

Deep Learning based Sentiment Analysis for Social Media Network

R. Murugeswari, T. Manjunath Kumar, D. Devaraj , P. Karthika



Abstract: Sentiment Analysis is the process of identifying opinions expressed in a piece of text. It determines whether the writer's attitude towards a product is positive, negative, or neutral. Sentiment evaluation addresses such need by way of detecting evaluations on the social media textual content. Product evaluations are valuable for upcoming shoppers in supporting them make choices. In recent, deep learning is loom as a powerful manner for fixing sentiment classification troubles. The neural network intrinsically learns a beneficial representation without the efforts of human. This paper presents the overall performance evaluations of deep learning classifiers for big-scale sentiment evaluation. In this system the reviews from the online shopping website called flipkart.com is analyzed and divided as positive, negative and neutral by Multilayer Perceptron (MLP) Neural Network depending on the aspect of the product. The proposed work is simulated by using SPYDER. In our system the accuracy, precision, F-measure and recall is calculated for Multilayer Perceptron (MLP) Neural Network, Random Forest and Support Vector Machine (SVM) algorithm. During comparison Multilayer Perceptron (MLP) Neural Network gives the best accuracy of 99% than other two algorithms.

Keywords: Multilayer Perceptron (MLP) Neural Network, Random Forest, Support Vector Machine (SVM), Sentiment Analysis.

I. INTRODUCTION

Rapid success in the internet makes us to the use of ecommerce. Ecommerce is selling merchandise on on-line. Many clients share their own opinions approximately products on online. The consumer opinion provides a brand-new way in selection making procedure to make effect on commercial enterprise version. Online buying is a simplest way to buy the goods from proprietor over the internet by the use of various forms of web sites and apps. The users additionally keep in mind the overview of other on-line buyers while shopping for the product on on-line web sites which enables the brand-new on-line shopper's lots.

Manuscript published on 30 December 2019.

* Correspondence Author (s)

Dr.R.Murugeswari*, Computer Science and Engineering, Kalasalingam Academy of Research and Education, KrishnanKoil, India. Email: r.murugeswari@klu.ac.in

T.ManjunathKumar, Computer Science and Engineering, Kalasalingam Academy of Research and Education, KrishnanKoil, India. Email: manjunathkumar.t@klu.ac.in

D.Devaraj, Electrical and Electronics Engineering, Kalasalingam Academy of Research and Education, KrishnanKoil, India. Email: deva230@yahoo.com

P.Karthika, Computer Science and Engineering, Kalasalingam Academy of Research and Education, KrishnanKoil, India. Email: karthikaagila@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Thus, we want to observe and analyze client opinions.

In general ratings are provided in star layout. Each and every product has range of rankings. It could be very difficult to read each rating in detail. Many researches proven pictorial representation is powerful and can be memorized, understood without problems instead of textual representations. Sentiment evaluation for product review is analyzed for diverse online web sites known as flipkart. Com, amazon. Com as well as for rediff.Com.

Sentiment evaluation helps each provider as well as the consumer. It enables the provider to introduce new products and the consumer to choose the authentic product by means of using the user rating on online websites. The consumer review is classified on three extraordinary foundation called positive, negative and neutral. In this paper, the user rating for products is extracted from the online buying website known as flipkart. Com and then the extracted dataset has been saved in CSV report. The information extracted is processed by using Multilayer Perceptron (MLP) Neural Network and the accuracy is measured.

II. RELATED WORK

In [1] proposed deep learning framework for product review sentiment class which employs scores as vulnerable supervision alerts. The framework consists of two steps called mastering high degree representation which captures the overall sentiment distribution of the sentences via score statistics and then adding a classification layer on top of the embedding layer. The author discovered two kinds of low-level community structure for modeling evaluation sentences, specifically, Convolutional function extractors and lengthy short-time period reminiscence. And examined the proposed framework, they constructed a dataset containing 1.1M weakly labeled overview sentences and eleven,754 labeled evaluation sentences from Amazon.

In [2] proposed Convolutional neural networks (CNN) for multimedia sentiment evaluation. CNN architectures were used for studying visible features and textual functions, CNN structure was used for exploiting the inner relation between textual content and image. Here the author evaluates sentiment by means of taking pictures the mixed records of texts and snap shots. The proposed approach exploits the inner relation between textual content and picture in photograph tweets and achieved good performance in sentiment prediction. In [3] described Probabilistic Neural community (PNN) with a self-adaptive approach to perform sentiment analysis on tweets. Probabilistic Neural Network as a multi-layered feed-ahead neural network was used for distinguished capabilities of adaptive studying, fault tolerance, parallelism and generalization which offer a superior overall performance.



The smoothing parameter of PNN performed a top-notch role for predicting a correct elegance of classifier. Self-adaptive algorithm was described to calculate and optimize the smoothing parameter. Two varieties of Probabilistic Neural Network fashions were implemented inside the proposed technique.

In [4] highlighted brand-new studies regarding the implementation of deep studying models including deep neural networks, convolutional neural networks and plenty of more for fixing different troubles of sentiment analysis which includes sentiment type, move lingual problems, textual and visible evaluation and product overview evaluation. Different researches have been discussed in this review to offer a deep view regarding developing of deep studying programs in the field of sentiment evaluation.

In [5] performed extraordinary configuration settings of convolutional neural network (CNN) and accomplished Sentiment Analysis of Hindi movie opinions gathered from on-line newspapers and Web sites. The dataset had been manually annotated via 3 native audio system of Hindi to prepare it for schooling of the model. Here the sentences of critiques were classified into three classes which includes fine, bad and neutral. All the experiments are done using various parameter settings for every CNN model.

In [6] focused distributed representations for words by means of the usage of word2vec to reduce the wide variety of dimensions of one-of-K representations. Two experiments had been conducted: (1) sentiment analysis the usage of a small information set, the IMDB dataset, and (2) sentiment analysis the use of a large-scale statistics set, Rakuten Data. By using Rakuten Data the neural networks were examined. Overall performance opinions of deep learning classifiers for massive-scale sentiment analysis were done by using Rakuten Data.

In [7] focused on data mining methods and techniques to extract and seize data for analysis the subjective opinion of a document or collection of files like opinions, social media, and e-commerce sites. In the sphere of sentiment evaluation there are many algorithms should tackle NLP issues to identify the wonderful and poor evaluations of the consumers to your products on on-line marketplace. Here the reviews were collected from the online websites called Amazon.com Flipkart.com, Redif.com. The evaluation of product reviews was classified into effective, poor and neutral categories which suggest the sentiment of a product.

In [8] proposed a technique of presenting automatic comments on the idea of statistics accrued from twitter. These information streams are filtered and then analyzed. The comments are acquired through opinion mining. Here the author analyzed the information for cellular telephones and accumulated the information's from twitter web page the usage of the crawler Twitter4J API. After that they stored the crawled records to a fashionable database and then preprocessed it through doing away with the prevent words. Finally, the facts from the database are labeled the usage of POS tagging method and the tag set is being created. By the usage of the unigram method the frequency of the words is calculated and the overall product rating was calculated.

In [9] focused sentiment evaluation which detects reviews or emotions at the social media textual content. Sentiment analysis can be accomplished in various domains together with social, clinical and commercial packages. This paper provided a survey about sentiment analysis addressing the distinctive standards on this place, problems and its answers and imparting a listing of open challenges on this vicinity. This survey supplied sentiment evaluation levels, strategies, enhancement techniques, applications, lexicons, tools and existing research gaps.

In [10] Projected on the basics of opinion mining. It is composed of different techniques which consist of Extraction, Clustering and Classification. The flipkart product critiques were extracted by using the use of product API. The writer fetched the emblem name, opinions, score and other related matters for product, the clustering is accomplished through the use of ROCK and using CART algorithm to classes evaluations as best and awful terms from comments and in the end, they come to recognize the product having more percent of nice reviews. They have been labeled as exceptional and negative words from opinions and from which the proportion of first rate and terrible terms turned into calculated. The end result provides the evaluation percent permits the client to finish based at the excessive nice evaluates percent of the product.

III. SENTIMENT ANALYSIS

Sentiment Analysis is contextual mining of text which identifies and extracts subjective records in deliver material, and supporting an agency to identify the social sentiment of their logo, products or services on the same time as monitoring on-line conversations. Sentiment analysis is also known as subjective analysis, it classifies the textual content in keeping with the polarity and orientation of the opinion expressed into positive, impartial and negative.

It defines sentiments and personal opinions from texts provided in social media network. It helps to identify how the individual users feel about the product of each and every particular brand. The system of sentiment analysis consists of Feature Selection, Sentiment Identification, Sentiment Classification and Sentiment Polarity.

IV. MULTILAYER PERCEPTRON (MLP) NEURAL NETWORK

Multilayer Perceptron (MLP) Neural Network is a category of feed forward artificial neural community. An MLP includes three layers of nodes: an enter layer, a hidden layer and an output layer. MLP utilizes a supervised learning method known as back propagation for education. A Multilayer Perceptron (MLP) is a deep, synthetic neural community. They input layer is to acquire the signal, an output layer that makes a decision or prediction about the input, and in among those two, an arbitrary number of hidden layers that are the real computational engine of the MLP. The algorithm is given in Table- I.

Table- I: Algorithm for Multilayer Perceptron (MLP) Neural Network

Multilayer Perceptron (MLP) Neural Network Algorithm
Step 1: Load the dataset taken from flipkart.com and then applying Multilayer Perceptron (MLP) Neural Network.
Step 2: The required records were picked from the dataset and then it is given to the input layer
Step 3: The data from the input layer is then transformed to the two hidden layer to tune the data.
Step 4: The tuned data from the hidden layer is given to the output layer.
Step 5: The performance of Multilayer Perceptron (MLP) is compared with Random Forest and Support Vector Machine (SVM) algorithm based on the Performance Metrics such as accuracy, pre cision, F-measure and recall.

A. Working principle of Multilayer Perceptron (MLP) Neural Network

1. The input information is given to the input layer. The extracted information set for products from flipkart has been imported into the input layer
2. Transformation of facts into the hidden layer 1
Then the data in the input layer is transformed to the hidden layer 1 in which an appropriate weight are assigned to the records and then the operation concerning the dataset had been executed.
3. The hidden layer 1 transforms the data to the hidden layer 2. The hidden layer 1 sends the data to the hidden layer 2. In which the records get finely tuned in the second layer.
4. The hidden layer 2 sends the finely tuned data to the testing and training process.
5. After completing testing and training process the data required by the user is sent to the output layer.
6. In final the user gains the required data from the output layer.

V. PROPOSED SYSTEM ARCHITECTURE

Review analysis enables the online consumers to benefit information about the product they need to shop for. Usually if the people want to shop for products from on line they may first take a look at critiques for the particular product that they want to shop. The consumer relies on the star provided for the product, depending on the stars they go for their decision.

This paper shows the assessment evaluation of flipkart merchandise to assist the online customers to make the quality choice. The accuracy stage of the critiques was additionally determined by the use of Multilayer Perceptron (MLP) Neural Network. The distinct work of the system is shown in Fig. 1. It consists of various method referred to as Data Collection, Data Preprocessing, Feature Selection, Detection Process and Sentiment Classification.

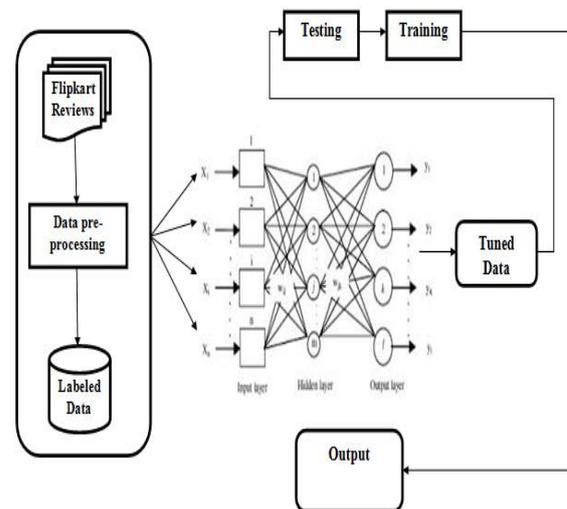


Fig .1. System Architecture

A. Data Collection

The proposed system collects the flipkart data set from the internet site referred to as kaggle.Com. It includes 20,000 records out of which 50% had been taken as training data and 50% were taken as testing data, i.e. 10000 records had been used for the process of training and 10000 dataset used for the process of testing and additionally the irrelevant dataset were eliminated from the samples as there's no utilization of those statistics.

B. Data Preprocessing

In data preprocessing the raw data is converted into the required layout. All the impurities had been left out from the dataset and then the unwanted attributes and products were additionally eliminated. The dataset consists of various attributes and products out which seven attributes and five products were selected randomly as samples.

C. Feature Selection

The feature selection is carried out after the completion of preprocessing of data. In feature selection process features of each and every product is analyzed, which helps the customer to compare one product with another product.

D. Detection Process

After the feature selection process, the next step to proceed is detection process, which detects the sentiment by using the following information.

True Positive: This defines the correct Real Positive Reviews in the testing data

False Positive: It defines the incorrect Fake Positive Reviews in the testing data.

True Negative: This defines the correct Real Negative Reviews in the testing data.

False Negative: It defines the incorrect Fake Negative Reviews in the testing data.

E. Sentiment Classification

The final process is called as sentiment classification. The sentiment of each and every buyer regarding the product is analyzed by using Multilayer Perceptron (MLP) Neural Network.

The main supply of facts used is the flipkart product opinions from the internet site known as kaggle.Com. Here first the raw information is cleaned and then saved as CSV. The CSV is loaded into SPYDER. The product critiques offered in the dataset is divided into 5 exceptional classes. Here the class 0, 1 denotes the poor critiques and class 2, 3 denotes the impartial reviews and class 4, 5 denotes the tremendous reviews. This class separation makes the system very simple and it end up very clean all through the type technique.

The flipkart dataset consists of reviews for numerous merchandises from which some of the products is randomly selected as pattern. For instance, Clothing, Toys and School Supplies, Tools and Hardware, Baby Care, Mobile and Accessories are selected. The dataset includes the subsequent attributes Product Name, Product ID, Retail Price, Discounted Price, Product Rating, Overall Rating, and Brand.

After acting all the above method, the class technique is made via the use of Multilayer Perceptron (MLP) Neural Network which detects the accuracy for the loaded facts as well the precision, recall and F-measure is also detected. Then the Receiver Characteristic Operator (ROC) curve and the confusion matrix are also evaluated. Finally, the Multilayer Perceptron (MLP) Neural Network is compared Random Forest and Support Vector Machine (SVM) to make sure Multilayer Perceptron (MLP) Neural Network gives the exceptional result.

VI. RESULT AND DISCUSSION

The flipkart product evaluations are analyzed with the use of the Multilayer Perceptron (MLP) Neural Network set of rules. The proposed work is carried out using SPYDER. In this module the following performance evaluations were delivered. It includes Real Positive Reviews, Real Negative Reviews, Fake Positive Reviews and Fake Negative Reviews.

Performance Metrics are

A. Accuracy

In general, overall performance of the classification algorithm is measured as accuracy. Accuracy is the ratio of wide variety of appropriately evaluated examples to the total range of predicted examples. The person can occur high accuracy after they label all of the examples as dominant magnificence.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Precision and Recall

Precision and Recall compares the overall performance of textual content mining. Precision assess correctness and recall evaluates the completeness. Precision is the proportion of wide variety of examples effectively labeled as positive to the wide variety of examples classified as positive labels. Recall is the proportion of range of examples effectively labeled as positive to the whole range of examples categorized as positive.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

C. F-measure

F-measure is the sum of precision and recall. F-measure acts as evaluation metric to calculate the views of sentiment classification.

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

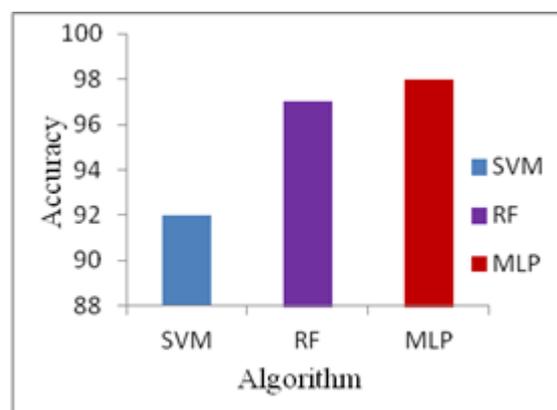


Fig. 2. Accuracy Comparison

In Fig. 2 the graphical representation for the accuracy comparison is determined between Multilayer Perceptron (MLP) Neural Network, Random Forest and Support Vector Machine (SVM).

Multilayer Perceptron (MLP) Neural Network gives the best accuracy while comparing with other classification algorithms with 98%, the Random Forest gives the accuracy of 97% and then the Support Vector Machine gives the accuracy of 92%.

In Multilayer Perceptron (MLP) Neural Network the implicit feature selection is carried out during the dataset evaluation. The Multilayer Perceptron (MLP) Neural Network insists unbalanced records into the system and it also controls over fitting without increasing the error charge. This leads Multilayer Perceptron (MLP) Neural Network to obtain the greater accuracy.

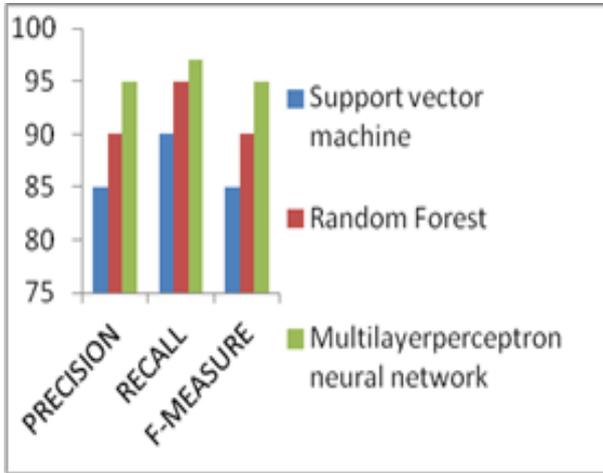


Fig. 3. Comparison for Precision, Recall and F-measure

Fig. 3 shows the comparison for precision, recall and F-measure between Multilayer Perceptron (MLP) Neural Network, Random Forest and Support Vector Machine (SVM).

The accuracy for Clothing, Toys and School Supplies, Tools and Hardware, Baby Care, Mobile and Accessories were determined using Multilayer Perceptron (MLP) Neural Network Random Forest and Support Vector Machine (SVM). The dataset contains many different products but here we have selected only 7 products randomly as samples. In which the Multilayer Perceptron (MLP) Neural Network shows the greater accuracy than the Random Forest and Support Vector Machine (SVM).

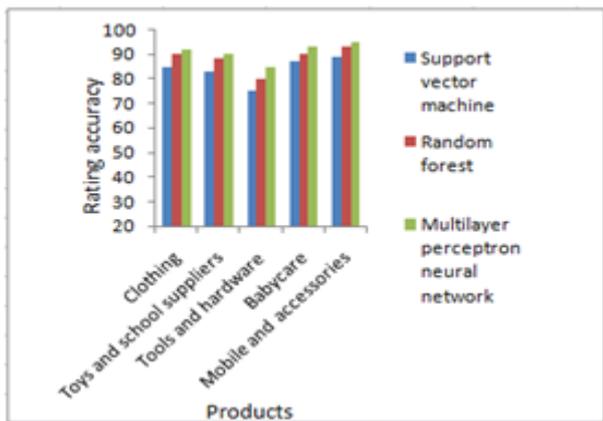


Fig. 4. Comparison of accuracy of product rating

The Fig. 4 shows the accuracy of product rating using Multilayer Perceptron (MLP) Neural Network.

D. Receiver Operating Characteristic (ROC) curve to evaluate classifier output

ROC curves are generally utilized in binary class to have a look at the output of a classifier. To increase ROC curve and ROC location to multi-class or multi-label

category, it is vital to binarize the output. ROC curve is drawn regular with label, ROC curve is drawn by way of using indicator matrix as a binary prediction also referred to as a micro-averaging. The possibilities are computed because the expected classifiers within the ensemble.

ROC curve represents true positive rate at the Y axis and false positive rate at the X axis. The top left corner of the plot is defined as the “best” point that represents fake positive rate as 0 and a true positive rate as 1. This module suggests the ROC reaction for flipkart product dataset. The ROC curve is created for class 0, class 1, class 2, class 3, class 4 and class 5. This class predicts the following comments, class 0 and class 1 is predicted as negative, class 2 and class 3 is predicted as neutral, class 4 and class 5 is predicted as positive.

The “steepness” of ROC curve increases the true positive rate while reducing the false positive rate. The Fig 4 indicates the ROC curve for multiclass for Multilayer Perceptron (MLP) Neural Network.

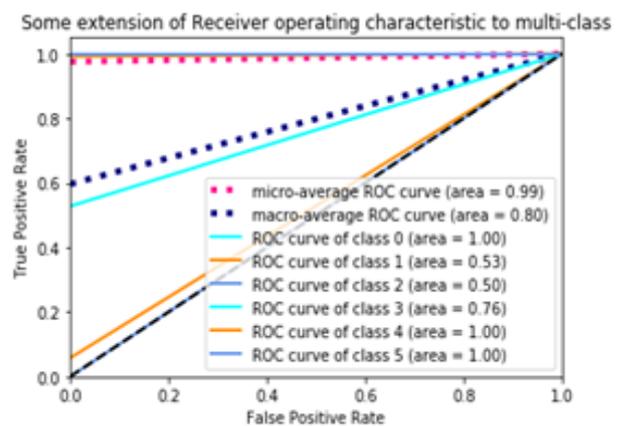


Fig. 4. ROC curve for Multiclass

The graphical representation of ROC curve in Fig 4 determines the mean area and variance by splitting the training set into different subsets.

E. Confusion Matrix

A confusion matrix consists of specified facts about the anticipated and actual values. Performance of the category values is evaluated by the usage of the data within the matrix. In confusion matrix the X axis includes expected labels and the Y axis includes real labels.

Here the matrix is done for each row and column by the usage of the expected value and real value. If the diagonal consists of no value then its far states that authentic label and the predicted label includes same value. If there's some fee inside the diagonal, it determines there are some differences among authentic labels and expected labels.

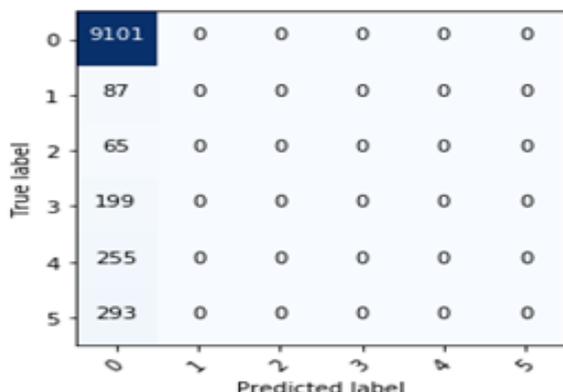


Fig. 5. Confusion matrix for Multilayer Perceptron (MLP) Neural Network

The Fig. 5 represents the confusion matrix for Multilayer Perceptron (MLP) Neural Network.

VII. CONCLUSION AND FUTUREWORK

Masses of consumer posts their comments (or) evaluations on social media, this enables the provider in addition to the consumer to the sentiment about the product. Using purchaser remarks the provider can improve their brand. It additionally helps the brand-new consumer to find more information about the product that they looking for. In this paper the category and analysis have been carried out for flipkart product critiques. This work classifies the star rating given by the customer for the product, right here the classification process is performed by Multilayer Perceptron (MLP) Neural Network. The accuracy comparison is made for the product among the Multilayer Perceptron (MLP) Neural Network, Random Forest and Support Vector Machine (SVM). In which the Multilayer Perceptron (MLP) Neural Network suggests the excellent accuracy and also the precision, F-measure and recall, Receiver Operating Characteristic (ROC) curve and confusion matrix is measured. Moreover, the proposed method is limited to the product domain which can be implemented in various domains such as educational Datamining, Recommendation system etc... and it also observed that a large dataset can be considered in future.

REFERENCES

- Ziyu Guan, Wei Zhao, Long Chen, Xiaofei He, Deng Cai, Member, Beidou Wang and Quan Wang, "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering, Vol. 30, No. 1, pg no 185-196, 2018.
- Mubashir Ali, Qurat Tul Ain, Amna Riaz, Amna Noureenz, Muhammad Kamranz, Babar Hayat and A. Rehman, "Sentiment Analysis Using Deep Learning Techniques: A Review", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, pg no 424-434, 2017.
- Parteek Kumar and Sujata Rani, "Deep Learning Based Sentiment Analysis Using Convolution Neural Network," Arabian Journal for Science and Engineering <https://doi.org/10.1007/s13369-018-3500-z>, pg no 1-10, 2018.
- Pranali Ramteke, Krutika Wase, Bhuvneshwar Kumar, Rushabh Bandewar, Nadim Badole, "Sentiment analysis of product review", International Journal of Innovations in Engineering and Science, Vol. 3, No.5, e-ISSN: 2456-3463, pg no 8-13, 2018.
- Binbin Xia and Guoyong Cai, "Convolutional Neural Networks for Multimedia Sentiment Analysis", Springer International Publishing Switzerland, NLPCC 2015, LNAI 9362, pg no 159-167, 2015.

- Nivya Johny, Minara P Anto, Mejo Antony, Vinay James, Muhsina K M and Aswathy Wilson, "Product Rating Using Sentiment Analysis", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 978-1-4673-9939-5/16/\$31.00, pg no 3458-3462, 2016.
- Ritu Rani and Akshi Kumar, "Sentiment Analysis Using Neural Network", 2nd International Conference on Next Generation Computing Technologies (NGCT-2016) Dehradun, India 14, 978-1-5090-3257-0/16/\$31.00, pg no 462-467, 2016.
- Takashi Ninomiya and Kazuhei Katoh, "Deep Learning for Large-Scale Sentiment Analysis Using Distributed Representations", SEMAPRO 2015: The Ninth International Conference on Advances in Semantic Processing, ISBN: 978-1-61208-420-6, pg no 92-96, 2015.
- Neamat El Tazi and Khaled Ahmed and Ahmad Hany Hossny, "Sentiment Analysis Over Social Networks: An Overview", IEEE International Conference on Systems, Man, and Cybernetics, 978-1-4799-8697-2/15 \$31.0, pg no 2174-2179, 2015.
- Rajasree R and Neethu M S, "Sentiment Analysis in Twitter using Machine Learning Techniques", 4th ICCCNT, pg no 4-6, 2013.
- Narsimha Reddy and U Ravi Babu, "Sentiment Analysis of Reviews for E-Shopping Websites", International Journal of Engineering and Computer Science, ISSN: 23197242, volume 6, Issue 1, 2017.
- P.Ajitha and R.Jenitha Sri, "Survey of Product Reviews Using Sentiment Analysis", Indian Journal of Science and Technology, vol 9(21), pg no 1-4, 2016.
- Aashutosh Bhatt, Ankit Patel, "Amazon Review Classification and Sentiment Analysis", International Journal of Computer Science and Information Technologies (IJCSIT), Volume 6(6), pg no 5107-5110, 2015.
- G.Angulakshmi and R.ManickaChezian, "An Analysis on Opinion Mining: Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering, Vol 3, Issue 7, 2014.
- R. Chandrasekaran and G. Vinodhini, "Sentiment Analysis and Opinion Mining: A Survey", International Journal, vol 2, pg no 1-6, 2012.

AUTHORS PROFILE



Dr.R.Murugeswari completed her Ph.D(CSE) from Kalasalingam University, Krishnankoil and M.E (CSE) from Arulmigu Kalasalingam College of Engineering, Krishnankoil under Anna University, Chennai. She is currently working as a Associate Professor in the Department of Computer Science and Engineering at Kalasalingam University, Krishnankoil. She has more than 18 years of Teaching Experience. Her research interests are wireless network, Evolutionary Algorithm, Big Data Analytics and Social Network Analysis. She has published more than 15 papers in various International journals and International Conferences.



T.Manjunathkumar currently pursuing Ph.D in the department of Computer Science and Engineering at Kalasalingam Academy of Research and Education, Krishnankoil. He Completed M.Tech Dergree in Kalaslingam university and Received B.Tech degree in Anna university.