International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 3177–3186

Sentimental Analysis on Amazon's Alexa Reviews using ML and DL approach: A Comparative study

Monika Kanojiya
Assistant Professor
monika.kanojiya@sakec.ac.in

Deepti Deepak Nikumbh
Assistant Professor
deepti.nikumbh@sakec.ac.in

Abstract

Sentimental analysis is a processes of mining textual data like reviews, statements and make prediction about emotions behind it. Businesses today use sentimental analysis to monitor their product popularity, customer's perception about product through feedback. Amazon's alexa is one of the virtual assistants increasingly adopted by people. Despite of its popularity, there are mixed reactions among people regarding alexa and its variations. In order to get good insight of alexa and its variants, it is important to go through the written reviews and proper analysis of these written review should be done, but doing it manually will be a time consuming and cumbersome task. Thus, an automatic sentimental analyzer is required which will benefit the buyer as well as a manufacturer. This work attempts to develop a sentimental analyzer which analyzes 5000 Alexa reviews using naïve bayes algorithm, Random forest, Long Short-Term Memory Networks (LSTM) and multilayer perceptron (MLP).It further evaluate performance of each algorithm with respect to various parameters like precision, recall, F1 score and ROC and a brief comparative study is presented.

Keywords- *Sentiment analysis, Natural Language Processing, Machine Learning, Amazon Echo, Naïve Bayes algorithm, Random Forest, Multilayer perceptron.*

1. Introduction

Today huge amount of data in the form of reviews are generated by e-commerce websites like amazon, flipkart etc. Machine learning algorithms helps to extract useful information from such huge volume of data and help in decision making [10]. Sentiment analysis is one such process which uses machine learning algorithms to automatically analyze huge volume of reviews and classify them as positive, negative or neutral. Written reviews are important as it refers to writer's or reviewer's emotion and attitude towards an entity or product. The opinions vary from person to person toward an entity so for doing sentiment analysis it is important to collect all kinds of reviews so that the analysis can be justified properly [1].

Major types of sentimental analysis include sentiment classification, sentiment lexicon, sentiment summarization, and quality of reviews. The various approaches for sentimental analysis is shown in figure 1.There are multiple fields which incorporate with sentiment analysis like information retrieval, machine learning natural language processing, computational linguistics.

The structure of the paper is as follows: First section is about introduction to sentiment analysis; second section consist of related works. Third section is dedicated to methodology which consist of sub-sections: proposed system architecture, data characteristics, Feature extraction and machine learning models. Fourth section shows performance evaluation metrices with training and validation accuracy plots. Fifth section gives the conclusion and future works followed by the references.

ISSN: 2233-7857 IJFGCN

This work consists a comparative study of sentiment classification by performing experiment on Alexa reviews using four machine learning techniques Multilayer perceptron, Naïve bayes and Random forest ensemble and LSTM. For the experiment following tasks are performed-

1. A Dataset of Alexa review is downloaded and prepared.
2. Data set is preprocessed using Natural language techniques.
3. On the preprocessed dataset, machine learning techniques namely Multilayer perceptron, Naïve bayes, LSTM and Random forest ensemble are applied.
4. After step 3 a computing model of each above-mentioned machine learning techniques are built and comparative study regarding performance of each model was done.

2. Related works:

There are related fields to sentiment analysis like exchange learning, emotion detection.[3] discusses about various sentiment analysis techniques and its related fields. They showed different methods to do sentiment analysis. According to medhat et. al article classification of sentiment analysis techniques are as per figure 1.

Sentiment analysis is considered as a classification problem thus machine learning techniques can be used.[2,5]

Y. Gao, Z. Pan, H. Wang and G. Chen[4] have applied feature mining and sentiment mining on Amazon Echo Reviews and carried out the emotion analysis. They defined the how Alexa plays an important role in people's life compared to other electronic devices by Emotion Analysis.

There is not much research done for Emotion Detection. Many major types of approaches for emotion detection on corpus are explained and brain mapping based on polarity of valence and arousal model taken into consideration to classify sentiment in corpus. A lot of survey has been done on emotion recognition and EEG tool for sentiment analysis on text reviews.[6]

An ensemble classifier is aiming for improvement in sentiment analysis technique. The classifier performs well as compared to traditional standalone classifiers. Also data preprocessing and feature representation plays a very important role in sentiment classification.[8]

D. Elangovan and V. Subedha[11] proposed an efficient technique od sentiment analysis for online reviews using the combination of feature extraction and classification. They used Firefly(FF) and Levy flights(FFL model for extracting the features from a review and Multilayer Perceptron (MLP) is used for classification.

Saleh et al. accomplished experiments on sentiment classification and polarity determination by using SVM with many different feature selection methods. They performed experiments on three benchmark datasets. [12]

A lot of investigation has done to identify the specificity of the statements given in reviews for automatically predicting the helpfulness of product reviews. Comparison done using most relevant features on datasets and results are visualized for ranking information.[16]

Using Machine learning classification reviews in Arabic classified using Ridge Classifier which have been proven best performer classifier in term of accuracy,recall,F1 score and precision.They have also defined that preprocessing improves the performance of ML classifier.[17]

Using statistical analysis feedbacks of three of top ranked brands Samsung, BLU and Apple were analyzed. This work gives various outcome for example higher cost products have better quality and

ISSN: 2233-7857 IJFGCN

customer satisfaction. For conduction of sentiment analysis on the three brands a built-in package 'Syuzhet' is used.[18]

Naïve bayes is one of the most classical probabilistic model which is used for sentimental analysis. It is highly scalable and requires lesser parameters[20]. Surya Prabha PM et. al used naïve bayes

classifier for sentimental analysis of amazon products .They work on small dataset of 600 records and achieved an accuracy of 89%.

Current state of art technologies like LSTM's are popularly used for analysing textual data.[19] demonstrates use of LSTM for sentimental analysis using keras library in python.

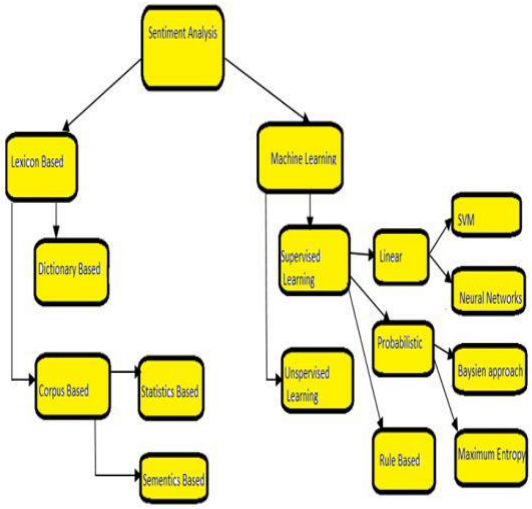
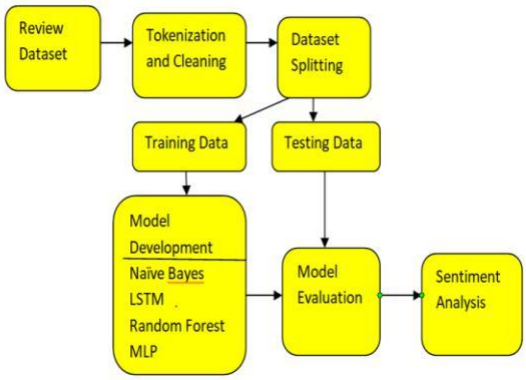


Figure 1 Sentiment Classification Techniques [3]

3. Methodology

3.1 Proposed System Architecture



ISSN: 2233-7857 IJFGCN

Figure 2 Proposed System architecture

3.2. Data Characteristics

The Dataset of Alexa reviews is extracted from kaggle.It contained 3150 reviews given by Amazon Alexa users. It includes reviews of various Amazon Alexa products like Alexa Fire sticks, Echo dots, Alexa Echo etc., Since the dataset was very small and unbalanced, having more number of positive

reviews, to increase the negative reviews web scrapping was done using Amazon review scrapper. The reviews are in text format, positive reviews are indicated by 1 and negative reviews by 0. Since the reviews are in text format they are further subjected to preprocessing. Following are the Data Preprocessing steps are performed-

1. Converting all alphabets into lowercase .

2. Split each review into individual word.

3. Replaced all non alphabets into ‘ ‘ (white spaces).

4. Stopwords like is,am,are etc. are removed by using nltk library

5. Stemming is performed using PorterStemmer.In this process if the word is not included as a stopword it will be converted into its original form. for example “learning” will be converted to “learn”.

3.3 Feature Extraction

After Preprocessing, feature extraction is performed using CountVectorizer class provided by scikit learn library in python. For input reviews ,in the form of text are considered and its corresponding label is stored as an integer where 0 represents negative review and 1 represents positive review. The output features denotes whether the customer liked or disliked the ALEXA product.

After Feature extraction the data is divided into training and testing set. -The ratio we took is 75:25 respectively.

3.4 Machine Learning Models

A machine learning model can be a mathematical model which represents a real-world process. The machine learning algorithm searches the patterns in training data set .The outcome of this process is machine learning model. For this work Learning Models are prepared by applying Gaussian Naive bayes, Long Short-Term Memory Networks, Random Forest algorithm on data set and Performance parameters are evaluated and compared.

3.4.1 Gaussian Naïve bayes

Gaussian naive bayes classifier,mainly works on normal distribution. It is based on bayes theorem that used to calculate conditional probability with no covariance between dimensions. It deals with input reviews and perform class univariate distribution on reviews.It predicts the output on test cases and calculate accuracy score. After data Preprocessing following Working steps are performed-

1. Importing the naïve bayes classifier.

2. Creation of Gaussian Classifier.

3. Training the model using training set.

4. Prediction using test set.

ISSN: 2233-7857 IJFGCN

5. Preparation Classification report and evaluation of Performance parameters.

3.4.2 Random Forest

Random forest algorithm can be used for classification and regression both but mostly it is used for classification. It is based upon the voting given by the decision trees made from the data samples. For this work the following steps are performed after data preprocessing.

1. Used binning for shaping of the data samples.

2. Trained the model using RandomForestClassifier.

3. Prediction performed using test data samples.

4.Performance parameters are evaluated using classification report.

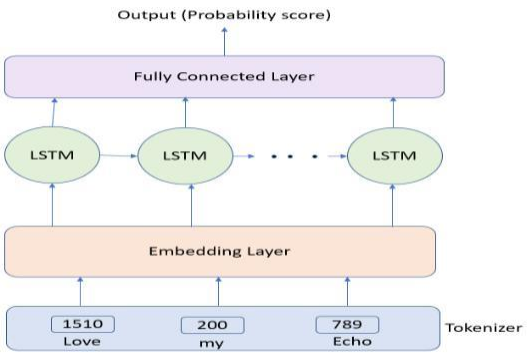
3.4.3 Multilayer Perceptron

MLP is a computing model which is used in the field of deep learning .It has a layered architecture which keeps input layer, hidden layers and output layer. In this work we have created a three layer MLP classifier with 10 neurons in input layer and 100 neurons in hidden layers using RELU activation. The output layer contains single neuron using sigmoidal activation.

3.4.4 Long and short term Memory(LSTM)

Recurrent Neural Networks faces the problem of vanishing gradient so at time of processing the long paragraph it might leave some important text or information. To remove this problem LSTM's were introduced. LSTM's are good at grasping long term dependencies in text. It has gates an internal structure which decides which information is important and should be carry forwarded and which information is less important and can be forgotten. LSTM's are good for sequential and time series data. Reviews are textual data which is nothing but sequence of words. LSTM can be used for such textual data. It not only considers the individual words but also keeps the track of the order in which they appear. Thus LSTM's are popularly used for sentimental analysis.

Figure 3 demonstrates the LSTM model used in the proposed work. After preprocessing tokenizer converts each word to integers. It maintains a dictionary which maps vocabulary to integers. The mapping is done by using an approach i.e frequently used words will have lower indexes. The output of tokenizer is given to embedding layer which converts tokens into embeddings of specified size[19]. In the work 10 LSTM units using RELU activation is used. Batch normalization and Dropout are used to reduce overfitting. Finally, a dense layer of single neuron with sigmoidal activation function is used which gives the final prediction probability of a review.



ISSN: 2233-7857 IJFGCN

Figure 3 LSTM architecture for proposed system.

4. Results Analysis

4.1 Performance Evaluation Metrics

Dataset used in the problem is slightly imbalanced, thus accuracy metric will not always give best model. To evaluate the effectiveness of a model other performance parameters should also be studied. The precision, recall and F1 score of all the models is provided in table 1. Precision gives the correctness of the classifier. Higher the precision means less false positives(reviews which are negative but still classified as positive).Recall gives the sensitivity of the classifier. Higher the recall means less false negative(reviews which are positive but still classified as negative). It is desired to have high values for both recall and precision[13].Precision is more focused on positive class than negative class. Since the dataset used in the problem is imbalanced wherein positive samples are more compared to negative samples. It can be observed from the table that precision values are higher than recall values . From the table it is clear that LSTM's give higher values for both precision and recall followed by MLP.

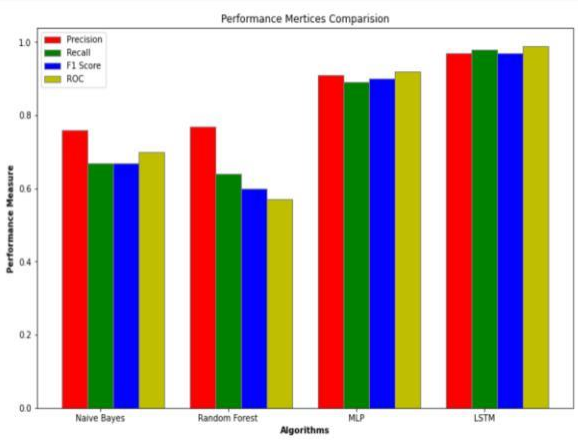
A classifier always tries to find a trade-off between precision and recall. Some models aim to get higher precision whereas some try to get higher recall. In our model both precision and recall are equally important, but precision and recall have conflicting goals. Increase in recall can come at expense of decrease in precision. In such scenario F1 score is very useful. F1 score is harmonic mean of precision and recall[13].Table below gives the F1 score of all the algorithms. It is clear from the table that LSTM's give highest F1 score.

Another interesting metrics which is used in case of imbalanced and binary classifier is ROC-AUC metrics. ROC is useful when predicting both the classes are equally important and imbalanced dataset contains majority of positive samples like in our case. Since positive samples are more the precision and recall will give prediction of positive class and not negative class as negative class samples are less[14].From the table it is clear that LSTM's give best ROC values i.e it is able to predict both positive and negatives reviews accurately.

Table 1 Performance evaluation metrics

Measure	Naïve Bayes	Random Forest	MLP	LSTM
Precision	0.76	0.77	0.91	0.97
Recall	0.67	0.64	0.89	0.98
F1 Score	0.67	0.60	0.90	0.97
ROC	0.70	0.57	0.92	0.99

The bar chart is also presented in figure 4 that gives a clear comparison of various performance metrices against algorithms under study.



ISSN: 2233-7857 IJFGCN

Figure 4 Bar chart of performance metrices with respect to algorithms

4.2 Training and validation accuracy plots :

Accuracy plots are very useful in understanding the progress of the Machine learning and deep learning algorithms. Following diagram shows the accuracy plots for MLP and LSTM's. The gap between the training and validation accuracy gives information about overfitting. Larger the gap, higher is the overfitting scenario[15].From the figure 3 and figure 4 it is clear that MLP and LSTM's are suffering from slight overfitting .

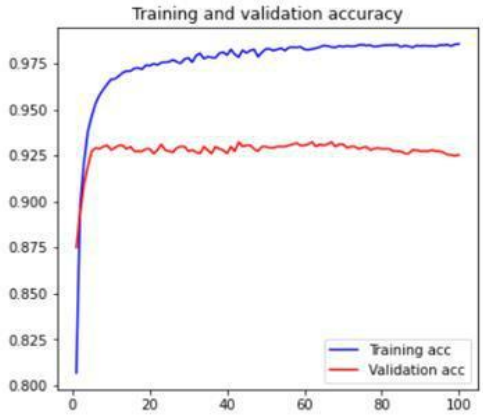


Figure 5 Accuracy plot for MLP

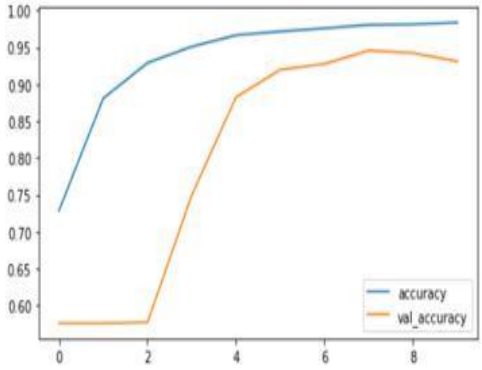


Figure 6 Accuracy plot for LSTM.

ISSN: 2233-7857 IJFGCN

5. Conclusion and Future Works

In this work, a sentimental analysis was performed on Amazon's alexa reviews dataset. Performance in terms of precision, recall, F1 score and ROC of four algorithms namely naïve bayes, random forest ,Multi layer perceptron and LSTM was evaluated. It is clear from the result analysis that LSTM's outperform all the models followed by MLP. LSTM's gives an F1 score of 97% and ROC score of 99%.The main reason behind LSTM's giving such high accuracy is that it has the ability to remember sequence of past words to make decision regarding sentiments of current word. The work also gives comparative evaluation of

performance metrices of all the algorithms. It also demonstrates the which metrices can be best used for imbalanced datasets.

In future work we would like to consider neutral reviews given by the customers and also study various other algorithms that can be used for sentimental analysis and measure their performance using metrics used in current study. Further we will try to reduce the overfitting scenarios in MLP and LSTM by increasing or providing strong regularization.

References

[1]R. B. Shamantha, S. M. Shetty and P. Rai, "Sentiment Analysis Using Machine Learning Classifiers: Evaluation of Performance," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 21-25, doi: 10.1109/CCOMS.2019.8821650.

[2] E. Aydoğan and M. A. Akcayol, "A comprehensive survey for sentiment analysis tasks using machine learning techniques," 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), Sinaia, 2016, pp. 1-7, doi: 10.1109/INISTA.2016.7571856.

[3] S. Singh and V. Singh, "Sentiment Analysis, Emotion Detection and Opinion Mining Algorithms, Applications and Challenges: An Exploratory Study," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, 2018, pp. 462-467, doi: 10.1109/CONFLUENCE.2018.8442631.

[4] Y. Gao, Z. Pan, H. Wang and G. Chen, "Alexa, My Love: Analyzing Reviews of Amazon Echo," 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, 2018, pp. 372-380, doi: 10.1109/SmartWorld.2018.00094.

[5]W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, 2014.

[6] K. Hulliyah, N. S. Awang Abu Bakar and A. R. Ismail, "Emotion recognition and brain mapping for sentiment analysis: A review," 2017 Second International Conference on Informatics and Computing (ICIC), Jayapura, 2017, pp. 1-5, doi: 10.1109/IAC.2017.8280568.

[7] E. Cambria, "Affective Computing and Sentiment Analysis," in *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102-107, Mar.-Apr. 2016, doi: 10.1109/MIS.2016.31.

[8] Y. Wan and Q. Gao, "An ensemble sentiment classification system of twitter data for airline services analysis," in Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, pp. 1318–1325, IEEE, 2015.

ISSN: 2233-7857 IJFGCN

[9] Z. Jianqiang and G. Xiaolin, "Comparison research on text preprocessing methods on twitter sentiment analysis," IEEE Access, vol. 5, pp. 2870–2879, 2017.

[10] S. Tokle, S. R. Bellipady, R. Ranjan, and S. Varma, "Energy-efficient wireless sensor networks using learning techniques," Case Studies in Intelligent Computing: Achievements and Trends, pp. 407–426, 2014.

[11] H. Kelly, "Why Amazon's Echo is the computer of the future," Nov. 2014. [Online]. Available: http://www.cnn.com/2014/11/12/tech/innovation/amazon-echoalways-listening/

ISSN: 2233-7857 IJFGCN

[12] M. Rushdi Saleh, , M.T. Martín-Valdivia, A. Montejo-Ráez, L.A. UreñaLópez, “Experiments with SVM to classify opinions in different domains”, Expert Systems with Applications, vol. 38, no. 12, pp. 14799– 14804, 2011.

[13] “BenchmarkingSentimentAnalysisSystems”,28thApr2017[Online], Available:https://8kmiles.com/blog/benchmarking-sentiment-analysis systems/#:~:text=Precision%20measures%20the%20exactness%20of,recall%20means%20more%20false %20negatives.

[14] Shir Meir Lador, “What metrics should be used for evaluating a model on an imbalanced data set? (precision + recall or ROC=TPR+FPR)”, Sept 05,2017 [Online], Available: https://towardsdatascience.com/what-metrics-should-we-use-on-imbalanced-data-set-precision-recall-roc-e2e79252aeba

[15] George V Jose,”Useful Plots to Diagnose your Neural Network”, Oct 3, 2019 [Online], Available:https://towardsdatascience.com/useful-plots-to-diagnose-your-neural-network-521907fa2f45

[16] B. Lima and T. Nogueira, "Novel Features Based on Sentence Specificity for Helpfulness Prediction of Online Reviews," 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), Salvador, Brazil, 2019, pp. 84-89, doi: 10.1109/BRACIS.2019.00024.

[17] A. A. Sayed, E. Elgeldawi, A. M. Zaki and A. R. Galal, "Sentiment Analysis for Arabic Reviews using Machine Learning Classification Algorithms," 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, 2020, pp. 56-63, doi: 10.1109/ITCE48509.2020.9047822.

[18] Z. Singla, S. Randhawa and S. Jain, "Statistical and sentiment analysis of consumer product reviews," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203960.

[19].Lamiae Hana, “A step by step Guide on Sentiment Analysis with RNN and LSTM”, Jan 22,2019 [Online], Available:https://medium.com/@lamiae.hana/a-step-by-step-guide-on-sentiment-analysis-with-rnn-and-lstm-3a293817e314

[20]. Surya Prabha PM, Subbulakshmi B,” Sentimental Analysis using Naive Bayes Classifier”,IEEE, ViTECoN,2019

ISSN: 2233-7857 IJFGCN