International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 7 Issue 7 Ver V || July 2018 || PP 107-114

Sentiment Analysis Using Recommender Systems and On Micro-Blog Textual Data

Shivanand Sidramappa, Dr. Rajeev Yadav
Research schoolar department of computer science Opjs University, churu. (raj)
Associate professor department of computer science Opjs University, church. (raj)

Abstract: *Textual data, in contrast to digitised books, frequently does not exist in isolation but rather is linked to a significant variety of other information about the behaviours and preferences of users. This information may be found elsewhere in the system. As a direct result of this, there has been a growth in interest in the field of research pertaining to the analysis of social media and the various uses it has. A sentiment analyzer and recommender system are both examples of these applications. Other tasks, such as opinion retrieval, opinion summarization, subjectivity classification, sarcasm/irony detection, and many more, have a tight connection to the two that have been covered thus far.*
keywords: *Recommender Systems ,Micro-Blog ,Textual Data*

I. Introduction

Recommender systems, which are used to make suggestions by processing information from actively gathered diverse sorts of data, aid users in making judgments about their preferences for the product or item. Sentiment analysis is essential to the creation of recommender systems and must be used extensively. The introduction of sentiment analysis technologies on the cloud platform served as a decision support system for the cloud-based recommender system.

The idea behind this architecture is the gathering of data from the numerous nodes. The cloud platform and social network that are utilised to gather data from nodes could include Twitter, Facebook, and other social networks. The geographic position, comprising the latitude and longitude, is part of the data collected from the nodes. applying a method for sentiment analysis to the acquired data using a range of deep learning models, such as the multilayer perceptron model, convolutional neural network model, and recursive neural network model. The framework for the recommender system, which is based on deep learning models for sentiment analysis, is shown in Figure 1, and the necessary neural network models, which are considered to be deep learning models, are shown in Figure 2.

Neural networks are tiny computers made up of several interconnected processors. Neural network models can be used to address issues in natural language processing, such as word embedding learning and sentiment categorization. In order to categorise how they made the reader feel, both the text and the photos were analysed. The deep learning models for identifying someone's emotion were divided into two categories: feed-forward neural network models and recursive neural network models. The multilayer perceptron model and the convolutional neural network models were included in the second division of the feed-forward neural network models into two distinct groups. Each of the three models was applied to text in order to categorise sentiment, but only a convolutional neural network can be used to categorise images. Consequently, this kind of neural network might be used for both text and images.

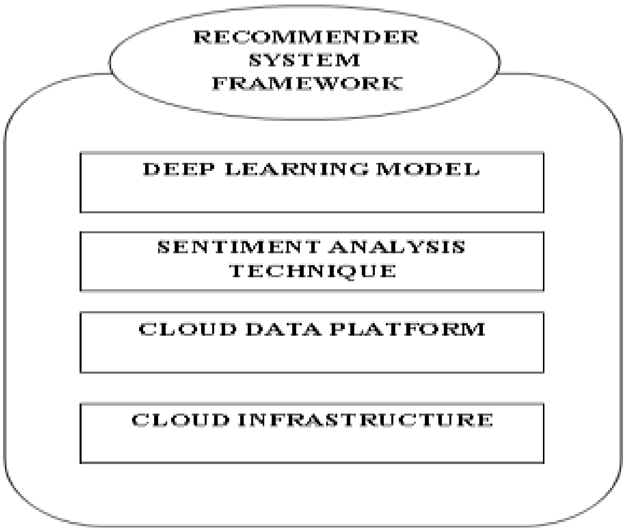

Figure 1 :Recommender System Framework.

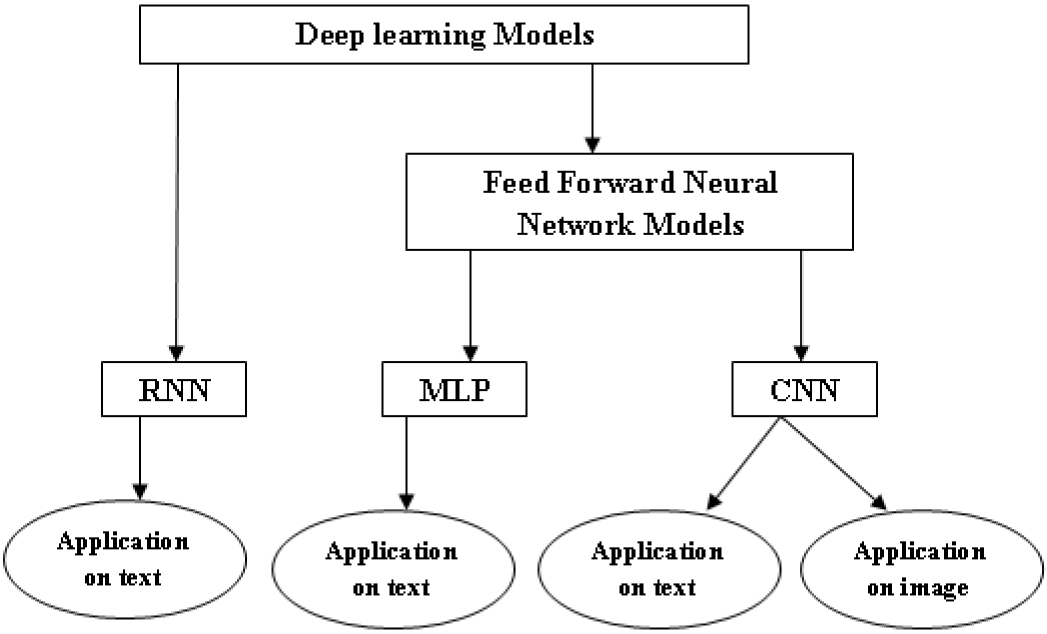

Figure 2: Deep learning Models.

RECURSIVE NEURAL NETWORK

Recursive neural networks can be viewed as an expansion of recurrent neural networks. A type of neural network is a recursive neural network. A particular kind of neural network that can learn tree architectures from input is a recursive neural network. Recursive neural networks have been successfully used to process data structures as neural net inputs in natural language processing. Recursive neural networks can construct parent representations by recursively creating child representations when given the structural representation of a sentence. Tokens are combined to create representations for phrases, and eventually the complete sentence, to achieve this.

Recursive neural networks work by retrieving information for a given text, representing those features in a high-dimensional vector space, and then feeding those features into the neural network. Recursive neural networks are able to understand the phrase's structure very effectively. The investigation of sentiment is thus effectively appropriate for this neural network model.

Think about a paragraph or document that says Di is "doing good job" as an example. It is a literary compilation that is being offered as the item. Words are dissected into their constituent pieces, which are then shown as vectors in a high-dimensional space. Then, as shown in figure 3 a decisive classification for the given input sequence may be produced using the representation of the supplied document at the sentence level.

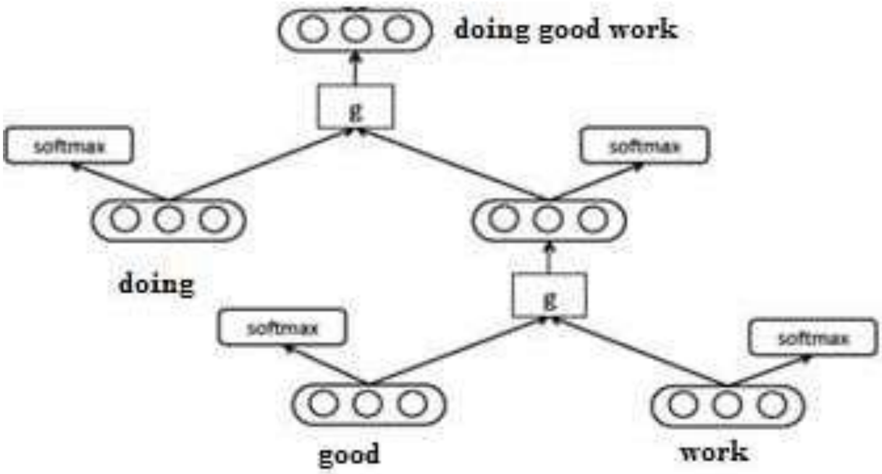

Figure 3: Sentiment Classification Using Recursive Neural Network

II. Method

Supervised Learning:

In this part, an overview of the supervised machine learning approach that was selected for the sentiment categorization utilising several deep learning models for recommender systems is provided.

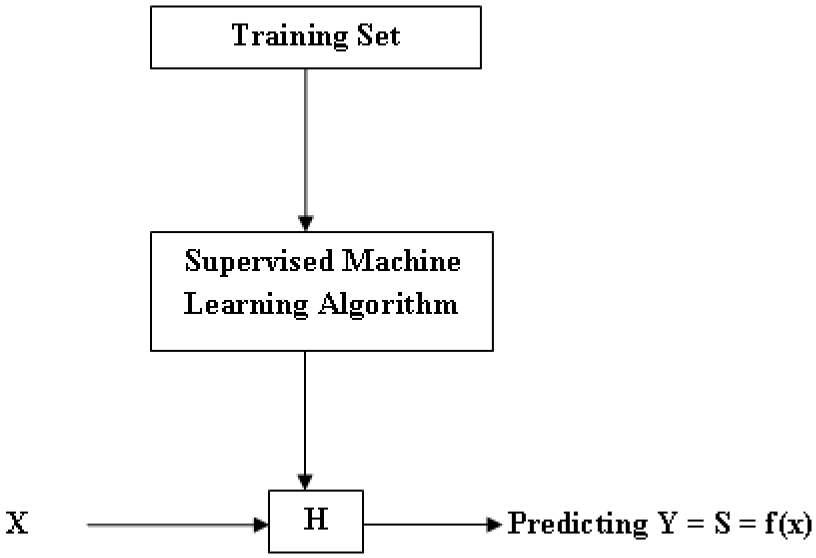

Figure 4: Supervised Learning.

Supervised learning is the process of predicting the values of the outputs based on the values of the inputs. This chapter used supervised machine learning to categorise people's attitudes toward a particular issue. The steps that were taken in order to do a supervised machine learning technique are shown in figure 4. Imagine a training data set as a collection of documents with the specified attitudes listed below for the purpose of providing examples:

Table 1: Training Data Set

Documents	Sentiments {0,1}
Document D1	0
Document D2	1
:	:
Document Dn	1

Table 2 Test Data Set

Documents DX	Sentiment SY-{0, 1}
Document D1	?
Document D2	?
:	:
Document Dn	?

Processing on the training set is carried out using the supervised machine learning algorithm. The document Di's numerous attributes or characteristics make up the training set. The goal of the supervised learning process is to develop the skill of learning a function denoted by H: XY using the available training data. In this example, X stands for an input document based on the query, Y for an effort to anticipate the sentiment of the input documents, and H for the hypothesis. The supervised learning approach is used to do classification and regression. The polarity Y of the target data or test data, also known as the sentiment variable S, is predicted by the supervised learning approach. The training dataset X, which comprises the hypothesis H of a small number of categories, such as positive or negative, is the foundation for this forecast. Depending on how the query of the input data X came out, the training data might be used to define the polarity of the given data as either being positive or negative. The terms that are used in the supervised learning approach are described as follows:

X – Input data based on the queryH– Hypothesis
Y=S=f(X)={0, 1}

This chapter provides an overview of the working of deep learning models that are built on feed-forward artificial neural networks in order to reach the hypothesis of the projected sentiment of the provided input data.

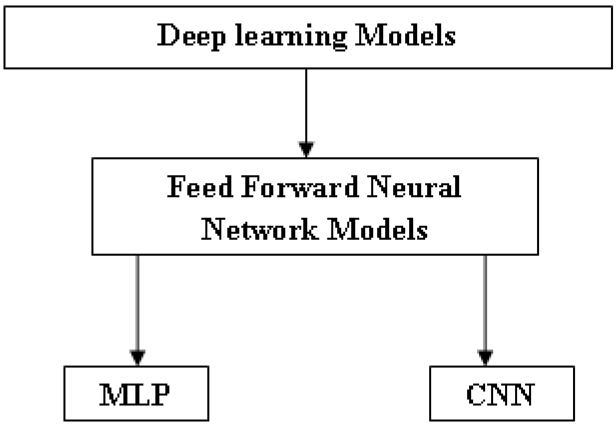

Figure 5: Deep learning Models with Feed Forward Neural Network.

Convolution Neural Network

The recommender system uses a feed-forward-like neural network design that additionally includes convolution layers and pooling operations. A region vector serves as the input to each computing unit in a convolution layer. Numerous computing units make up a convolution layer. The name of this region, which contains a vector, stems from the fact that it is made up of a small portion of a picture or piece of text that has had a nonlinear function applied to it. Convolutional neural networks were initially created with image processing in mind. Because of the convolution layers, these networks may utilise the two-dimensional structure of the image data. Written text can be categorised using convolutional neural networks, which can benefit from the one-dimensional nature of text data thanks to their convolution layers. The convolutional layer's processing units each respond to a distinct chunk of the document differently. Convolutional neural networks are used to analyse or handle data that has a known grid-like design.

Convolutional Neural Network for Image:

This method was first used in the area of computer vision, where it won praise for its success. A convolutional neural network is composed of several convolution layers, which serve as the visual environment's cells, in accordance with and The network's functionality is carried out by these tiers. An area vector serves as the input to each of the various computer units that make up a convolution layer. The vector representation of this region is made up of the pixels in that area. applying a nonlinear function on the subset of pixels that comprise this area. When convolution is applied to data, a picture that can be compared to an m-dimensional weight vector is created. The convolution layer, which makes the network's computations become easier, comes after the pooling layers. Pooling units, which are used in the execution of various merging algorithms like maximum pooling and average pooling, are made up of pooling layers. The pooling layers are the ones that come after the convolution layer and receive the output. By combining nearby pixels into a single, bigger pixel, pooling layers can reduce the number of individual pixels.

SENTIMENT ANALYSIS ON MICRO-BLOGTEXTUAL DATA

A wide range of industries that analyse data on linguistics and natural languages to supply the required text are increasingly using research that is based on sentiment analysis. Since the goal of sentiment analysis is to find polarity categorisation, the term "opinion mining" is frequently used interchangeably with "sentiment analysis". Extremity grouping is the process of giving a label to a record of evaluation to indicate whether it is possible, whether it is positive, or whether it is negative. At many levels, such as the term level, the expression level, and the sentence level or record level, recognition of the sequence of extreme evaluation should be possible. The word level of analysis is frequently covered by techniques that use n-gram classifiers or vocabularies. One of the jobs that is thought to be among the most crucial in the process of opinion investigation is identifying the semantic introduction of a phrase. The social media platforms of today have changed from being merely a place for information to include text expressing opinions and a forum for dialogue. This type of text has developed into the main source for analysing user behaviour and mining opinions. Product feedback, user engagement, and the generation of lead opportunities are all given insights. Sentiment analysis played a key role in the area of study that was being undertaken by numerous experts. Finding the result of an assumption test can be done in a variety of ways.

The machine learning strategy, the lexicon based approach, and the hybrid approach can be used to classify the methodologies for slant arrangement. supervised learning and unsupervised learning are the two main divisions of machine learning methods. Using supervised learning, which accomplishes this by comparing the test dataset to the trained dataset, the polarity of the emotion in the test dataset is predicted. Unsupervised and semi-supervised classification algorithms are suggested when it is not possible to provide a prior training dataset or labelled documents or views to classify the remaining objects. The two primary phases in performing sentiment analysis were creating a dictionary of keywords with positive and negative polarity and determining the client's perspective by splitting phrases in the content using the dictionary. Tokenizing the content is a stage in the typical calculating method that incorporates information extraction and classification utilising methods like Naive Bayes or Support Vector Machines.

Systems are available that can be used in the cloud to perform sentiment analysis, such as Learning Automata based Sentiment Analysis. One system that can perform sentiment analysis is this one. Sentiment analysis is performed on the data provided by the users by Learning Automata based Sentiment Analysis, which then provides a response that is either positive or negative depending on the results.

III. Experimental Results:

A customised and tested logistic regression model is used to tweets that have been downloaded from Twitter and utilised as the data source in order to perform sentiment analysis. In the tweets that were gathered, the hashtags "Agriculture INDIA" and "@AgriGoI" were mentioned. This is the official Twitter account for the Ministry of Agriculture and Farmers Welfare of India's Department of Agriculture, Cooperation, and Farmer Welfare. To find out more about our work, follow us! The Twitter API allows for the retrieval of all 1,000 tweets in response to a single request. Retweets are included in the extraction of textual data from microblogs, such as tweets, which affects accuracy and results in the development of more neutrally polarised tweets overall. This results in a collection of tweets that is pure because the retweets are not included in the list of testing data. Additionally, we are eliminating modified and older-style tweets with manual tweets. Plots that include retweets and those that do not are graphically depicted in Figures 6 and 5 respectively.

The dummy variable technique was used to analyse the data, which resulted in the elimination of retweets that did not contain altered or manually-written tweets. The answer that was produced after applying the raw data to the linear classifier was the subsequent response, which suggested that the likelihood of tweets having a positive sentiment rate was between 0 and 1. The outcome of the dummy variable technique will be either 0 or 1, depending on the likelihood that the sentiment rate of the tweets is positive.

When the dummy variable strategy was used, the accuracy of the model's performance on the testing data for the unigram model and the bigram model with normalisation of DTM using the TF-IDF transformation technique was 0.89 and 0.88, respectively. Figures 6 and 7, respectively, show this. Four-fold cross validation was used in the trials' execution. Bigram model outperformed the unigram model in terms of accuracy of model performance on the training dataset. On the other hand, utilising the dummy variable strategy, the unigram model performed better on the testing dataset than the bigram model. The experimental results of the training data and testing data.

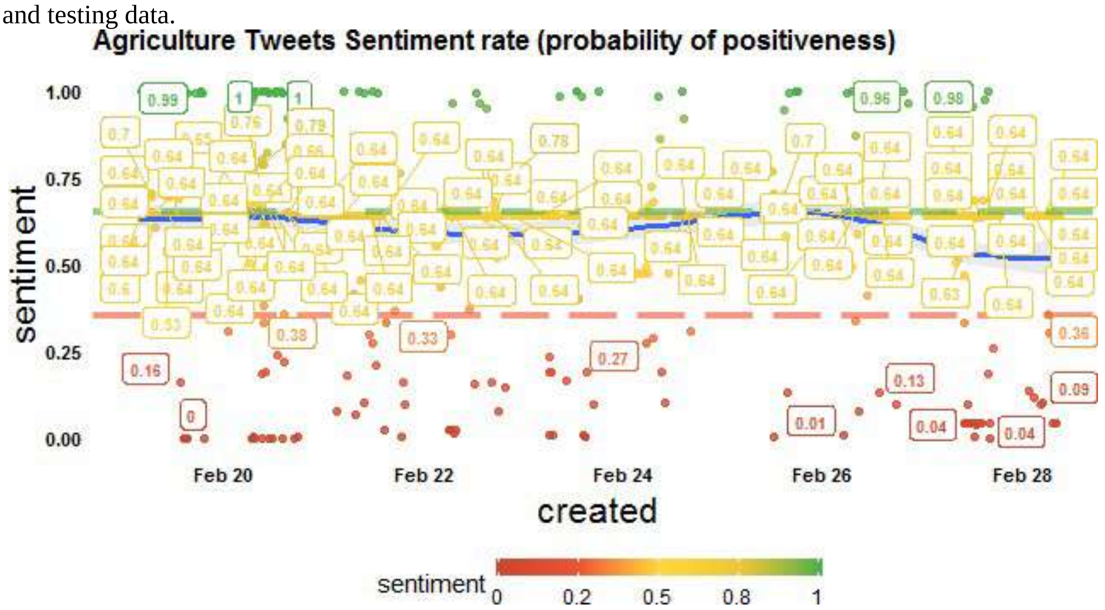

Figure 6: Probability of Positiveness of Tweets Sentiment Rate With Re tweets.

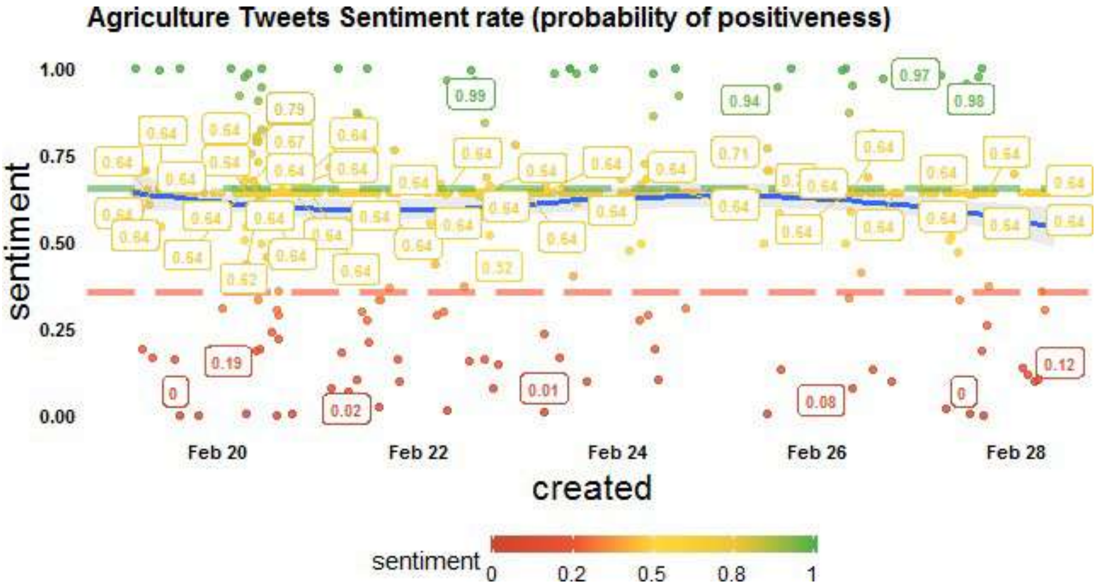

Figure 7: Probability of Positiveness of Tweets Sentiment Rate With Exclusion of Re tweets.

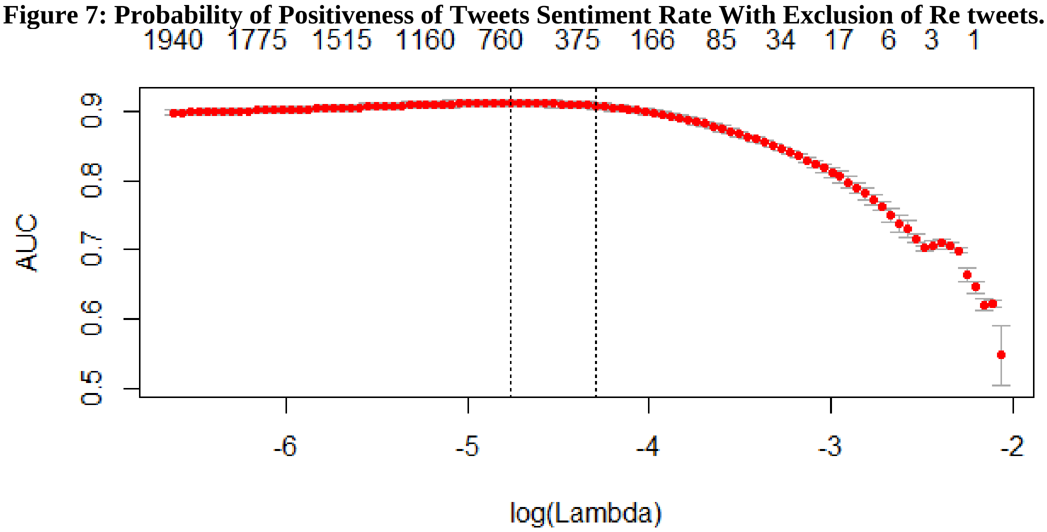

Figure 8: Accuracy of the Model Performance on the Trained Data of Unigram Approach.

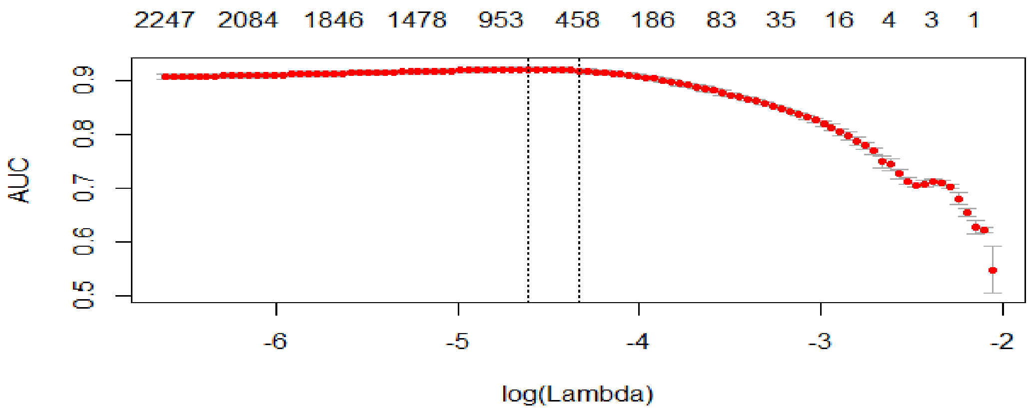

Figure 9: Accuracy of the Model Performance on the Trained Data of Bigram Approach.

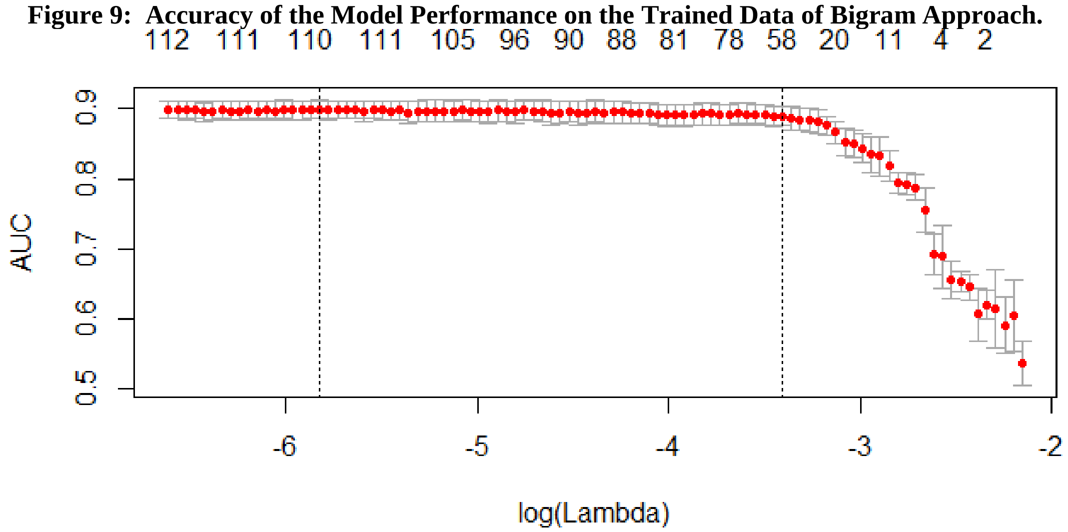

Figure 10: Accuracy of Model's Performance on the Tested Data of Unigram Model.

IV. Conclusion

Build a framework for a recommender system that analyses sentiment using deep learning models. All of the objectives specified in the different chapters of the literature review must be accomplished by the framework. Propose a way for doing sentiment analysis on the textual content of micro blogs using supervised machine learning methods and a deep learning-based methodology. Also, present a technique for using a dummy variable strategy to enhance and confirm the precision and model performance of real-time micro blog textual data. Evaluation of the built DLSARS framework for a recommender system in the cloud using supervised learning technique with unigram and bigram approach and a linear classifier that conducts sentiment classification using a dummy variable approach based on a multilayer perceptron model.

References

[1]. Y. Zhang, and B. Wallace, "A Sensitivity Analysis of (and Practitioners' Guide to)ConvolutionalNeuralNetworksforSentenceClassification",IJCNLP,2017.
[2]. R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolution neural networks", Proceedings of the 2015 Conference of the NorthAmericanChapteroftheAssociationforComputationalLinguistics:HumanLanguageTechnologies, 2015, pp. 103-112.
[3]. P. Georgios and T. Mike, "A Study of Information Retrieval Weighting Schemes for Sentiment Analysis", Proceedings of the 48th Annual meeting of the Association for Computational Linguistics, 2010, pp. 1386-1395.
[4]. Zhang, X.; Zhao, J.; LeCun, Y. Character-level convolutional networks for text classification. In Advances in Neural Information Processing Systems 28; Neural Information Processing Systems Foundation, Inc., Montreal,Quebec, Canada, 2015; pp. 649–657.
[5]. Shen, Y.; Xiaodong, H.; Gao, J.; Deng, L.; Mensnil, G. Learning Semantic Representations Using Convolutional Neural Networks for Web Search. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 373–374
[6]. Alm, C.O.; Roth, D.; Sproat, R. Emotions from text: Machine learning for text-based emotion prediction. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; pp. 579–586
[7]. Heraz, A.; Razaki, R.; Frasson, C. Using machine learning to predict learner emotional state from brainwaves.In Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), Niigata, Japan, 18–20 July 2007; pp. 853–857.
[8]. Kim, Y. Convolutional neural networks for sentence classification. arXiv 2014, arXiv:1408.5882
[9]. Chen-Yu Lee, Saining Xie, Patrick W Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeplysupervised nets.In AISTATS, volume

2, pp. 5, 2015.
[10]. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In ´ ECCV, pp. 740–755. Springer, 2014.
[11]. Tolga Bolukbasi, Joseph Wang, Ofer Dekel, and Venkatesh Saligrama. Adaptive neural networks for fast testtime prediction. arXiv preprint arXiv:1702.07811, 2017.
[12]. A. R. Zamir, T.-L. Wu, L. Sun, W. Shen, B. E. Shi, J. Malik, and S. Savarese. Feedback Networks. ArXiv e-prints, December 2016.
[13]. A Recurrent Neural Network Based Recommendation System by David Zhan Liu Gurbir Singh Stanford University.
[14]. A Traffic Flow Prediction Method Based on Road Crossing Vector Coding and a Bidirectional Recursive Neural Network by Shuanfeng Zhao Qingqing Zhao YunruiBai and Shijun Li.
[15]. Deep Learning Sentiment Analysis For Recommendations In Social Applications