International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Sentiment Analysis using Naïve Bayes, CNN, SVM

Ankita Yekurke[1], Gayatri Sonawane[2], Harshada Chavan[3], Pradhyumn Thote[4], Anuja Phapale[5]

[1-5]*Department of Information Technology, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, India*

Abstract - *Thousands and millions of customers express their opinions on a brand, product, or service in major businesses. In the big picture, manually evaluating every review is difficult, and as a result, Sentiment analysis has received a lot of attention lately. Sentiment analysis is the mining of text which identifies and extracts subjective information from the source material and helping businesses to understand the social sentiment of their brand, product, or service while monitoring online conversations. The proposed paper focuses on different approaches and algorithms that could potentially help in sentiment analysis.*

Key Words: **Sentiment Analysis, Naïve Bayes, CNN, SVM**

1.INTRODUCTION

Human uses natural language to communicate with each other and the same thing computer should replicate (Copy) while communicating with the user and for this purpose only we use NLP. we know that NLP has a wide domain like Speech Recognition, sentimental analysis, machine translations, Chatbots are some applications of NIP. But in this paper, we are going to work on one of the specific applications of NLP named Sentimental analysis. Sentiment analysis models focus on polarity - positive, negative, neutral review but also feelings and emotions and even on intentions of users.

Sentiment analysis is a procedure that analyses a review and evaluates if the writer's attitude towards a given topic or product is positive, negative, or neutral using natural language processing and a text categorization tool. Analyzing many evaluations without a properly trained system is a difficult process. Sentiment analysis is a solution to this problem that involves training a system on millions of texts to determine if a message is good, negative, or neutral. It works by breaking down a review into separate sets of words, cleaning all non-essential material, and then defining the type of review based on the evaluation score (positive, negative, or neutral).

2. ALGORITHM

After reviewing the literature, we found that for sentiment analysis three main algorithms were used. In this section, we are discussing those algorithms.

1. *Naïve Bayes Algorithm:*

Classification means the grouping of data based on common characteristics. A Naive Bayes classifier is a probabilistic classifier that works by figuring out the probability of different attributes of data being associated with a certain class. The Naive Bayes classification algorithm utilizes the probability theory proposed by British scientist Thomas Bayes, which predicts future probabilities based on experience.

Bayes Theorem: $P(A/B) = \frac{P(B/A)*P(A)}{P(B)}$

This Theorem states that "the probability of A on condition that B is true equals to the probability of B providing A is true times the probability of A being true, divided by the probability of B being true."

In Naïve Bayes, the classification process is often divided into two phases: learning and testing. a number of the information that has been known for the information class is fed into the educational phase to construct an approximate model. Then within the test phase, the model that has been formed is tested with another data to see the accuracy of the model.

a) Working:

Let us consider,
x -> the text data or say a bunch of reviews
y -> the classes (let's say 0 for negative and 1 for positive)

$P\left(y=\frac{1}{x}\right) = \frac{p(x/y=1)*p(y=1)}{p(x)}$ for positive reviews

$P\left(y=\frac{0}{x}\right) = \frac{p(x/y=0)*p(y=0)}{p(x)}$ for negative reviews

Now we are attempting to see whether a review is positive or negative due **to** which the denominator will be ignored for the once

One of the foremost important assumptions we've considered in Naive-Bayes is termed the bag-of-words. It implies that everyone the algorithm cares about are that the word and its frequency i.e., what number times the word has appeared in our data. The position of the word in our sentence(document) doesn't matter in the least. We

only should keep track of the quantity of times a specific word appeared in our document. That's it.

b) Implementation:

Let's take an example- "Fabric is very good".

Initially, a word count is done.

words	count
fabric	1
is	1
very	1
good	1

This is a bag-of-words concept. It does not matter where the words have been used in the sentences, all that matters is its frequency.

If you may recall, our main goal was to find the class (whether positive or negative sentiment) given a particular sentence(document). So, we can approach this problem this way:

1. Suppose we have a set of possible classes C.

2. We find the probability of a document being in a particular class.So essentially conditional probability of a class given a document.

3. We iterate over all the classes and find which class has the maximum conditional probability, giving us our answer.

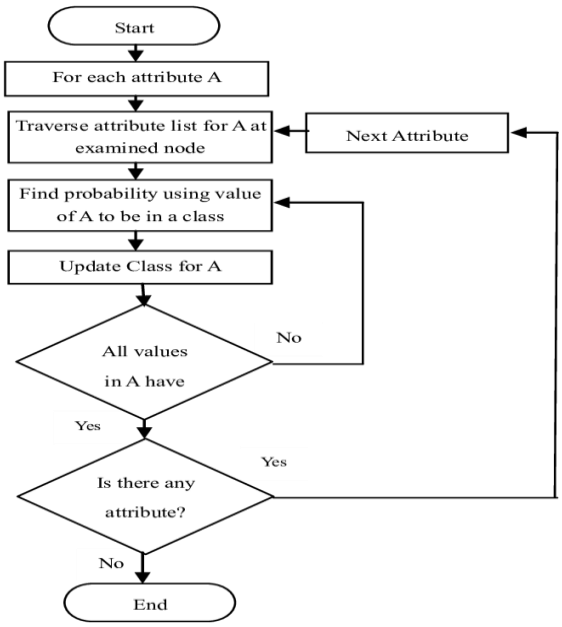

Fig 1: Basic System Architecture of Naïve Bayes Algorithm

2. SVM (Support Vector Machine)

SVM (support vector machine) is a Supervised learning method algorithm, which is used for classification worldwide. It separates new inputs with the help of its hyperplane and generates output.

a) Pre-processing:

Let us plot the n-dimensional space & take the value of each new feature as the value of a specific coordinate. Then, by differentiating those two classes we can find the ideal hyperplane. [Fig 2]

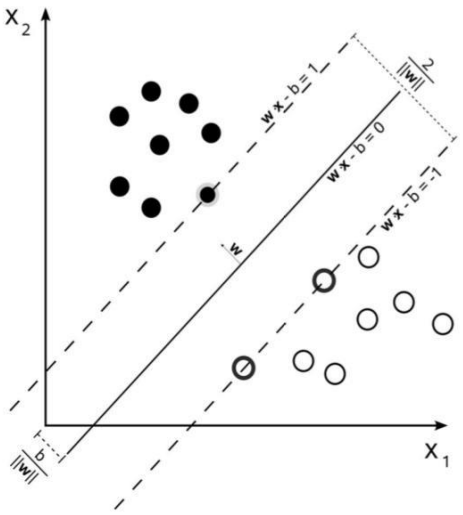

Fig 2: SVM model.

b) Working:

SVM set aside the hyperplane that maximizes the margin between two classes. [Fig 2]

$(\vec{x}.\vec{y}) + b = \sum_i yi\ (\vec{x}.\vec{y}) + b = 0$

$\vec{x}$ = n-dimensional input vector,

yi = its output values

$\vec{w}$ = weight of vector (the normal vector)

ai = Lagrangian multiple.

If $\vec{w}.\vec{x} + b \geq 0 \rightarrow$ positive class

Else → negative class

a) Implementation:

Steps for implementation of SVM -

Step 1: Import the important libraries which are required for the implementation of the project.

Step 2: Download the dataset and apply SVM to that dataset for further analysis.

Step 3: Dataset will get splitter into two parts, using one user-defined function Split ().

Step 4: Now we can visualize and observe data that one of the classes is linearly separable.

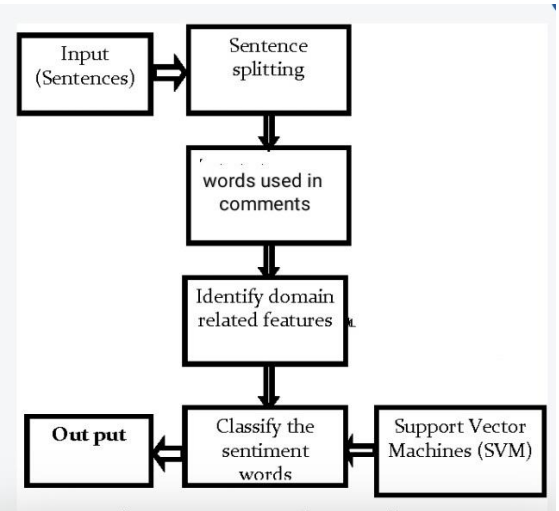

Fig 3: Basic System Architecture of SVM

SVM has been proved one of the most efficient learning algorithms for text Tokenization. This SVM algorithm can be used to perform an evaluation measure on comments or feedback received from customers.

3.CNN (Convolutional Neural Network):

CNN stands for "Convolutional Neural Network" and is largely utilized in image processing. The number of hidden layers employed between the input and output layers determines the CNN's strength. A set of features is extracted by each layer.

A series of filters are applied to the input to create feature maps. Each filter multiplies its weights by the input values after going over the full input. The result is sent to a Rectified Linear Unit (ReLU), sigmoid, or the activation function. The set of weights is evaluated using a loss function. The feature maps that the filters produce highlight various aspects of the input.

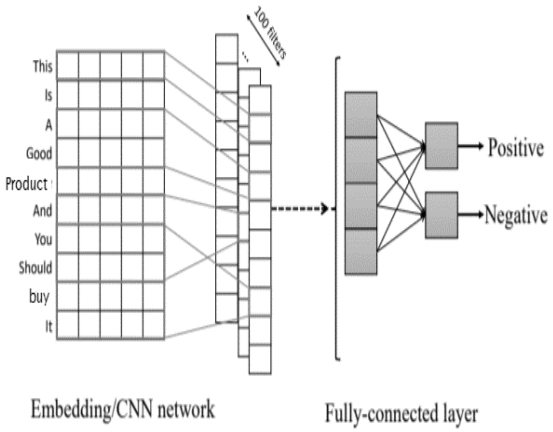

Fig 4: CNN Model

Even though CNN is commonly utilized in image and video processing, recent approaches to NLP utilize CNN. A pre-processing step in NLP converts the text input to a matrix representation. Sentence characters are used as rows, and alphabet letters are used as columns, in the matrix form. In NLP, a filter is slid across the matrix's words. As a result, the words are detected using the sliding window technique.

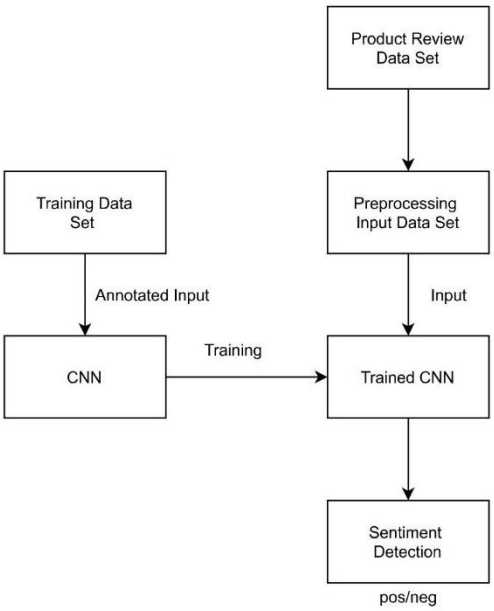

Fig 5: Basic Architecture of CNN

3. LITERATURE SURVEY

Sr. No.	Title	Methodology	Conclusion	Limitation
1.	Machine Learning-Based Customer Sentiment Analysis for Recommending Shoppers, Shops based on Customers Review	This paper is motivated towards applying Machine Learning Algorithm Hybrid Recommendation.	The performance of this Hybrid Recommendation System was evaluated using three metrics namely MAE, MSE, and MAPE.	The regression model used to implement the recommending system is complex and requires expertise.
2.	Experimental Investigation of Automated System for Twitter Sentiment Analysis to Predict the Public Emotion using MLA	This paper introduces an automated Twitter sentiment analysis system using MLA such as SVM, Naïve Bayes Classifier, MLP, Clustering	In this paper, the sentiment analysis of Twitter data by using different MLA is discussed.	The test case algorithm used in this paper is giving a contrast result with the existing algorithm.
3.	A Literature Review on Application of Sentiment Analysis using Machine Learning Techniques	Common techniques of analyzing sentiments such as Supervised Learning, Unsupervised Learning, and hybrid-based approach is used.	This paper summaries the techniques for machine learning used in the analysis of emotions in the latest period.	Does not determine accurate and less time feature extraction.
4.	The Impact of Preprocessing steps on the Accuracy of Machine learning Algorithms in Sentiment Analysis	In this paper, different MLAs and feature extraction techniques are used. These algorithms are NB, MaxE, and SVM.	We studied the impact of different preprocessing steps on the accuracy of three MLA for sentiment analysis.	accuracy of Maximum Entropy (MaxE) remains the same even after applying preprocessing steps.
5.	Multimodal Sentiment Analysis: Addressing Key Issues and Setting up the Baseline	For unimodal features extraction, the procedure by bc-LSTM is followed and for textual features extraction, CNN is followed.	The useful baseline for multi-model sentiment analysis and multimodal emotion recognition.	Does not include contextual dependency learning within the model.
6.	A Survey of Sentiment Analysis based on Transfer Learning	This paper focuses on the algorithms and application of transfer learning in sentiment analysis	This review makes a comprehensive investigation and discussion on sentiment analysis in a different direction.	Out of three transfer models Parameter transfer model is not easy to converge, within the Instance transfer model application scope is laid low with the similarity that exists between two domains. Within the Feature representation transfer model application scope is more extensive.
7.	The Essential of Sentiment Analysis and Opinion Mining in social media	It makes use of the supervised learning method and unsupervised learning method	Different methods using different approaches and algorithms and has been tested in a various dataset.	It is challenging to identify negation in the sentence, sarcasm, and others.
8.	Comparison of Naïve Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset	This research paper contains supervised learning which is under the machine learning approach.	In this paper, we use machine learning algorithms of Naïve Bayes and support vector machines for sentiment classification of airline reviews.	In this paper, the SVM algorithm is found to be more accurate than Naïve Bayes. Whereas in some cases exactly opposite is observed.
9.	Survey of Sentiment Analysis Using Deep Learning Techniques	Deep learning methods such as CNN, RNN, LSTM are used in sentiment analysis.	A detailed Survey of various deep learning architecture used in sentiment analysis at both sentence-level and aspect level is presented.	many deep learning algorithms were studies but no algorithm is found to capture the deep and implicit relation information.
10.	A Survey of Sentiment Analysis techniques	Features selection and sentiment classification is done using sentiment classification	This survey concluded that there is a lot of scope in sentiment analysis and it is found that SVM and Naïve Bayes algorithms are used for sentiment analysis.	deeper analysis is required when it comes to social networking sites and context consideration is very important.

3. CONCLUSIONS

The impact of performing data transformations on different classification algorithms is discussed in this work, although the type of transformation depends on the dataset and the language it contains.

Determining which categorization methods will yield the greatest results is indeed tough. Different strategies have been evaluated in various datasets using various methodologies and algorithms. The accuracy of Naïve Bayes and Support Vector Machines (SVM) was found to be comparable, whereas Convolutional Neural Network (CNN) was shown to be less accurate for textual sentiment analysis.

REFERENCES

[1] Shanshan Yi Xiaofang Liu (2020) "Machine learning-based customer sentiment analysis for recommending shoppers, shops based on customers' review"

[2] Priti Sharma, A.K. Sharm (2020) "Experimental Investigation of Automated System for Twitter Sentiment Analysis to Predict the Public Emotion using MLA"

[3] Anvar Shathik J., Krishna Prasad K. (2020) "A Literature Review on Application of Sentiment Analysis using Machine Learning Techniques"

[4] Saqib Alam, Nianmin Yao (2018) "The Impact of Preprocessing steps on the Accuracy of Machine learning Algorithms in Sentiment Analysis"

[5] Soujanya Poria, Navonil Mujumdar, Devamanyu Hazarika, Erik Cambria, Alexander Gelbukh, Amir Hussain (2018) "Multimodal Sentiment Analysis: Addressing Key Issues and Setting up the Baseline"

[6] Ruijun Liu, Yuqian Shi, Changjiang Ji, Ming Jia (2019) "A Survey of Sentiment Analysis based on Transfer Learning"

[7] Hong-Cheng Soong, Norazira Binti A Jalil, Ramesh Kumar Ayyasamy (2019) "The Essential of Sentiment Analysis and Opinion Mining in Social Media"

[8] Abdul Mohaimin Rahat, Abdul Kahir, Abu Kaisar Mohammad Masum (2019) "Comparison of Naïve Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset"

[9] Indhra om Prabha M, G. Umarani Srikanth (2020) "Survey of Sentiment Analysis Using Deep Learning Techniques"

[10] Harpreet Kaur, Veenu Mangat, Nidhi (2017) "A Survey of Sentiment Analysis techniques"