Comparative Study Of Algorithms For Sentiment Analysis

[1]Mr. Akshay Agrawal, [2]Mr.Vishal Kamlapure, [3]Mr.Ravi Pai, [4]Mr.Prasad Mahimkar

[1]Asst.Professor ,UCoE,Kaman,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharshatra, India.

[1]akshay1661@gmail.com, [2]vishalkamlapure.vbak@gmail.com, [3]ravi983365@gmail.com, [4]shree.mahi19@gmail.com

Abstract- Sentiment analysis or opinion mining is the study of people's opinions, sentiments. attitudes, and emotions expressed in written language. In recent years natural language processing and text mining are some of the most active research fields. On this day there is tremendous increase in number of people using online social networks and services to not only connect and communicate but also to voice their opinions. Sentiment Analysis is the identifying and categorizing of these opinions to determine the public's opinion towards a particular topic, problem, product etc. The importance of Sentiment analysis is increasing day by day. Machine learning is a field of computer science that gives computers the ability to learn without being manually programmed.

Keywords: Natural Language Processing, Sentiment analysis, Deep learning, Machine learning

I. INTRODUCTION

Sentiment Analysis refers to the use of Natural Language Processing and Information Extraction task that aims to obtain the writer's feelings expressed in positive or negative comments, questions, and requests, by analyzing a large number of documents. As sentiment analysis aims to determine the attitude of a speaker or a writer concerning some topic or the overall tonality of a document. In recent years, the increase in Internet usage and the exchange of public opinion is the driving force behind Sentiment Analysis today. The Web is a sea of structured and unstructured data. The analysis of this data to extract public opinion and sentiment is a challenging task. The analysis of sentiments may be either document-based where the sentiment in the entire document is summarized as positive, negative or objective. or it can be sentence based where individual sentences, bearing sentiments, in the text are classified.

The analysis of sentiments may be document based where the sentiment in the entire document is summarized as positive, negative or objective. It can be sentence based where individual sentences, bearing sentiments, in the text are classified SA can be phrase based where the phrases in a sentence are classified according to polarity.

Sentiment Analysis identifies the phrases in a text that has some sentiment. The person may speak about some objective facts or subjective opinions. It becomes necessary for an organization to distinguish between the two. A text may contain many entities but it is necessary to find the exact entity towards which the sentiment is directed. It identifies the polarity and degree of sentiment. Sentiments are classified as objective positive (denotes a state of happiness, bliss or satisfaction on part of the writer) or negative (denotes a state of sorrow, dejection or disappointment on part of the writer). The sentiments can further be given a score based on their degree of positivity, negativity or objectivity to be used in the industries. For the purpose of providing better customer support.

II. AIMS AND OBJECTIVE

a) Aim

The aim of making this project is to make it easier for companies to make sense of sea of unstructured text by automating business processes, getting actionable insights, and saving hours of manual data processing from the reviews provided by the customers towards their products.

b) Objective

- Adjust marketing strategy
- To improve product quality
- Improve customer service

III. LITERATURE SURVEY

Paper 1: Yilin Chen and HI Zhang, "Research on Text Sentiment Analysis based on CNNs and SVM",

 DOI : 10.35291/2454-9150.2020.0236

13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018.

Chen and Zhang proposed a model that combines SVM and Convolution Neural Network (CNN) for sentiment analysis The results of their experiment show that their method gives better accuracy than CNN or SVM alone for text sentiment analysis. NLPCC2014 emotional evaluation data set was used by them. This data set was processed such that that it is first filtered and then the filtered set is trained by Word2vec.The CNN learns the features whereas the SVM does the classification.

Paper 2: Abdalraouf Hassan and AusifMahmood, "Convolutional Recurrent Deep Learning Model for Sentence classification" IEEE Access, vol:6, 2018

The authors propose a model for sentence classification using a joint framework of a CNN and RNN. In the first step the words from sentences are changed to feature vectors. A 'Black Propagation through Time' algorithm is used to control the neural net. Finally, a classification layer is present which is a logistic regression classifier to give the final result with an accuracy between 89.2% and 93.

Paper 3: Min-Yuh Day and Yue-Da Lin, "Deep Learning for sentiment Analysis On Google Play Consumer Review", IEEE International Conference on Information Reuse and Integration (IRI),2017.

The authors perform sentiment analysis on Google play consumer reviews that are in the Chinese language. The data was then pre-processed and integration of dictionaries was performed. LSTM, Naive Bayes and SVM classifiers were used and the results were compared. 196,651reviews were collected for the experiment through a WebCrawler. The results suggest that the LSTM with an accuracy of 94% gave far better results than SVM and Naive Bayes with an accuracy of 76.46% and 74.12% respectively.

IV. EXISTING SYSTEM

Sentiment analysis is the analysis of the emotions or reviews of the consumers or users to determine whether a product, service, news, article, etc. are generating positive, neutral or negative responses. Data can be taken from any of the websites or many data sets of opinions and reviews available online. The traditional text sentiment analysis method is mainly based on an emotion dictionary or machine learning. However, its dependence on emotion dictionary construction and artificial design and extraction features make the generalization ability limited.

Drawbacks:

- Word2vec is not suitable for online reviews.
- Emotion classification is the core problem of sentiment analysis technology.

V. COMPARATIVE STUDY

Table: Comparative Study

Sr no	Paper Name	Author/ Publication	Technology	Advantage	Disadvantage
1	A Comparative Study of ML and DL Techniques for Sentiment Analysis	Kruttika Jain Shivani Kaushal	Naive Bayes , Lexicon Based	Flexibility of algorithms for different types of problems.	
2	Research on Text Sentiment Analysis based on CNNs and SVM	Yulin Chen and Zhi Zhang IEEE(2018)	CNNs and SVM	Find more accurate data	Algorithm requires more data preprocessing Hard to implement

 DOI : 10.35291/2454-9150.2020.0236

3	Convolutional Recurrent Deep Learning Model for Sentence classification	Abdalraouf Hassan and AusifMahmood IEEE(2018)	LSTM recurrent neural network	Combination of 4 algorithms to produce more accurate results	Fails to remove maximum misleading comments..
4	Deep Learning for sentiment Analysis On Google Play	Min-Yuh Day and Yue-Da Lin IEEE(2017)	Naive and Bayes SVM	helps users to discover valuable information	System is developed for Chinese language it needs data preprocessing for other languages .

VI. PROBLEM STATEMENT

1) The problem in sentiment analysis is classifying the polarity of a given text at the document, sentence or feature/aspect level.

2) Whether the expressed opinion in a document, a sentence or an entity feature or aspect is positive, negative or neutral.

3) Ambiguous negative words: Their context needs to be thoroughly understood and tagged accordingly. Example: "That backflip was so sick" is really a positive statement.

VII. PROPOSED SYSTEM

Sentiment analysis is carried out on these reviews given on various micro blogging sites such as Facebook, twitter, etc. The reviews may also belong to a product being sold on an online retail store such as amazon. A lot of jargon and mixed comments can be misleading and this is where sentiment analysis comes into place. It allows the businesses to better understand the general reaction to their products and their standing in the market. Sentiment analysis is done at many levels. Depending on the amount of time that can be dedicated to sentiment analysis and its importance for a proceeding task, the level of sentiment analysis is chosen.

VIII. ALGORITHM

The general idea of working of proposed system algorithm is given as follow:

Naive Bayes Classification

Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and classes.

Step 1: Start
Step 2: Input parameter = review
Step3: Initialize s, sum
Step4: Accuracy = Naive Bayes
Step5: While within bound.
Step6: Predict sentiment.
Step 7: Transaction
if transaction <ledbalance then
transaction=sum-ledbalance
Step8: Update Value of sentiment
Step 9: Update Node
Step10: Update Ledger **Step11:** Transaction successful
Step12: End.

SVM classification

SVM is a **supervised(feed-me)** machine learning algorithm that can be used for both classification or regression challenges. Classification is predicting a label/group and Regression is predicting a continuous value. SVM performs classification by finding the hyper-plane that differentiate the classes we plotted in n-dimensional space.

Step 1: Start
Step 2: Sample data= Review data **Step 3:** Input parameter= Symptoms **Step 4:** Initialize s
Step 5: Analyze = SVM (Trained data)
Step 6: Identify the sentiment
Step 7: Update Admin
Step 8: Prediction successful
Step 9: End

Random Forest

Random forests or **random decision forests** are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees

Step 1: Randomly select **"k"** features from total **"m"** features.
Where **k << m**
Step 2: Among the **"k"** features, calculate the node **"d"**

 DOI : 10.35291/2454-9150.2020.0236

using the best split point.

Step 3: Split the node into **daughter nodes** using the **best split**.

Step 4: Repeat **1 to 3** steps until "l" number of nodes has been reached.

Step 5: Build forest by repeating steps **1 to 4** for "n" number times to create **"n" number of trees**.

.IX. MATHEMATICAL MODEL

1. NAÏVE BAYES:

Naive Bayes Classifier is an easy probabilistic classifier based on applying Bayes Theorem with independence assumption. the more expressive term for the underlying quantity model would be

independent.

2. SVM CLASSIFICATION:

$y=\text{sign}(\sum^{m} K(x,x_i))$

Here,

K: Is a kernel function which measures the similarity between two reviews x and xi.

x: Predict the review by polarity.

3. NEURAL NETWORKS:

The interval activity of the neuron can be shown to be The skewness of aconveyance is characterized as,

$$v_k = \sum_{j=1}^{p} w_{kj} x_j$$

X. SYSTEM ARCHITECTURE

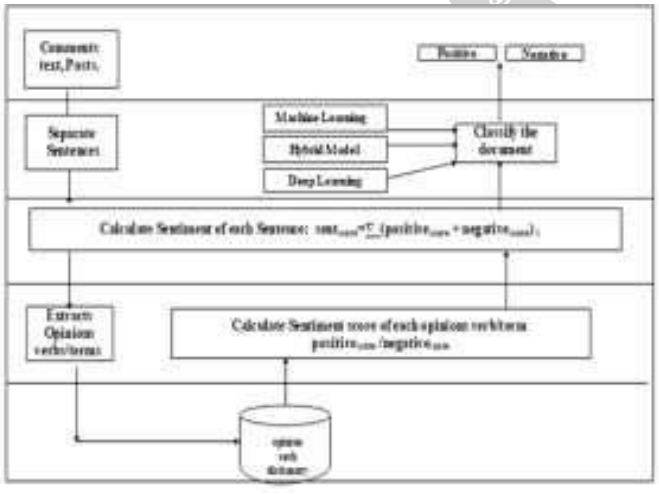

Fig.1: System Architecture

Output Phase: The data after the pre- processing enters the prediction model which outputs the sentiment score

XI. ADVANATGES

- First is sentence level analysis. A sentence has a subject and an object. Subjectivity gives us the subjective views of the object. Thus, the subjective part of the sentence is classified into positive negative or neutral.
- Another level is the feature level. This is the best level as it takes into account words and phrases. It takes a word with an emotion and determines the target of the word. It then classifies the word into positive, negative or neutral. Thus, maximum misleading comments are removed.

XII. DESIGN DETAILS

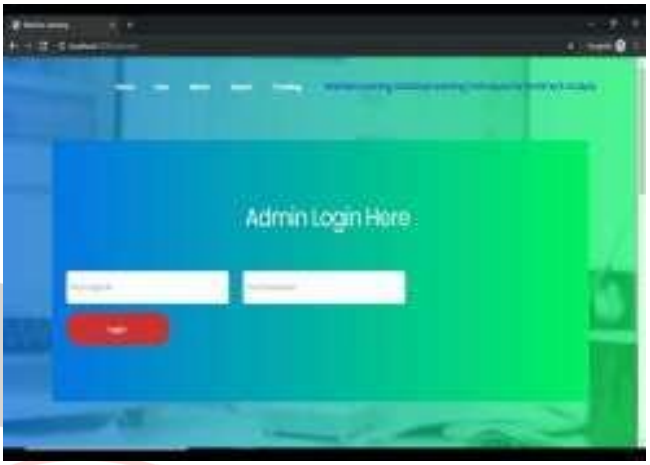

Fig.2: System Architecture

In this section , The Admin login through the interface , Admin gets options to manage users and analysis section.

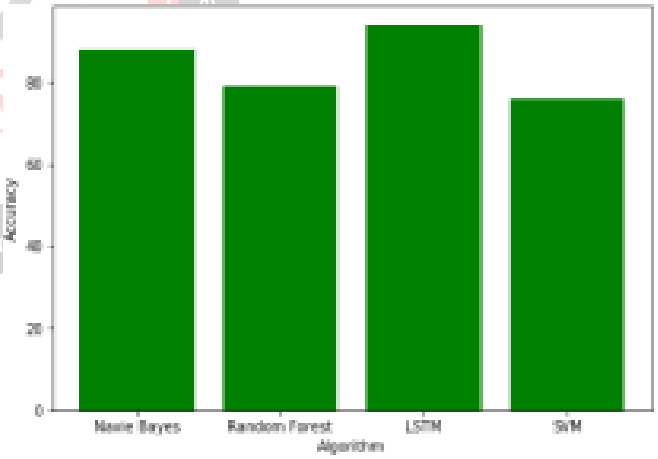

Fig 3 Result Set

In this section The user can compare the different algorithms for there accuracy as a graphical view , the analysis is based on the data collected at each search which helps user to decide which algorithm is best.

XIII. CONCLUSION

Thus, we have tried to implement the paper "Krutika Jain, Shivani Kaushal", "*A comparative study of machine learning and deep learning techniques or sentiment analysis*" IEEE 2018. Machine Learning and Deep learning Techniques avails great scope in sentimental analysis using different-different machine learning algorithm and deep

 DOI : 10.35291/2454-9150.2020.0236

learning algorithm to identify the algorithm which provide better efficiency and better performance during sentimental analysis. According to the analysis LSTM performs the best in all algorithm with accuracy from 85-91% and then naïve bayes with 83-85% followed by descison tree 81-85%, svm 78-82% to random forest 78-80%. This can be used for analyzing reviews and feedback on most ecommerce and movie website which can be used by company to increase the accuracy of their review systems.

REFERENCE

[1] Yulin Chen and Zhi Zhang, "Research on Text Sentiment Analysis based on CNNs and SVM", 13th IEEE Conference on Industrial Electronics and Applications(ICIEA), 2018.

[2] Abdalraouf Hassan and AusifMahmood, "Convolutional Recurrent Deep Learning Model for Sentence classification" IEEE Access, vol:6, 2018

[3] Min-Yuh Day and Yue-Da Lin, "Deep Learning for sentiment Analysis On Google Play Consumer Review", IEEE International Conference on Information Reuse and Integration (IRI),2017.

[4] Rahul Ghosh, Kumar Ravi and VadlamaniRavi,"A Novel Deep learning Architecture for Sentiment Classification", 3rd International Conference on Recent advances in Information Technology(RAIT- 2016).

[5] Arman S. Zharmagambetov and Alexander A. Pak, "Sentiment Analysis of a document using deep learning approach and decision trees", 12th International Conference on Electronics Computer and Computation (ICECCO), 2015

[6] Zhongkai Hu, Jianqing Hu, Weifeng Ding and XiaolinZheng, "Review Sentiment Analysis Based on Deep Learning", IEEE e-Business Engineering,2015.

[7] AlperKursatUysal and Yi Lu Murphey, "Sentiment classification: Feature selection based approaches versus deep learning", IEEE International Conference on Computer and Information Technology, 2017

 DOI : 10.35291/2454-9150.2020.0236