ZKG INTERNATIONAL ISSN: 2366-1313

Machine Learning and Deep Learning Techniques for Sentiment Polarity Detection: A Comprehensive Study

[1]Shahwar Fatima, [2]Mir Abdul Mateen Ali, [3]Raza Ahmed, [4]Syed Farhan Haji Abbas

[1]Assistant Professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

[2,3,4]B.E Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

shahwar012@gmail.com, mateenali1202@gmail.com, razanet11@gmail.com, hajisyed@hotmail.com

Abstract: *As e-commerce has grown in latest years, on line shopping has improved at the side of the number of product opinions on line. Customer pointers or proceedings have an effect on customers and their purchasing decisions. Polarity evaluation is the translation and type of facts. The fundamental goal of our paintings is to categories each patron evaluation into a class that represents its nice (appropriate or awful). Our polarity view includes the subsequent: pre-processing, function extraction, education, category, and generalization. First, the analysis is transformed into a vector illustration using the unique techniques Tf-Idf and Tokenize. Then, we studied with system mastering fashions of Linear SVM, RBF, Sigmoid kernel and LSTM deep learning fashions. Then, we compare the version using accuracy, f1 rating, and precision and don't forget. Our LSTM model predicts 86% accuracy for Amazon-based totally customer evaluations and 85% accuracy for Yelp consumer reviews.*

Keywords- Natural Language Processing (NLP), Support Vector Machine (SVM), Long Short Term Memory (LSTM), Machine Learning, Deep Neural Network

I. INTRODUCTION

Online buying has emerged as a daily exercising for lots humans round the arena. The value of client devices and online offerings has improved extra than ever. One of the motives why on line shopping has grown so quick over time is the satisfaction the organization can offer its clients. It is predicted that 2.14 billion people international will shop on-line in 2021. Global e-trade revenue is expected at $4.891 billion over the equal length. If the evaluation of on line purchasing isn't considered enough, estimates display that global e-commerce sales will reach $6.Four trillion through 2024[1]. Online buying is becoming an increasing number of famous for plenty reasons, along with

ISSN: 2366-1313

comfort and decrease charges. Retailers are constantly looking to get customers to keep in a physical area. You can behave as you desire even if buying online.

In addition, online stores do the whole lot to make shopping at the Internet similar to buying. Online purchasing offers many options that aren't essential when purchasing in individual. You can browse one-of-a-kind websites at the Internet and pick out the products according to your desires. Although you have got the option to pay with cash on shipping, on-line shopping now not calls for you to physically supply the product first. Alternatively, costs may be paid with a debit or credit card. Online purchasing debts for sixty-3 percent of all purchases [2]. Many human beings purchase matters on-line, especially for food, and day by day desires have improved due to intense climate measures implemented through international locations round the sector to prevent the spread of disease. During the Covid-19 pandemic, it isn't feasible to go to stores, so people are inclined to buy matters online. Due to Covid-19, on-line merchandise almost doubled in March 2020 in comparison to 2018, or 22% [3]. In this situation, on-line purchasing gives us control whilst we want and rewards us with a “non-poisonous” shopping enjoy. When you appearance, if you discover the opinion of this product is affordable or bad, and you could get thoughts from the evaluate, then you should buy a very good product. Then if we can pick out the advantageous and bad assessment then the acquisition will now not be clean for clients. To do that, our newspaper will recognise consumer evaluations and divide the scores to avoid the trouble of knowledge the reviewer's factor of view.

Amazon, one in every of the largest e-trade web sites, gives a wide range of merchandise including books, medicines, fitness equipment, and so on. In order to get rid of patron wishes or reviews, our layout might be in comparison with Amazon product evaluations and enlarge the structure of Amazon, Yelp products and wishes can be evaluated and put notes on associated schooling. For example, if a purchaser critiques a product, our layout could be rated as terrific or terrible.

The reason of this newsletter is to separate user evaluations on various products into right and horrific. More than 88% of customers surveyed in the article determined many private opinions, consistent with analysis conducted on Amazon over the past twelve months. Every article at the net with large critiques has solid data approximately its validity. The type of purchaser rating and assessment elimination is the primary

ISSN: 2366-1313

cognizance of this newsletter. There are many high-quality algorithms in system getting to know, our affirmation is on SVM with Linear, RBF, Sigmoid kernel and LSTM to see which classifier is pleasant on Amazon patron score and Yelp purchaser opinions test facts. Most of the fashions created within the clinical literature many researchers have targeted on information measured with the aid of an unmarried export, so it is not viable to offer more self belief on what to do, how to model can be executed at the identical time as trying out with information from a. - a very good product.

II Related Work

In latest years, a few researches has been done to perceive the polarity of feelings, the use of devices to recognize and study customers' feelings extensive.

[4] They used a superb product for us on one of the product critiques. With the aim of improving all the extraction techniques and the first step, they have got chosen the most correct technique for their look at. It has been verified that each one the documents are most beneficial whilst there are TF-IDF capabilities and the bag of phrases is used as part of the unique choice. Using a vector book system is a higher option due to the reality that the information is large and does now not want to be an excessive amount of. Based on these results, the satisfactory degree of accuracy has elevated to ninety four.02%.

In this look at, Sanjay Dey, Sarhan Wasif, Dhiman Sikder Tonmoy, Subrina Sultana, Jayjeet Sarkar and Monisha Dey make two gear to get pattern recognition: Support Vector Machine (SVM), NB [5]. This article provides a specific system learning strategy for reading user critiques on Amazon products. In this graph, their model is educated using almost 2,250 functions with almost 6,000 post-processing statistics. During this time, almost 4000 measuring devices have been surpassed using the measurement gadget. The tool gives 82.Eighty five percentage accuracy, 82.88 percent healing, eighty-4 percentage, and eighty two.662 percent f1 index for the SVM classifier. .

Atiqur Rahman and Md. Sharif Hossen uses SVM approach, the information incorporates 2000 video evaluations of which one thousand is bad and the remaining one is wonderful [6]. According to this text, the accuracy of the SVM model is 87.33%, the accuracy is 80 5. Ninety%, taken into consideration to be 89.33% and f1 score is 87.58%.

Arwa S. M. Al Qahtani reviewed Amazon overview statistics and studied the principle using different models [7]. The

ISSN: 2366-1313

classifiers used include BERT, NB, BiLSTM, RF and Logistic Regression. With an accuracy of 90-4 percentage and 90-eight percentage in binary and multiclass, the BERT model suggests particular consequences that offer an accuracy of 90-8 4 %, recall 98. Four percentages, f1 get ninety eight. 4 percentages. The author used RF broadly speaking with Gloves. RF with gloves suggests an accuracy of 90%.

In this paper, Akanksha Halde, Aditi Uttekar and Amit Vishwakarma carried out numerous BERT fashions for RF, NB and SVM sides [8]. The BERT classifier is the most accurate in predicting the perception of seeing values with approximately 90% accuracy after numerous iterations and assessments of the model.

Naveen Kumar Gondhi, Chaahat, Eishita Sharma, Amal H. Alharbi, Rohit Verma and Mohd Asif Shah used LSTM, CNN, SCA, NB fashions for sentiment analysis [9]. After approximately 10 intervals of look at, they calculated the gaining knowledge of loss and accuracy and validity in their model. It turned into set up that 0. Seventy 8 have become the endpoint of the distribution speculation, specially based at the ROC curve. They use the f1 index as a good indicator of the overall performance of the design as the statistics become unbalanced. The reason of their studies is to evaluate the overall performance of the version with a huge quantity of statistics.

Roobaea Alroobaea proposed long time reminiscence (LSTM), gated recurrent unit (GRU) and convolution neural network (CNN), RNN type [10]. The creator uses Arabic records and tools to research the mind and make certain that the quit end result of many deep know-how models of our data of all types. It makes use of information equations and models to be expecting massive statistics higher than shorter records. He compared Palsah's BiLstm method and showed the accuracy of his RNN version which became higher for three datasets.

Shiva prasad TK and Jyothi Shetty reviewed many articles from diverse researchers [11]. They tried to combine their thoughts approximately diverse classifications of questioning on the identity of various gadgets. They have defined many standards these days that describe the assessment

III DATASET

Most of the 43,000 (Table I) user reviews were used. 36,500 insights from Amazon customer reviews 1. The dataset of 6,500 reviews was taken from Yelp 2. Here, the opinions of one celebrity and two celebrities are considered in negative

ISSN: 2366-1313

opinions and ratters of four famous people and five famous people are considered. Remarkable thoughts, however, three-star reviews are not considered in Amazon customer reviews. Reviews from one megastar and two superstars, three superstars and four stars were considered to determine negative and positive reviews for the Yelp dataset. Positive and negative reviews are labelled 1 and zero. Each data set is equal. The source code for this document is also available on github3.

TABLEI
DATASET

Positive (1)	Negative (0)	Total
13500	13500	27000(Training set)
1500	1500	3000 (Validation set)
3250	3250	6500(Test set)
3250	3250	6500(Yelptestset)

IV. CLASSIFICATION APPROACH

In the Fig 1 we've got visualized our classification approach for detecting sentiment polarity. After finishing pre-processing of each dataset, we trained our model and confirmed it with the validation dataset. Then we used two one-of-a-kind test datasets and got predicted output from the skilled model and evaluated the consequences.

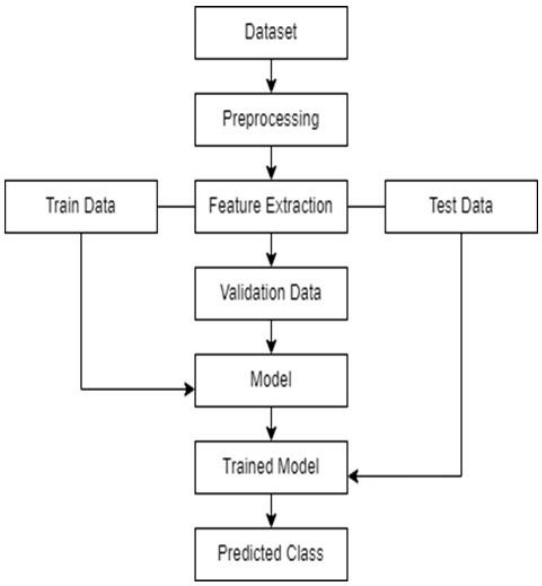

Fig.1.ClassificationFramework

1) We have used Support Vector Machine (SVM) with RBF, Linear, Sigmoid kernel and LSTM in our proposed machine.

2) Support Vector Machine (SVM): At first in SVM to extract the competencies from textual facts and redesign it to numerical we have used the Tf-Idf victimizer and n-gram variety from 2 to three. Then for RBF, Linear and Sigmoid kernel SVM fashions are run and particular predicted output is calculated for every check datasets. We changed the n-gram range, implemented don't forget victimizer, and evaluated the predicted outputs to similarly first-class-music the model. The SVM classifier performed pleasant the usage of the version we advised that included a Tf-Idf victimizer.

3) 1) LSTM: In Fig three we've visualized our LSTM version architecture.

ISSN: 2366-1313

In LSTM to extract the competencies from textual statistics, we have used Tokenize. Our model is sequential. The model is built with an Embedding layer, LSTM layers and dense layers.

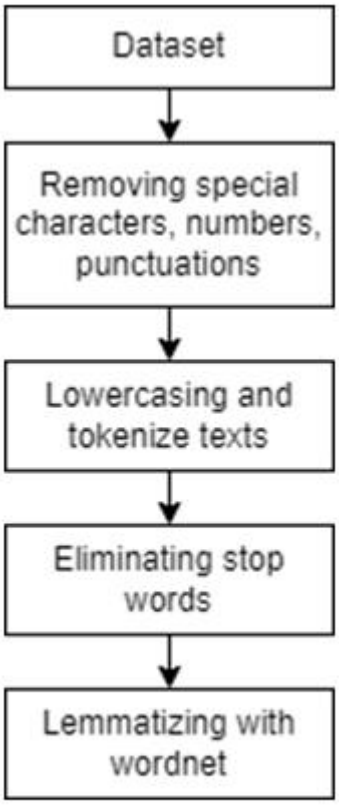

Fig.2.DataPreprocessing

Length is calculated from the pre-processed textual content files which include the very best number of phrases. LSTM layers include a complete of 640 hidden nodes. Dense layers comprise a total of forty three hidden nodes and Elu, Sigmoid are used as activation features. To hyper-track the LSTM version we've got used distinct activation capabilities with distinct optimizers. We have experimented through converting the wide variety of layers of LSTM and Dense. With RMSprop optimizer we've got the quality-anticipated outcomes. We have used binary move-entropy as the loss characteristic of our LSTM model.

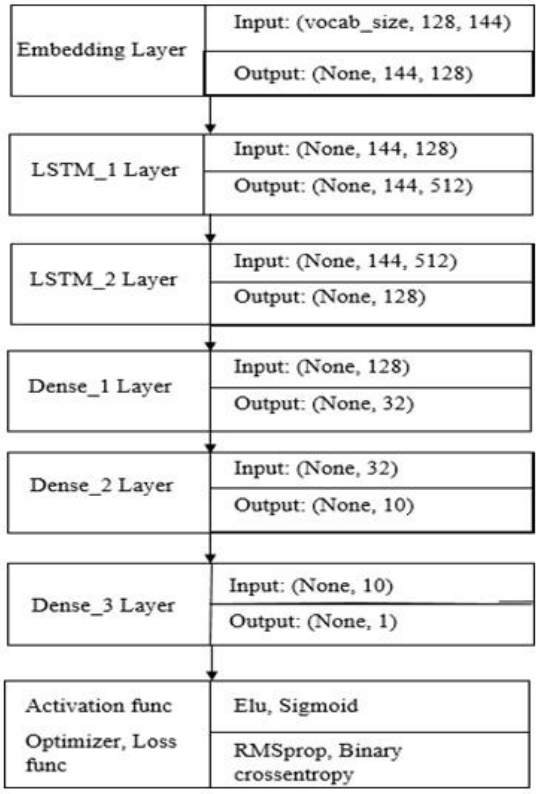

Fig.3.LSTMClassifierModel

V EXPERIMENTANDEVALUATION

A. SVM Performance

According to Table II and Table III, we're able to see that RBF SVM classifier showed the very first-class accuracy eighty four% for the Amazon take a look at dataset and eighty two% for the Yelp test dataset than superb sigmoid and linear SVM classifier. Precision is the very amazing 86p. C for excessive excellent (1)

ISSN: 2366-1313

reviews and 83percent for negative (0) reviews in RBF SVM and Linear SVM on the Amazon test dataset. For the Yelp test dataset, precision is the satisfactory 88p.C for fantastic

(1) Opinions and seventy six% for negative (0) evaluations in RBFSVM. We moreover see that keep in thoughts and f1-score parameters are nearly identical for the RBF, Linear, and Sigmoid kernel of SVM classifier for Amazon test dataset. After evaluating common performance parameters for each take a look at datasets with Linear, Sigmoid, RBF kernel of SVM we see that each SVM classifier kernel not unusual overall performance parameters expect properly for super (1) customer critiques than awful (0) client evaluations except consider typical performance parameter.

TABLEII
RBF, LINEAR, SIGMOIDSVMWITHAMAZONTESTDATASET

(a)Performance Parameters

	RBFSVC		Linear SVC		Sigmoid SVC	
Class/Label	0	1	0	1	0	1
Precision	0.83	0.86	0.83	0.86	0.82	0.85
Recall	0.85	0.83	0.85	0.83	0.85	0.83
F1Score	0.84	0.84	0.84	0.84	0.83	0.84
Accuracy (%)	84		84		84	

TABLEIII
RBF, LINEAR, SIGMOIDSVMWITHYELPTESTDATASET

(a)Performance Parameters

	RBFSVC		Linear SVC		Sigmoid SVC	
Class/Label	0	1	0	1	0	1
Precision	0.76	0.88	0.75	0.88	0.75	0.87
Recall	0.86	0.78	0.86	0.78	0.85	0.78
F1Score	0.81	0.83	0.80	0.83	0.80	0.82
Accuracy(%)	82		82		81	

VI CONCLUSION

After evaluating the application, we found that a very good accuracy of 86% is shown with the useful capacity of the LSTM model for the Amazon control dataset. The RBF, Linear, and Sigmoid kernels of the SVM model test accuracy is the same for the Amazon test data, which is 84%. For the Yelp test data, we obtained the highest accuracy from the LSTM version. For both test data, the performance of the LSTM model is higher than that of the SVM model.

We have noticed that the performance of our model is greatly increased when we add hyper-tune models. In the future, we will be able to increase the time of our dataset and improve all the performance according to one of the release. We will use the BERT model. We will continue to research the extraction of data resources and use lemmatization's and special models to improve the system.

REFERENCES

[1] "Online shopping statistics, "last accessed 18August, 2022.[Online]. Available: https://www.oberlo.com/blog/online-shopping-statistics

[2] "Online shopping statistics second," last accessed 18 August, 2022. [Online]. Available:

ISSN: 2366-1313

https://www.oberlo.com/blog/online-shopping-statistics

[3] T.U.Haque, N.N.Saber, and F.M.Shah, "Sentiment analysis on large scale Amazon product views,"*IEEE International Conference on Innovative Research and Development*, 2018.

[4] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and naive bayes classifier for sentiment analysis on Amazon product reviews," *International Conference on Contemporary Computing and Applications (IC3A)*, 2020.

[5] A.Rahman and M.S.Hossen, "Sentiment analysis on movie review data using machine learning approach," *International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019.

[6] A. S. M. AlQahtani, "Product sentiment analysis for Amazon reviews,"*InternationalJournalofComputerScienceandInformationTechnology*, 2021.

[7] A.Halde, A.Uttekar, and A.Vishwakarma, "Sentimental alyssum Amazon product reviews," *International Research Journal of Modernization in Engineering Technology and Science*, 2022.

[8] N. K. Gondhi, Chaahat, E. Sharma, A. H. Alharbi, R. Verma, and M. A.Shah, "Efficient long short-termmemory-based sentiment analysis of e-commerce reviews," *Hindawi Computational Intelligence and Neuro-science*, 2022.

[9] R. Alroobaea, "Sentiment analysis on Amazon product reviews using the recurrent neural network (rnn)," *International Journal of Advanced Computer Science and Applications*, 2022.

[10] S. T. K and J. Shetty, "Sentiment analysis of product reviews: A review," *International Conference on Inventive Communication and Computational Technologies (ICICCT 2017)*, 2017.

[11] Z. Singla, S. Randhawa, and S. Jain, "Sentiment analysis of customer product reviews using machine learning,"*2017InternationalConference on Intelligent Computing and Control*, 2017.

[12] Prasadu Peddi and Dr. Akash Saxena (2014), "EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE", International Journal of Emerging Technologies and Innovative Research

ISSN: 2366-1313

(www.jetir.org), ISSN:2349-5162, Vol.1, Issue 6, page no.314-318, November-2014, Available: http://www.jetir.org/papers/JETIR1701B47.pdf

[13] Prasadu Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.