www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 02, Issue 06, June 2022, pp : 11-14

e-ISSN : 2583-1062

Impact Factor : 2.265

AMAZON PRODUCT REVIEWS USING WORD2VEC (A Sentiment Analysis)

Shivani Joshi[1], Jiten Prasad[2], Divakar Sinde[3]

[1,2,3]Student, Department of Computer Science & Engineering

Anil Neerukonda Institute of Technology and Sciences

Sangivalasa, Visakhapatnam, Andhra Pradesh

ABSTRACT

Sentiment analysis on amazon product reviews, here we want to analyze between Amazon product reviews and rating of the products given by the customers. This gives the result of the amazon reviews dataset and studies sentiment classification with different machine learning approaches. We use word2vec model in this paper. Our experimentation contains two parts: First, the reviews are transformed in to vector representation using Continuous Bag of Words and Skip gram. And then, we trained the machine learning algorithms such as Random Forest and Long Short-term Memory (LSTM). The results shows that LSTM with Word2Vec gives best accuracy.

Keywords: Word2Vec, Natural Language Processing, Random Forest.

1. INTRODUCTION

E-commerce websites consist of consumer reviews of items with information removed. People are depending on customer reviews before they buy some products online. These reviews are also helpful in defining the product. Customers compare different items whether they have a positive or negative review. Sentiment classification is a method which separates positive reviews from negative reviews. Sentiment analysis is a form which finds positive and negative in a text or a sentence. Sentiment analysis helps the customers by knowing the reviews of others and also helpful for the companies to develop their products. Our dataset contains reviews & ratings of customers and we remove the quality of our dataset and build a supervised learning model. The Machine Learning algorithms which we will use are Random Forest and Long Short-term Memory. These techniques will help us to find out whether the sentence or a text is a positive or negative review. By using these techniques, we will predict which algorithm gives the best accuracy. We will perform both the algorithms and determine which is the best algorithm.

2. METHODOLOGY

MACHINE LEARNING ALGORITHMS

RANDOM FOREST-

In random forest, the decision tree contains Less Similarities & High Differences. Less similarities ends outcome with higher accuracy and have very less training error. High differences have higher priority to give more training error. If the decision tree has a high difference, then the random forest works well. High difference can be converted to low difference when the new test data of each of the trees will be joined together. Suppose if we have 780 unique products in a dataset. And if we change minimum 250 products the output is Low difference with the help of combination of two or more decision trees. Coming to training, the random sample data are trained and the decision trees are constructed for samples and the prediction result is found for decision trees. Coming to testing, the priority checked for each prediction result. The prediction result that is more prioritized or more voted is final prediction result.

LONG SHORT-TERM MEMORY-

The LSTM Layer is described as a number of hidden state layers and dimensions. Fully Connected Layer shows the output of LSTM layer to an expected output. Sigmoid Activation function converts the output values into 0 and 1. The result of this network is observed from the final process of the Sigmoid activation function. With the help of NLTK (Natural Language Tool Kit) the priority of words like "wonderful" is checked and the overall sentiment of the sentence is found and declares whether the sentiment is positive, negative, or neutral. Lstm is a solution to the short term for a long sequence with the help of its states (input state, output state, cell state and forget state).

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 02, Issue 06, June 2022, pp : 11-14

e-ISSN : 2583-1062

Impact Factor : 2.265

3. LITERATURE REVIEW

Barkha el al., [2] used word2vec model converting reviews into vectors for classification. They used consumer reviews in the cellphone from the e-commerce website named "Amazon". Including word2vec they have also CBOW (continuous bag of words) and skip-gram models and for further computation machine learning algorithms like Logistic Regression, Random Forest, SVM and Naïve Bayes were implemented. By combining the classification algorithms with these models i.e., word2vec & CBOW the results performed very well on balanced and unbalanced datasets. And on comparing each algorithms performance Random Forest with CBOW achieved 90.6622% accuracy and Naïve Bayes with skip gram gained less accuracy i.e., with a percentage of 51.68%.

Dipti el al., [4] aimed to construct a model by focusing on NLP (Natural language processing) in accordance with Stanford Library such that the capability of the machine can be improved. Neural network concept of RNN (Recurrent Neural Network) rather than using conventional machine learning algorithm like Naïve Bayes. During the pre-processing stage Google translator app was used for translating the overseas sentences into English. The model achieved an accuracy of 90% using single algorithm i.e., RNN.

Wanliang el al., [1] proposed a model using traditional algorithms i.e., SVM (Support Vector Machine), KNN (K-nearest neighbor) and deep networks such as RNN (Recurrent Neural Network). And they have observed that LSTM outperformed other algorithms in terms of accuracy. They also used the feedback from the people in real-time and tried to include resampling and different weighting techniques to increase the efficiency of the model.

4. MODELLING AND ANALYSIS

SYSTEM ARCHITECTURE

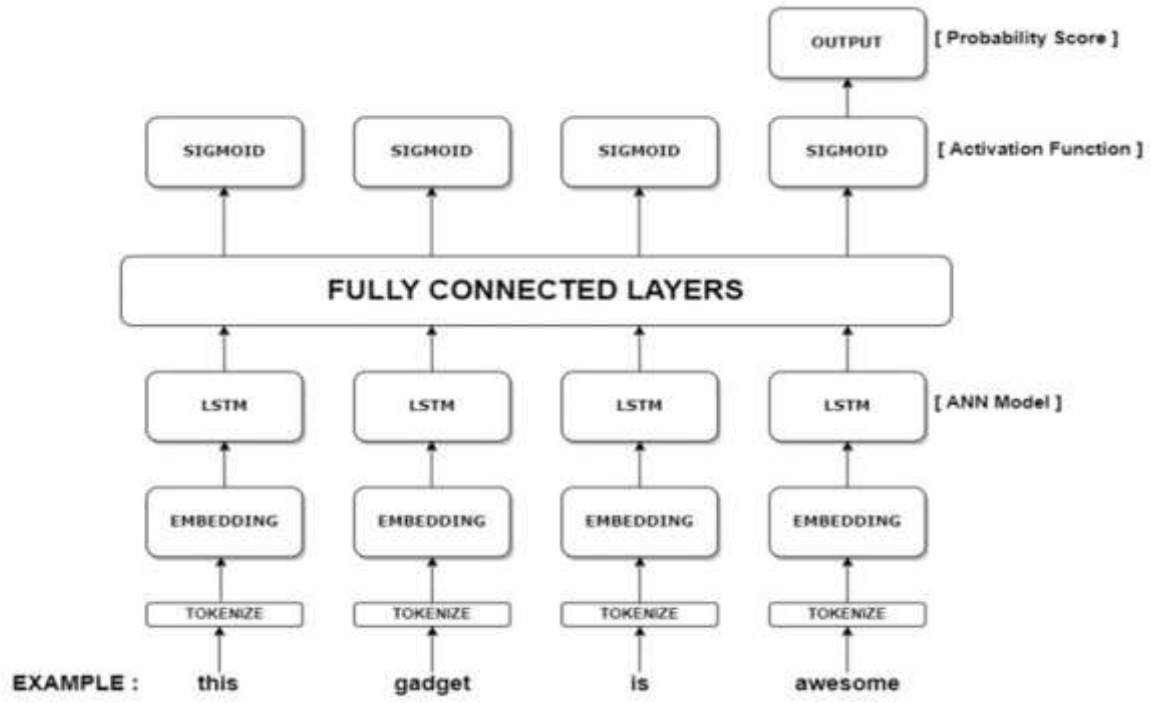

Figure-1: Block Diagram Of The Model

The above fig. represents the architecture of the entire system. It shows the implementation from beginning to end i.e., Word embedding, Tokenizer, calculating the likelihood, Long-short term memory, & Activation functions.

DATASET

The datasets consist of 4410 unique product data and their corresponding brand names, it also includes 162492 reviews of the individual users. The dataset also incorporates the price of the product, users review in the form of text, their ratings, and the number of people found these reviews as helpful. These attributes will help and increase the efficiency by training and testing the model. Using this vast amount of data our model is prepared for future proof values.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 02, Issue 06, June 2022, pp : 11-14

e-ISSN : 2583-1062

Impact Factor : 2.265

4.4 EXPERIMENTAL WORK

4.4.2 PRE-PROCESSING

It is the first step while creating a machine learning model. It is a process of making a raw review into cleaned review and checks whether it is suitable for the machine learning model. By doing this, it also increases the accuracy and efficiency of a mchine learning model. There are few steps which are used in our project are:

Step-1: Conversion of raw review into cleaned review by removing html and non-characters. Conversion of upper case to lower case will be done in this step.

Step-2: Stop words like “the”, “in”, “a”, etc., will be removed to get a dictionary with limited dimensions.

Step-3: Stemming converts the words in to their actual root forms by using Porter Stemmer & Snowball Stemmer. But Snowball Stemmer is slightly rapid and intelligent when compared to Porter Stemmer.

Step-4: Split text is used to split the review text into parsed sentences using Natural Language Tool Kit (NLTK) and later the parsed sentences will be fitted in to Word2vec model.

5. RESULT & DISCUSSION

RANDOM FOREST-

```
Accuracy on validation set: 0.9262

Classification report :
              precision    recall  f1-score   support

           0       0.87      0.83      0.85       778
           1       0.94      0.96      0.95      2311

    accuracy                           0.93      3089
   macro avg       0.91      0.89      0.90      3089
weighted avg       0.93      0.93      0.93      3089


Confusion Matrix :
 [[ 642  136]
 [  92 2219]]
```

Figure-2: Classification And Accuracy Report Of Random Forest

LONG SHORT-TERM MEMORY-

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, None, 128)         2560000

lstm (LSTM)                  (None, 128)               131584

dense (Dense)                (None, 2)                 258

activation (Activation)      (None, 2)                 0

=================================================================
Total params: 2,691,842
Trainable params: 2,691,842
Non-trainable params: 0
_________________________________________________________________
Epoch 1/3
869/869 [==============================] - 311s 353ms/step - loss: 0.2651 - accuracy: 0.8923
Epoch 2/3
869/869 [==============================] - 307s 353ms/step - loss: 0.1457 - accuracy: 0.9477
Epoch 3/3
869/869 [==============================] - 305s 351ms/step - loss: 0.1012 - accuracy: 0.9659
97/97 [==============================] - 4s 37ms/step - loss: 0.1595 - accuracy: 0.9463
Test loss : 0.1595
Test accuracy : 0.9463
```

Figure-3: Report Of Lstm

6. CONCLUSION

After the complete analysis we have found that the parsed sentences that are in Word2vec model uses both Random Forest as well as Long Short-term Memory algorithm. Among two, Random Forest with Word2vec gives an accuracy of 92.62%. Whereas LSTM with Word2vec gives an accuracy of 94.63%. LSTM shows important results with best accuracies by using three iterations or epochs.

7. REFERENCES

[1] Al Qahtani, Arwa S. M., "Product Sentiment Analysis for Amazon Reviews (2021)". International Journal of Computer Science & Information Technology (IJCSIT) Vol 13, No 3, June 2021.

[2] D. Mahajan and D. Kumar Chaudhary, "Sentiment Analysis Using Rnn and Google Translator," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2018, pp. 798-802, doi: 10.1109/CONFLUENCE.2018.8442924.

[3] Najla M. Alharbi, Norah Saleh Alghamdi, Eman H. Alkhammash, Jehad F. Al Amri, "Evolution of Sentiment Analysis via Word Embedding and RNN variants for Amazon online Reviews.

[4] Yash Inaniya, "Amazon product review Sentiment analysis using BERT".

[5] Aman Kharwal, "Amazon Product Reviews Sentiment analysis with python".