International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

A Hybrid Model for Sentiment Analysis Based on Movie Review Datasets

Kirti Jain[1], Pinaki Ghosh[2], Shital Gupta[3]
Associate Professor, School of Advanced Computing, SAGE University Bhopal, Madhya Pradesh, India[1]
Professor, School of Advanced Computing, SAGE University Bhopal, Madhya Pradesh, India[2]
Assistant Professor, Computer science & Engineering Department,Bansal Institute of Science and Technology,Bhopal[3]

Abstract—The classification of sentiments, often known as sentiment analysis, is now widely recognized as an open field of research. Over the past few years, a huge amount of study work has been carried out in these disciplines by utilizing a wide variety of research approaches. Due to the possibility that the performance of sentiment analysis may be impacted by the high-dimensional feature set, text mining demands careful consideration during in the construction and selection of features.The process of recognising and extracting subjective information from written data is referred to as sentiment analysis. Sentiment analysis enables companies to understand the social sentiment around their brand, product, or service by monitoring the conversations that take place in internet chat rooms. In order to categorise people's attitudes or sentiments, this study provides a hybrid model (Support Vector Machine, Convolutional Neural Network, and Long Short-Term Memory). The findings of using the network model to sentiment analysis on the movie review or amazon review datasets reveal that it is possible to gain a good classification impact by using the model. The preprocessing is used for text mining, the removal of punctuation, and the generation of vocabulary, also uses GLOVE for vectorization and TF-IDF algorithms for better feature extraction. The results that were proposed were compared with various base models such as KNN, and MNB, amongst others, which demonstrates that the hybrid model performs better than other models.

Keywords- Text Classification, Sentiment Analysis,Natural Language Processing, TF-IDF Algorithm, GloVe, Deep Learning, Machine Learning.

I. INTRODUCTION

Text categorization has become important now because the amount of data being processed is growing at an exponential pace[1]. Reasonable data handling has become more important as the number of users has grown. Data should be uploaded and retrieved in the shortest amount of time possible for optimal handling. Many real-time applications, such as web apps, banking servers, scientific literature, and digital libraries, may make use of the uploaded and retrieved data. Filtering and organising large amounts of data based on their perceived usefulness is another function served by certain applications. The mining of public opinion is another crucial use case, alongside data organisation. Therefore, work has been done to broaden the categorised data available for sentiment analysis. Finally, text classification is also used to great effect in the realm of email classification, with the aim of sorting out spam emails [2]. As was mentioned before, there are obstacles in the applications that must be overcome critically and with exactitude.

The Sentiment Analysis feature of natural language processing (NLP) is used to determine the polarity of user- generated information. The growth of social media sites such as Facebook, etc. has led to a constant flow of information. In the world, according to Domo's estimation, there are approximately 2.4 billion social media users as of Information Never Sleep 2.0's release. Social media platforms like Twitter alone generate about 300,000 tweets every minute. Over 26,000 reviews have simultaneously been published to Yelp, an internet user review platform.Such a massive volume of semi-structured data presents a significant barrier in efficiently processing it for any given purpose. Due of its unstructured and loud character (e.g. gr8, g8, etc. for excellent) and spelling and grammatical problems, web-generated content sentiment analysis is a challenging problem. Many authors have suggested sentiment analyzers for Twitter data and/or online reviews [3], despite the difficulties that have been identified. Many studies, however, have only been conducted in languages with abundant resources, like English[4].

The solution we have presented is built on deep learning, which has proven its efficacy in a number of NLP issues, including sentiment analysis. Numerous modifications of its design have been developed by authors from all over the globe [5][6], and these variants have shown effectiveness in addressing issues pertaining to a variety of fields. The majority of these studies make use of a conventional method that consists of using softmax as an activation function on top of a hybrid deep neural network (DNN). On the other hand, in the context of our study, we train sentiment embedded vectors using DNN and then do final classification using a robust classifier[7].

The remaining parts of the paper may be summed up in the following manner: In the second part, We'll give a succinct examination of the state-of-the-art sentiment analysis research employing deep learning and machine learning techniques. After that, in the third part, we detailed the recommended approach in conjunction with the proposed model. In the fourth part of this article, we will discuss the outcomes of putting the suggested approach into practise. In conclusion, the study discusses more research that should be done.

II. LITERATURE SURVEY

To develop hybrid sentiment classification algorithms that can produce more accurate results, this study's goal is to do so. We previously looked at the approaches suggested and used in many other studies; these will now be covered in the following discussion.

There are a variety of approaches to creating hybrid models. The authors of the research [8] paired a CNN model with an SVM, which has the potential to increase the accuracy of image recognition. The extraction of features is handled by a layer of convolutional networks, while support vector machines serve as the recognizer. The original CNN is used in conjunction with SoftMax operations. Alsukhni[9] demonstrates the efficacy of using deep learning algorithms to Arabic texts' identification issue with so many categories. We have created two Python models utilising the Multi-Layer Perceptron (MLP) implementation and Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM). All of the data had to be cleaned up, which improved the quality of the experimental data. The LSTM model's test data result was 82.03, whereas the MLP model's test data result was 80.37. The F1 score was used to compare the two models.

Rajpoot et al. [10] have used sentiment analysis methods, such as Naive Bayes classifier, K-Nearest Neighbour, & Random Forest, on a dataset of movie reviews. In this research [11], The Doc2vec word embedding method was used to classify Turkish news texts from the Turkish Text Summarization 3600 (TTC-3600) dataset and English news texts from the BBC-News dataset. CNN, which is based on deep learning, is used in addition to more conventional classification machine learning algorithms such the Gauss Naive Bayes classifiers (GNB), Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM). The suggested model worked best using CNN classification, with a success rate of 96.41% just on English dataset and 94.17 percent on the Turkish dataset alone.

In this work [12], Traditional Machine Learning algorithms and Deep Learning models (Convolutional Neural Networks CNN and or Recurrent Neural Networks RNN) have been applied to a corpus gathered from social media platforms (Facebook, YouTube, and Twitter) about the Hirak 19 in order to analyse sentiments and gain a deeper understanding of opinions (a popular protest in Algeria during 2019).The corpus discusses the 2019 public protest in Algeria. The corpus, which contains 7800 remarks about various political Hirak issues, was created using a variety of materials published in dialectal Arabic. In this work, CNN and RNN exhibit their capability for recognising and categorising views relevant to a societal problem as well as doing sentiment analysis. A wide range of applications have used CNN and RNN. The cross-validation tests' accuracy methodology demonstrates the validity of the results. It demonstrates that 63.28% (CNN) & 60.97% are the accuracy schemes (RNN).

Srinidhi et al. [13] created a hybrid model for textual categorization of positive and negative emotions that includes LSTM, SVM, and even a kernels with such a radial base function. The hybrid approach was examined and evaluated using datasets that included IMDb movie reviews.These models are constructed by combining individual deep learning models with SVM so that they may be classified. Image recognition is an important application for several of them. In their study, they integrate two different types of deep learning models, and then classify the data using SVM or ReLU.For the purpose of this study, Parveen et al. [14]select movie reviews from the corpora community. Corpora have a lot of datasets inside them, and here is where they choose movies to evaluate. This dataset has two categories: positive and negative and each category comprises 1000 files. They employ a variety of classification techniques, including Naive Bayes, SkLearn, & Support Vector Machine. They employed many different classifiers, such as MultinomialNB, GaussianNB, BernoulliNB, and numerous others. We employ specific features to better understand the accuracy level of different algorithms (in count or numbers). The authors used an approach known as ensemble learning, combining various classification algorithms and utilising a try-and-vote strategy to decide which algorithm offered the maximum accuracy.

The work [15] gives a machine learning solution to the sentiment analysis problem. The collection contains data that was collected from numerous different travel review websites. In this study, they examined the effectiveness of Count Vectorization with TF-IDF Vectorization, two alternative feature extraction methodologies. The Naïve Bayes classifier (NB), Support Vector Machines (SVM), & Random Forest are more examples of classifiers (RF). The range of data that the algorithms have evaluated includes accuracy, recall, precision, or f1-score, among other metrics.For the dataset under evaluation, they discovered that the accuracy of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

classification method was higher when the TFIDF Vectorization feature extraction approach was used instead of the Count Vectorization algorithm. TF-IDF Vectorization + RF has the greatest accuracy (86%) in sentiment categorization of tourist site reviews (as measured by a standard dataset).

In this study, Tariyalet al. [16] use a number of machine learning approaches to categorise the reviews; we, in turn, develop a number of classification models Identify the most promising ones, evaluate their relative efficacy, & select them. They will combine straightforward linear methods (LDA), nonlinear methods (CART, KNN), and complex nonlinear approaches (SVM, RF, C5.0).

The purpose of this research [17] is to find a way to mechanically categorise user opinions on mobile apps. In this investigation, a machine earning based automatic classification method was used. Some of the attributes gleaned from user reviews are the unigram, bigram, star rating, comment length, as well as the proportion of words with both positive and negative emotion. They used the classification methods NB, SVM, LR, and DT. The experiment's findings show that when combined with unigrams, phrase length, and sentiment score, Logistic Regression yields the highest F-Measure of 85%. The unigram was discovered to be the most important feature because the F-measure only increased by about 1% when parameters like sentence length & emotion score were added. The bigram and star rating both have a negative impact on the classifier's effectiveness.

III. RESEARCH METHODOLOGY

This section provides a research methodology with the different research methods, and also provides a flowchart of the proposed work, and a proposed algorithm with some research steps.

A. Proposed Methodology

Our research assesses a hybrid model, taking into account all of its benefits and possibilities, with the goal of enhancing the effectiveness of sentiment analysis methods. The methodology is concentrated on three key elements: feature extraction, data preprocessing, and building a hybrid model for an optimal sentiment analysis solution. Each of these elements is detailed below individually.

a) **Data Collection**

Data collection is the initial step for implementation. So, in this step, I have collected IMDb andCornell Movie Reviews Datasets.

b) **Data Preprocessing**

Data preprocessing is a crucial component of text analytics systems, acting as a tool in text mining as well as a means of reducing noise from the corpus. Data processing is the process of gathering data and transforming it into usable information. Data processing, which is frequently done by such a data science or team several data scientists, must be done accurately in order to prevent harming the outcome, also known as that of the data output. Several preprocessing processes are carried out in this study.Firstly, remove the punctuations, and generation of vocabulary, and also uses GloVe for vectorization.

c) **Feature Extraction**

After preprocess the dataset, feature extraction is performed for extracting the features. For this purpose, TF-IDF vectorizer is applied which is discussed below in detail.

- **TF-IDF**

The Term Frequency-Inverse Document Frequency (TF-IDF) analysis illustrates the significance of a term. The expression is the result of computing the dot product of TF and IDF. Before we go on to this, let us first have a grasp on each of the words on their own.

Term Frequency (TF): It provides evidence of the significance of the word to a doc, supporting the notion that the term's significance increases in proportion to the breadth of the doc's interpretation.

$$\text{TF (m)} = \frac{\text{Number of times term m in doc}}{\text{Total terms in doc}} \quad \text{……………..(1)}$$

Inverse Document Frequency (IDF):Describes the ways in which a phrase is truly applicable. It is not required that a phrase that appears often in certain documents, such as stopwords, might be significant (the that, of, etc). Because stopwords do not provide any information about the situation, it is best to avoid using them. The IDF operates in such a manner that it is able to disregard them.

- It gives a negative weighting to the term that appears several times throughout the texts.
- The score for the relevant phrase on the IDF is greater, but the weight given to the stop word is lower.
- The natural logarithmic function, often known as log e, is taken into consideration.

$$\text{IDF (m)} = \log \frac{\text{Total number of docs}}{\text{Number of docs with term m}} \quad \text{…………(2)}$$

In a nutshell, the TF-IDF value is dependent on the document, while the IDF relies on the corpus.

Raw data are not transformed into useful features by Tf*Idf in a straightforward manner. To begin, it transforms raw strings or datasets into vectors, with a separate vector being assigned to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

each each word. Then, in order to get the feature, we will implement a specific method such as the Cosine Similarity algorithm, which operates on vectors, etc.

B. Proposed Hybrid Model

The proposed model consists of following sub models for the analysis as shown in Figure 4.

1) Support Vector Machine (SVM)

One of the most prominent supervised learning techniques, Support Vector Machine (SVM), is also helpful for regression issues. However, its main use is for categorization issues in computer vision.

To effectively categorise new data points inside the future, the SVM approach seeks to identify the optimal line of decision border that splits that space into n separate classes. A hyperplane is this optimal decision boundary.The hyper-extreme points & vectors that are employed to build the plane are specifically chosen by SVM. Because these extreme situations are represented by support vectors, the appropriate method is known as a supported vector machine. Look at the figure below, which classifies two groups using a decision boundary called hyperplane:

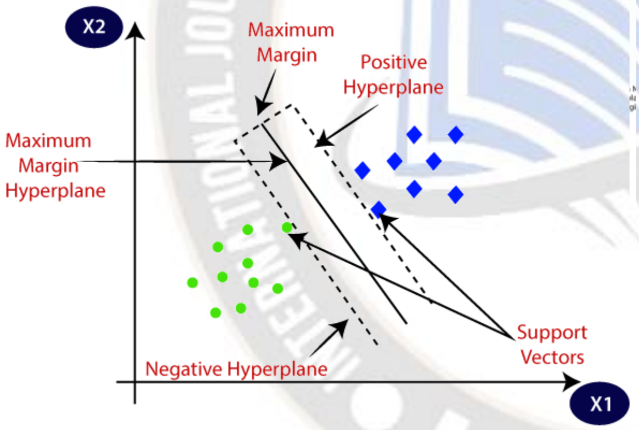

Figure 1. Support Vector Machine

2) Convolutional Neural Network (CNN)

A unique type a multi-layer neural network called a CNN, commonly referred to as a ConvNet, is made exclusively for extracting patterns from unprocessed image data. Convolution is the term used to describe the mathematical operation employed in CNN.It is a linear operation in which one function is multiplied by another to produce a third function that represents the transformation that the first function experienced. Two images, each represented by a matrix, are multiplied to get an output that may be utilised for data extraction. While comparable to other neural networks, CNN adds a new level of complexity due to the use of a series of convolutional layers. Convolutional layers are essential to the operation of CNN.

In a number of computer vision tasks, CNN artificial neural systems has demonstrated improved performance. It has aroused interest among people in a variety of contexts.

A convolutional neural network is constructed of numerous layers, including convolution, pooling layers, or fully connected layers, and uses a backpropagation algorithm to automatically & able to adapt learn spatial hierarchies of input.

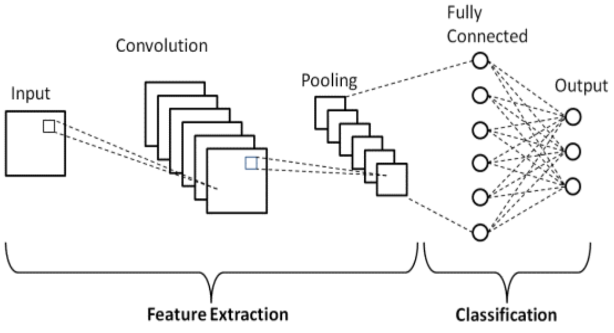

Figure 2. Architecture of Convolutional Neural Network

The three types of layers that make up the CNN are the convolutional layer, the pooling layer, as well as the fully connected (FC) layer. Consequently, a convolutional network is the final structure (CNN).Furthermore, the dropout layer and the activation function are also critical components of the network architecture.

3) Long Short-Term Memory (LSTM)

A type of recurrent neural networks is the LSM (RNN). In RNN, the result from one stage serves as the input for the following one. Hochreiter & Schmidhuber were in charge of the initial design of LSTM. We overcame the RNN's inability to identify a word from its long-term memory & demonstrate that it now makes more correct estimates based on the additional data that is available.RNN's performance declines as spacing is increased. Long-term storage of information is a strength of LSTM. This approach can then be used to process, forecast, and categorise time-series data.

Structure of LSTM:

The four neural networks that make up LSTM's chain structure each have their own unique memory blocks (cells).

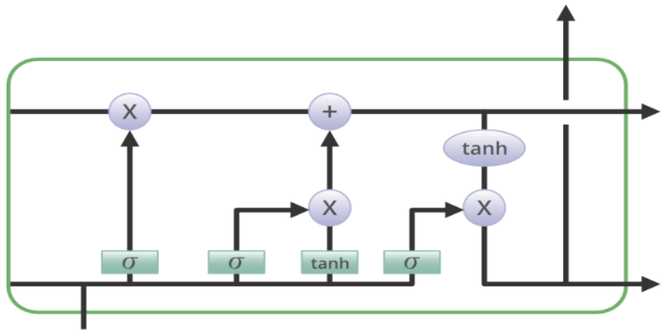

Figure 3. Structure of LSTM

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

Data is stored in the cells, while the gates manipulate the data in memory. A total of three gates are available:

- Forget Gate
- Input Gate
- Output Gate

A. Proposed Flowchart

With the use of these three models, a hybrid model is designed which is given below with different steps.

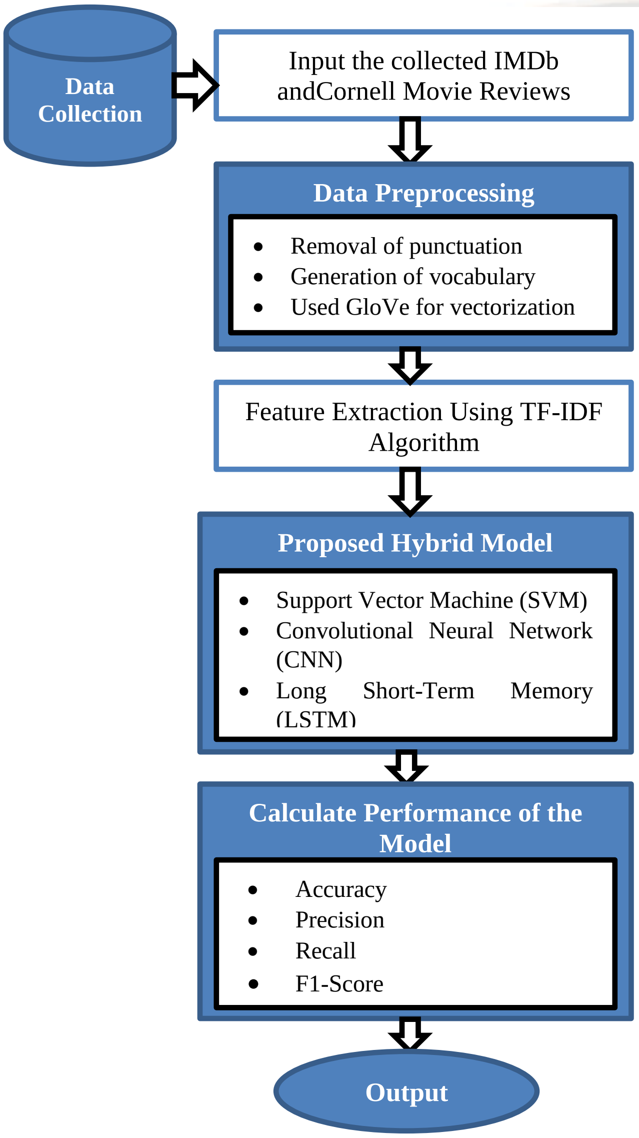

Figure 4. Flowchart of the Proposed Hybrid Model

B. Proposed Algorithm

Here, proposed algorithm is presented with different research steps, which is given below:

Algorithm 1: Proposed Hybrid Model

Input: IMDb and CornellMovie Review Dataset

Output: Classify the Sentiments, and Get Predicted Output

Strategy:

Step 1: Initialization
Step 2: Import required python libraries, and NLP.
Step 3: Collect, and load the IMDb movie review and Cornellmovie review dataset.
Step 4: Next, preprocess the data in data preprocessing that is used for text mining.

- Removal of punctuation
- Generation of vocabulary
- Used GloVe for vectorization

Step 5: Perform feature extraction using TF-IDF algorithm.
Step 6: Create a hybrid model for sentiment classification using the SVM, CNN, and LSTM methods..
Step 7: Calculate the model's performance using different performance metrics i.e.,

- Accuracy
- Precision
- Recall
- F1-Score

Step 8: Finally, obtain the accurate outcome
Step 9: Terminate the model.

IV. RESULTS AND DISCUSSION

The suggested model's viability was verified by extensive experimental testing. The process of creating a deep learning model is facilitated by a plethora of libraries and tools. One of the most popular software frameworks is called Keras. Keras relies on Tensorflow in the background to function in either a CPU or GPU setting. Both Keras and Tensorflow were implemented in this research using a graphics processing unit. The deep learning research made use of the GeForce GTX Titans Black 6 GB GPU, Intel Processor i-7 CPU, 24 GB RAM, with SSD hard drive. Google COLAB was used when the existing computer resources weren't enough for an experiment. For this analysis, we focused on the accuracy value. We also utilised measures like F1-score, Precision, and Recall to assess the efficacy of our models[18].

A. Dataset Description

Our research does not aim to address a specific issue but rather to provide a benchmark for models applicable across domains. Our research made use of data sets including both IMDb and Cornell movie reviews. The capacity to avoid privacy problems [19], widespread acceptability in the academic community, a wide range of sources and subjects, and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

a manageable size were all taken into account. Because of the variety and depth of the datasets used, this research is able to provide a thorough comparison of the sentiment analysis methods considered. The project aims to learn whether the models provide reliable findings across a variety of datasets. IMDb movie reviews [20]and Cornell movie reviews were used in the tests. Both the IMDb and Cornell movie reviews may be found in the first dataset, while the latter has reviews from both sites. The aggregate user-generated commentary in these review databases for movies exceeds 125,000 words.

Table 1: Number Samples of Dataset

S. No.	Datasets	Number of Samples
1.	IMDb movie review	25.000
2.	Cornell movie review	10.662

B. Performance parameters

Measurements from of the confusion matrix will be compared against classification successes gained using sentiment classification from related research to show the method's accuracy. The values for accuracy, precision, and F1 can be derived using the data in the confusion matrix. In the confusion matrix, the letters TP, FP, FN, & TN, respectively, stand to True Positive, Fake Positive, False Negative, and True Negative[18]:

- **TP:**Positive samples are those when both the expected and the actual class labels are right.
- **FP:**The amount of times a positive class label was anticipated when a negative label was accurate.
- **FN:**There have been a negative number of these cases when the anticipated class label is wrong.
- **TN:** the overall number of predicted classes that are wrong when the real classes are correct.

The accuracy, precision, recall, & F1 values are calculated using the confusion matrix. Estimates of accuracy are created using formula (3).

$$Accuracy = (TN + TP)/(TN + TP + FN + FP) \quad (3)$$

Precision is the proportion of instances in which a given class label was correctly anticipated. The equation (4)used to determine the precision.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (4)$$

The recall value counts the number of examples of each class that were successfully identified. This value may be obtained using Equation (5).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FP}} \quad (5)$$

The F1 metric is used to aggregate the recall and accuracy scores. Measured between 0 and 1, with a value of 1 indicating that all samples were properly categorised by the classifier. Measured by equation (6), a good F1 score is near to 1 and indicates successful classification.

$$F1 - \text{score} = \frac{2\times(\text{Precision}-\text{Recall})}{(\text{Precision}+\text{Recall})} \quad (6)$$

C. Results

In this section discusses about the obtained outcome of the proposed hybrid model that are implemented using Python programming.

TABLE I. ACCURACY COMPARISON FOR BOTH DATASETS

Datasets	SVM	CNN	LSTM
IMDb Movie Reviews	82.8	87.3	85.1
Cornell Movie Reviews	67.7	72.4	76.1

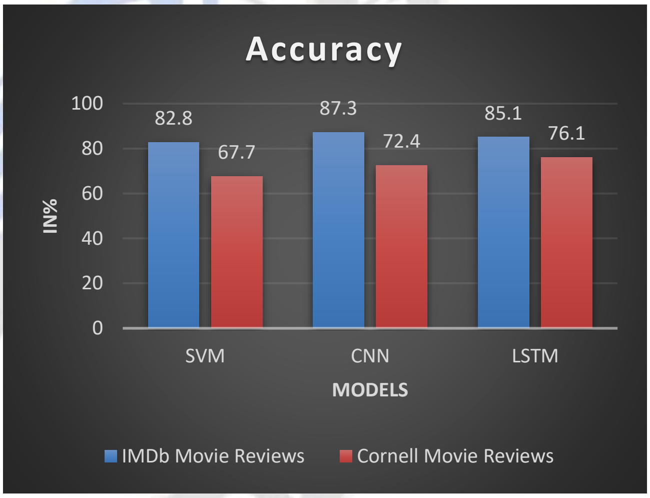

Figure 5. Accuracy Graph for Proposed Hybrid Models

The accuracy statistics for suggested models employing both datasets are displayed in the aforementioned table 1 and figure 5. Figure 2 displays the proposed models on the x-axis and the percentage accuracy scores on the y-axis. These values are representing in a bar graph. According to this figure, SVM achieves the 82.8%, and 67.7% of accuracy for IMDb and Cornell movie review dataset, CNN achieves the 87.3%, and 72.4% of accuracy for IMDb and Cornell movie review dataset, and LSTM achieves the 85.1%, and 76.1% of accuracy for IMDb and Cornell movie review dataset, respectively.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

TABLE II. PRECISION COMPARISON FOR BOTH DATASETS

Datasets	SVM	CNN	LSTM
IMDb Movie Reviews	82.7	87.9	84.8
Cornell Movie Reviews	69.8	70.8	78.4

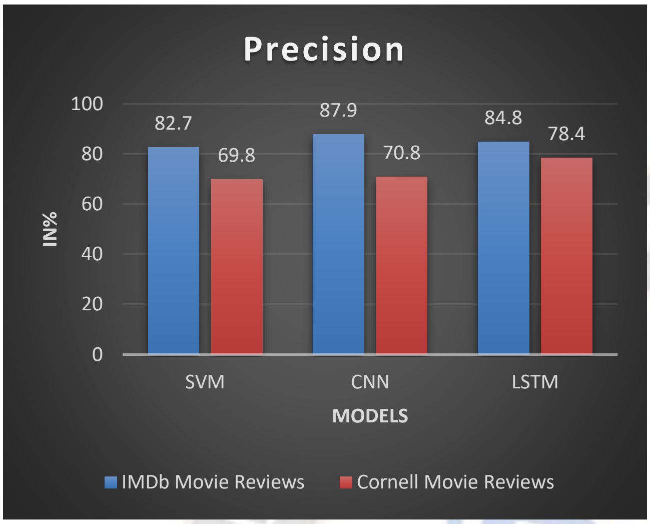

Figure 6. Precision Graph for Proposed Hybrid Models

THE ABOVE TABLE 2, AND FIGURE 6 SHOWS THE PRECISION RESULTS FOR PROPOSED MODELS USING BOTH DATASETS. IN FIGURE 3, X-

axis shows the proposed models, and y-axis shows the precision scores in percentage. These values are representing in a bar graph. According to this figure, SVM achieves the 82.7%, and 69.8% of precision for IMDb and Cornell movie review dataset, CNN achieves the 87.9%, and 70.8% of precision for IMDb and Cornell movie review dataset, and LSTM achieves the 84.8%, and 78.4% of precision for IMDb and Cornell movie review dataset, respectively.

TABLE III. RECALL COMPARISON FOR BOTH DATASETS

Datasets	SVM	CNN	LSTM
IMDb Movie Reviews	82.9	86.9	85.4
Cornell Movie Reviews	67.1	73.6	75.1

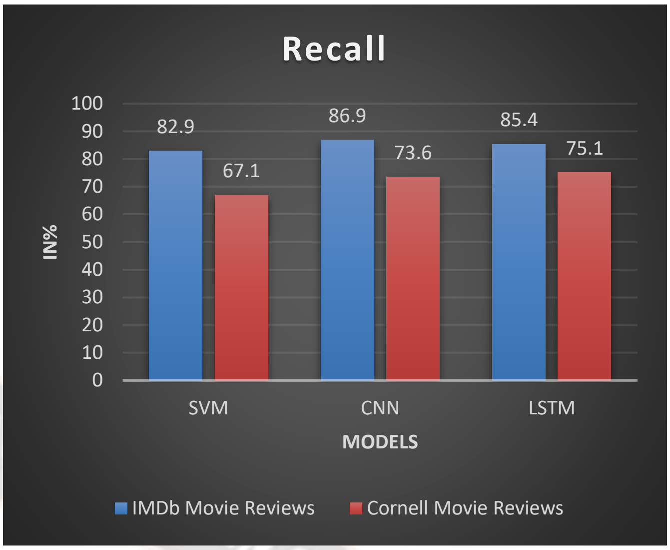

Figure 7. Recall Graph for Proposed Hybrid Models

The recall outcomes for suggested models employing both datasets are displayed in the aforementioned table 3 and figure 7. Figure 4 depicts the proposed models on the x-axis and the recall rates in % on the y-axis.These values are representing in a bar graph. According to this figure, SVM achieves the 82.9%, and 67.1% of recall for IMDb and Cornell movie review dataset, CNN achieves the 86.9%, and 73.6% of recall for IMDb and Cornell movie review dataset, and LSTM achieves the 85.4%, and 75.1% of recall for IMDb and Cornell movie review dataset, respectively.

TABLE IV. F1-SCORE COMPARISON FOR BOTH DATASETS

Datasets	SVM	CNN	LSTM
IMDb Movie Reviews	82.8	87.4	85.0
Cornell Movie Reviews	68.3	71.5	76.6

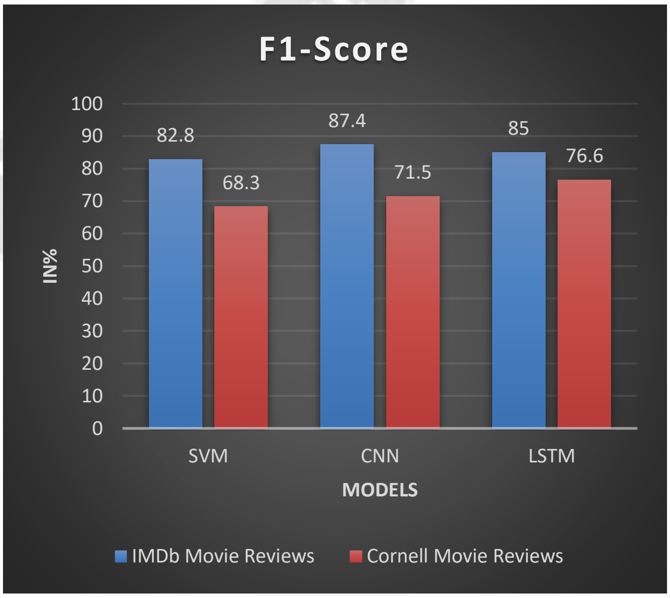

Figure 8. F1-Score Graph for Proposed Hybrid Models

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.7082
***Article Received**: 12 March 2023 **Revised**: 30 April 2023 **Accepted**: 21 May 2023*

The results of the f1-score for proposed models employing both datasets are shown in the aforementioned table 4 and figure 8. Figure 5 displays the proposed models on the x-axis and the percentage values of the f1-scores on the y-axis. From this graph, SVM obtains the 82.8%, and 68.3% of f1-score for IMDb and Cornell movie review dataset, CNN achieves the 87.4%, and 71.5% of f1-score for IMDb and Cornell movie review dataset, and LSTM achieves the 85%, and 76.6% of f1-score for IMDb and Cornell movie review dataset, respectively.

V. CONCLUSION

In the present study, we suggested applying hybrid deep learning algorithms to the issue of sentiment extraction from social media data. We tested the effectiveness of integrating SVM, CNN, & LSTM on two review datasets, notably the Cornell and IMDb movie reviews. This study investigates various text preparation techniques, including punctuation removal, vocabulary expansion, and vectorization. The feature extraction is done using the TF-IDF vectorizer, while the vectorization is done using GloVe. These experiments are done to find out how adaptable hybrid systems are, specifically if hybrid methodologies can support a wide range of dataset types and sizes. By examining the results of employing two data, two feature extraction techniques, with two deep learning models, the reliability of sentiment polarity assessment was examined. Our experiments' findings indicate that the hybrid version is more reliable than competing products when it come to sentiment polarity analysis. According to this study, CNN has an accuracy rate of 87.3% for the IMDb film reviews data and 76.1% for the Cornell film reviews dataset. Future plans include investigating recurrent neural networks (RNN) for aspect term extraction and expanding our CNN-based method for multi-label classification. In addition, as the quality of word representation is a crucial aspect of any neural network design, we want to apply approaches such as distance supervision to improve the quality of word representations.

REFERENCES

[1] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” Information (Switzerland). 2019, doi: 10.3390/info10040150.

[2] P. Sunagar and A. Kanavalli, “A Hybrid RNN based Deep Learning Approach for Text Classification,” vol. 13, no. 6, pp. 289–295, 2022.

[3] S. M. Mohammad, S. Kiritchenko, and X. Zhu, “NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets,” 2013.

[4] D. K. Gupta, K. S. Reddy, Shweta, and A. Ekbal, “PSO-asent: Feature selection using particle swarm optimization for aspect based sentiment analysis,” 2015, doi: 10.1007/978-3-319-19581-0_20.

[5] C. N. Dos Santos and M. Gatti, “Deep convolutional neural networks for sentiment analysis of short texts,” 2014.

[6] Y. Kim, “Convolutional neural networks for sentence classification,” 2014, doi: 10.3115/v1/d14-1181.

[7] S. Poria, E. Cambria, and A. Gelbukh, “Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis,” 2015, doi: 10.18653/v1/d15-1303.

[8] C. N. Dang, M. N. Moreno-García, and F. De La Prieta, “Hybrid Deep Learning Models for Sentiment Analysis,” Complexity, 2021, doi: 10.1155/2021/9986920.

[9] B. Alsukhni, “Multi-Label Arabic Text Classification Based on Deep Learning,” 2021, doi: 10.1109/ICICS52457.2021.9464538.

[10] V. Rajpoot, R. Agrawal, A. Chaturvedi, and K. Goyal, “An Empirical Study of Sentiment Analysis on Movie Review using Machine Learning based Classification Approach,” 2021, doi: 10.1109/ISCON52037.2021.9702471.

[11] H. B. Dogru, S. Tilki, A. Jamil, and A. Ali Hameed, “Deep Learning-Based Classification of News Texts Using Doc2Vec Model,” 2021, doi: 10.1109/CAIDA51941.2021.9425290.

[12] A. C. Mazari and A. Djeffal, “Deep Learning-Based Sentiment Analysis of Algerian Dialect during Hirak 2019,” 2021, doi: 10.1109/IHSH51661.2021.9378753.

[13] H. Srinidhi, G. M. Siddesh, and K. G. Srinivasa, “A hybrid model using MaLSTM based on recurrent neural networks with support vector machines for sentiment analysis,” Eng. Appl. Sci. Res., 2020, doi: 10.14456/easr.2020.26.

[14] R. Parveen, N. Shrivastava, and P. Tripathi, “Sentiment classification of movie reviews by supervised machine learning approaches using ensemble learning voted algorithm,” 2020, doi: 10.1109/IDEA49133.2020.9170684.

[15] A. A. Wadhe and S. S. Suratkar, “Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques,” 2020, doi: 10.1109/I4Tech48345.2020.9102673.

[16] A. Tariyal, S. Goyal, and N. Tantububay, “Sentiment Analysis of Tweets Using Various Machine Learning Techniques,” 2018, doi: 10.1109/ICACAT.2018.8933612.

[17] Y. Ekanata and I. Budi, “Mobile application review classification for the Indonesian language using machine learning approach,” 2018, doi: 10.1109/CATA.2018.8398667.

[18] M. U. Salur and I. Aydin, “A Novel Hybrid Deep Learning Model for Sentiment Classification,” IEEE Access, 2020, doi: 10.1109/ACCESS.2020.2982538.

[19] S. Kumar, M. Gahalawat, P. P. Roy, D. P. Dogra, and B. G. Kim, “Exploring impact of age and gender on sentiment analysis using machine learning,” Electron., 2020, doi: 10.3390/electronics9020374.

[20] A. Sadeghian, “Bag of Words Meets Bags of Popcorn,” CS224N Proj., 2013.