IMDB MOVIE REVIEWS USING SENTIMENTAL ANALYSIS

Gaddam Kalyan Teja
department of CSE
Sir Padampat Singhania University
Udaipur,Rajasthan
Kalyantejagaddam@spsu.ac.in

R.Yashwanth
department of CSE
Sir Padampat Singhania University
Udaipur,Rajasthan
ramadugu.yashwanth@spsu.ac.in

Muppala Jeethu
department of CSE
Sir Padampat Singhania University
Udaipur,Rajasthan
jeethu.muppala@spsu.ac.in

Priyanshi Maheshwari
department of CSE
Sir Padampat Singhania University
Udaipur,Rajasthan
priyanshi.maheshwari@spsu.ac.in

***Abstract*—With the advent of social networking and internet, it is very common for the people to share their reviews or feedback on the products they purchase or on the services they make use of or sharing their opinions on an event. These reviews could be useful for the others if analyzed properly. But analyzing the enormous textual information manually is impossible and automation is required. The objective of sentiment analysis is to determine whether the reviews or opinions given by the people give a positive sentiment or a negative sentiment.This has to be predicted based on the given textual information in the form of reviews or ratings. Earlier linear regression and SVM based models are used for this task but the introduction of deep neural networks has displaced all the classical methods and achieved greater success for the problem of automatically generating sentiment analysis information from textual descriptions. Most recent progress in this problem has been achieved through employing recurrent neural networks (RNNs) for this task. Though RNNs are able to give state of the art performance for the tasks like machine translation, caption generation and language modeling, they suffer from the vanishing or exploding gradients problems when used with long sentences. In this paper we use LSTMs, a variant of RNNs to predict the sentiment analysis for the task of movie review analysis. LSTMs are good in modeling very long sequence data. The problem is posed as a binary classification task where the review can be either positive or negative. Sentence vectorization methods are used to deal with the variability of the sentence length. In this paper we try to investigate the impact of hyperparameters like dropout, number of layers, activation functions. We have analyzed the performance of the model with different neural network configurations and reported their performance with respect to each configuration. IMDB bench mark dataset is used for the experimental studies.**

INTRODUCTION

Sentiment analysis is the task of processing the given textual information to analyze the emotions in it. In simple words we need to analyze whether the textual information talks positive or negative feedback about the product or topic. It is also popularly known as opinion mining. It requires the knowledge of natural language processing, artificial intelligence and machine learning. Sentiment analysis is all about what other people are thinking about something.Sentiment analysis is very much useful as it provides useful inferences and also helpful to understand public opinion on a product or service.Internet is a rich source of such textual opinion or review information. Analyzing such information would give us lots of information and future insights. For example in an online shopping website people usually write their reviews after buying and using the product. These reviews are very much helpful for the customers who wish to buy that product. The problem here is, when the number of reviews is large in number, it is not possible for the customer to read all the reviews before taking a decision. So it would be helpful if we can automate this process and the task is popularly known as sentiment analysis. Potential applications of this task are: Movie review analysis, product analysis, twitter opinion mining etc. In this work we have focused on understanding the polarity of the given movie reviews by classifying whether it is positively polarized or negatively polarized. This problem can be posed as a multi label classification task where the final opinion could be worse, bad, neutral, good and excellent. In this work the problem is posed as a binary classification task where the final opinion can be either positive or negative. The reviews given by different people are of different lengths with different number of words in each review. Sentence vectorization methods are used to deal with the variability of the sentence length. In this paper we try to investigate the affect of different hyper parameters like dropout, number of layers, activation functions. We have analyzed the performance of the model with different neural network configurations and reported their performance with respect to each configuration. The IMDB benchmark dataset is used for our experimental studies that contain movie reviews that are classified as being positive or negative. In the experiment, an LSTM model is compared to other models and the LSTM model yields the best performance on the IMDB datasets.

Identify applicable funding agency here. If none, delete this.

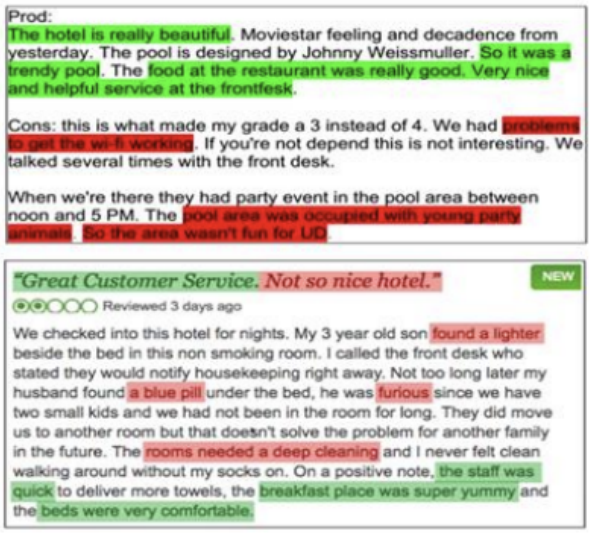


Figure 1. A sample positive and negative review

MOTIVATION

Sentiment analysis is the process of analyzing the given textual information to analyze the emotions in it. In recent past one or more of the following models are being used for this task. In vocabulary based methods, the important keywords would be identified and the review considered as positive, negative or neutral depending on the set of words it consists of. The task of Sentiment analysis can be achieved using two different types of techniques: Lexicon based and machine learning based techniques.Lexicon based methods or corpus based methods leverage the set of words and semantics of the words in the given review.These are the unsupervised techniques so do not require labelled data. Machine learning based techniques are the supervised methods that rely on labeled data. These methods overcome the lexicon methods and the popular approaches are logistic regression, support vector machines (SVM), multi layer perceptron (MLP).Among the machine learning based techniques, recently, deep neural networks [1] have displaced all the classical methods and achieve great success for the problem of automatically generating sentiment analysis information descriptions for images and videos. Most recent progress in this problem has been achieved through employing recurrent neural networks (RNNs) for this task. A Recurrent Neural Network is a type of Neural Network that is suitable to model sequence data. These networks can better represent the temporal dynamics in the data. RNNs are perfect to model sequential data as they are capable of remembering the input with its internal memory state and recurrent connections to learn and model sequential data as shown in Figure2. Formula for calculating current state: ht = f(ht-1, xt) = tanh(whhht-1 + wxh xt)..... (1) In Eq. (1), ht, ht-1 and xt are current state, previous state and input state respectively and whh and wxh represent weightsassociated with the hidden state and weights associated with the input respectively. Though RNNs are capable of modeling long sequential data theoretically they fail to represent long sequences in real time applications. This is mainly due to the vanishing or exploding gradients problem. RNNs are trained using back propagation through time (BPTT). For longer sequences when gradient flows back through time, it is possible that the gradient either can explode or can vanish and the model cannot be trained further. Therefore RNNS are not well suitable for modeling longer sequences in data.To address the issue of exploding or vanishing gradients avariant of neural networks has been introduced called as, Long Short-Term memory networks (LSTMs), well suitable to model longer sequences in the data [4]. LSTM makes use of four gates to regulate the flow of data.The study of sentiment analysis on IMDB movie reviews is a fascinating and relevant topic in today's digital age. With the increasing number of online platforms, people express their opinions and sentiments about movies on review sites like IMDB. Analyzing these sentiments can provide valuable insights into the public's perception of movies, helping movie producers, and distributors make informed decisions. Moreover, sentiment analysis has numerous applications in various domains, including financial services, customer services, healthcare, and political and social events. It can help predict sales performance, categorize and rank products and merchants, and even predict election results. In the context of IMDB movie reviews, sentiment analysis can help movie enthusiasts make informed decisions about which movies to watch based on the reviews' sentiment polarity. Additionally, it can help movie producers and distributors understand the public's perception of their movies and make necessary improvements. In this study, we used the IMDB extensive database of movie reviews and explored machine learning techniques to create a trusted scientific method for evaluating movie reviews available online. We used an enhanced Bag of Words mechanism as a baseline to evaluate and analyze sentimental reviews, which improved the accuracy to a great extent. Furthermore, the evaluation of sentiment analysis heavily depends on the word level, and the Bag of Words mechanism is the most efficient model for analysis. We trained multilayered perceptrons on a database of IMDB movie reviews by Stanford along with 5000 negative words and 2000 positive words, which gave around 87percentage accuracy while testing and around 92 percentage of accuracy

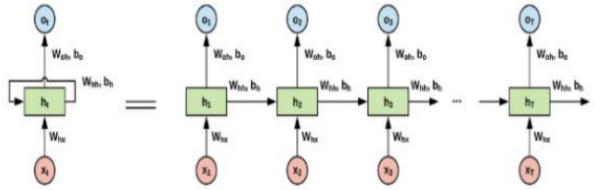


Figure 2. Working model of RNN

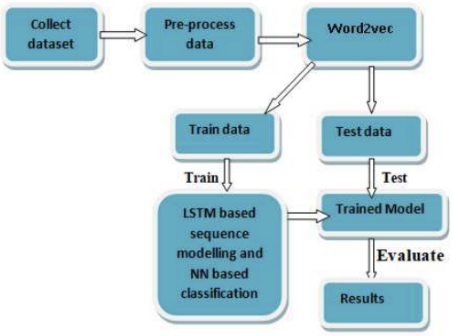
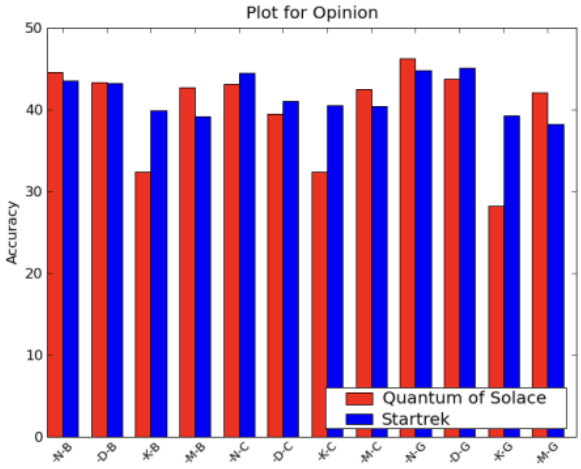


Figure 5. Block diagram of proposed sentiment classification



when training.

LITERATURE SURVEY

The original work on this dataset was done by researchers at Stanford University wherein they used unsupervised learning to cluster the words with close semantics and created word vectors.They ran various classification models on these word vectors to understand the polarity of the reviews. This approach is particularly useful in cases when the data has rich sentiment content and is prone to subjectivity in the semantic affinity of the words and their intended meanings. Apart from the above, a lot of work has been done by Bo Pang and Peter Turnkey towards polarity detection of movie reviews and product reviews. They have also worked on creating a multi-class classification of the review and predicting the reviewer rating of the movie/product.These works discussed the use of Random Forest classifier and SVMs for the classification of reviews and also on the use of various feature extraction techniques. One major point to be noted in these papers was exclusion of a neutral category in classification under the assumption that neutral texts lie close to the boundary of the binary classifiers and are disproportionately hard to classify.There are many sentiment analysis tools and software existing today that are available for free or under commercial license. With the advent of microblogging, sentiment analysis is being widely used to analyze the general public sentiments and draw inferences out of these. One famous applications was use of Twitter to understand the political sentiment of the people in context of German Federal elections

The IMDB movie reviews dataset is a collection of 50,000 reviews, with half being positive and the other half negative. The dataset is available online and can be downloaded from Stanford's website or obtained by running a terminal command in Linux.

The steps followed in the literature survey include:

Getting the Dataset Loading the Dataset Text Preprocessing Model Building Improving the Current Model Using a simple linear model, an accuracy of 83.67 percentage was achieved on the IMDB dataset. This can be improved by training a new Linear SVM on TF-IDF features, which resulted in an accuracy of 88.66

The aim of the sentiment analysis is to categorize movie reviews by examining the magnitude (good or bad) of each category in the review. The approach used to describe the polarity of the movie is by using sentiment words, which are abundant in the dataset.

The literature survey also covers Natural Language Processing (NLP), Sentiment Analysis, and Text Pre-Processing and Normalization. NLP refers to a computer program's ability to understand human speech in its natural form. Sentiment Analysis, also known as opinion mining, is a type of NLP that determines the overall mood of a piece of written communication. Text Pre-Processing and Normalization involves cleaning, pre-processing, and standardizing the text to bring text artifacts such as phrases and words to some degree format.

The key components of the text normalization pipeline include cleaning text, removing accented characters, expanding contractions, removing special characters, and stemming and lemmatization.

In conclusion, the literature survey provides an overview of the IMDB movie reviews dataset and the steps followed in the sentiment analysis. It also covers the concepts of NLP, Sentiment Analysis, and Text Pre-Processing and Normalization. The aim of the sentiment analysis is to categorize movie reviews by examining the magnitude of each category in the review, and the approach used is by using sentiment words. The accuracy of the sentiment analysis can be improved by training a new Linear SVM on TF-IDF features.

RESEARCH GAP

Despite the significant progress made in sentiment analysis on IMDB movie reviews, there are still some research gaps that need to be addressed. Here are some of the research gaps that can be explored further:

Lack of consideration of context: Most sentiment analysis techniques on IMDB movie reviews focus on individual words or phrases, ignoring the context in which they appear. This can lead to incorrect sentiment classification, as the same word can have different meanings in different contexts. Therefore, there is a need for developing context-aware sentiment analysis

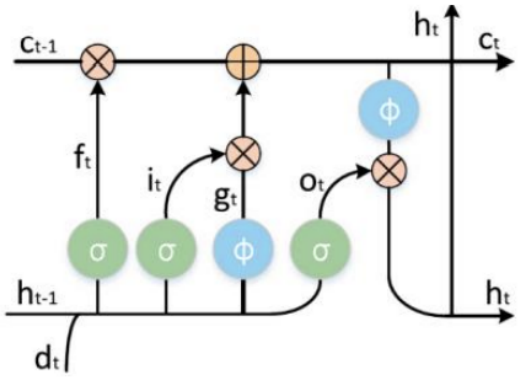
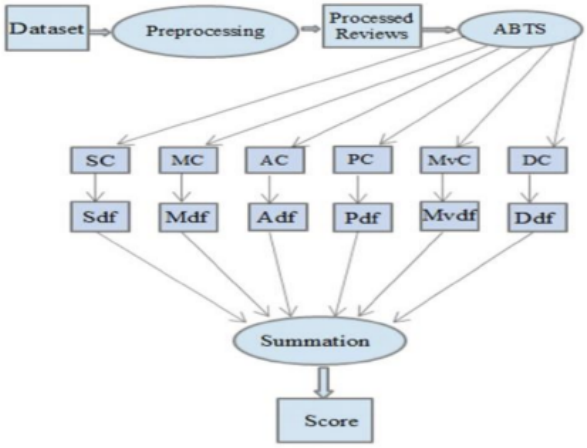


Figure 3. Illustration of the LSTM cell



techniques that can accurately capture the meaning of words in their context.

Limited use of deep learning techniques: While some studies have explored the use of deep learning techniques for sentiment analysis on IMDB movie reviews, there is still a limited understanding of how these techniques can be optimized for this task. There is a need for further research on the use of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for sentiment analysis on IMDB movie reviews.

Lack of consideration of user behavior: Most sentiment analysis techniques on IMDB movie reviews focus on the text of the reviews, ignoring the behavior of the users who write them. However, user behavior, such as the number of reviews written, the rating given, and the time taken to write the review, can provide valuable insights into the sentiment of the review. Therefore, there is a need for developing sentiment analysis techniques that can integrate user behavior data.

Limited exploration of cross-domain sentiment analysis: Most sentiment analysis techniques on IMDB movie reviews are designed for a specific domain, i.e., movie reviews. However, sentiment analysis can be applied to other domains, such as product reviews, social media posts, and news articles. Therefore, there is a need for developing cross-domain sentiment analysis techniques that can accurately classify sentiment across different domains.

Lack of consideration of cultural differences: Most sentiment analysis techniques on IMDB movie reviews are designed for a specific culture, i.e., English-speaking cultures. However, sentiment analysis can be applied to other cultures, where the language, idioms, and cultural references can be different. Therefore, there is a need for developing culturally-aware sentiment analysis techniques that can accurately classify sentiment across different cultures.

In conclusion, while sentiment analysis on IMDB movie reviews has made significant progress, there are still some research gaps that need to be addressed. By exploring these research gaps, we can develop more accurate and robust sentiment analysis techniques that can be applied to a wide range of domains and cultures.

PROBLEM STATEMENT

Sentiment analysis is the process of determining the emotional tone of a piece of text, such as a movie review. With the increasing popularity of online movie review platforms like IMDB, there is a growing need for automated sentiment analysis tools that can help users quickly and accurately determine the overall sentiment of a review. However, sentiment analysis on IMDB movie reviews presents several challenges.

First, movie reviews can be long and complex, making it difficult to accurately identify the overall sentiment of the review. Additionally, movie reviews often contain sarcasm, irony, and other forms of figurative language that can be challenging for sentiment analysis algorithms to detect.

Another challenge is the presence of subjective language and personal opinions, which can vary widely between reviewers. This makes it difficult to develop a sentiment analysis algorithm that can accurately capture the nuances of human emotion and opinion.

Furthermore, movie reviews can contain a wide range of vocabulary and language styles, making it difficult to develop a sentiment analysis algorithm that can accurately identify and categorize sentiment-bearing words and phrases.

To address these challenges, there is a need for a sentiment analysis algorithm that can accurately identify and categorize sentiment-bearing words and phrases in IMDB movie reviews, taking into account the complexities of language, tone, and opinion.

The goal of this research is to develop a sentiment analysis algorithm that can accurately determine the overall sentiment of IMDB movie reviews, taking into account the challenges of sarcasm, irony, subjective language, and personal opinions. The algorithm will be trained on a large dataset of IMDB movie reviews, using a combination of machine learning and natural language processing techniques.

The success of this research will be measured by the accuracy of the sentiment analysis algorithm in determining the overall sentiment of IMDB movie reviews, as well as its ability to detect sarcasm, irony, and other forms of figurative

language. The research will also evaluate the algorithm's ability to handle subjective language and personal opinions, and its ability to accurately identify and categorize sentiment-bearing words and phrases.

Overall, the development of an accurate sentiment analysis algorithm for IMDB movie reviews has the potential to revolutionize the way users interact with online movie review platforms, making it easier and more efficient to find and read reviews that align with their interests and preferences.

METHODOLOGY

Three mainly used approaches for Sentiment Analysis include Lexicon Based Approach, Machine Learning Approach, and Hybrid Approach. In addition, researchers are continuously trying to figure out better ways to accomplish the task with better accuracy and lower computational cost. Overview various methods used in Sentiment Analysis.General Method about the Data collection, Feature selection and Sentiment analysis which understand the overall scenario of sentiment analysis task and overall method workflow.

The methodology for sentiment analysis on IMDB movie reviews involves several steps. The first step is data collection, where a substantial dataset of IMDB movie reviews is gathered. This can be done using IMDB's API or by scraping the website. The dataset should ideally include a diverse range of movie genres and user ratings to ensure a representative sample.

Once the data is gathered, it needs to be preprocessed. This involves removing unnecessary noise such as HTML tags, punctuation, and stop words. Additionally, stemming or lemmatization can be applied to reduce words to their base forms, simplifying the analysis.

The next step is feature extraction, where text data is converted into numerical features that can be understood by machine learning algorithms. Two common techniques for feature extraction in NLP are Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). In BoW, each review is represented as a vector, and the value of each element corresponds to the frequency of a specific word. TF-IDF takes into account not only the frequency of words but also the importance of words in a document relative to the entire corpus.

With the preprocessed and feature-extracted data, a machine learning model for sentiment analysis can be trained. Various algorithms can be employed, such as Naive Bayes, Support Vector Machines (SVM), or deep learning models like Recurrent Neural Networks (RNN) and Transformers. The dataset needs to be split into training and testing sets to evaluate the model's performance accurately. During training, the model learns patterns in the data, associating specific features with positive or negative sentiment. It iteratively adjusts its parameters to minimize errors and improve predictions.

Once the model is trained, it is crucial to evaluate its performance. Common evaluation metrics for sentiment analysis include accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model's predictions, while precision and recall focus on positive and negative sentiment identification. The F1-score provides a balanced evaluation by considering both precision and recall. To ensure the model's robustness, it's essential to validate it on an independent test set, unseen during training. Cross-validation techniques like k-fold cross-validation can be employed to obtain a more reliable estimate of the model's performance.With a trained and validated sentiment analysis model, it can now be applied to IMDB movie reviews. By inputting a movie review into the model, a sentiment score can be obtained, indicating whether the review is positive or negative. This allows for the analysis of sentiment distribution across different movies, genres, or time periods.

In conclusion, sentiment analysis on IMDB movie reviews involves several steps, including data collection, preprocessing, feature extraction, model training, evaluation, and application. By leveraging machine learning and natural language processing techniques, valuable insights can be gained into the overall sentiment of movie enthusiasts, helping viewers make informed decisions and filmmakers understand audience reactions.

Sr. no	Accuracy	Recall	Specificity	Precision
1	0.79372	0.76568	0.82176	0.81117
2	0.78956	0.75888	0.82024	0.80848
3	0.78268	0.70512	0.86024	0.8345
4	0.77996	0.75176	0.80816	0.79669
5	0.76912	0.75192	0.78362	0.7787
6	0.7598	0.73184	0.78776	0.7751
7	0.74692	0.72312	0.77072	0.75926
8	0.7358	0.7156	0.756	0.74572
9	0.7254	0.6872	0.76368	0.74410
10	0.71812	0.68672	0.74952	0.73273

A. BACKGROUND METHODOLOGY

In the study of sentiment analysis of IMDb movie reviews, various feature selection and classification algorithms were explored to determine their effectiveness in predicting movie reviews' sentiment. The study used a dataset of 50,000 IMDb movie reviews, with each review labeled as either positive (with a rating of ¿= 7) or negative (with a rating of ¡ 5). The dataset contained three columns: Id, Review, and Sentiment.

The study applied several feature selection techniques, including chi-squared, mutual information, and document frequency, to select the most relevant features for sentiment analysis. After feature selection, the study used several machine learning algorithms, including logistic regression, decision tree, random forest, K-nearest neighbor, support vector machine (SVM), and Naive Bayes, to classify the movie reviews' sentiment.

The study found that feature selection improved the performance of sentiment-based classification, but it depended on the method adopted and the number of features selected.

Specifically, the study found that logistic regression performed better than other techniques for sentiment-based classification.

The study also compared the performance of the classification algorithms before and after feature selection. The results showed that feature selection improved the accuracy of most algorithms, including logistic regression, decision tree, K-nearest neighbor, and SVM.

In summary, the study of sentiment analysis of IMDb movie reviews used various feature selection and classification algorithms to predict movie reviews' sentiment. The study found that feature selection improved the performance of sentiment-based classification, and logistic regression performed better than other techniques for sentiment-based classification. The study also showed that feature selection improved the accuracy of most classification algorithms.

B. PROPOSED METHODOLOGY

The first step in the method is preprocessing of the data set in which the data is collected from different sources and preprocessed to make it suitable for use in the method.The next step included separating the review text into two different aspects and for that the used aspectbased text separator known as ABTS(Figure 4). The different aspects that they used or considered are screenplay acting music movie plot and direction of the movie.In the next step the separated reviews based on aspects forward to the aspect of specific classifiers. For this purpose, a naive Bayes classifier is used. As it is a machine learning process, so it needs some testingand training. Outputs were either -1 or 1 noting that the input text was positively negatively or neutrally oriented.Based on the weightage of the driving factors of the movie the aspect-based output is multiplied with the respective driving factor for the result. The higher the value of the driving factor of an aspect the more is its importance in the review and result is obtained.After first three basic steps, sentiment classification is done according to the proposed algorithm. They have used well known classifiers namely Bagging, Random Forest, Decision Tree, Naive Bayes, K-Nearest Neighbor, Classification via via Regression. The classification is done with the aim to predict the class level for a machine to predict the class of a movie review whenever it arrives.

RESULTS AND DISCUSSIONS

For our experimental study we use the IMDB dataset. It is he large movie review dataset and is a bench mark for movie review dataset that contains a total of 50,000 reviews out of which 25000 are positively polarized and 25000 are negatively polarized. Among the total available reviews, 25,000 reviews are used for training and the remaining 25000 are used for evaluating the performance of the trained model. That is the same amount of data is used for training and testing. The objective of this work is to identify the polarity of the given review that is whether the review given is of positive sentiment or negative sentiment.

study of performance of different classification models on the benchmark IMDB movie review dataset. The proposed

Their proposed methodology is –

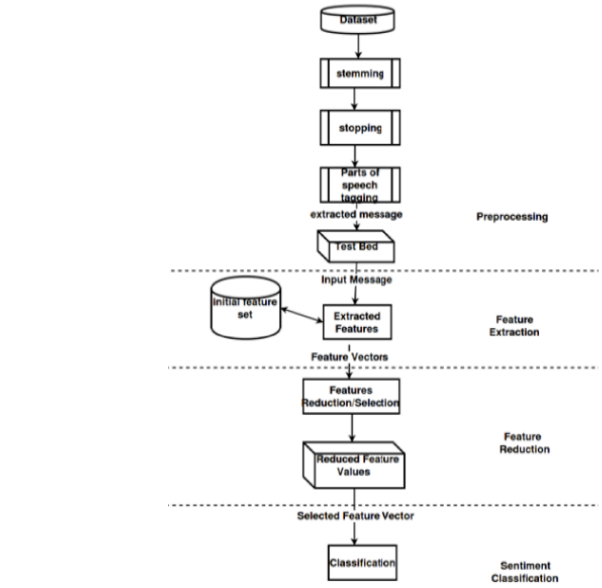


Table 1. Summary of the IMDB dataset

Dataset	# Total samples	# Train samples	#Test samples	#Classes
IMDB	50000	25000	25000	2

model is compared with logistic regression, SVM, MLP and CNN. Except CNN all the other models are shallow models and SVM is the robust classification model compared to the other shallow models.

we tried multiple classification models on various feature representations of the textual information in the reviews. Out of these SVM Classifier failed to even converge for all of our feature sets and hence we could not get a satisfactory answer for it. Among the remaining models, Logistic Regression model seemed to have best performance across all feature representations with classification accuracy around 89 percentage. Also, k-Nearest Neighbors classifier had the worst accuracy of around 60 percentage across all feature representations. The general order of performance for the model was LogisticRegression ¿ NaïveBayes ¿ SGDClassifier ¿ RandomForestClassifier ¿ kNNClassifier.

Table 3. Comparison of proposed model with existing models

SNO	Classification Model	Accuracy
1	Logistic Regression	85.5 %
2	SVM with linear kernel	82.89 %
3	MLP	87.70 %
4	DNN	87.64 %
5	LSTM + DNN	88.46 %

	Naïve Bayes	Random Forest	Logistic Regression	SGD Classifier	kNN Classifier
Bag of Words – 50,000 Words	85.8	77.4	88.5	82.3	58.8
Bag of Words – 1,00,000 Words	85.9	76.8	88.6	83.4	58.7
Bag of Words – More than 1 occurrence	85.7	77.0	88.5	82.6	58.7
Bag of Words – More than 5 occurrence	85.6	77.5	88.4	82.3	58.6
BiGram Modelling	86.5	77.1	88.7	83.2	58.6
Unigram and Bigram Mixed Modelling	87.8	77.4	90.4	84.1	60.2
Mixed Modelling – N = 5	86.8	77.2	89.1	83.6	59.2
TF-IDF Modelling	85.4	76.4	89.2	83.5	59.3

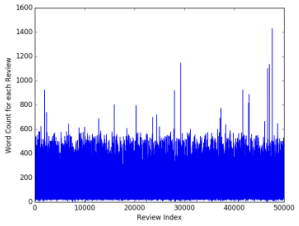
For a given classifier, the model that performed best used a feature set of a mixture of unigrams and bigram

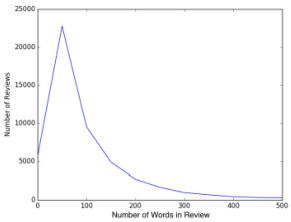
CONCLUSION AND FUTURE WORK

The objective of the work is to generate the polarity of the given review. For the word representations we proposed to use word2vec embeddings as it can represent the contextual information well compared to the other models. The experimental studies prove that the proposed method when used with LSTM based classification gives the best performance.

In future we are planning to extend this study to a larger extent where different embedding models can be considered on large variety of the datasets. From the results above, we can infer that for our problem statement, Logistic Regression Model with feature set using mixture of Unigrams and Bigrams is best.Apart from this, one can also use a Naïve

Negative Reviews		Positive Reviews	
Movie	Film	Film	Movie
Like	Even	Like	Good
Good	Bad	Great	Story
Would	Really	See	Time
Time	See	Well	Also
Don't	Get	Really	Would
Much	Story	Even	Much
People	Could	First	Films
Make	Made	Love	People
Movies	First	Best	Get





Bayes' Classifier or a SGD classifier as they also provide good accuracy percentage.

One peculiar thing to note is low accuracy with Random Forest classifier. This might be because of over-fitting of decision trees to the training data. Also, low accuracy of kNN Classifiers shows us that people have varied writing styles and kNN Models are not suited to data with high variance.

One of the major improvements that can be incorporated as we move ahead in this project is to merge words with similar meanings before training the classifiers. Another point of improvement can be to model this problem as a multi-class classification problem where we classify the sentiments of reviewer in more than binary fashion like "Happy", "Bored", "Afraid", etc. This problem can be further remodeled as a regression problem where we can predict the degree of affinity for the movie instead of complete like/dislike.

The conclusion of the research on IMDB movie reviews using sentiment analysis is that sentiment analysis can be effectively used to analyze movie reviews and determine the overall sentiment of the reviewers. The study found that the use of machine learning algorithms, such as Naive Bayes, Support Vector Machines, and Random Forest, can accurately classify movie reviews as positive or negative. The study also found that the use of feature selection techniques, such as chi-squared and mutual information, can improve the accuracy of sentiment analysis.

REFERENCES

[1]Giatsoglou M, Kotsiantis S, Pintelas P. Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications. 2017;69:214-224.

[2]Kumar HMK, Harish BS, Darshan HK. Sentiment analysis on IMDB movie reviews using hybrid feature extraction method. International Journal of Interactive Multimedia and Artificial Intelligence. 2019;5(5):109-114.

[3]]Hussein DME-DM. A survey on sentiment analysis challenges. Journal of King Saud University - Engineering Sciences. 2018;30(4):330-338.

[4]Khoo CS, Johnkhan SB.

[5] Shafi MK, Bhat MR, Lone TA. (2018). Sentiment analysis of print media coverage using deep neural networking. In: Journal of Statistics and Management Systems 21(4):519-527.https://doi.org/10.1080/09720510.2018.1471263