**MINOR PROJECT REPORT**

Submitted in partial fulfillment for the award of the degree of

**BACHELOR OF TECHNOLOGY**

**(Department of Computer Science and Engineering)**

Submitted to

**INDIAN INSTITUTE OF INFORMATION TECHNOLOGY**

**BHOPAL (M.P.)**



**Submitted by**

SANSKAR ASATI (20U02010)

HARSH VARDHAN BADERIYA (20U02055)

**Under the supervision of**

Dr. GAGAN VISHWAKARMA

(Department of Computer Science and Engineering, IIIT Bhopal)

Dr. YATENDRA SAHU

(Department of Computer Science and Engineering, IIIT Bhopal)

|  |  |
| --- | --- |
|  | **INDIAN INSTITUTE OF**  **INFORMATION TECHNOLOGY**  **BHOPAL (M.P.)** |

**Certificate**

This is to certify that the minor project report entitled “**COMPARITIVE ANALYSIS OF VARIOUS**

**ML AND DEEP LEARNING ALGORITMS FOR IMDB SENTIMENT ANALYSIS**” is

submitted by **SANSKAR ASATI** and **HARSH VARDHAN BADERIYA** of **INDIAN INSTITUTE**

**OF INFORMATION TECHNOLOGY** in fulfilment of the requirements for the degree of **Bachelor**

**of Technology in Department of Computer Science and Engineering**. This project is an authentic

work done by them under my supervision and guidance.

This project has not been submitted to any other institution for the award of any degree.

Date: 19th April 2023

**Minor Project Supervisor**

Dr. GAGAN VISHWAKARMA

Department of Computer Science and Engineering, IIIT Bhopal

**Minor Project Co-ordinator**

Dr. YATENDRA SAHU

Department of Computer Science and Engineering, IIIT Bhopal

|  |  |
| --- | --- |
|  | **INDIAN INSTITUTE OF**  **INFORMATION TECHNOLOGY**  **BHOPAL (M.P.)** |

**Student Declaration**

I hereby declare, that the work presented in the project report entitled **“COMPARITIVE ANALYSIS OF VARIOUS ML AND DEEP LEARNING ALGORITMS FOR IMDB SENTIMENT ANALYSIS”** in partial fulfilment of the requirement for the award of degree of **“Bachelor of Engineering”** from **INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, BHOPAL** is record of my own work.

I, with this, declare that the facts mentioned above are true to the best of our knowledge. In case of any unlikely discrepancy that may occur, we will be the ones to take responsibility.

Date: 19th April 2023

Place: Bhopal

HARSH VARDHAN BADERIYA

20U02055

SANSKAR ASATI

20U02010

|  |  |
| --- | --- |
|  | **INDIAN INSTITUTE OF**  **INFORMATION TECHNOLOGY**  **BHOPAL (M.P.)** |

**Acknowledgement**

With immense pleasure I, Mr/Miss **SANSKAR ASATI** presenting “**COMPARITIVE ANALYSIS OF**

**VARIOUS ML AND DEEP LEARNING ALGORITMS FOR IMDB SENTIMENT ANALYSIS**”

minor project report as a part of curriculum of “**Bachelor of Engineering**”. I wish to thank all the

people who gave me unending support.

I express my profound thanks to my project supervisor “**Dr. GAGAN VISHWAKARMA**”, coordinator “**Dr. Yatendra Sahu**” and all those who have indirectly guided and helped me in preparation of the report.

( SANSKAR ASATI: 20U02010

HARSH BADERIYA:20U02055)

|  |  |
| --- | --- |
|  | **INDIAN INSTITUTE OF**  **INFORMATION TECHNOLOGY**  **BHOPAL (M.P.)** |

**Acknowledgement**

With immense pleasure I, Mr **HARSH VARDHAN BADERIYA** presenting “**COMPARITIVE**

**ANALYSIS OF VARIOUS ML AND DEEP LEARNING ALGORITMS FOR IMDB**

**SENTIMENT ANALYSIS**” minor project report as a part of curriculum of “**Bachelor of**

**Engineering**”. I wish to thank all the people who gave me unending support.

I express my profound thanks to my project supervisor “**Dr. GAGAN VISHWAKARMA**”, coordinator “**Dr. Yatendra Sahu**” and all those who have indirectly guided and helped me in preparation of the report.

( SANSKAR ASATI: 20U02010

HARSH BADERIYA:20U02055)

**Contents**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Particulars** | **Page. No.** |
| 1. | Abstract | 1 |
| 2. | Introduction | 1 |
| 3. | Related work | 3 |
| 4. | Background Knowledge | 5 |
| 5. | Methodology | 6 |
| 6. | Implementation | 7 |
| 7. | Performance Evaluation | 14 |
| 8. | Results and Analysis | 15 |
| 9. | Comparisons | 20 |
| 10. | Conclusion and Future Works | 23 |
| 11. | References | 24 |

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Figure** | **Page No.** |
| 1. | Methodology | 6 |
| 2. | Sentiments of Dataset | 7 |
| 3. | Pre-Processing | 8 |
| 4. | Confusion Matrix for Logistic Regression | 16 |
| 5. | Confusion Matrix for SVM | 16 |
| 6. | Confusion Matrix for MNB | 16 |
| 7. | Confusion Matrix for SNN | 17 |
| 8. | Confusion Matrix for CNN | 17 |
| 9. | Confusion Matrix for CNN-OP | 17 |
| 10. | Confusion Matrix for LSTM | 17 |
| 11. | Confusion Matrix for BI-LSTM | 18 |
| 12. | Confusion Matrix for CNN-LSTM | 18 |
| 13. | Confusion Matrix for CNN-BILSTM | 18 |
| 14. | Confusion Matrix for HYBRID(HNN) | 18 |
| 15. | F1 Comparison for ML | 20 |
| 16. | F1 Comparison for Deep Learning | 21 |
| 17. | Accuracy and Loss graphs for Deep Learning | 19 |

**List of Tables**

|  |  |  |
| --- | --- | --- |
| S.No. | Tables | Page No. |
| 1. | Accuracy for ML(BOW) | 20 |
| 2. | Accuracy for ML(TF-IDF) | 20 |
| 3. | Accuracy for Deep Learning | 21 |

**COMPARITIVE ANALYSIS OF VARIOUS ML AND DEEP LEARNING ALGORITMS FOR IMDB SENTIMENT ANALYSIS**

**Abstract—Due to the sheer volume of opinion-rich web resources, a lot of current researches are focusing on sentiment analysis. Sentiment analysis is the study, to classify the text based on customer reviews which can provide valuable information to improve business. Previously the analysis was carried out based on the information provided by the customers using natural language processing and machine learning. The goal is to create new models that are able to recognize and classify the opinions or sentiments expressed in an electronic text to evaluate and improve a given system.**

**ML algorithms are the traditional algorithms that work in a single layer while deep learning algorithms work on multilayers and gives better output. Currently, Deep Learning is a rapidly growing field, it has proven its effectiveness to solve many complex problems due to its ability to learn and extract meaningful information from data.**

**In this paper, sentiment analysis on IMDB movie reviews dataset is implemented using Machine Learning (ML) and Deep Learning (DL) approaches to measure the accuracy of the model. This paper helps the researchers to identify the best algorithm for sentiment analysis. The comparison of the machine learning and deep learning approaches shows that DL algorithms provide accurate and efficient results.**

**Keywords— Sentiment Analysis, Machine Learning, NLP, Deep Learning, IMDB reviews, Prediction.**

**I. INTRODUCTION**

A movie is a spectacle that can be done at a relaxed time. Currently, there are many movies that can be watched via the internet or cinema. Movies that are watched on the internet are sometimes charged to watch so that potential viewers before watching a movie will read comments from users who have watched the movie. The website that is often used to view movie comments today is IMDB. On this website, one can search for the film they want to watch by reading the comments first to determine which film to watch based on the most positive or negative comments. Movie comments are many and varied on the IMDB website, you can see comments based on the star rating aspect. This causes users to have difficulty analysing other users' comments

Sentiment analysis [1] is a perspective, thought, or judgment of a specific feeling. Most of the analysis is done based on online reviews. It is troublesome to analyse using forum discussions because most of the information provided is irrelevant. To overcome this, aspect-based sentiment analysis can be used as it is a text analysis technique used to categorize data and identify the sentiment attributed to specific review.

Aspect-based sentiment [2] analysis can be used to analyse customer feedback through sentiments with different aspects of a product or service. By using this analysis, the sentiment of the reviews can be identified easily.

Natural language [3] explanations have become challenging due to the complexity of human languages. The main reason is that textual data doesn't have a proper structure. There is a need to parse the data as it helps the machine to understand and utilize the data. Sentiment analysis helps companies to analyse large amounts of data within less time. Sentiment analysis helps to analyse and prepare strategies based on customer reviews. Reviews are considered as short text by which the opinion of a person can be identified.

As everything is digitalized [4], people check blogs for reviews before they watch a movie or purchase a product. These reviews drive people to theatres and customers towards products. This analysis of sentiment helps to implement a strategic plan and attract more customers.

As per the previous work [5], sentiment analysis implementation was through NLP and machine learning algorithm. Sentiment analysis of movie reviews includes pre-processing and implementation of modelling techniques. These modelling techniques are used to identify the reviewer sentiment that is categorized as either positive or negative. Model accuracy can also be known by machine learning algorithm which provides the accuracy percentage of the sentiment identified on a particular review dataset.

In this paper, sentiment analysis [6] is done using ML and Deep learning approach on the considered IMDB review dataset. The text reviews categorised as negative and positive reviews.

Machine Learning (ML) techniques, which learn with experiences to discover knowledge or make decisions (predictions), without being humanly guided or explicitly programmed to handle a particular data. Indeed, the learning process is usually supervised which begins with labelled words, sentences, or phrases to find the best model which can be able to make the best decisions in the future.

Deep Learning (DL)[1] is the emerging technique of Machine Learning. Its basic concepts and models have been derived from the Artificial Neural Network which mimics the activity of the nervous system of the human brain to intelligent algorithms and avoids tedious human labour. Currently, Deep Learning is extremely active in the sentiment analysis area.[7][8] In fact, it is the most adapted facing to the rapid growth of the Internet and websites containing user reviews. Deep learning [4] is used to reduce human intervention and the issues of big data are dealt effectively to provide a better model for handling the data efficiently.

Deep learning can also be used in NLP to analyse the customer's point of view. [9] In this paper, deep learning and machine learning approaches are used to analyse the sentiment of the movie reviews and to identify whether the reviews are negative or positive. These reviews can be utilized as a tool for recommending movies. Filmmakers can also utilize this information to make marketing decisions and attract customers.

The rest of this paper is organized into nine main parts. Section II presents some related works. Section III describes the background knowledge. Next up Section IV describes the methodology of our project while the Section V describes its implementation along withs its performance evaluation in Section VI. Results and analysis are shown by Section VII and their comparison by Section VIII. Section IX concludes the project and is rounded up by Section X detailing the references.

**II. RELATED WORK**

Sentiment Analysis is based on feature extraction and sentiment classification. It is achieved by applying the statistical approach using machine learning [2]. IMDB reviews were analysed using sentiment analysis and show reviews by n-gram approaches.

Different classifiers are used to train the model and the unigram approach is performed well in comparison to others. The n-gram approach of experimentation was finished and tried to produce the simplest results. Similar work was done by Tripathy et al. [10], where TF, TF-IDF was used for the conversion of the text file to a numerical vector. Experimentation was done with n-gram approaches and its combination are tried to get the best results.

Not only the word features, but special symbols present with words can also be considered as features [11]. The emotions with the word features can be utilized. Classifier ensemble is used to classify the results obtained by different ML algorithms and produce a good result.

Research is done to extract features that include the parts of the speech by using a tagger. The unigram model is used to extract adjectives that describe the positive or negative sentence. Emotion recognition had always been attracted attention in multiple fields that include NL processing, psychology [12].

Aspect-based sentiment analysis requires two conceptual tasks, syntactical information to explore the grammatical methods and address this problem by utilizing the effective encoding of the syntax.

Analysing the sentiment of the customers is essential as a customer plays an important role to improve business. The work of sentiment recognition mainly focused on semantic processing to understand the customer better and analyse through the reviews. This approach, shows the graphs, similarity measures, algorithms using graph theory to make the process simplified and easy by considering the texts that could promote to understand the language better so that the sentiment analysis tasks can be done easily.

Islam and Sultana [13] proposed a new system to classify the reviews of amazon and IMDB datasets. They used ngrams process to split the sentence into words and transform them in to numerical values. Then, they used six Machine Learning techniques to find review classes, these techniques are Naïve Bayes (Multinomial and Bernoulli), Logistic Regression, Stochastic Gradient Descent (SGD), Linear Support Vector Machine (SVM) and Random Forest (RF). The experiment results have shown that SGD, SVM and RF give the best results.

Preethi et.al.[14] introduced a new application (RNN) with deep learning system for sentiment analysis using pic reviews. Qian et.al.[15] projected a model to train with sentence-level annotation, they conjointly did an effort to come up with linguistic coherent representations of the model using regularizes.

Nguyen et.al.[5] Designed a model that predicts stock worth movement by mistreatment of the sentiment collected from social media. It shows the analysis of stock prediction task and the authors adopted the neural network methodology to demonstrate that ML methodology is

beneficial for finding out the variations and commonalities of various quantizing strategies of quantum correlation [16].

The study of Recursive Neural Tensor Network was accustomed to determine sentences based on the user sentiment, and they used a dataset that contains 11,855 movie reviews. The accuracy achieved by RNTN is 80.7%.

Socher et.al.[6] projected that a text cannot be analysed in isolation which states that sentences are closely associated with the words.

The invention [17] expressed that (RNN) model will subsume the short dependence during a sequence of information however, includes a drawback of gradient explosion as they will have a major impact on overall polarity to unravel this semipermanent dependence drawback, the LSTM model is proposed by Vu et al. [18]

Remote sensing applications are not fully engaged in the use of CNN. In order to address this issue, they introduced a novel CNN so that there is an increase in the performance of detectors [19]. The research study performs an analysis by considering a dataset obtained from online social media, where the detection is based on a ML algorithm. [4]

Mathapati et al. [20] tested and compared four Neural Network classifiers on 50,000 reviews of the IMDB dataset considering 5,000 topmost words. The proposed classifiers are Long Short-Term Memory (LSTM), Multilayer perceptron Neural Network (MNN), Convolution Neural Network (CNN), and a hybrid LSTM\_CNN network. The experiment results have shown that CNN is the most efficient technique with an accuracy of 89%.

Hassan and Mahmood [21] proposed a new neural language model ConvLstm, based on CNN and LSTM models. They exploit LSTM as a substitute for the pooling layer in CNN to reduce the loss of detailed local information and capture long-term dependencies in the sequence of sentences. The Pre-trained word vectors were obtained using the word2vect tool. This proposal was tested on two sentiment datasets IMDB and Stanford Sentiment Treebank (SSTb).

Yenter and Verma [22] presented and compared several combinations of CNN and LSTM models. Each combination is composed of parallel branches, which use the same layers but different parameters. Each branch is composed of Conv layer, ReLU activation layer, pooling layers, batch normalization layer, dropout layer, and finally LSTM layer. A concatenation layer is added to combine the output of each LSTM layer to find the sentiment class using the sigmoid function. This proposal was tested using the IMDB dataset of 25000 reviews.

Nahal et al. [23] developed a classification sentiment analysis approach using different deep learning networks namely: Multilayer Perceptron (MLP), LSTM recurrent neural network, CNN, and a hybrid CNN\_LSTM (where LSTM is the last layer of CNN). Initially, the data was pre-processed using word2vec and the word embedding processes then the finding vectors are sent to the four classifiers. The experimental results tested on IMDB dataset have shown that the hybrid CNN\_LSTM model outperformed the three other networks.

**III. BACKGROUND KNOWLEDGE**

**Natural language processing:** NLP is taken into account as a field of applied science and is additionally concerned with the interactions between machines and human languages. It helps to spot the sentiment of the reviewer and consists of many pre-processing techniques to convert the information into simpler text. So, that it will be easily understood.

**Machine Learning approaches:**

**Logistic Regression:** Logistic regression is a statistical method that is used for building machine learning models where the dependent variable is dichotomous: i.e., binary. Logistic regression is used to describe data and the relationship between one dependent variable and one or more independent variables. The independent variables can be nominal, ordinal, or of interval type.

**Naive Bayes:** Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes’ theorem with the “naive” assumption of conditional independence between every pair of features given the value of the class variable.

**SVM:** The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

**Deep Learning Approaches:**

**SNN:** A simple neural network, also known as a feedforward neural network or a multilayer perceptron (MLP), is a type of artificial neural network (ANN) that consists of multiple layers of interconnected nodes (neurons) that process input data in a feedforward manner. The nodes in the input layer receive the input data, and the nodes in the output layer generate the network's output predictions. The nodes in the hidden layers perform intermediate computations on the input data.

**CNN**: CNN often used for identifying objects inside images and for text classification by using word embeddings. It's been found effective for text in search query retrieval, sentence modelling, and NLP tasks.

**RNN**: RNN takes a sequence of information as input; recursive process is performed in the evolution direction of the sequence. It is the study on nonlinear characteristics of the sequence and has many advantages. RNN is applied on NLP, like speech recognition, language modelling, and other fields and it is a deep learning approach that may be used for sentiment analysis. It produces the output supported by previous computation and taking sequential information.

**LSTM (Long Short-Term Memory):** It is used to overcome the RNN problem for memorizing data for a longer time. LSTM works as a part of long-term dependence. This can be used for text classification and it produces long-term memorizing of the data compared to RNN. Thus, LSTM is also used for the implementation and analysis of the sentiment based on reviews.

**BI-LSTM:** BILSTM stands for Bidirectional Long Short-Term Memory, which is a variant of the LSTM neural network architecture that processes input data in both forward and backward directions. This allows the network to capture dependencies in both past and future input sequences, which can be particularly useful for sequence labelling tasks such as named entity recognition, speech recognition, and sentiment analysis.

**CNN-LSTM:** The CNN LSTM architecture involves using Convolutional Neural Network (CNN) layers for feature extraction on input data combined with LSTMs to support sequence prediction. A CNN-LSTM is a model architecture that has a CNN model for the input and an LSTM model to process input time steps processed by the CNN model.

**CNN-BILSTM:** A CNN BiLSTM is a hybrid bidirectional [LSTM](https://paperswithcode.com/method/lstm) and CNN architecture. In the original formulation applied to named entity recognition, it learns both character-level and word-level features. The CNN component is used to induce the character-level features.

**HNN:** A hybrid neural network is a type of neural network architecture that combines multiple types of neural network models to leverage their strengths and overcome their limitations. By combining multiple neural network models, a hybrid neural network can be customized to fit specific needs and perform well on a wider range of tasks.

**IV. METHODOLOGY**

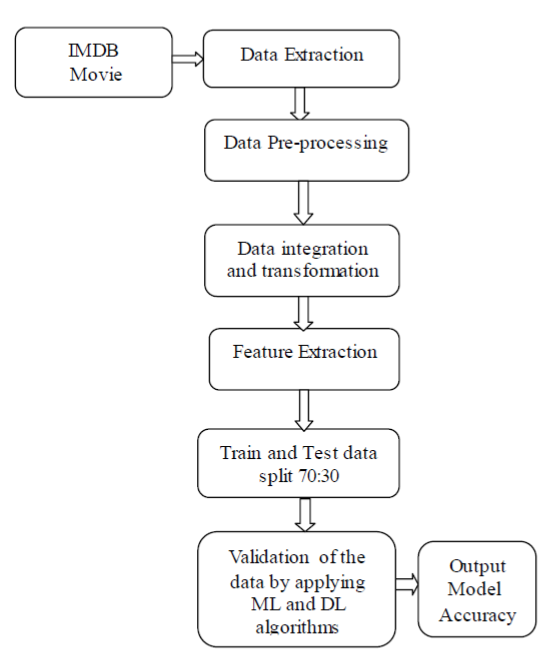


Figure 1 Methodology

The input data set is applied as a set of movie reviews and the expected output is the accuracy of the model. In this paper, both the ML and DL approaches are implemented and a comparison of these approaches is shown.

From the IMDB movie reviews database, the reviews along with its sentiment is extracted to create the base table for analysis. The reviews extracted are then processed to various steps of Data Cleaning (i.e., Translation into English language, removing HTML tags, Special Characters, Stopwords etc.). The Pre-processed text is now integrated with respective sentiment thus Transforming the table having both reviews and sentiments. The reviews are then flowed through for feature extraction using BOW and TFIDF. Tokenizer and Embedding Layer (by using GloVe word embeddings) also comes in place for sentiment prediction model.

Now the dataset is split into two parts: TRAINING and TESTING dataset. Training dataset constitute 70% of the total whereas the rest 30% comprises the Testing dataset.

The first implementation is performed on this dataset by applying ML algorithms for the prediction of accuracy on the model. It evaluates the accuracy on both BOW and TFIDF features. The second implementation is performed using deep learning techniques which resulted in better accuracy for sentiment analysis. To keep a check on both predicted and testing data, Validation Data is allotted 20% of the dataset for each epoch.

Under ML; Logistic Regression, SVM and Random Forest Classifier helps to build a model for prediction. Under Deep Learning; SNN, CNN, LSTM, BI-LSTM, and HNN models predicts the possibility of a sentiment to be Positive or Negative and also provide the expected rating for the same.

Output consists of Accuracy of the models along with Classification matrix, Confusion matrix and graphs for both Accuracy and Losses. Finally, there is a comparison among various models of both ML and DL on the basis of their ability to correctly predict the sentiment and approximately provide its rating.

**V. IMPLEMENTATION**

**Data set:**

To identify the better model that can be used for sentiment analysis, a public IMDB dataset that contains 50,000 reviews is considered, of which 25,000 are used for training and the remaining 25,000 are used for testing.

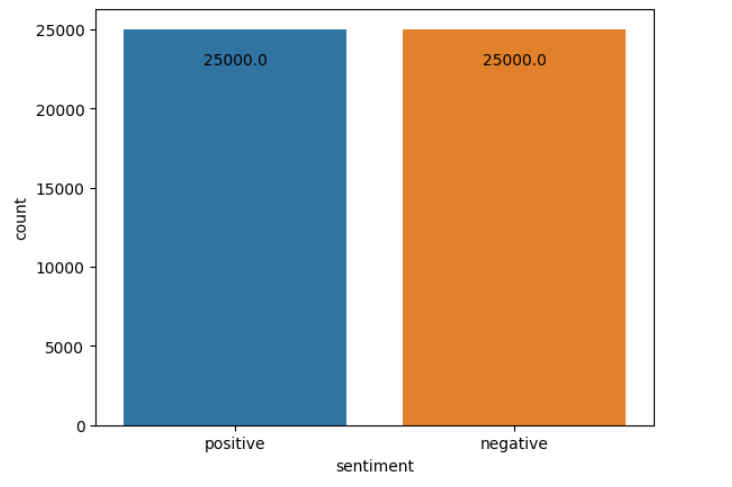


Figure 2 Sentiments of Dataset

**Pre- Processing Stage:** As the dataset is required to be clean to apply and create and train model. In this stage, it includes removal of attribute missing values, Standard Scalar, Min-Max Scalar has been applied to the dataset to clean the data and obtain required clean data.

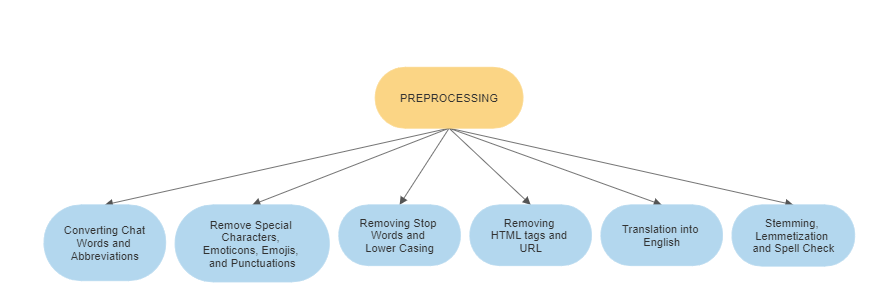


Figure 3 PRE-PROCESSING

Chat word conversion refers to the process of converting words or phrases commonly used in online chat or messaging platforms into their standard English language equivalents. For example, common chat words like "u" (you), "r" (are), "gr8" (great), "lol" (laugh out loud), "thx" (thanks), "btw" (by the way), etc., can be converted into their standard English language equivalents like "you", "are", "great", "laugh out loud", "thanks", "by the way", etc.

Converting abbreviations refers to the process of expanding or replacing abbreviations commonly used in text data with their full form or standard English language equivalents. For example, common abbreviations like "Mr." (Mister), "Dr." (Doctor), "e.g.," (for example), "i.e.," (that is), "etc." (et cetera), "w/ " (with), "w/o" (without), "yr" (year), etc., can be expanded or replaced with their full form equivalents like "Mister", "Doctor", "for example", "that is", "et cetera", "with", "without", "year", etc.

Removing punctuation marks: Punctuation marks such as commas, periods, question marks, etc., can be removed during pre-processing by using regular expressions, such as [^\w\s]. Removing newline characters: Newline characters (\n) can be removed during pre-processing by replacing them with a space or an empty string.

Converting text to lowercase refers to the process of changing all the characters in a text data to their lowercase form. This is a common step in text data pre-processing. Converting text to lowercase helps in standardizing the text data and reducing its complexity. It also makes the text data consistent and easier to handle for various NLP tasks. For example, the sentence "This is a Sample Text" can be converted to "this is a sample text" by converting all the uppercase characters to their lowercase form.

Remove any text inside square brackets, punctuation marks, newline characters, URLs, HTML tags, alphanumeric strings. Removing text inside square brackets: Square brackets are commonly used in text data to indicate additional information or annotations. However, they can be removed during pre-processing by using regular expressions, such as \[.\*?\].

Removing URLs: URLs can be removed during preprocessing by using regular expressions, such as https?://\S+. Removing HTML tags: HTML tags can be removed during preprocessing by using libraries such as Beautiful Soup or regular expressions, such as <.\*?>.

Removing Stopwords: Stop words are words that are commonly used in a language but do not carry significant meaning or contribute to the understanding of a sentence. Examples of stop words in English include "the," "and," "of," "to," etc. “This is a sample sentence, showing off the stop words filtration." ['This', 'is', 'a', 'sample', 'sentence', ',', 'showing', 'off', 'the', 'stop', 'words', 'filtration', '.'] After stop words removal: ['This', 'sample', 'sentence', ',', 'showing', 'stop', 'words', 'filtration', '.']

It is sometimes necessary to convert text in one language to another language, such as translating non-English text to English for analysis or processing.

Google Translate is a popular machine translation service that can be used to convert text from one language to another. Google provides an API that allows developers to integrate the translation service into their applications and programs.

Stemming refers to the process of reducing inflected (or derived) words to their base or root form, known as a stem. For example, the word "running" can be stemmed to its base form "run." Stemming is a common step in text data preprocessing that can help reduce the dimensionality of the text data and improve the efficiency.

Lemmatization is the process of reducing a word to its base or dictionary form, called the lemma, by applying morphological analysis to the word. For example, the words "am," "are," and "is" would all be reduced to the lemma "be." This process requires knowledge of the part of speech of the word and can be performed using various tools, such as the WordNet lemmatize.

Spelling check is an essential part of text preprocessing. Text data from social media platforms like Twitter often contain spelling errors, typos, and other forms of noise that can affect the accuracy of sentiment analysis. Therefore, it is necessary to identify and correct these errors before analysing the text data. Spelling check involves detecting and correcting misspelled words in a text document.

**Tokenization**: separates a piece of text into smaller components to count easily the number of words in the text and the frequency of the word. Tokenization is a way of separating a piece of text into smaller units called tokens. Here, tokens can be either words, characters, or subwords. Hence, tokenization can be broadly classified into 3 types – word, character, and subword (n-gram characters) tokenization.

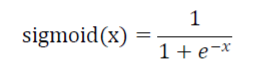
**Vectorization and embedding**: This process refers to a set of learning methods aiming at representing the words of a text as real-valued vectors. An embedding is a mapping of a discrete — categorical — variable to a vector of continuous numbers*.* [In the context of neural networks, embeddings](https://www.tensorflow.org/guide/embedding) are low-dimensional,learned continuous vector representations of discrete variables. Neural network embeddings are useful because they can reduce the dimensionalityofcategorical variables andmeaningfully representcategories in the transformed space. In our system, we have used GloVe Word embeddings.

**Classification phase:**

We used ReLU (Rectified Linear Unit) function as an activation layer because its gradient is simple to compute which allows the model to train easier, faster and perform better.



The function used to calculate the probability of the output classes is Sigmoid it is given by the following formula:



Where x is the output value calculated by the model.

− So, if sigmoid(x) ≈ 1, then the predicted sentiment is positive,

− Otherwise, sigmoid(x) ≈ 0, the predicted sentiment is negative.

The process is shown, firstly a Movie reviews data set is applied and then pre-processing of data that includes text normalization, removing noisy text, special characters, text stemming, removing stop words, word embedding etc. Then modelling techniques in NLP are applied that include count vectorizer and TF-IDF. The data here was split into train and test set after that ML algorithms were applied to check the accuracy of the model.



**Deep Learning approach:**

The **sequential model** allows us to specify a neural network, precisely, sequential: from input to output, passing through a series of neural layers, one after the other.

An **embedding layer** is a type of hidden layer in a neural network. In one sentence, this layer maps input information from a high-dimensional to a lower-dimensional space, allowing the network to learn more about the relationship between inputs and to process the data more efficiently.

The **Dropout layer** randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by 1/(1 - rate) such that the sum over all inputs is unchanged.

**Dense Layer** is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer.

**Flattening** is used to convert all the resultant 2-Dimensional arrays from pooled feature maps into a single long continuous linear vector.

**Conv1d** layer creates a convolution kernel that is convolved with the layer input over a single spatial (or temporal) dimension to produce a tensor of outputs.

**Max Pooling** is a pooling operation that calculates the maximum value for patches of a feature map, and uses it to create a down sampled (pooled) feature map.

**Batch normalization** applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1.

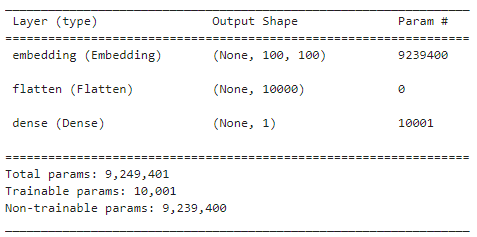
**GlobalMaxPooling1D** down samples the input representation by taking the maximum value over the time dimension.

An **LSTM layer** is an RNN layer that learns long-term dependencies between time steps in time series and sequence data. The layer performs additive interactions, which can help improve gradient flow over long sequences during training.

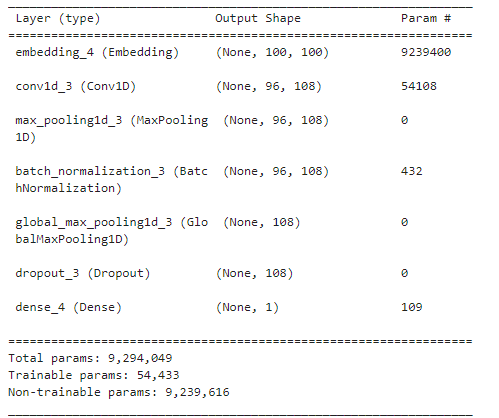
**Bidirectional** recurrent layers are defined as connecting two hidden layers of the opposite directions to same output. Because of this generative deep learning, output layer gets the information from past or backwards and the future or forward states simultaneously.

Following are the Deep Learning Models with their layered structure:

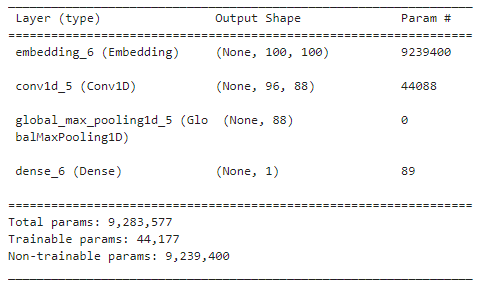
1. Simple Neural Network



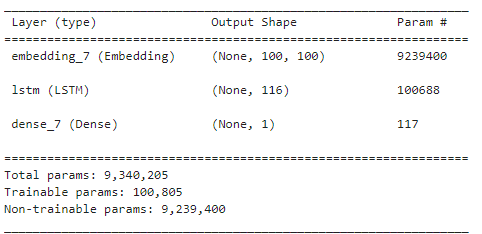
2. Convolutional Neural Network

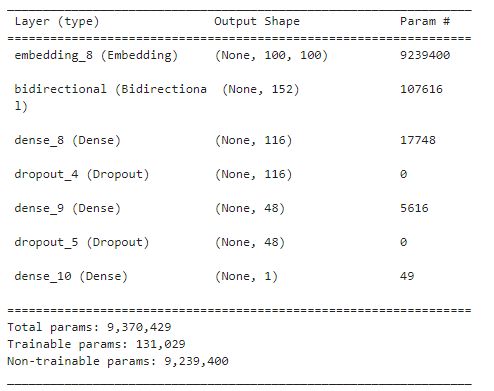


3. CNN-OPTIMISED

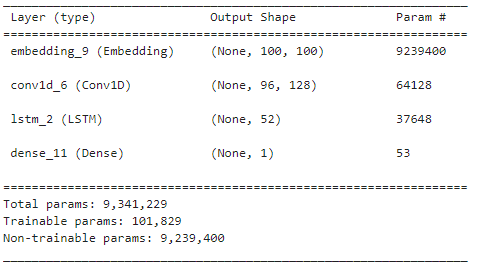


4. LSTM

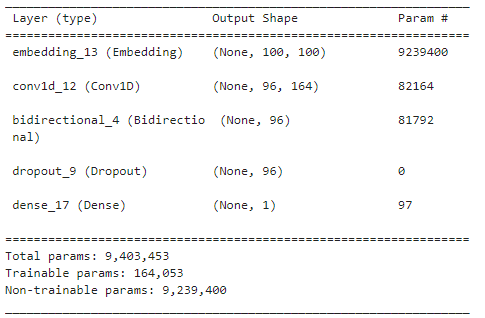


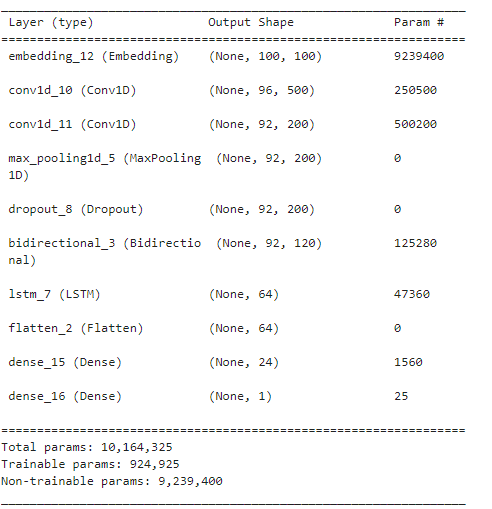
5. BILSTM

6. CNN-LSTM



7. CNN-BILSTM



8. HYBRID (HNN)

**VI. PERFORMANCE EVALUATION**

**Performance Metrics and its Evaluation:**

**Epochs**: An epoch is when all the training data is used at once and is defined as the total number of iterations of all the training data in one cycle for training the machine learning model.

**Batch size** is a term used in machine learning and refers to the number of training examples utilized in one iteration.

**Evaluation metrics** are used for classifiers to know the performance. Metrics for a model on a binary classification problem are listed with the equations:

**Confusion Matrix** is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. The matrix itself can be easily understood, but the related terminologies may be confusing. Since it shows the errors in the model performance in the form of a matrix, hence also known as an error matrix.

**Recall**: Recall should ideally be 1 (high) for a good classifier.  Recall becomes 1 only when the numerator and denominator are equal i.e., *TP = TP +FN*, this also means *FN* is zero. As *FN* increases the value of denominator becomes greater than the numerator and *recall* value decreases (which we don’t want).

RECALL=TP / (TP + FN)

**F1 Score**: F1-score is a metric which takes into account both precision and recall

F1=2 TP / (2 TP + FP + FN)

**Accuracy**: Accuracy represents the number of correctly classified data instances over the total number of data instances.

Accuracy= (TP + TN) (TP + TN + FP +FN) × 100

**Precision**: Precision should ideally be 1 (high) for a good classifier. Precision becomes 1 only when the numerator and denominator are equal i.e., TP = TP +FP, this also means *FP* is zero. As FP increases the value of denominator becomes greater than the numerator and precision value decreases (which we don’t want).

Precision= TN / (TP + FP)

**VII. RESULTS AND ANALYSIS**

In this approach ML and DL algorithms are applied to the IMDB movie reviews dataset to analyse positive and negative sentiment by detecting the emotion of the reviewer through text that includes some emotional key words which determine the emotion of the reviewer. Some positive emotions include “good”, “like”, “best”, “great” and negative emotions include “worst”, “sadly”, “disappointed”, “uncomfortable”, “bad”.

The different approaches are compared with Recall metrics, F1 Score metrics, Accuracy metrics and Precision metrics. The Accuracy comparison of ML and DL algorithms is also represented as Bar graph. The accuracy and loss of Deep Learning Algorithms are shown in graphical representation.

The first algorithm considered by us is logistic regression which is a classification technique that serves to solve the binary classification problem. Movie reviews were considered as positive and negative thus logistic regression can be used and was applied on both count vectorizer and TF-IDF to find out the accuracy score. The accuracy score obtained is around 74.9% for TF-IDF and 74.8% for the count vectorizer.

The second algorithm considered is the support vector machines and was applied on TF-IDF and count vectorizer to find out the accuracy score and the accuracy score obtained is 51.1% for TF-IDF and 57.4% for count vectorizer.

The third algorithm considered is Multinomial Naive bayes algorithm that can be used for text analysis and this technique is accustomed to find the possibilities of classes assigned to texts by considering the joint probabilities of the words and classes. This algorithm was also applied on both TF-IDF and count vectorizer to understand the accuracy score. The obtained accuracy is 74.9% in both cases.

The First Deep learning approach is SNN, by applying this model the accuracy achieved is 75.0%. The Second approach is the CNN model and the accuracy achieved is 83.8% and when optimised the accuracy is 84.4%.

Third approach is the LSTM model and the accuracy achieved by this model is 85.2% and fourth approach of Bi-LSTM gives the accuracy 85.3%.

Next up the compound models such as CNN-LSTM gives 84.5% accuracy, CNN-BILSTM gives 85.5% accuracy and finally the Hybrid model gives us 86.0% accuracy.

**LINEAR LOGISTIC MODEL**

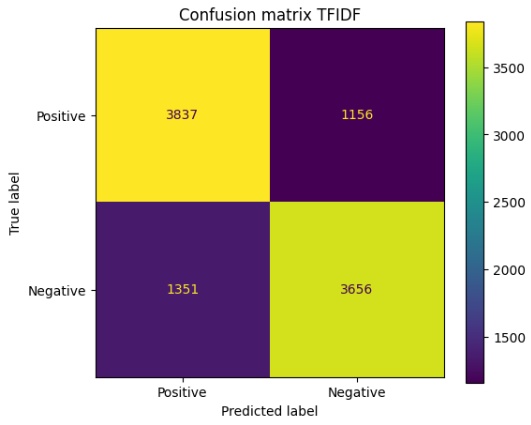
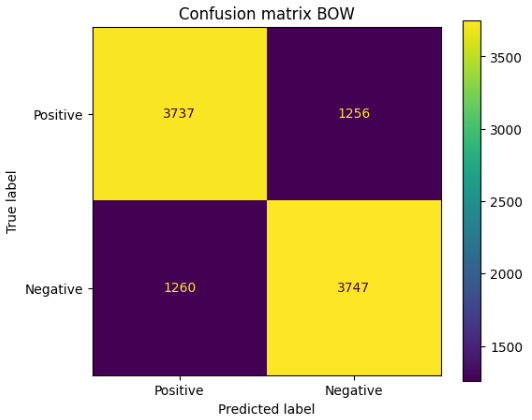
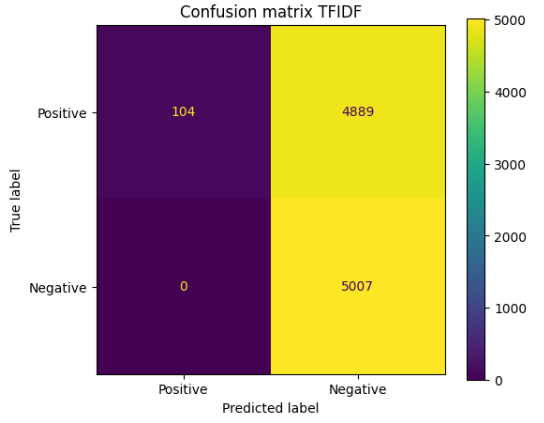
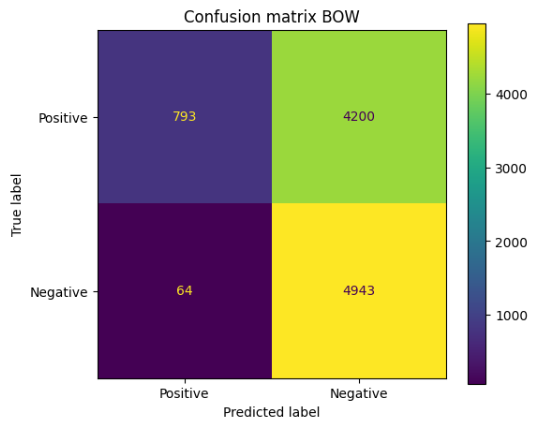


Figure 4 Confusion Matrix for Logistic Regression

**LINEAR SUPPORT VECTOR MACHINES**

Figure 5 Confusion Matrix for SVM



**MULTINOMIAL NAÏVE BAYES MODEL**

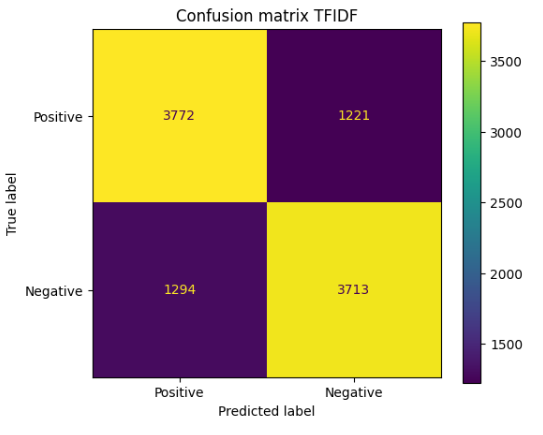
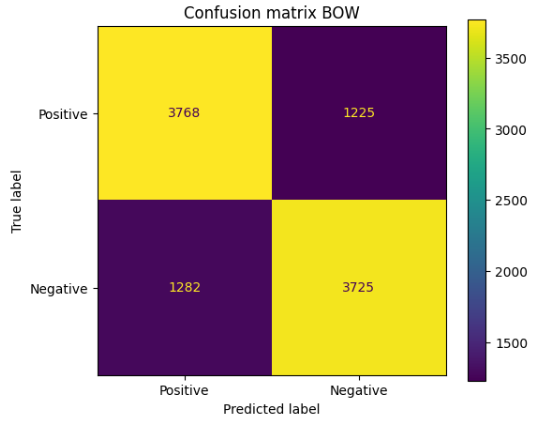


Figure 6 Confusion Matrix for MNB

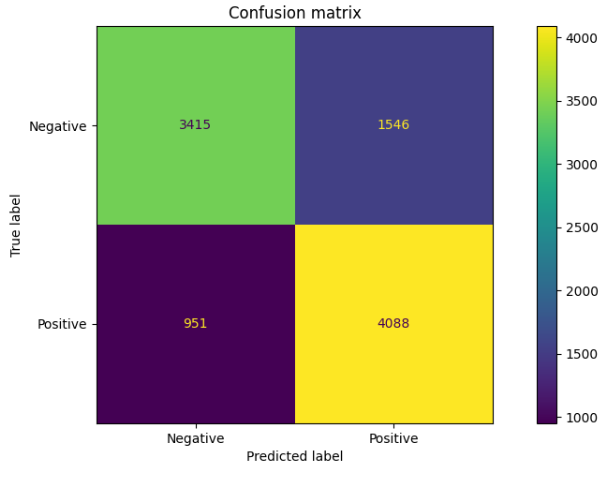
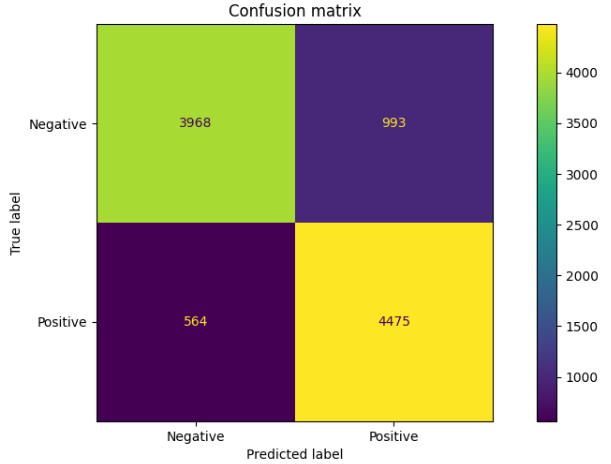
**SNN**

Figure 7 Confusion Matrix for SNN

****

**CNN**

Figure 8 Confusion Matrix for CNN

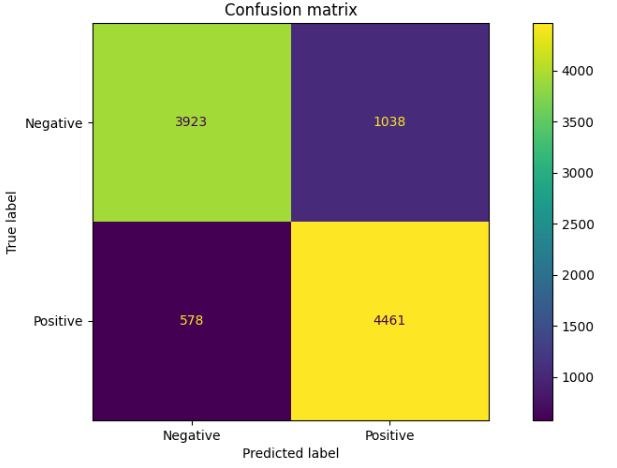
**CNN-OP**

Figure 9 Confusion Matrix for CNN-OP

**LSTM**

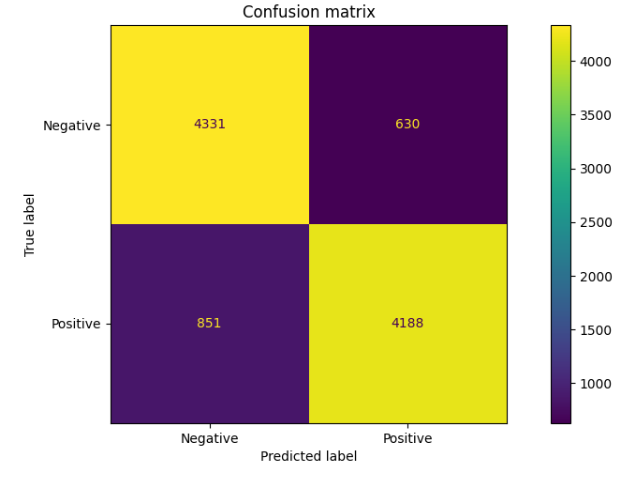
****

Figure 10 Confusion Matrix for LSTM

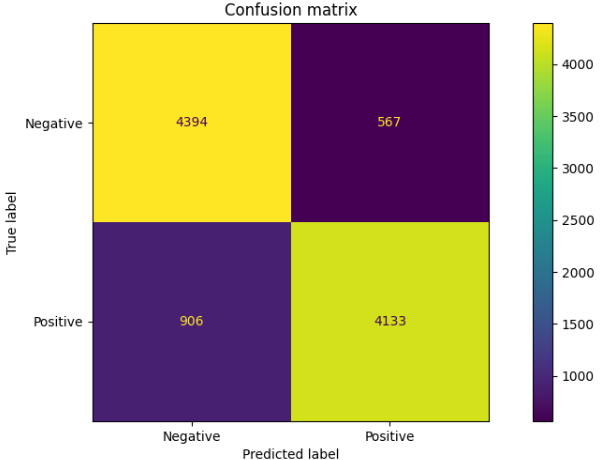
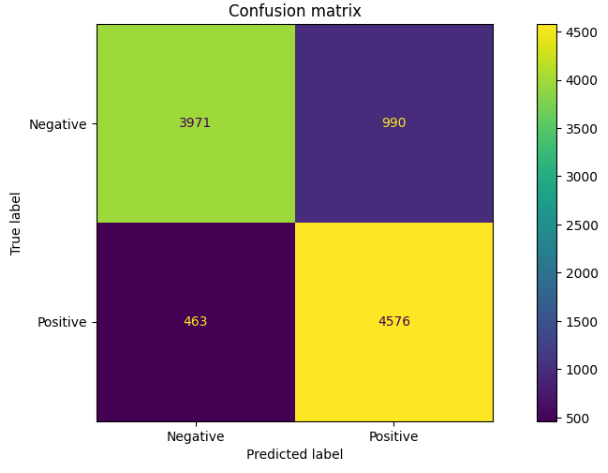
**Bi-LSTM**

Figure 11 Confusion Matrix for BI-LSTM

**CNN-LSTM**

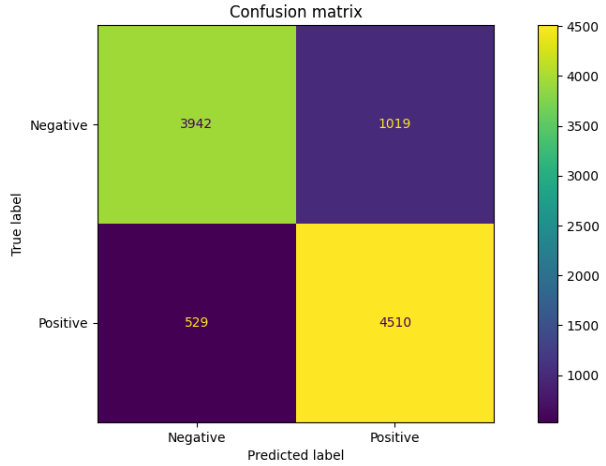
**CNN-BILSTM**

Figure 13 Confusion Matrix for CNN-BILSTM

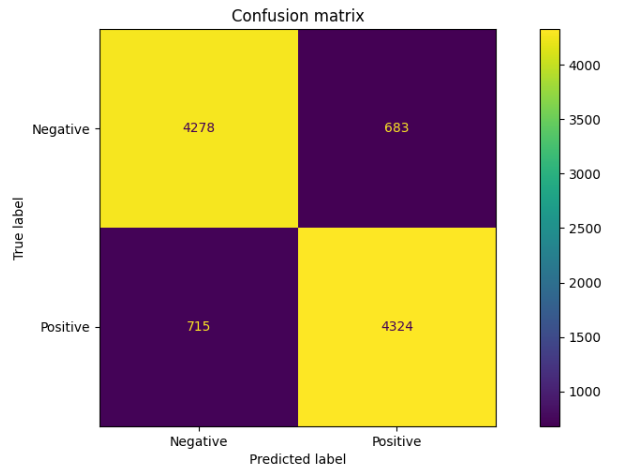
**HNN**

Figure 14 Confusion Matrix for HYBRID (HNN)

A higher accuracy value indicates that the model is making accurate predictions. However, accuracy may not be the best metric in situations where classes are imbalanced, as it can be misleading if the majority class dominates the dataset. It is important to consider other metrics like precision, recall, and F1-score in conjunction with accuracy for a comprehensive evaluation of model performance.

Figure 12 Confusion Matrix for CNN-LSTM

A higher precision value indicates that the model has fewer false positives, which means it is making fewer incorrect positive predictions.

A higher recall value indicates that the model has fewer false negatives, which means it is capturing a higher proportion of the actual positive instances.

A higher F1-score indicates that the model has a good balance between precision and recall, and it is commonly used when both false positives and false negatives are important considerations.

The confusion matrix also provides information about the misclassified instances, which includes both false positives and false negatives. False positives are instances that are predicted as positive but are actually negative, while false negatives are instances that are predicted as negative but are actually positive.

Analysing the misclassified instances can provide insights into the types of errors the model is making and help identify areas for improvement. For example, if the model is consistently misclassifying instances of a certain class, further investigation and model refinement may be needed for that specific class.

It is evident from the above results that both Logistic Regression and Multinomial Naïve Bayes are able to achieve to produce good results in comparison to SVM Machines. These both have a higher Accuracy score which eventually means higher F1 score and thus higher value of Recall and Precision.

In comparison to these ML algorithms, Deep learning models eventually possess higher value for these attributes. Among them Hybrid model has the highest F1 score. Following HNN is CNN-BILSTM, BI-LSTM and LSTM models. Worst among these models is the SNN model providing the least value for F1, Recall, Precision and thus Accuracy.

Deep learning algorithms tend to achieve higher accuracy, precision, and recall due to their ability to learn complex patterns and representations from large amounts of data. Interpretability of the confusion matrix is also different in machine learning and deep learning algorithms. Machine learning algorithms often have simpler models that are more interpretable, while deep learning algorithms are characterized by complex architectures that

may lack interpretability.

**VIII. COMPARISION**

**(MACHINE LEARNING MODELS)**

**BAG OF WORDS(BOW)**

|  |  |
| --- | --- |
| MODEL | ACCURACY |
| LR | 74.8 |
| SVM | 57.4 |
| MNB | 74.9 |

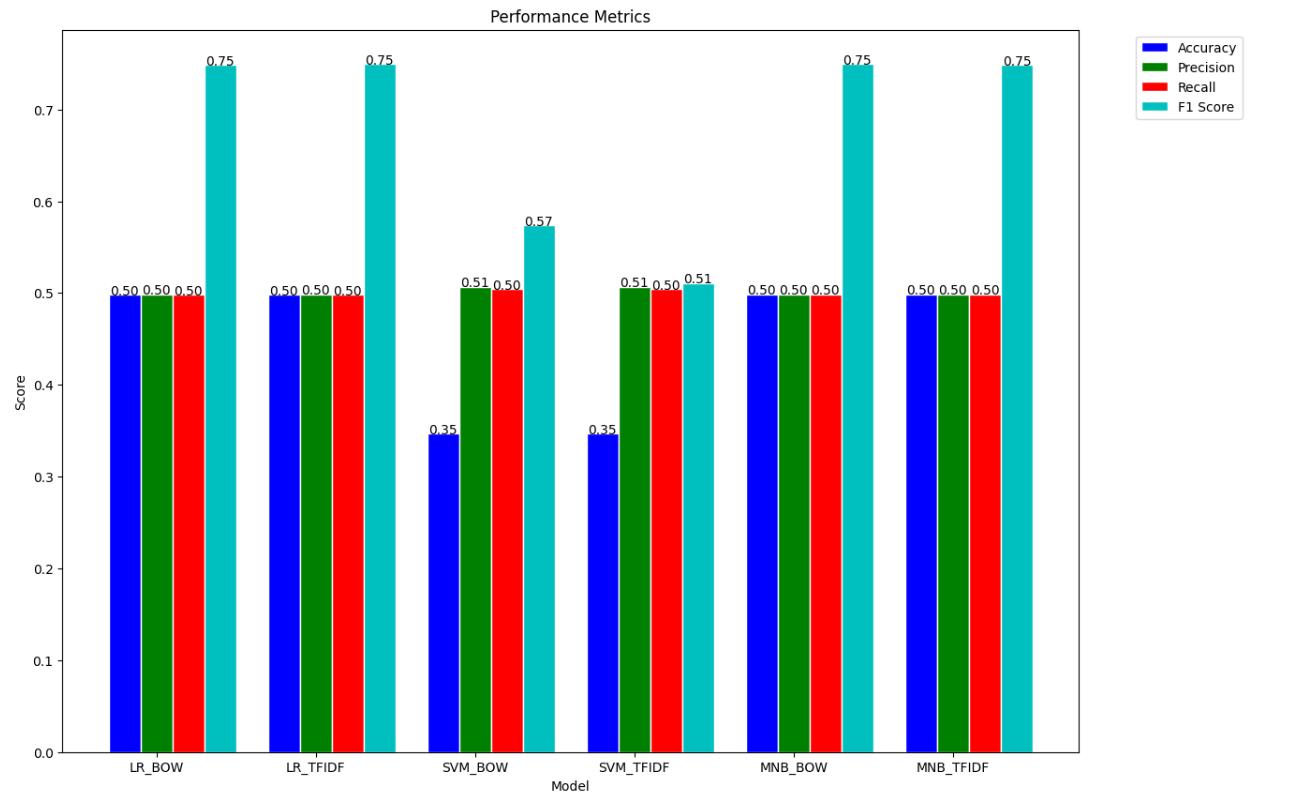
|  |  |
| --- | --- |
| MODEL | ACCURACY |
| LR | 74.9 |
| SVM | 51.1 |
| MNB | 74.9 |

**TF-IDF FEATURES**

*Table 2 Accuracy for ML(TF-IDF)*

*Table 1 Accuracy for ML(BOW)*

MNB and LR is very superior in predicting the sentiments and SVM is relatively worst in predicting. Both Accuracy and F1 scores for both BOW and TFIDF features are higher for MNB.

**F1, ACCURACY, RECALL AND PRECISION**

**DEEP LEARNING MODELS**

Figure 15 F1 Comparison for ML

Following depicts the comparison among different Deep Learning Models such as SNN, CNN, CNN-OP, LSTM, BILSTM, CNN-LSTM, CNN-BILSTM, HYBRID(HNN).

|  |  |
| --- | --- |
| MODEL | ACCURACY |
| SNN | 75.0 |
| CNN | 83.8 |
| CNN-OP | 84.4 |
| LSTM | 85.2 |
| Bi-LSTM | 85.3 |
| CNN-LSTM | 84.5 |
| CNN-BILSTM | 85.5 |
| HYBRID | 86.0 |

*Table 3 Accuracy for Deep Learning*

Hybrid model stands on top in achieving the highest Accuracy and F1 score. Next in line is the CNN-BiLSTM and LSTM models. The accuracy is still considerable for CNN-LSTM, CNN-OP and CNN model but the Accuracy and F1 score provided by SNN is highly not acceptable.

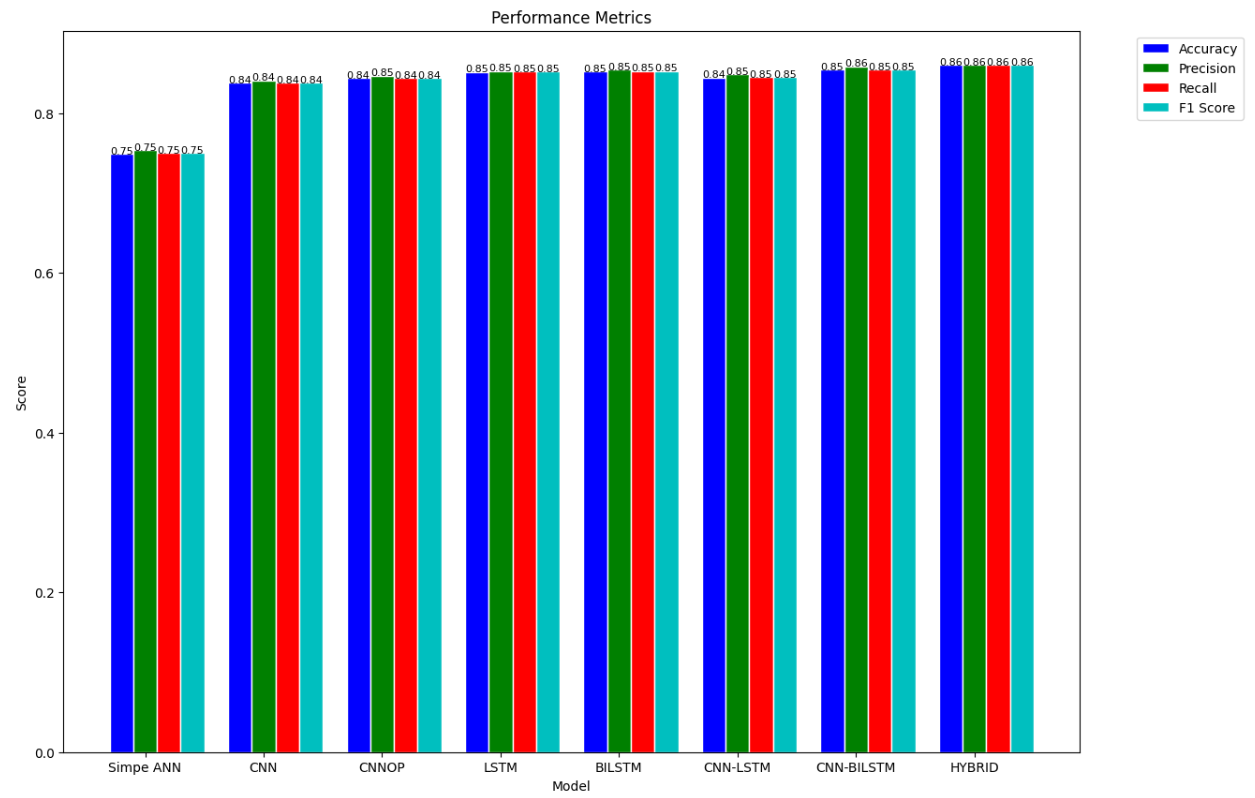
**F1, ACCURACY, RECALL AND PRECISION**

Figure 16 Comparison for Deep Learning

**GRAPHS**

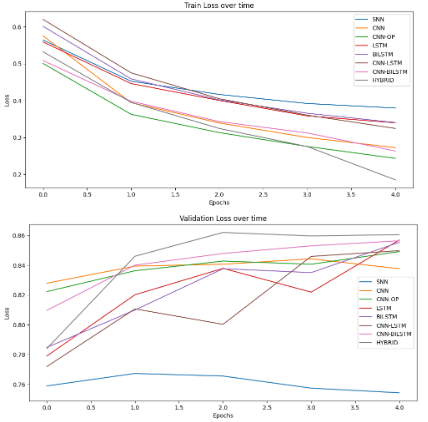
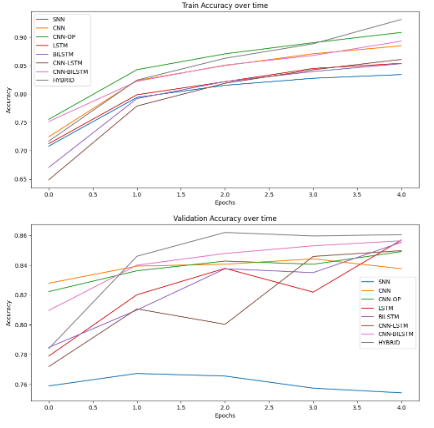


Figure 17 Accuracy and Loss Graphs for Deep Learning

These graphs provide insights into how the model is learning from the data over time and can help in understanding the behaviour of the model during training.

Training Accuracy vs. Epochs: This graph shows the accuracy of the model on the training dataset as the number of training epochs (iterations through the entire dataset) increases. It helps in understanding how well the model is learning from the training data over time. Generally, as the number of epochs increases, the training accuracy should improve, indicating that the model is learning to make better predictions on the training data.

Validation Accuracy vs. Epochs: This graph shows the accuracy of the model on a separate validation dataset (not used during training) as the number of training epochs increases. It helps in understanding how well the model is generalizing to new, unseen data. If the validation accuracy improves with increasing epochs, it indicates that the model is generalizing well and is likely to perform well on new data.

Training Loss vs. Epochs: This graph shows the loss (error) of the model on the training dataset as the number of training epochs increases. The loss is a measure of how well the model is fitting the training data. Lower values of loss indicate better model performance. Ideally, the training loss should decrease with increasing epochs, indicating that the model is fitting the training data better over time.

Validation Loss vs. Epochs: This graph shows the loss of the model on the validation dataset as the number of training epochs increases. It helps in understanding how well the model is generalizing to new, unseen data. If the validation loss decreases with increasing epochs, it indicates that the model is generalizing well and is likely to perform well on new data.

These graphs are useful for monitoring the progress of model training and can help in identifying issues such as overfitting (when training accuracy is high but validation accuracy is low) or underfitting (when both training and validation accuracies are low.

**IX.CONCLUSION AND FUTURE WORK**

Sentiment Analysis is a useful issue that needs to be evaluated more thoroughly. In this paper, the NLP approach using Machine Learning algorithms and the Deep Learning methods is used to classify reviews of the data set taken into positive and negative categories.

Comparison of ML and DL approaches is done by considering IMDB movie reviews. From the observations it is found that DL approaches provided accurate results than ML (Linear Logistic, Support Vector Machine, Multinomial Naïve Bayes and Random Forest Classifier) algorithms. Among the DL algorithms (SNN, CNN, CNN-OP, LSTM, BI-LSTM, CNN-LSTM, CNN-BILSTM and HYBRID(HNN)), HNN gives more accuracy of 86.0%.

SVM is very suitable for use in text data problems. When it comes to data analytics, neural networks are unrivalled. Long-term dependence on sentences that help in effective classification is discovered by LSTM. Convolutional Neural Networks can recognise objects in considerably less time and with greater accuracy. The Hybrid CNN-BiLSTM sentiment analysis was proposed in this research study.

The results obtained concludes that the Deep learning algorithms are far more impactful and effective in sentiment analysis than Machine Learning Algorithms. F1 score can be called to be the overall summary of the model, itself being the combination of both precision and recall and thus providing support to the overall accuracy of the model.

In future work, better models are hoped to be identified using deep learning to achieve better accuracy and to improve the effect of movie reviews by using sentiment analysis. Data pre-processing plays an important role in such large data sets. The aim is to identify better data pre-processing methods to achieve improved accuracy for movie review sentiment analysis. Also new customised Hybrid Models (Deep Learning) can be constructed to achieve higher accuracy.

**X. REFERENCES**

1.“Sentiment Analysis for IMDb Reviews Using Deep Learning Classifier” by Sara Sabba, Nahla Chekired, Hana Katab, Nassira Chekkai, Mohammed Chalbi.

2.B. P ang, L. Lee, and S. Vaithyanathan. “Thumbs up? Sentiment classification using machine learning techniques.” In Proceedings of the ACL-02 conference on Empirical methods in natural language processing Volume 10, Association for Computational Linguistics, pp. 79-86. 2002.

3.G. Gautam, and D. Yadav. “Sentiment analysis of twitter data using machine learning approaches and semantic analysis.” In Contemporary computing (IC3), 2014 seventh international conference on, pp. 437-442. IEEE, 2014.

4.Smys, S., and Jennifer S. Raj “Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study” Journal of t rends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 24-39.

5.T. H. Nguyen; K. Shirai; J. Velcin (2015): “Sentiment analysis on social media for stock movement predict ion”. Expert Systems with Applications, vol. 42, no. 24, pp. 9603-9611.

6.R. Socher; A. Perelygin; J. Wu; J. Chuang; C. D. Manning (2013): Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631-1642.

7.L. Zhang, S. Wang and B. Liu, “Deep learning for sentiment analysis: A survey,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, https://doi.org/10.1002/widm.1253.

8.D. Tang,B. Qin and T. Liu, “Deep learning for sentiment analysis: successful approaches and future challenges,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2015, <https://doi.org/10.1002/widm.1171>.

9.Kumar, T. Senthil “Construction of Hybrid Deep Learning Model for Predicting Children Behaviour based on their Emotional Reaction” Journal of Information Technology 3, no. 01 (2021): 29-43.

10.A. Tripathy, A. Agrawal, and S.K. Rat h. “Classification of sentiment reviews using n-gram machine learning approach.” Expert Systems with Applications, Vol. 57, pp. 117-126. 2016.

11.M. S. Mubarok, Adiwijaya, and M. D. Aldhi. “Aspect -based sentiment analysis to review product s using Naïve Bayes.” In AIP Conference Proceedings, vol. 1867, AIP Publishing, no. 1, pp 1-8.2017.

12.J.B. Delbrouck, N. Tits, M. Brousmiche, and S. Dupont, ‘‘A transformer based joint-encoding for emotion recognition and sentiment analysis”, in Proc. 2nd Grand-Challenge Workshop Multimodal Lang. (Challenge HML), 2020, pp. 1–7.

13.I.M. Mohaiminul et S. Naznin, “Comparative study on machine learning algorithms for sentiment classification,” International Journal of Computer Applications, vol. 182, no 21, pp. 1-7, 2018.

14.G. Preethi; Krishna, P. V.; Obaidat, M. S.; Saritha, V.; Yenduri, S. (2017): “Application of deep learning to sentiment analysis for recommender system on cloud”. International Conference on Computer, Information and Telecommunication Systems, pp.93–97.

15.Q. Qian.; M. Huang.; J. Lei; X. Zhu (2016): “Linguistically regularized LSTMs for sentiment classification. arXiv preprint arXiv:1611.03949. Sak, H.; Senior, A.; Beaufays, F. (2014): “Long short -term memory based recurrent neural network architectures for large vocabulary speech recognit ion”. arXiv preprint arXiv:1402.1128.

16.Q. Li, X.; Zhu; Q. Meng; You, C.; M. Zhu (2019): “Researching the link between the geometric and rènyi discord for special canonical initial states based on neural network method”. Computers, Materials Continua1, vol. 60, no. 3, pp. 1087-1095.

17.T. Lin;B.G. Horne; P.Tino; C.L.Giles(1996): Learning longterm dependencies in narx recurrent neural networks. IEEE Transactions on Neural Networks, vol.7, no. 6, pp. 1329-1338.

18.N.T. Vu; H. Adel; P. Gupta.; H. Schütze, (2016): “Combining recurrent and convolutional neural networks for relation classification”.arXiv preprint arXiv: 1605.07333.

19.P. Karuppusamy “Building Detection using Two-Layered Novel Convolutional Neural Networks” Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 29-37.

20.S. Mathapati, A. K. Adur, R. Tanuja, S. H. Manjula and K. R. Venugopal, “Collaborative Deep Learning Techniques for Sentiment Analysis on IMDb Dataset,” Tenth International Conference on Advanced Computing (ICoAC), pp. 361-366, 2018, doi: 10.1109/ICoAC44903.2018.8939068.

21.A. Hassan and A. Mahmood, “Deep Learning approach for sentiment analysis of short texts,” 3rd International Conference on Control, Automation and Robotics (ICCAR), pp. 705-710, 2017, doi:10.1109/ICCAR.2017.7942788.

22.A. Yenter and A. Verma, “Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis,” IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 540-546, 2017, doi:10.1109/UEMCON.2017.8249013.

23.A.N Mohamed et al., “Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models,”. International Journal of Data Mining & Knowledge Management Process (IJDKP) vol.9, no.2/3, May 2019, Available at SSRN: https://ssrn.com/abstract=3403985