and Text Sentiment Analysis through Hierarchical Deep Learning Networks", 2019 available on www.bokklubben.no.

[9] Joseph M. De Guia, Madhavi Devaraj, "Methods and Trends in Natural Language Processing Applications in Big Data", International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-6S5, April 2019, pp. 234-249.

[10] Pankaj Verma, Sanjay Jamwal, "Mining Public Opinion on Indian Government Policies using R", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-3, pp. 1310-1315, January 2020.

[11] Mong Shen Ng, "People Analytics & Text Mining with R", (2019, March 21), available on www.dymocks.com.au

AUTHOR PROFILES

Dr. Nilesh Shelke is a senior faculty at CSE Department of PIGCE, Nagpur. He is MCA, M. Tech., M.Phil and also acquire PhD in Computer Science & Engineering. He is Microsoft Certified Solution Developer and has more than 20 years of experience of imparting IT Training which includes learners from different streams, faculties and Microsoft Certifications to corporate employees. He has published patents, copyrights and sellable technical articles in the renowned journals. He has successfully completed many certifications from IIT including Machine Learning, Deep Learning, Data Science, Databases etc. He is also mentoring several students in student's centric activities where many of his students is getting outstanding performances including excellent placements, IIT Certifications, paper publications etc.

Dr. S. P. Deshpande is currently working as Associate Professor at Post Graduate Department of Computer Science & Technology, MCA at Shree H. V. P .Mandal's DCPE, Amravati since last 22 years. He has guided more than 100 students at Post Graduate level. His interest of research is DBMS, Data Mining, Web based technologies, AI and NLP. He holds Doctorate Degree from Sant Gadge Baba Amravati University, Amravati