International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 10 | Oct 2024 www.irjet.net p-ISSN: 2395-0072

Analysis of movies Reviews using Seq2Seq Pattern in Different Machine Learning

Pradeep [1], Prof. Priyanka Bhatele[2], Prof. Ajit Kumar Shrivastava[3]

M.Tech Scholar, Department of CSE, SISTec-R, Bhopal (M.P), India[1]
Assistant Professor, Department of CSE, SISTec-R, Bhopal (M.P), India[2]
Professor & Head of Department of CSE, SISTec-R, Bhopal (M.P), India[2]

ABSTRACT

Here, Authors did their Experiment and show their Experimental work. In this scenario they choose a dataset of movies where reviews are recorded. When we talked about sentiment Analysis corpora contains different dataset from where we selected movies data. As we know that in natural language processing, we have different categories like positive, negative and neutral. In this dataset we have only 2 categories positive and negative. As data scientist we have to extract the polarity against the given reviews by thousands of users. Here Author's used Seq2Seq and Word2Vector functionality for extraction the real and true values. As we know that for creating a model we need to pick different Machine Learning Algorithms. Here we have taken NaiveBayes with different flavour, Regression and support vector machine. After creating model, we applied different performance parameters and found that Voted Algorithms gives some better result in comparison to others. In terms of Accuracy this gives us nearly 90% of Accuracy.

Keywords: *Sentiment Analysis, Plaintext Corpora, Tokens, Features, Word2Vector,* Seq2Seq, Sklearn libraries, Performance parameters like Accuracy.

I INTRODUCTION

According In Recent Era Internet play vital role in every sector from Education to Entertainment, From Research to Agriculture, from space to sea at every place we found that presence of Internet is available. Day by social media is also playing very important roles in business analysis as well as product analysis. When we start exploring the social media from every corner of the world have some involvement in every Industries or sector. Now social media is replacing the news media as well as print media. when we are observing the social media nowadays reels play a vital role where people show their views in terms of reels, text messages, images or video's. If we observe a big player in E-Commerce like Amazon, Flipkart and many more all totally depends upon user's review, means reviews are driving force today in every business [1].

Famous websites like imdb.com, Carwale.com and many more are available with millions of reviews of thousands of users. Take an example carwala.com where every user is totally depends of reviews of a given Car or Model. The positive reviews can make positive to buy a car by many customers on daily basis. In movies section one best example is Box-Office which gives us clear cut message of any movie's impression on initial 2-3 days. By these reviews we come to consideration whether movies are going to direction of Hit or flop. In many bloggings site we see that in day-to-day life at many topics Millions of pass their positive or Negative comments. which can be exploited for emotion. Many times, they contain more factual or relevant information and they are not considered to the desired extent.

In day-to-day life we have different scenario that can be solved with the help of following mechanism.

Predictive models: In this mechanism given model will predict the result for near future. The model needs to train itself by huge data available for that domain. Then they are able to predict a value for a given new data. The predictive model is very important for upcoming business Analysis.

Descriptive models: This model is using for deep analysis which is used further for doing segmentation-based analysis. Here Relationship is available with different parameters of given domain. By these data we can able to create an equation in mathematical form which gives us consolidate information in deep analysis. It is generally used to help to understand what the system is Right now and what it can do and how it is doing it.

1.1 Machine Learning Techniques

Learning algorithm play important roles in different industries like E-Commerce, Insurance, Education, Engineering etc. In this section we worked for sentiment analysis or finding polarity from movies reviews dataset. We know that we have number of algorithms to solve our problem out of them we are explain some algorithms:

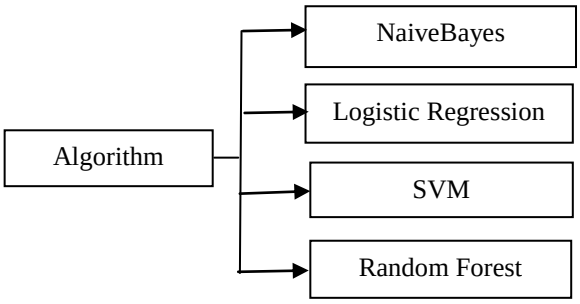

Figure 1: Used Algorithms in Sentiment Analysis [1]

1.2 Outline of Sentiment Analysis

Sentiment Analysis is the process to find the view in terms of Positive, Negative & Neutral. Here by the comments given by different audience in terms of Text, Image, Audio, Video & reels we as Data Scientist able to detect their opinion for a specific product or domain. In the pictorial diagram we are trying to explain where Sentiment Analysis can be used:

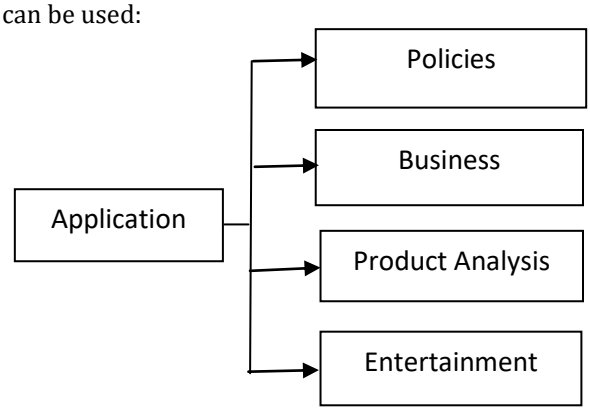

Figure 2: Area where Sentiment Analysis works [2]

Note: By using Sentiment analysis, we can take better decision in terms of product launching and others activities.

1.3 Steps Followed in Sentiment Analysis

For Implementing sentiment analysis, at first stage we need the required libraries that is nltk module. In nltk module we have hundreds of functions that will help Data Analyst to explore information from a given Data set. In next section we will explain in details.

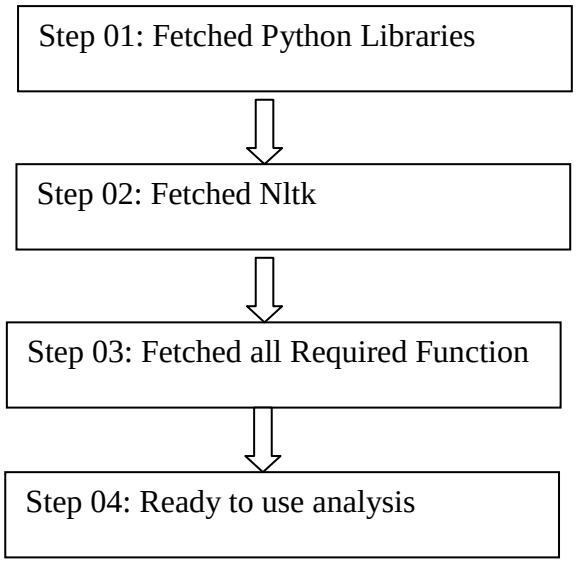

Figure 3: Steps to use Sentiment Analysis Module [3]

1.4 Used Functionalities

We know that by nltk module we can perform sentiment analysis process. For doing this we have to understand how they work. Major component in this is given below:

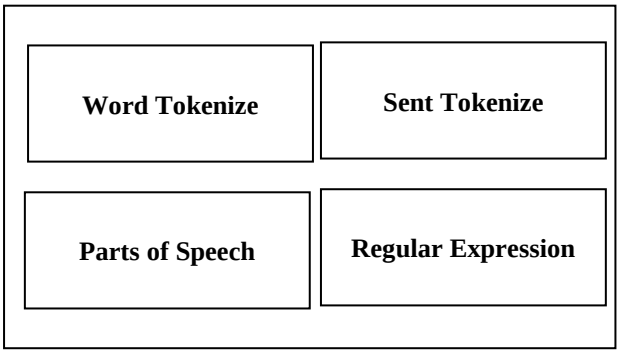

Figure 3: Used Functions in Sentiment Analysis [4]

Word Tokenize: In In this section, we divide the Given sentences in smallest token that is word tokenize. This helps during individual's analysis of word.

Sent Tokenize: Here we will divide our message into useful sentences probably it divided on the basis of dot.

POS: Here we will analyse that which word comes under which category like Noun, Adj, Verb etc.

Regular Expression: It is very important part in our analysis by this mechanism we can find many useful information.

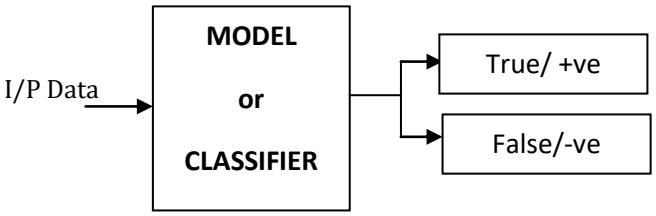

Figure 4 - class classifier [5]

II RELATED WORK

Many different mechanisms are used for finding emotion analysis. The purpose is to build a model or machine that can learn and further predict the result. Today's we have many statistical approaches that will give domain related analysis. This kind of approaches will provide us beneficial approaches for result finding. The sentiment analysis can be done in many different languages, but very small work has been done in Hindi language. The very first work was done in Hindi, Marathi and Bengali. But at this time the level of work in Hindi is not very appreciable [8]. Therefore, the requirement of the same as the result of various surveys has been felt.

In Here authors used the lexicon method for classification so that the proposed algorithm could be compared with the UGGram presence method. In this scenario a large collection of Data or files is converted into unigrams, bigrams, trigrams, fourgrams and fivegrams. we are explaining this concept in experimental way [9] :

text = "Dumy Text Dumy Text Dumy Text Dumy "

token = nltk.word_tokenize(text)

bigrams = ngrams(token,2)

trigrams = ngrams(token,3)

fourgrams = ngrams(token,4)

fivegrams = ngrams(token,5)

In Here authors explains that the better news spirit analysis method. In this, the news sentiment is analysed by cutting out the title and text separately and through intensive study of Chinese news, and two different algorithms are applied in both areas. A neutral news assessment method has been proposed for the title part and a subjective sentence recognition algorithm is used for news lessons [3].

III PROBLEM IDENTIFICATION

Many research work has been done in the field. Authors have learned several things from this study (work). We found that many authors did their work in small Data Set and they took small chunk of data for Training and Testing. and they took small chunk of data for Training and Testing. Here Authors try to explore on the basis of big Data Set.

IV FLOW DIAGRAM OF SYSTEM

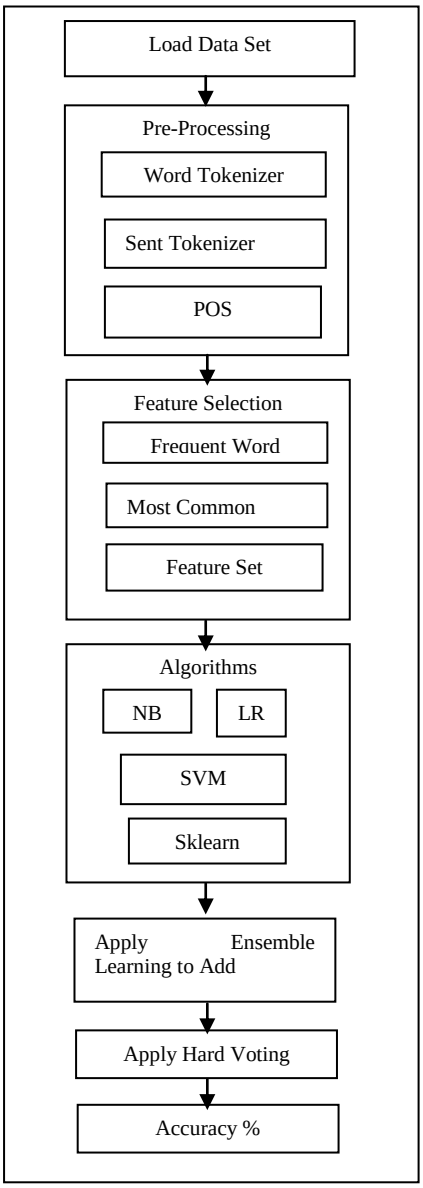

Figure 6: Flow Diagram

Explanation: In Step 01 we fetched data from Data source where we have thousands of files where reviews are listed. In Step 02 we pre-process the given Data using given functionalities. In Step 03 we picked the useful words that will help you for finding sentiment. In step 04 we implement different ML and find the performance parameter like accuracy for every taken models. As we know that we have many performance parameters in Machine Learning but here we are taking some of them. Accuracy plays very important roles for claiming best Model here we are explaining Accuracy in short:

Accuracy: it refers to the closeness of a measured value to a known value.

Accuracy = True Positive + True Negative / Total

V RESULT ANALYSIS

In this part we will represent our numerical values which I found during process we explained in terms of table.

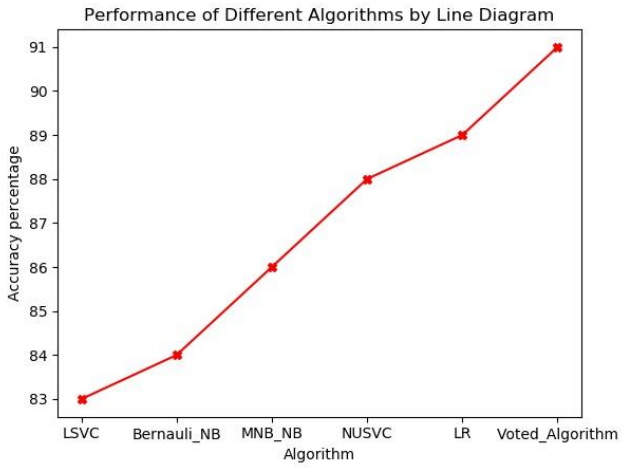

Figure 7: Line Graph Between Algorithms

Explanation: After analysing the above Figure 6 we can say that with Voted Algorithm gives better result. Results also depends upon pre-processing mechanism Here we used Word2Vec & Seq2Seq mechanism for data pre-processing. We analyse the results on Different Algorithms. it may show a sharp and definite pattern of our analysis.

VI CONCLUSION

Author's studied number of papers and finally they give their own implement did much work but still there is much scope in performance improvement. And during model implementation Author's found that features finding play very vital roles here.

VII FUTURE SCOPE

During study we find that sentiment analysis has major three component Positive, Negative and Neutral. In future we can include neutral sentiment also in our Analysis.

REFERENCES

[1] Charu Nanda, Mohit Dua and Garima Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning", International Conference on Communication and Signal Processing, April 3-5, 2018, India,978-1-5386-3521-6 2018 IEEE

[2] V.K. Singh, R. Piryani, A. Uddin,P. Waila ,"Sentiment Analysis of Movie Reviews and Blog Posts",978-1-4673-4529-3 2018 IEEE .

[3] Mr. B. Narendra, Mr. K. Uday Sai, Mr. G. Rajesh, Mr. K. Hemanth, Mr. M. V. Chaitanya Teja, Mr. K. Deva Kumar,"Sentiment Analysis on Movie Reviews: A Comparative Study of Machine Learning Algorithms and Open-Source Technologies", I.J. Intelligent Systems and Applications, 2016, 8, 66-70 Published Online August 2016 in MECS (http://www.mecs-press.org/)

[4] Rachana Bandana, "Sentiment Analysis of Movie Reviews Using Heterogeneous Features",978-1-5386-5550-4 2018 IEEE

[5] Tirath Prasad Sahu,Sanjeev Ahuja,"Sentiment Analysis of Movie Reviews: A study on Feature Selection & Classification Algorithms",978-1-4673-6621-2/16 2016 IEEE

[6] upma kumari, Dr.Arvind k. sharma, Dinesh soni,"sentiment Analysis of smart phone product Review using SVM Classification Techniques",International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).

[7] Rasika Wankhede,Prof. A.N.Thakare,"Design Approach for Accuracy in Movies Reviews Using Sentiment Analysis",International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[8] Kamil Topal,Gultekin ozsoyoglu,"Movie Review Analysis: Emotion Analysis of IMDb Movie Reviews",2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

[9] Akshay Amolik, Niketan Jivane, Mahavir Bhandari, Dr.M. Venkatesan,"Twitter Sentiment Analysis of Movie Reviews Using Machine Learning Techniques'-ISSN: 0975-4024 Akshay Amolik et al. / International Journal of Engineering and Technology (IJET)

[10] Nagamma P, Pruthvi H.R, Nisha K.K and Shweta NH, "An Improved Sentiment Analysis of Online Movie Review Based on Clustering for Box Office Prediction".