International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 12, Volume 6 (December 2019) www.irjcs.com

SENTIMENT ANALYSIS OF CUSTOMER SATISFACTION USING DEEP LEARNING

Shreyas Rajesh Labhsetwar
Department of Computer Engineering,
Fr. C. Rodrigues Institute of Technology,
Navi Mumbai 400703, India
shreyas.labh@gmail.com

Manuscript History
Number: **IRJCS/RS/Vol.06/Issue12/DCCS10083**
Received: 13, December 2019
Final Correction: 19, December 2019
Final Accepted: 24, December 2019
Published: **December 2019**
Citation: Shreyas (2019). Sentiment Analysis of Customer Satisfaction using Deep Learning. International Research Journal of Computer Science (IRJCS), Volume VI, 709-715. **doi://10.26562/IRJCS.2019.DCCS10083**
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India

Abstract—The rapid advancement of web technology has led to an exponential increase in the volume of data present online for internet users. Travellers book their hotels only after extensive scrutinisation of hotel reviews on online websites. Hence, it is an absolute necessity for hotel's management board to gain insights from customer reviews and feedbacks to improve upon their services for a better customer satisfaction index. This research explores the use of Artificial Neural Networks (ANN) powered by Google's Word2Vec skip-gram algorithm for customer sentiment analysis and review classification. The proposed model achieves a high test accuracy of 0.9248, with an average F1-Score of 0.925. Unsupervised sentiment clustering effectively classifies the reviews into four distinct categories and enables the Hotel Management to work out the major problems experienced by the customers.

Keywords— Deep Learning; Artificial Neural Networks; Sentiment Analysis; Customer Satisfaction;

I.INTRODUCTION

With a rapid increase in the volume of data present online, computationally intelligent methods are proving to be extremely important to analyse the data in many industries. For example, in telecommunication industry, data is mined for extracting patterns which would help predict and pre-empt potential customer churn, and improve customer retention and marketing [1].The advancement of Internet has changed the way millions of people express their views and opinions, which is now mainly through online forums, booking websites, blog posts, and social media. The online communities are a source of interactive media enabling consumers to inform and influence others. People mostly make their decisions based upon online content, making it extremely essential for companies to gain insights to improve upon their user experience and fetch good reviews [2]. The huge amount of data that warrants analysis is actually present in written form as reviews and feedback instead of mathematical data. The written text is subject to interpretation, and converting and representing the data in an absolute mathematical syntax (such as a binary system) is cumbersome. However, computationally intelligent methods allow for such fuzziness, and successfully extract patterns from such data [3]. Sentiment analysis consists of various computational areas such as natural language processing, data mining and text mining, and is rapidly becoming crucial to organisations as they strive to integrate computationally intelligent algorithms into their operations, and attempt to improve upon their products and services [4]. Established organizations as well as individuals can make use of sentiment analysis and opinion mining. When an individual wishes to buy a product or decides whether to use a service or not, he/she can go through a large number of user reviews, but reading manually and analyzing all of them would be a lengthy and time consuming process.

Also, when an organization seeks to understand the public opinion about their products, it needs to consider an overwhelmingly large data of available customer reviews [5]. With sentiment analysis algorithms, large amounts of data can be analysed automatically to gain insights of opinions and patterns. Also, the bulk of data can be classified into a definite number of classes. Thereafter, the classes critical to an organisation's growth can be scrutinised with a higher priority to improve upon the shortcomings. The effectiveness of such measures can be clearly identified with shrinking number of reviews in critical classes.

II. RELATED WORKS

B. Pang and L. Lee used the most popular approach for sentiment analysis called the bag of words approach consisting of two modes of summarization; single-document opinion-oriented summarization and multi-document opinion-oriented summarization. They also delved into the application of unsupervised machine learning approaches and classification, thereby evaluating the economic impact of reviews and implications of manipulation [6]. V. Rentoumi, S. Petrakis, M. Klenner, G. A. Vouros, and V. Karkaletsis proposed a hybrid model comprising of machine learning system and rules-based system in which the issue of erroneous evaluations due to lack of decision making capability pertaining to polarity orientation of sentences was tackled . The approach is evaluated on the rotten tomatoes movie reviews dataset and compared with other state of the art approaches [7]. In Research [8], social media analytics is performed on twitter data using support vector machine algorithm to classify tweets into positive and negative categories, achieving a classification accuracy of 87%. Stanimira Yordanova and Dorina Kabakchieva presented an approach for prediction of customer opinion using decision tree algorithm for binary classification of hotel reviews as positive or negative. They demonstrated that most accurate classification is achieved after removal of rear words and applying the model generated from the balanced training set [9].

III.DATA SET

The dataset used in this research [10] was scraped from Booking.com, and is publicly available. It consists of 515,000 customer reviews and scoring of 1493 luxury hotels across Europe, with their geographical locations. Table 1 highlights all the 17 attributes contained in the dataset.

TABLE 1

FEATURE	DESCRIPTION
Hotel Address	Address of the hotel
ReviewDate	Date of posting the review.
AverageScore	Average score of the hotel, calculated based on the latest comment in the last year.
HotelName	Name of the hotel
ReviewerNationality	Nationality of the reviewer
NegativeReview	Negative Review the reviewer gave to the hotel
Review_TotalNegativeWord_Counts	Total count of words in NegativeReview
PositiveReview	Positive Review the reviewer gave to hotel
Review_TotalPositiveWord_Counts	Total count of words in PositiveReview
ReviewerScore	Score given by the reviewer to the hotel
Total_Number_of_Reviews_Reviewer_Has_Given	Number of reviews the reviewer has given in the past
TotalNumberOfReviews	Total number of valid reviews a hotel has.
Tags	Tags given by reviewer to the hotel
DaysSinceReview	Duration between the review and scrape date
AdditionalNumberofScoring	Numbering of scorings (not reviews)
HLatitude	Latitude of the hotel
HLongitude	Longitude of the hotel

IV.METHODOLOGY

A. Deep Learning

Deep learning is a type of machine learning methodology that empowers computers to imitate the way humans perceive items and gain knowledge. Deep learning is an intricate element of data science, which includes statistics and predictive modelling.It is being applied in various domains including agriculture, health-care, cyber security, robotics, computer vision, and natural language processing [11]. Deep learning involves building up a neural network capable of extracting valuable insights from any form of data. A neural network is made up of multiple neurons interconnected with each other, such that each connection of the network is associated with a weight, which gets multiplied to the input value, thereby generating a weighted output.

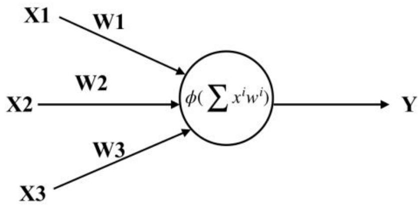

Fig. 1 An Artificial Neuron

Each neuron consists of an activation function that introduces non-linearity in the modelling capabilities of the network, and its output consists of the activated-weighted sum of inputs. There exist several types of activation functions including Sigmoid, Tanh, ReLu and Leaky-ReLu functions [12],[13].

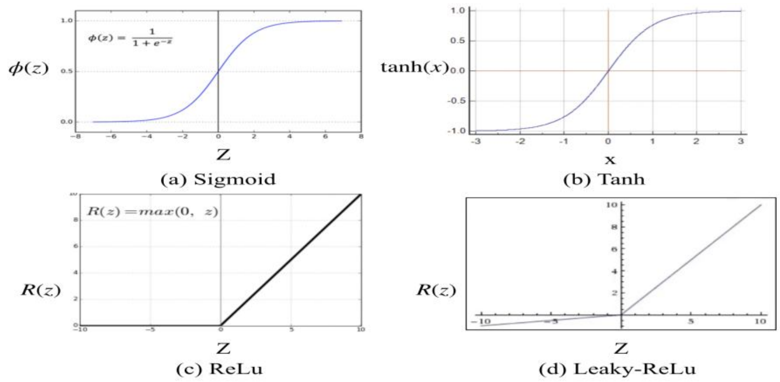

Fig. 2 Activation Functions

The first phase of training a neural network involves forward-propagation of the training data [14], which propagate the entire neural network and their predictions (labels) are calculated. During this phase, each neuron of each layer applies its transformation to the information it receives from the neurons of the previous layer and sends it to the neurons of the next layer. Next, an error function computes the error between test data and predicted data to measure how accurate the prediction result is compared to the correct result [15].Ideally, the error/loss must be equal to zero, that is, no divergence between estimated and expected value. Therefore, as the model gets trained, the weights associated with each inter-neuron interconnection are gradually adjusted until the error function reaches a minima. After the calculation of loss, it is propagated backwards in the neural network, and this phase is called back-propagation [16]. Starting from the output layer, the loss information propagates through all the neurons in the hidden layers, based on which the weight associated with each inter-neuron connection is adjusted. This process is repeated, layer by layer, until all the neurons in the network have adjusted their weights.

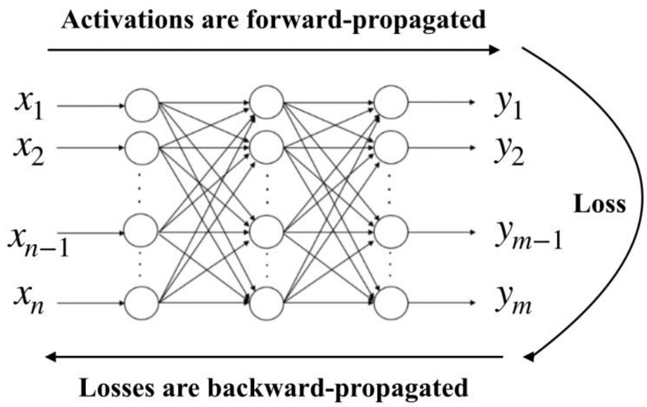

Fig. 3 Training a Neural Network

Optimization algorithms are used to minimise the loss by finding the minima of Objective function/Loss function which is a mathematical function dependent on the model's internal learnable parameters which are used in computing the target values (y) corresponding to the set of predictors (X) used in the model.

The internal parameters of a model play a vital role in effectively and efficiently training the model and produce most accurate results.Hence, various Optimization strategies and algorithms are used to update and compute appropriate and optimum values of such model's parameters.Optimization algorithms are mainly classified into two categories:

1. First Order Optimization Algorithms
2. Second Oder Optimization Algorithms

B. Word Embedding

Word embeddings is a class of methodologies [17] which provide a dense vector representation of words and documents, and their relative meanings.They are an improvement over the traditional sparse representations employed in simpler bag of word models.Graphically, a dense vector represents the projection of the given word into a continuous vector space [18]. The position of a word within the vector space, also called as embedding, is learnt from the input text based upon the set of words surrounding the given word. Keras, Python's deep learning library, provides an Embedding layer implementation which can be used for neural networks to process text data [19].

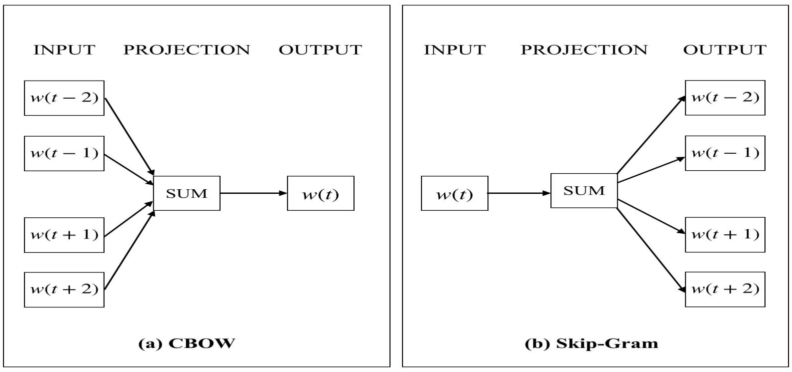

Fig. 4 Complete Neural Network

C. Natural Language Processing

Natural Language Processing (NLP) is a discipline that focuses on the interaction between data science and human language, thereby giving computers the ability to parse, understand and derive insights from human languages [20]. NLP is extensively used in various industries including healthcare, agricultural industry, fake news detection, virtual assistants, and many more. NLP involves applying algorithms to categorise and extract the natural language rules such that the unstructured data gets converted into a computer parsable form.

The two main techniques utilised in NLP are syntactic analysis and semantic analysis. Syntactic analysis involves techniques including lemmatization, morphological segmentation, word segmentation, part-of-speech tagging, parsing, sentence breaking and stemming. Semantic analysis involves techniques including named entity recognition, word sense disambiguation, natural language generation, and many more [21].

D. Word2Vec

Word2Vec is a cluster of related models which can be used to generate word embeddings [22]. These models are shallow neural networks that are trained to convert linguistic context of words into vectors. Word2vec converts a large chunk of text into a vector space, of the order of a several hundred dimensions, with one vector assigned for each unique word in the word space. Word2Vec offers two modes of analysis; Continuous Bag-Of-Words (CBOW) and Continuous Skip-Gram (SG) [23]. CBOW attempts to predict the output label (target word) from its neighbouring words (context words) , thereby achieving a better accuracy for frequent words.In contrast, CSG predicts the context words from a given target word, working well even for infrequent words.

Neural Probabilistic models are generally trained using the principle of Maximum Likelihood which maximizes theprobability of the next word wt (target) given the previous words h (history) by applying softmax activation as shown [24]:

$$P(wt|h)=\text{softmax}(\text{score}(wt,h)) \quad (1)$$

$$P(wt|h)= \frac{\exp(\text{score}(wt,h))}{\sum_{Word \in Vocab} \exp(\text{score}(w,h))} \quad (2)$$

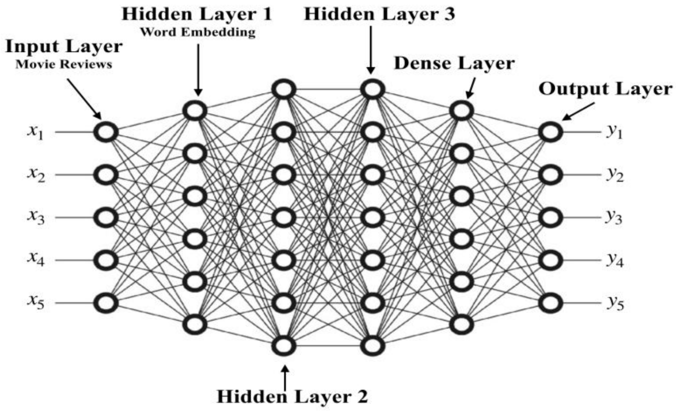

Fig. 5 Word2Vec Models

Where,
Score (wt,h) computes the compatibility of the target word wt in the context h.
This model is trained by maximising its log-likelihood on the training set.Its loss function is given as:

$$JML = \log(P(wt\,|h)) \quad (3)$$

$$JML = score(wt,h) - \log(\sum \exp(score(wt,h))) \quad (4)$$

The Skip-Gram model defines the probability as follows:

$$P(wi|wt\,;\theta) = \exp(\theta wi)\sum \exp(\theta wt) \quad (5)$$

Where,
wi is a one-hot encoded N-dimensional vector
θ represents N*K embedding matrix for N words in K dimensions.

V.RESULTS

The research is performed on 515K Hotel Reviews Data in Europe Dataset [10] consisting of 515000 customer reviews for 1493 luxury European hotels, across 17 attributes. It is observed that Deep Learning using Artificial Neural Networks provides excellent capability for sentiment classification on text data. The reviews are classified into the four classes mentioned in Table 2.

Table 2

CLASS	CONDITION
Best	sentiment_score >= 0.7
Good	0.7 > sentiment_score >= 0.5
Bad	0.5 > sentiment_score >= 0.3
Worst	0.3 > sentiment_score

The Word2Vec Skip-Gram Neural Network consists of 256 neurons in the hidden layer, hence, each word has a 256-dimensional vector representation. The Neural Network achieves a Test Accuracy of 0.9248 and a Test Loss of 0.2374. The Confusion Matrix of the classification is shown in Fig 6.

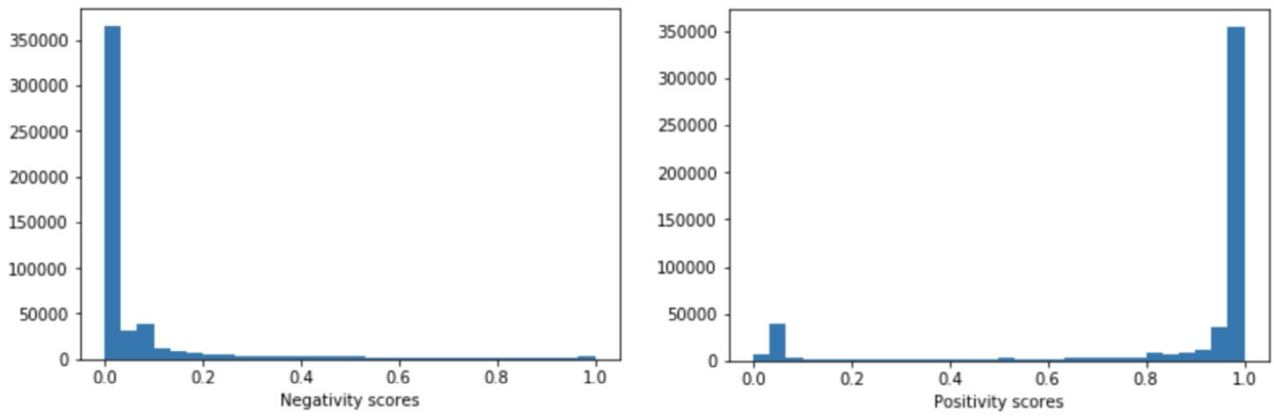

Fig. 6 Confusion Matrix

The Neural Network achieves a very high classification accuracy of 0.9813 and the precision, f1-score, recall and support values are mentioned in Table 3.

Table 3

TYPE	PRECISION	SPECIFICITY	F1-SCORE	SUPPORT
CLASS 0	0.89	0.95	0.92	2428
CLASS 1	0.96	0.90	0.93	3122
MICRO AVG	0.92	0.92	0.92	5550
MACRO AVG	0.92	0.93	0.92	5550
WEIGHTED AVG	0.93	0.92	0.93	5550

The Confusion Matrix, Positivity Scores and Negativity Scores show that only a small number of reviews are misclassified.

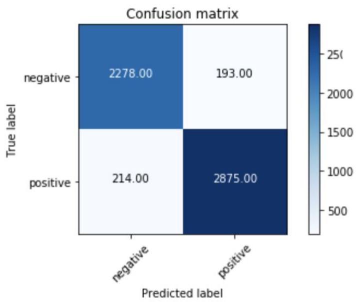

Fig. 7 Positivity and Negativity Score Plots

Thus, all the reviews were accurately classified into the above mentioned four categories. The proportion of reviews classified in each category are shown in Fig 8.

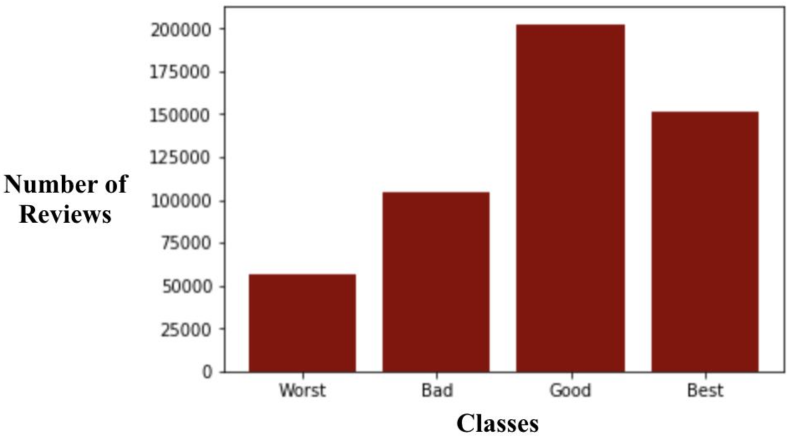

Fig. 8 Number of reviews vs Class Distribution

VI. CONCLUSION

The research presents a deep learning methodology for sentiment analysis of customer satisfaction from hotel reviews. The research delves into the fundamentals of Deep Learning ,Word Embedding and Natural Language Processing. The proposed Artificial Neural Network powered by Google's Word2Vec skip-gram algorithm accurately and successfully classifies the customer reviews into four categories of Best, Good, Bad and Worst. Hotel managements can utilise such algorithms to gain insights from online reviews and improve upon their shortcomings for a better customer satisfaction and online ranking. A better ranking can boost their revenues and provide incentives for further betterment of services. In general, service oriented companies and institutions can utilise such technologies to keep track of their customer feedback.

REFERENCES

1. B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications." Expert Systems with Applications, 39(1), 2012, pp. 1414-1425
2. Pennington, Buddy &Chapman, Suzanne &Fry, Amy &Deschenes, Amy &McDonald, Courtney. (2016).Strategies to Improve the User Experience. Serials Review. 42. 00-00. https://doi/10.1080/00987913.2016.1140614.
3. Gullo, Francesco. (2015). From Patterns in Data to Knowledge Discovery: What Data Mining Can Do. Physics Procedia. 62. 18–22. 10.1016/j.phpro.2015.02.005.

4. Farhadloo,Mohsen & Rolland,Erik.(2016).Fundamentals of Sentiment Analysis and Its Applications. https://doi/10.1007/978-3-319-30319-2_1
5. Sun, Shiliang &Luo, Chen &Chen, Junyu. (2016). A Review of Natural Language Processing Techniques for Opinion Mining Systems. Information Fusion. 36. https://doi:/10.1016/j.inffus.2016.10.004.
6. B. Pang and L. Lee. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1135.
7. V. Rentoumi, S. Petrakis, M. Klenner, G. A. Vouros, and V. Karkaletsis. 2010. United we stand: Improving sentiment analysis by joining machine learning and rule based methods. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta.
8. Alotaibi, Shaha & Alnassar, Allulo & Alshahrani, Asma & Al-Mubarak, Amany & Albugami, Sara & Almutiri,Nada &Albugami, Aisha. (2018). Customer Satisfaction Measurement using Sentiment Analysis. International Journal of Advanced Computer Science and Applications. https://doi:/9.10.14569/IJACSA.2018.090216.
9. Stanimira Yordanova and Dorina Kabakchieva, "Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning". International Journal of Computer Applications (0975 – 8887) Volume 158 – No 5. January 2017
10. Kaggle, "515K Hotel Reviews Data in Europe". Available at: https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe Accessed on 28 November, 2019.
11. Vargas, Rocio & Mosavi, Amir & Ruiz, Ramon. (2017). DEEP LEARNING: A REVIEW. Advances in Intelligent Systems and Computing. 5.
12. Nwankpa, Chigozie & Ijomah, Winifred & Gachagan, Anthony & Marshall, Stephen. (2018). Activation Functions: Comparison of trends in Practice and Research for Deep Learning.
13. Olgac, A &Karlik, Bekir. (2011). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. International Journal of Artificial Intelligence And Expert Systems. 1. 111-122.
14. Sazli, Murat. (2006). A brief review of feed-forward neural networks. Communications, Faculty Of Science, University of Ankara. 50. 11-17. https://doi:/10.1501/0003168.
15. Nie, Feiping & Zhanxuan, Hu & Li, Xuelong. (2018). An investigation for loss functions widely used in machine learning. Communications in Information and Systems. 18. 37-52. 10.4310/CIS.2018.v18.n1.a2.
16. Hossain, Md &Rahman, Md. Mijanur &Prodhan, Dr. Uzzal &Khan, Md. (2013). Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition. International Journal of Information Sciences and Techniques. 3. https://doi:/10.5121/ijist.2013.3401.
17. Li,Yang & Yang,Tao.(2017).Word Embedding for Understanding Natural Language:A Survey. https://doi:/10.1007/978-3-319-53817-4.
18. Teofili, Tommaso. (2017). par2hier: towards vector representations for hierarchical content. Procedia Computer Science. 108. https://doi:/10.1016/j.procs.2017.05.077.
19. Zebin, Tahmina. (2017). Training Deep Neural Networks in Python Keras Framework(Tensorflow Backend) with Inertial Sensor Data for Human Activity Classification.
20. Verspoor, Karin & Cohen, Kevin. (2013). Natural Language Processing. 10.1007/978-1-4419-9863-7_158.
21. Dr.Michael J.Garbade,"A Simple Introduction to Natural Language Processing".Available at: https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32.Accessed on:20 December,2019.Dhingra, Atul. (2017). How Does Word2Vec Work ?. 10.13140/RG.2.2.18448.92163.
22. Derek Chia, "An implementation guide to Word2Vec using NumPy and Google Sheets". Available at: https://towardsdatascience.com/an-implementation-guide-to-word2vec-using-numpy-and-google-sheets-13445eebd281.Accessed on: 20 December, 2019.
23. Chen, Ding-Geng &Lio, Yuhlong. (2009). A Note on the Maximum Likelihood Estimation for the Generalized Gamma Distribution Parameters under Progressive Type-II Censoring. International Journal of Intelligent Technology and Applied Statistics. 2. 57-64.