****

**REVIEW - II**

**CSE4022 - NATURAL LANGUAGE PROCESSING**

**TITLE : Sentiment Analysis of IMDB Movie Reviews using TF-IDF**

**Under the guidance of**

**Dr. Vetriselvi T**

**By**

**B1 Slot Batch 18**

**Johann Kyle Pinto - 20BKT0009**

**Naman - 20BKT0046**

**Rishabh Kumar Rai - 20BKT0076**

**Abstract**

Movie reviews are an important way to analyze and make sense of the performance of a movie. It helps users decide if the movie is worth their time. Movie-rating websites are often used by critics to post comments and rate movies which help viewers decide if the movie is worth watching.

Sentiment analysis of a movie review helps determine how positive or negative a movie review is and hence predict the overall rating for a movie. Therefore, since the machine learns through training and testing the data, it is possible to automate the process of determining whether a review is positive or negative.

In this project we aim to use Sentiment Analysis on a set of IMDB movie reviews given by some reviewers for a movie and try to understand what their overall reaction to the movie was, i.e., if they liked the movie or they hated it.

**Keywords**

Natural Language Processing (NLP) , Sentiment Analysis, Opinion Mining, Text Classification, Term weighting, Tokenization, Stemming, Term Frequency-Inverse Document Frequency, Naive Bayes, Logistic Regression

**Introduction**

The most practical form of entertainment for most people is watching movies. However, only a small number of films are well-liked and profitable. Rating Websites provide the movie lovers with a convenient way with which they can choose which movies they would want to see and which movies they could skip. The most popularly known of these rating websites are websites like IMDB, Rotten Tomatoes, etc. where movie critics put up their opinions. By awarding a movie a score out of 10 based on the stars that the viewers have given it, these websites measure the success of the film. However, no technology exists that can make a prediction based on movie reviews. Sentiment analysis is thus used to assess a film's popularity based on its reviews. IMDB reviews on movies are being used as part of this project to predict how the users have rated the movies and predict the movies that have a positive or negative review. Models like Naïve Bayes and Logistic regressions are proposed to be evaluated which include different sentiment analysis methods that would help us to extract useful information from the data and predict which is the most suitable classifier for this particular domain by looking at accuracy. Tokenization is used to transfer the input string into a word vector, stemming is used for extracting the root of the words, while the feature selection technique namely Term Frequency - Inverse Document Frequency is used to fetch the essential word and lastly classification algorithms are used to classify the movie as positive or negative.

Movie reviews help users decide if the movie is worth their time. Users can save time by not reading all of the reviews for a movie by using a summary of all the reviews to make this decision. Critics frequently publish reviews and ratings on websites that assess movies, which aids audiences in deciding whether or not to watch the film. Based on their reviews, sentiment analysis can infer the attitude of the critics. A movie review's sentiment can be analyzed to determine if it is good or negative, and this affects the movie's total score. Therefore, since the machine learns through training and evaluating the data, it is possible to automate the process of determining whether a review is positive or negative. Sentiment analysis is frequently used to get insightful data about reviews and categorize them as positive or negative. Various sentiment analysis models can be developed for textual movie reviews utilizing various feature extraction methods like count vectorization, TF-IDF, and Word2Vec etc. and feature selection methods like Mutual Information Gain and Chi-square etc. , which are very popular in the field of Natural Language Processing.

Sentiment analysis can not only be restricted to movie reviews though. It can find its applications on various other types of data. User-centric design web servers store a lot of user-generated data. Instead of being passive consumers, users now actively contribute to the creation of web content. The Web currently includes a significant amount of social media. According to the data, social media is used by four out of every five Internet users. Social media user contributions include tweets, blog entries, reviews, and uploads of photos and videos, among other things. Unstructured text makes up a sizable portion of the material on the Internet. Reviews and postings that reflect opinions on social media are a significant and intriguing topic worth investigating and using. As the availability of opinion resources such as movie reviews, product evaluations, blog reviews, and social network tweets has increased, a new difficult issue has emerged: mining massive volumes of texts and developing appropriate algorithms to interpret other people's opinions. Companies looking to learn customer opinions about their goods or services might benefit greatly from this information. This input enables them to make wise selections. The reviews and opinions gleaned from them are valuable for people in addition to being useful for businesses. For instance, evaluations of hotels in a city may assist a visitor in finding a suitable hotel there. Similar to this, movie reviews assist other users in determining whether or not to see a certain film.

The History of the Sentiment Analysis problem is quite fascinating as given by the authors in [11]. The need to know what others think is perhaps almost as old as vocal communication. In the past, leaders have been curious in the viewpoints of their followers in order to either anticipate resistance or boost their popularity. There are instances of attempts to identify internal dissension dating back to Ancient Greece. These topics are found in both Eastern and Western ancient writings. The first decades of the 20th century saw attempts to measure and quantify public opinion using surveys , and a scholarly magazine on public opinion was founded in 1937. Many issues that affect the application of sentiment analysis, such as irony identification and multi-lingual support, have been addressed through research that crosses the boundaries of sentiment analysis and natural language processing. A further development in the study of emotions is the endeavor to distinguish between negative emotions like rage and sadness and move beyond simple polarity recognition.

With the introduction of Web 2.0, a number of sites such Facebook, Twitter, LinkedIn, IMDB and Instagram enable users to share their opinions on a wide range of subjects, from entertainment to education. These platforms have a massive quantity of information in the form of tweets, blogs, status updates, articles, etc. Sentiment analysis analyses text, reviews, and postings that are available online on various platforms to detect the polarity of emotions including happiness, sadness, grief, hatred, rage, and affection as well as opinions. Opinion mining determines the text's attitude toward a specific content source. The usage of slang terms, misspellings, abbreviations, repeated characters, regional language, and new, emerging emoticons hamper sentiment analysis. The challenge of determining the suitable sentiment for each word is therefore important. One of the most active study fields is sentiment analysis, which is also extensively researched in data mining. Due to the fact that opinions are at the heart of most human actions and behaviors, sentiment analysis is used in practically every economic and social area.

A variety of methodologies, methods, and tools are used in sentiment analysis to identify and extract subjective information from language, such as opinions and attitudes. Sentiment analysis has traditionally focused on opinion polarity, or whether a person has a favorable, neutral, or negative view about something. A product or service with a public online review has traditionally been the subject of sentiment analysis. This may account for the frequent conflation between sentiment analysis with opinion mining, even though it is believed that it is more realistic to conceive of feelings as emotionally charged opinions.

Authors suggested a method in [13] for word vector representation-based sentiment lexicon creation. The use of SVM in Sentiment Analysis with various data sources was described by the authors of [14]. An approach based on a sentiment score vector for SVM was suggested by the authors in [15]. Hu and Liu's research produced the product features and a product-based overview in [16]. The feedback data from the worldwide support services survey was used by the authors of [17].

Sentiment analysis is used in a variety of ways on a dataset of movie reviews. Authors of [18] divided Machine Learning- and lexical-based Sentiment Analysis approaches into broad categories. Sentiment Analysis has been completed by Yonas Woldemariam using Machine Learning and lexicon-based methodologies. He made advantage of the lexicon-based model in the Apache Hadoop framework [19]. The entity-level sentiment analysis of issue comments was explored by the authors of [20]. The enrichment of the sentiment lexicon dictionary using word2vec was discussed by the authors of [21]. They use SentiwordNet to expand their opinion. Authors discuss view-level Sentiment Analysis on ecommerce data in [22]. [23] asserts that Sentiment Analysis may be used to determine the polarity of customer feedback across a variety of characteristics.

Because of its effectiveness, sentiment analysis is fairly common. For sentiment analysis, tens of thousands of documents can be analyzed. It has a variety of uses since it is an effective method that offers excellent accuracy:

* Making a Good Decision When Buying Goods or Services: Making a good decision when buying goods or services is no longer a tough process. People can quickly assess reviews and opinions of any good or service and compare rival companies by using sentiment analysis.
* Product or service quality improvement: By using opinion mining, manufacturers may gather consumer feedback on their goods or services, whether positive or negative, and then improve and raise the bar on their level of quality.
* Making a decision: The opinions, feelings, and sentiments of other people are crucial considerations. Users read comments and reviews of a product before purchasing it, whether it be a book, outfit, or electronic gadget. These reviews have a significant influence on users' decisions.
* Purchasing Goods or Service: While purchasing a merchandise or service we must take a proper option which is not a tough process anymore. By sentiment analysis, consumers may readily analyze evaluations and opinions of any product or service and can effortlessly compare the rival companies.
* Marketing research: The findings from sentiment analysis methods can be applied in this field. This method may be used to study customer perceptions of certain goods, services, or new government regulations.
* Flame detection: Sentiment analysis makes it simple to monitor newsgroups, blogs, and social media. This method may identify rude, haughty, or overly emotional language in tweets, postings, forums, and blogs on the internet.

The various different Machine Learning algorithms that were normally chosen by the researchers in [3] and to find a solution for this problem-

1. K- Nearest Neighbor-

The simplest machine learning algorithm is KNN. The underlying idea of this method is to predict the label from a set of training samples that are closest to the new point. The number of samples may depend on the local point density or may be a user-defined constant. Any metric measure of distance is acceptable. The most popular method for determining the separation between two points is the standard Euclidean distance. In several classification and regression issues, such as the processing of satellite images or handwritten numbers, the Nearest Neighbors have proved effective.

1. Naive Bayes-

It is a method built on the Bayes Theorem. The Naive Bayes classifier makes the assumption that a certain feature's existence in a class has no bearing on the presence of any other feature. This model is simple to create and is very beneficial for really big datasets. In addition to being straightforward, Naive Bayes is known to perform better than extremely complex classification techniques.

1. Random Forest-

The learning technique for classification and regression is called Random Forests. At training time, it creates a number of decision trees. It sends the new case to each of the trees to categorize it. Each tree produces a class after doing categorization. The majority voting method is used to determine the output class, which is the largest group of classes that are comparable to one another and are produced by different trees. Both professionals and layman people may easily learn how to utilize Random Forests as minimal programming or research is needed. It is simple to use even for those without a background in statistics.

1. Logistic Regression-

In contrast to linear regression, which fits data to a line, logistic regression attempts to fit a "S"-shaped logistic function to the data (Sigmoid Function). Despite the word "regression" being in the name, this method is really utilized for classification. Its ability to categorize data using both continuous and discrete metrics makes logistic regressions a well-liked machine learning technique. Maximum Likelihood is a technique used in logistic regression to fit data. It may be used to classify samples and can classify samples using many types of data. Additionally, it may be used to evaluate whether factors are effective in classifying samples.

1. Support Vector Machine-

SVM is a classification and regression algorithm. To do the classifications, it creates a hyperplane or collection of hyperplanes in infinite dimensional space. It examines the data set's extremes and delineates a decision boundary (Hyperplane). SVM is renowned for its effective operation. It determines the separation between the two provided observations, after which it searches for a decision boundary to determine the separation between the nearest individuals in each class. SVM are resilient when the model is overfit. SVM from the scikit-learn toolkit was used in this case. Both dense and sparse sample vectors are acceptable inputs for scikit-support learn's vector machines.

This article will be majorly focusing on the the conductive research and analysis of several research papers to conclude a study on the Sentiment Analysis of IMDB movie reviews using TF-IDF. The solution proposed by us includes the use of the two well known Machine Learning algorithms Logistic regression and Naive Bayes Classification Algorithm. The IMDB dataset which would be used contains 50,000 reviews. It contains 25000 positive reviews and 25000 negative reviews. The data goes through preprocessing for Data Extraction. This includes Tokenization, Removal of Stop words, Stemming, Lemmatization etc. The proposed solution uses the existing Natural Language Algorithm known as Term Frequency-Inverse Document Frequency (TF-IDF) for Feature Extraction. This method counts the number of words in a collection of documents. Each word is typically given a score to indicate how important it is to the document and corpus. Information retrieval and text mining applications frequently use this method. Various other research studies were analyzed to compare existingly used models and how well they work. After this the data is divided into training and testing sets, the training sets being used to develop the model while the testing set being used to evaluate the model.

The flow of the rest of the paper starts from the New Terms discussed ahead. Various new terms that would be discussed in the paper and are quite common with the domain problem of Sentiment Analysis are explained. Following that is the Architecture of the proposed solution. This contains the structured flow diagram of the various steps in which the project will be carried out to create the model and evaluate it. The next section is the Contribution where several base research papers are discussed with their Methods proposed and the evaluation of the methods proposed. Each base paper in this section is analyzed in detail. After this section comes the Evaluation methods that contain the most well-known evaluation techniques applied to the given Sentiment Analysis problem. Along with the evaluation methods, this section also contains the widely used datasets that support the evaluation methodology. The comparison of base papers is done in the following part. Different components are compared between the approaches/methods that are used in the respective base paper. The discussion and finding of the base papers are also included along with their respective drawbacks. The Conclusion and Future Works section contains the conclusion of our paper. Followed by this section is the final section of the paper that includes all the references that were utilized in preparing this research paper.

**New Terms**

Stop words: Stop words are the words in a stop list (or stoplist or negative dictionary) which are filtered out (i.e. stopped) before or after processing of natural language data (text) because they are insignificant.[1] There is no single universal list of stop words used by all natural language processing tools, nor any agreed upon rules for identifying stop words, and indeed not all tools even use such a list.

Opinion orientation or polarity: The orientation of an opinion on a feature f represents whether the opinion is positive, negative or neutral .

Lexicon-Based Approaches: Lexicon based method uses sentiment dictionaries with opinion words and matches them with the data to determine polarity. They assign sentiment scores to the opinion words describing how Positive, Negative and Objective the words contained in the dictionary are. Lexicon-based approaches mainly rely on a sentiment lexicon, i.e., a collection of known and precompiled sentiment terms, phrases and even idioms, developed for traditional genres of communication, such as the Opinion Finder lexicon;

Support Vector Machine: Support vector machine analyzes the data, define the decision boundaries and

uses the kernels for computation which are performed in input space. The input data are two sets of

vectors of size m each. Then every data which is represented as a vector is classified into a class. Nextly we find a margin between the two classes that is far from any document. The distance defines the margin of the classifier, maximizing the margin reduces indecisive decisions. SVM also supports classification and regression which are useful for statistical learning theory and it also helps recognize the factors precisely that need to be taken into account, to understand it successfully.

Long Short-Term Memory (LSTM) Classifier :It is based on the Recurrent Neural Network (RNN)

algorithm. RNN has memory but it falls short when data has longer dependencies. In contrast, LSTM uses loops with the addition of gates to maintain a level of relevancy , and keeps pertinent data from vanishing in very long sequences. It elegantly addresses the vanishing gradient problem intrinsic in the RNN model using linear memory units, and certain gates which control the flow of information. LSTM mitigates the short-term memory abound in RNNs and indeed solves a number of problems with long-range pathological temporal dependencies.

Preprocessing: Data that are not well cleaned and organized might lead to false identifications. Hence, data preprocessing is a crucial task in the data mining process. It refers to cleaning up the data from useless information that will not help in the training process and might cause confusion during the classification process.

Vectorization: Vectorization or text embedding is the process of extracting features from text and passing it as an input to the classifier.

**Architecture**

The architecture is given as follows :

IMDB Movie Review Dataset

Data Extraction and Pre-processing

Feature Extraction Using TF-IDF

Training and Building the Model for Classification

Tokenization

Removal of HTML Tags and Noise Text

Text Normalization using Stemming

Lower Casing and Removal of Stopwords

Splitting the Dataset into Training and Testing data

Testing and Evaluating the Model

The IMDB dataset contains 50000 reviews, of which 25000 reviews are positive and the rest are negative. The data is first preprocessed. Data Extraction and Preprocessing includes Tokenization (sentences are divided into meaningful tokens), Removal of HTML tags and noise text (Brackets, numbers etc.), Text Normalization using Stemming ( inflected words are reduced back to their root form ) and finally Lower Casing and Removal of Stopwords ( All the words are converted to lowercase and the duplicates are removed. Stopwords such as ‘I’, ‘the’, ‘an’ etc. are removed from the data as they have no meaning). The next step is Feature Extraction using TF-IDF (Term Frequency-Inverse Document Frequency). This method counts the number of words in a collection of documents. Each word is typically given a score to indicate how important it is to the document and corpus. The data is now split into Training dataset and testing dataset. The training dataset is used to build the models using the two machine learning algorithms Naive Bayes Classification and Logistic Regression. After building and training the models, they are evaluated using the testing data and are evaluated based on their accuracy, precision, recall, and f-measure.

**Contribution**

**Paper : Sentiment Analysis and Classification Based on Textual Reviews**

**Method Proposed :**

A new algorithm called Sentiment Fuzzy Classification algorithm is proposed to improve classification accuracy on the benchmark dataset of Movies reviews.

Text preprocessing-

Text pre-processing techniques are divided into two subcategories.

Tokenization: Textual data comprises a block of characters called tokens. The documents are separated as tokens and used for further processing.

Removal of Stop Words: The term "stop-list" is frequently used to refer to a group or list of stop words. Typically, it is language-specific, however it could include terms. One stop-list for each language may be present in a search engine or other natural language processing system, or there may be one stop-list that is multilingual. The following are some of the more commonly used stop words in English: "a," "of," "the," "I," "it," "you," and "and." These words are classified as 'functional words,' meaningless words. By eliminating the functional terms while analyzing the contents of natural language, the message may be communicated more effectively.

Therefore, it is useful to eliminate terms that appear excessively frequently yet provide no information for the job. All of the stop words in the specific text file won't be loaded if the stop word removal is used. When the dataset is loaded, the stop word removal procedure will be deactivated if it is not applied.

Text transformation-

The total weight of each phrase in the associated sentences is added to determine each sentence's score in the source material. Each term's weight is determined by multiplying its TF and IDF based on adjective words that were taken from Parts of Speech tags. It is said that the TF and IDF are-

TF (t) = Number of times the adjective term occurs in document(d)/Total Number of adjective in document(d)

IDF(t)=log{ND/DF(t)}

Here ND is the total number of documents in the document collection and DF (t) is the number of documents in which the adjective term (t) occurs in the document collection.

Feature selection-

Sentiment analysis utilizes a variety of statistical feature selection techniques for document level categorization. The most straightforward statistical method for feature selection uses the most common words in the corpus as polarity indicators. The bulk of sentiment analysis methods follow a two-step procedure:

Identify the parts of the document to contribute to the positive or negative sentiments.

Join these parts of the document in ways that increase the odds of the document falling into one of these two polar categories.

Sentiment fuzzy classification-

Regarding its conceptual extent, sentiment polarity is ambiguous. The distinction between "positive," "neutral," and "negative" ideas is ambiguous. Then employed the fuzzy set theory to sentiment categorization in order to more effectively manage this inherent fuzziness in sentiment polarity. To do this, first recast sentiment classes as three fuzzy sets, then design membership functions for using already in existence fuzzy distributions. A membership function describes a fuzzy set. These shapes can be anything at all, however they are usually trapezoidal or triangular.

**Conclusion :**

It is challenging for a person to forecast the movie review in sentiment analysis. The current approach uses document-level sentiment categorization to address this. It establishes the positivity or negativity of an opinion document (movie review), as well as its neural emotion. It is possible to roughly categorize the sentiment using the Bag of Words. Parts of speech can be utilized to ensure that the categorization is accurate. On the benchmark dataset of movie reviews, a novel approach dubbed Sentiment Fuzzy Classification is suggested to increase classification accuracy.

**Paper : Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory**

**Method proposed :**

In this paper, a Python based application is utilized to analyze the IMDb dataset. The dataset is tactfully divided into training and testing parts. The training part is used for preparing the LSTM classifier. Afterwards, the testing set is used to quantify the classification precision. Confusion matrix and accuracy results are outlined. The model consists of five main blocks along with few minor components integrated in the system.

Data Division

It is common to divide the dataset into training and testing vectors . The training vector is the set of data that trains the considered classifier. The validation vector is a subset of the training vector that does not necessarily train, but is used to give some insight on the classifier performance. Test data is to evaluate the model accuracy. The split of the training and validation, testing, or both can occur in many ways. However, there is a rule-of-thumb that training gets the most data. A recurrent ratio encountered in various eclectic ML settings is the 80-20 split which gives 80% to the training and 20% to the testing. This ratio finds interesting roots in what is referred to as the Pareto Principle or the law of the vital few and arises in finance and economic theories . In this paper 10k reviews are considered. In order to avoid any biases an equal representation of positive and negative reviews is employed. 80% of this data is used for the training purpose and the 20% is used for the testing purpose. To evade the biases in classification results 10-fold cross-validation is utilized in this

study.

Preprocessing

Data that is not well cleaned and organized might lead to false identifications. Hence, data preprocessing is a crucial task in the data mining process. It refers to cleaning up the data from useless information that will not help in the training process and might cause confusion during the classification process. For the IMDb dataset, several data preprocessing steps are utilized.

* all the symbols such as “?”, “!” are removed.
* all letters in the dataset are converted into lowercase letters.
* all hybrid links are removed from the text.
* stop words such as, me, you, and we are evicted.
* stemming techniques are applied on the text to present the word in its original form after removing prefixes and suffixes.

Vectorization

Vectorization or text embedding is the process of extracting features from text and passing it as an input to the classifier.

Each movie review is encoded “vectorized” into a numeric value. It is achieved by utilizing the genism, a Python library for topic modeling and NLP . Compared to other techniques, the Doc2Vec model proved to deliver high accuracy results with lower computational cost .

Long Short-Term Memory (LSTM) Classifier:

It is based on the Recurrent Neural Network (RNN) algorithm. RNN has memory but it falls short when data has longer dependencies. In contrast, LSTM uses loops with the addition of gates to maintain a level of relevancy , and keeps pertinent data from vanishing in very long sequences. It elegantly addresses the vanishing gradient problem intrinsic in the RNN model using linear memory units, and certain gates which control the flow of information. LSTM mitigates the short-term memory abound in RNNs and indeed solves a number of problems with long-range pathological temporal dependencies.

The LSTM is employed with an adaptive moment estimation (Adam) optimizer. It is an adaptive learning rate optimization algorithm that is particularly designed for training the deep neural networks. The algorithm leverages the power of adaptive learning rates methods to find individual learning rates for each parameter. It uses estimations of first and second moments of gradient to adapt the learning rate for each weight of the neural network . Three layers are employed. In this case, the first layer employs 50 nodes to present the intended

words. The second layer is the LSTM with 101 units of memory and the final layer creates 2 outputs corresponding to the considered classes.

Classification Accuracy

The classification accuracy is used to measure how well the devised model is able to automatically identify the data. It is the percentage of labels that have been correctly classified.

**Evaluation of the proposed method :**

In this paper, a Long Short-Term Memory classifier is used with Adam optimizer to automatically categorize the preprocessed IMDb movie reviews. In total 10k reviews are considered, 5k for positive and 5k for negative sentiments. Results have concluded that the highest accuracy attained by the devised approach is 89.9%.

**Paper : Deep learning for sentiment analysis of movie reviews**

**Method proposed :**

**Binary classification approaches :**

For all binary classification approaches, perform a preprocessing step to clean up the data. This includes removing the HTML tags, unnecessary punctuation, converting to lowercase, and removing stop words if needed.

To convert a cleaned sequence of words to numerical feature vectors following methods were used:

Bag of words (BOW):

The easiest approach to mathematically represent texts is probably using a bag of words. Given a vector vT ∈ Nd to a text T, where vTi  represents the frequency with which the i'th word in the vocabulary appears in the text ‘T’. Our vocabulary, which consists of all terms in the collection of reviews with the exception of very uncommon words (used the 5000 most frequent words), is measured by the letter d. Then fit a classifier to the data after learning the BOW vectors for each review in the labeled training set.

Word2Vec:

Texts may be represented quantitatively by converting each word to a vector. The semantics of words should be preserved by this transformation, therefore if two words have similar meanings, then their vectors should also be similar. The word2vec task's independence from the main goal and lack of dependence on a labeled dataset are both significant features. In order to train word vectors, used the entire corpus of the 75 000 reviews in this instance. In addition to the standard preprocessing processes on the raw reviews, since the word2vec method only accepts sentences as input, so divide paragraphs into sentences in order to train word vectors (since words not in the same sentence as the current word are not part of its relevant context)

Note that for the sentiment analysis a feature vector for each review is needed. This can be done through a variety of methods.

Words to reviews: Averaging

Averaging perhaps the simplest way to assign a feature vector to a set of words (a review) is to average the word vectors of all words. Then ignore the stop words in the averaging process, since they do not contribute to the polarity of the whole paragraph. After assigning a feature vector to all reviews, train a binary classifier using the labeled training set.

Words to reviews: Clustering

Having the words transformed to vectors, find similarities (distances) among different words. This allows us to cluster similar words together. Using the K-means algorithm form clusters of words (with the average size of 5 words per cluster) that are represented by a centroid. Now, each review can be presented as a set (bag) of centroids, similar to the bag of words idea. After learning the bag of centroids vectors for every review in the labeled training set, fit a classifier to the data

Once you have the feature vectors for each review ,use a binary classifier to learn the sentiments.

* Random forest
* Logistic regression
* SVM

Recursive Neural Networks-

A hidden vector of length d is used in RNTN to represent each node in the parse tree (similar to word2vec). The word-vectors are really the hidden vectors at the leaf level, or the vectors that correspond to single words. The hidden vector of a non-leaf node is derived from the hidden vectors of its left and right children using the following procedure:



where assume p is the hidden vector of the parent node, and b and c are the hidden vectors of the children. W has size (d,2d) , and the tensor V has a (d,2d,2d) shape. Function f is the activation function of the neurons, which is assumed to be tangent hyperbolic tanh here. The term involving the tensor V is designed to produce non-linear (quadratic) features. This can help to overcome the issue of negated phrases.

**Evaluation of the proposed method :**

In the first part, a dataset provided by Kaggle was used and applied to the bag of words, and skip gram word2vec models to represent words numerically. Then used several classifiers, including random forest, SVM, and logistic regression to perform the binary classification task. Vector averaging, and clustering was used to produce the aggregated feature vectors. However, these suffer from losing the order of words in sentences. This motivated the second part of our work.

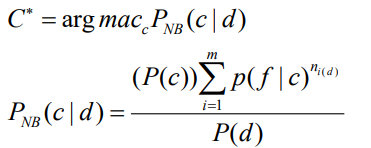
In the second part, implemented the recursive neural tensor networks to train a multi-class sentiment analyzer. The training of standard RNTN was computationally very expensive. So low-rank RNTN was introduced. Then showed that the low-rank RNTN can achieve comparable accuracies to that of standard RNTN much faster. This better training performance of the low-rank RNTN enables it to train several different models, and use them for ensemble averaging. A 1.5% accuracy improvement was achieved by ensemble-averaging.

**Paper : Sentiment Analysis of Twitter Data: A Survey of Techniques**

**Method proposed:**

In this paper, provide a survey and a comparative analysis of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, provide research on twitter data streams , also discuss general challenges and applications of Sentiment Analysis on Twitter.

Naive Bayes: It is a probabilistic classifier and can learn the pattern of examining a set of documents that has been categorized . It compares the contents with the list of words to classify the documents to their right category or class. Let d be the tweet and c\* be a class that is assigned to d, where



From the above equation, „f‟ is a „feature‟, count of feature (fi) is denoted with ni(d) and is present in d which represents a tweet. Here, m denotes no. of features.

Parameters P(c) and P(f|c) are computed through maximum likelihood estimates, and smoothing is utilized for unseen features. To train and classify using Naïve Bayes Machine Learning technique Python is used with the NLTK library .

Machine Learning Approaches : Machine learning based approach uses classification techniques to classify text into classes. There are mainly two types of machine learning techniques

1. Unsupervised learning: It does not consist of a category and they do not provide with

the correct targets at all and therefore rely on clustering.

1. Supervised learning: It is based on labeled dataset and thus the labels are provided to

the model during the process. These labeled dataset are trained

to get meaningful outputs when encountered during decision making.

The success of both these learning methods mainly depends on the selection and extraction of the specific set of features used to detect sentiment.

The machine learning approach applicable to sentiment analysis mainly belongs to supervised classification. In a machine learning techniques, two sets of data are needed:

1. Training Set

2. Test Set.

A number of machine learning techniques have been formulated to classify the tweets into classes. Machine learning techniques like Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have achieved great success in sentiment analysis.

Machine learning starts with collecting training dataset. The classifier is then trained using the training data. Once a supervised classification technique is selected, an important decision to make is feature. can tell us how documents are represented.

Lexicon-Based Approaches: Lexicon based method uses sentiment dictionaries with opinion words and matches them with the data to determine polarity. They assign sentiment scores to the opinion words describing how Positive, Negative and Objective the words contained in the dictionary are.

Lexicon-based approaches mainly rely on a sentiment lexicon, i.e., a collection of known and precompiled sentiment terms, phrases and even idioms, developed for traditional genres of communication, such as the Opinion Finder lexicon;

**Evaluation of the proposed method :**

In this paper, a survey is provided and comparative study of existing techniques for opinion mining including machine learning and lexicon-based approaches, together with cross domain and cross-lingual methods and some evaluation metrics. Research results show that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases, which require few effort in human-labeled document .The effects of various features on classifier were also studied. It can be concluded that with cleaner data, more accurate results can be obtained. The use of bigram model provides better sentiment.

**Paper : Sentiment Analysis of Movie Reviews using Machine Learning Techniques**

**Method proposed :**

Naive Bayes :

It is a method built on the Bayes Theorem. The Naive Bayes classifier makes the assumption that a certain feature's existence in a class has no bearing on the presence of any other feature. This model is simple to create and is very beneficial for really big datasets. In addition to being straightforward, Naive Bayes is known to perform better than extremely complex classification techniques.

K-Nearest Neighbor :

The simplest machine learning algorithm is K-NN. The underlying idea of this technique is to predict the label using a set of training samples that are physically closest to the new location. The number of samples may depend on the local point density or may be a user-defined constant. Any metric measure of distance is acceptable. The method used to determine the separation between two places that is most frequently used is the standard Euclidean distance. Many classification and regression issues, such as those involving handwritten numbers or the processing of satellite images, have been solved successfully using the Nearest Neighbors algorithm.

Random forest :

The learning technique for classification and regression is called Random Forests. At training time, it creates a number of decision trees. It sends the new case to each of the trees to categorize it. Each tree produces a class after doing categorization. The majority voting method is used to determine the output class, which is the largest group of classes that are comparable to one another and are produced by different trees.

Both professionals and laypeople may easily learn how to utilize Random Forests as minimal programming or research is needed. It is simple to use even for those without a background in statistics.

Steps used for performing experiment-

Step 1: Importing the dataset in WEKA. The first step performed was to import the dataset into the WEKA tool. To perform this step a simple import procedure for textual datasets called TextDirectoryLoader component was used.

Step 2: After importing the dataset it is converted and saved in the ARFF format.

Step 3: After that, a relation is created by containing 2000 instances and two attributes “Text” and “Class”.

Step 4: Then the StringToWordVector filter is applied.

Step 5: Then the AttributeSelection filter is applied.

Step 6: After applying the AttributeSelection filter, the results are obtained.

Step 7: Three algorithms are performed on the data generated from above steps. The three algorithms are Naïve Bayes, K Nearest Neighbor, and Random Forest.

**Evaluation of the proposed method :**

To determine the polarity of the tweets in this study, a variety of methodologies were employed. Random Forest, K-Nearest Neighbor, and Naive Bayes were the algorithms used. The Naive Bayes classifier produced the best results. The accuracy of the Naive Bayes classifier was 81.45%, that of the Random Forest classifier was 78.65%, and that of the K-Nearest Neighbor classifier was 55.30%.

Since just a few algorithms were evaluated, it is necessary to test other algorithms or develop hybrid approaches in order to improve the results' accuracy.

Finding the review polarity may be helpful in a variety of areas. Without the need for the user to read through individual reviews, intelligent systems can be developed that can give users comprehensive reviews of movies, products, services, etc. The user can then make decisions based on the results that the intelligent systems provide without having to read individual reviews.

**Paper : Sentiment Analysis of Movie Reviews A new Feature-based Heuristic for Aspect-level Sentiment Classification**

**Method proposed :**

Implemented the SentiWordNet based algorithmic formulation for both document-level and aspect-level sentiment classification.

Document level sentiment classification-

The goal of the document-level sentiment classification is to assign a positive or negative sentiment to the complete document (such as a review). SentiWordNet-based strategies aim to extract terms with the necessary POS label from the review document's term profile (such as adjectives, adverbs or verbs). This demonstrates unequivocally the need to apply the review text to a POS tagger that tags each phrase that appears in the review text before using the SentiWordNet-based formulation. The emotion score of each extracted phrase is then acquired from the SentiWordNet library once a few selected terms (with the necessary POS tag) are extracted. The results are then combined using a weighting and aggregation approach to account for the ratings for all retrieved words in the review. Thus two key issues are to decide (a) which POS tags should be extracted, and (b) how to decide the weightage of scores of different POS tags extracted while computing the aggregate score.

Aspect level sentiment analysis-

A fair indicator of the degree of positivity or negativity indicated in a review is the document-level sentiment categorization. However, it could be a good idea in some domains to investigate the reviewer's emotion towards different features of the item in that area as indicated in that review. It may be challenging and improper to insist on an overall document-level sentiment polarity indicated in a review for the item because the majority of evaluations contain a mix of positive and negative emotion about various components of the item. The sentiment categorization at the document level is not a full, appropriate, or comprehensive metric for a deep study of the positive and negative features of the item under evaluation. Examine an item's advantages and disadvantages using aspect-level sentiment analysis. But this sort of study frequently has a domain-specific focus. The aspect-level sentiment analysis involves the following: (a) identifying which aspects are to be analyzed, (b) locating the opinionated content about that aspect in the review, and (c) determining the sentiment polarity of views expressed about an aspect

**Evaluation of the proposed method :**

Their experimental work makes two important contributions. First, it explores the use of ‘Adverb+Verb’ combined with ‘Adverb+Adjective’ combined for document-level sentiment classification of a review. Second, it proposes a new feature-based heuristic scheme for aspect-level sentiment classification of a movie. The aspect level sentiment classification produces an accurate and easy to understand sentiment profile of a movie on various aspects of interest. Interestingly, the aspect-level sentiment profile result is congruent to the document level sentiment classification of reviews of a movie. Though, the aspect-level sentiment profile produces a more focused and accurate sentiment summary of a particular movie and is more useful for the users.

The aspect-level sentiment analysis algorithmic formulation designed in this research paper is a novel and unique way of obtaining a complete sentiment profile of a movie from multiple reviews on different aspects of evaluation. The resultant sentiment profile is informative, easy to understand, and extremely useful for users. Moreover, the algorithmic formulation used for aspect-level sentiment profile is very simple, quick to implement, fast in producing results and does not require any previous training. It can be used on the run and produces very useful and detailed sentiment profiles of a movie on different aspects of interest. This part of the implementation can also be used as an add-on step in movie recommendation systems that use content-filtering, collaborative-filtering or hybrid approaches. This aspect level sentiment profiling is a valuable form of sentiment analysis and subsequent exploitation of information expressed by a large number of users about a particular movie.

**Paper : Sentiment Analysis on Movie Review Data Using Machine Learning Approach.**

**Method proposed :**

In this study, information on movie reviews has been gathered and analyzed using five different types of machine learning classifiers. Therefore, Bernoulli Naive Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), and Multinomial Naive Bayes(MNB) are the classifiers that are taken into consideration. With this method, the dataset is split into the train set and the test set. A data set is first gathered from the movie review website. Next, preprocess the data using an NLP tool. Then, after creating a feature vector the data set is trained using the five ML classifiers.

Text Preprocessing :

Text Preprocessing on the dataset used included removal of URLs and HTML tags, Brackets, Numbers; Tokenization, Omission of Punctuation Marks and Stopwords, and Stemming.

The reviews in the dataset contain HTML tags like <br>, <b> etc. along with some links too. This needs to be removed as they do not play any role in contributing to the sentiment of a review. Similarly brackets and numbers have no meaning in sentiment classification. They are treated as noise in the dataset and thus need to be cleaned. Thus each review gets cleaned to obtain pure language text which would be divided into smaller components using Tokenization. The text is turned into a number of sentences and the sentences in turn are divided into words . Punctuation marks like quotation , semi-colon etc. also don’t play any role in Sentiment Analysis and are thus cleaned off. There may be the same words that are treated differently due to their case (uppercase or lowercase). Thus all the words are converted to lowercase and the duplicates are removed. Stopwords such as ‘I’, ‘the’, ‘an’ etc. also are removed from the data as they have no meaning. Finally, the text is normalized using Stemming. In this process the inflected words are reduced back to their root form.

Feature Extraction :

Each review is categorized as either good or negative. Every review is regarded as a straightforward document. Every document is identified in the unigram model by its primary words, with positive and negative primary words being used for the corresponding positive and negative documents. To obtain a more precise sentiment, additionally incorporate parts of speech (PoS) tags. Opinion words influence emotions. The relative emotion scores maintain both positive and negative sentiments. This makes these terms a crucial component of sentiment analysis. Dimensionality is a concept used to describe the number of features. Positive and negative keywords are signed in each document as pos and neg, accordingly. After that, the PoS are added to the sentence. Every word has a feature as a result.

ML approaches are for categorization after constructing the feature vector. As training and testing sets, 1400 and 600 movie reviews are used. The first set is used to train a classifier, and the accuracy of that classifier is determined using the second set. Five different major classifier types that are used are: MNB, BNB, SVM, ME, and DT. Python is used to implement each and every classifier. 2000 movie reviews are included in the dataset, 1000 of which are unfavorable and the rest good. True positive (U), false positive (V), true negative (X), and false negative (Y) are the terminology utilized for analysis. The first and second terms in this sentence imply that the review is actually good and negative, respectively, yet both are highlighted as positive terms. The third and fourth terms suggest that the review is actually negative or positive depending on whatever is presented as the third or fourth term. From the terms mentioned above, accuracy, recall, precision, and F-score are derived.

A few lines on the above used Classifiers are given below .

Naïve Bayes Classifier :

It considers that every feature is distinct from one another. NB classifiers are a collection of classifications algorithm which is not a standalone algorithm but a family of algorithm

Support Vector Machine :

Support-vector machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.

Maximum Entropy Classifier :

It uses a probabilistic concept and does not guess that the characteristics are conditionally autonomous of each other.

Decision Tree Classifier :

DT differentiates those records having many features checking the property from the root and vertex in a tree. All terminal vertices are assigned a class label positive or negative.

**Evaluation of the proposed method :**

According to the research conducted, MNB outperforms the competition in terms of accuracy, precision, and F-score, while SVM displays stronger recall. Additionally, it demonstrates that the BNB Classifier outperforms earlier experiments using this classifier in terms of accuracy. However, MNB has been found to get worse results in case of smaller training datasets. Comparatively, multinomial NB has higher accuracy. It achieves accuracy of 88.50% while Bernoulli NB, SVM, Maximum Entropy, and Decision tree all achieve accuracy of 87.50%, 87.33%, 60.67%, and 80.17% respectively. Although the SVM has a higher recall, Multinomial NB also has good precision and F-score. Additionally, it demonstrates that the Bernoulli Naive Bayes Classifier outperforms earlier research on this classifier in terms of accuracy. It has strong recall, precision, and f-score. The Maximum Entropy classifier performs poorly compared to other classifiers. It is less precise and has a lower f-score than the others. But compared to Multinomial NB and Decision Tree, the precision is better. The tuning of parameters affects all classifiers. Multinomial Naive Bayes classifier is superior to SVM, according to the results, although this is only true for a subset of parameters because multinomial NB performs poorly when the training dataset is small. The aforementioned result demonstrates the caliber of the features vector used for the movie review data.

**Paper : A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis**

**Method proposed :**

Convolutional neural networks (CNN) and the Long Short-Term Memory (LSTM) model have been applied to a variety of Natural Language Processing (NLP) tasks with impressive and successful outcomes. Using convolutional layers and max-pooling layers, the CNN model effectively recovers higher level information. The long-term dependencies between word sequences are capable of being captured by the LSTM model. In this study, a hybrid CNN-LSTM model is suggested called the Hybrid CNN-LSTM Model to solve the sentiment analysis issue. First, the first word embeddings are trained using the Word to Vector (Word2Vc) technique. Word2Vc converts text strings into a vector of numerical values, calculates the distance between words, and creates word groups based on similarity in meaning. The proposed model includes a collection of features that are retrieved by convolution and global max-pooling layers with long-term dependencies afterward embedding. In order to increase accuracy, the suggested model also makes use of dropout technology, normalizing, and a rectified linear unit.

Word2Vec:

Word2Vec is a deep learning model and it was proposed by Google in 2013. The Word2Vec model creates vector numeric values using sentences of words. Word2Vec technique was used to initialize the words as vector space and word2vec use skip gram and bag-of-words technique to convert the words in vector representation.

Word2Vc can provide incredibly precise estimations about a word's meaning given the vast quantity of data, usage, and content. As a result, Word2Vc operates quickly even for a large dataset. It trains using a dataset from Google News. The Applications and Multimedia Tools Pre-trained vectors can be found in the Google News dataset. Our model uses the dropout strategy to avoid overfitting and to remove the unnecessary data from the network to improve performance. It chooses the top words from the Google News dataset that are connected to "good," "poor," and "awful" using the dropout technique. On one side of the graph, words like "awful," "bad," and "horror" vanish, whereas words like "excellent" and "great" appear in the second group.

Convolutional neural network model(CNN) :

A unique kind of neural network used in the field of image processing is the CNN. However, text classification has successfully used the CNN model. CNN layers are known as feature maps because in a CNN model, a portion of the input to its previous levels is connected via a convolutional layer. The polling layer in the CNN model is used to lessen the computational complexity. The CNN polling techniques minimize the output size from one stack layer to the next while preserving critical information. Although there are other polling methods available, max-polling is most frequently employed when the pooling window has the max value element. The output of the polling layer is fed into the flattened layer, which then maps it to the following layers. CNN normally has a completely connected final layer.

Proposed model using CNN :

Word2Vc converts the text into vector numerical values before the CNN model is used to train the vector numerical values. A fully connected layer, three pooling layers, and three convolutional layers are all used. Tensor flow is an open-source Python package that is utilized for experimentation. For accuracy increase in CNN, pooling layers, convolutional layers, dropout out layers, and RLU are used. Dropout is a crucial deep learning technique because it stops machine learning algorithms from over-fitting. Dropout algorithms omit the neurons that do not participate during backpropagation. To stop neurons from co-adapting, the dropout strategy eliminates neurons during training. Each hidden neuron has a 0.5 probability of producing an output.

Preprocessing Layer : The first layer focuses on PreProcessing the data which converts the raw dataset into a useful and organized dataset for further use.

Embedding Layer : The second layer is the Embedding layer. A distinct and meaningful sequence of words is provided by the preprocessed dataset, and each word has a unique ID. The learning layer for embedding initializes the words with random weights and learns the embedding to embed all of the words in the training dataset. It is generally used to learn how to embed words so they can be utilized in other models later. This layer is used in a variety of ways. In this study, a pre-trained Word2Vc model was employed to embed words.

Convolution layer : Three convolutional layers, three pooling layers, and a fully linked layer make up the final three layers of the CNN model. Convolutional layers receive the word from the embedding layer in the form of sentences. Pooling layers are used in the convolution layer to convolve the input. By reducing the representation of input sentences and input parameters, pooling layers also aid to minimize network computation and control overfitting.

Global max-pooling : Global max-pooling has been applied at the end of network layers. After using many convolution layers, it offers the overall best results throughout the entire network.

Activation Function : RELU activation function is used in the model. At negative numbers, RELU returns zero, while it rises with positive values.

Dense layer : The convolutional layers' collected features are classified using the dense layer, also known as the fully connected layer. Every current input (or neuron) in the layer of the network is connected to every input (or neuron) in the layer that comes after it using a dense layer.

SoftMax : A function called SoftMax is frequently included in the neural network's top layer. It converts the random results' average into forms of 1 and 0.

Proposed Model using LSTM :

This model includes the Embedding Layer, LSTM layer, Dropout Techniques, Dense Layer and SoftMax. The Embedding layer, Dense Layer and SoftMax perform the same as in the proposed model using CNN that has been explained in detail above. The LSTM model is used for layers of RNN after embedding . Three types of gates and cells are used by LSTM for handling information flow in the network.

**Evaluation of the proposed method :**

The two common deep learning models (CNN, LSTM) and proposed Hybrid CNN-LSTM Model were applied on two datasets IMDB and Amazon. Many experiments on the IMDB sentiment analysis dataset were performed to attempt a candid comparison with competitive techniques. The f-measure score is improved by the proposed hybrid model by up to 4-8% when compared to CNN and LSTM separately, according to the preliminary findings on the IMDB dataset. Many experiments were also performed on the amazon movie reviews dataset and the results were compared with the traditional models. The results showed that performance of the proposed deep learning models is better than traditional machine learning techniques. The dropout technique used in the model improved execution time. Learning how to extract characteristics from the data with the aid of CNN. To capture long-term dependencies, a neural network also needs many convolution layers, and the difficulty of doing so increases as the duration of the input sequence rises. In essence, it leads to a convolution neural network layer that is extremely deep. The long-term dependencies between word sequences are capable of being captured by the LSTM model. A hybrid CNN-LSTM model was suggested for sentiment analysis in this study. In terms of accuracy, the proposed hybrid CNN-LSTM model outperformed the single CNN and LSTM models on two benchmark datasets of movie reviews. Comparing the proposed hybrid CNN-LSTM model to more established machine learning and deep learning models, it achieved 91% accuracy.

**Paper : Sentiment analysis using product review data**

**Method proposed :**

This paper tackles a fundamental problem of sentiment analysis, namely sentiment polarity categorization . Our contributions mainly fall into Phase 2 and 3.

In Phase 2:

1) An algorithm is proposed and implemented for negation phrases identification;

2) A mathematical approach is proposed for sentiment score computation;

3) A feature vector generation method is presented for sentiment polarity categorization.

In Phase 3:

1) Two sentiment polarity categorization experiments are respectively performed based on sentence level and review level;

2) Performance of three classification models are evaluated and compared based on their experimental results.

Algorithm : Negation Phrases Identification

Require: Tagged Sentences, Negative Prefixes

Ensure: NOA Phrases, NOV Phrases

1: for every Tagged Sentences do

2: for i/i + 1 as every word/tag pair do

3: if i + 1 is a Negative Prefix then

4: if there is an adjective tag or a verb tag in next pair then

5: NOA Phrases ← (i, i + 2)

6: NOV Phrases ← (i, i + 2)

7: else

8: if there is an adjective tag or a verb tag in the pair after next then

9: NOA Phrases ← (i, i + 2, i + 4)

10: NOV Phrases ← (i, i + 2, i + 4)

11: end if

12: end if

13: end if

14: end for

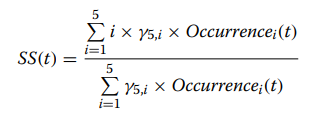
15: end for

16: return NOA Phrases, NOV Phrase

Sentiment score computation for sentiment tokens

A sentiment token is a word or a phrase that conveys sentiment. Given those sentiment words proposed in , a word token consists of a positive (negative) word and its part-of-speech tag. In total, 11,478 word tokens were selected with each of them occurring at least 30 times throughout the dataset. For phrase tokens, 3,023 phrases were selected of the 21,586 identified sentiment phrases, which each of the 3,023 phrases also has an

occurrence that is no less than 30. Given a token t, the formula for t’s sentiment score (SS) computation is given as



Occurrence i(t) is t’s number of occurrences in i-star reviews, where i = 1, ..., 5. The dataset is not balanced indicating that different numbers of reviews were collected for each star level. Since 5-star reviews take a majority amount through the entire dataset, a ratio, γ5,i was introduced, which is defined as:



In the equation , the numerator is the number of 5-star reviews and the denominator is the number of i-star reviews, where i = 1, ..., 5. Therefore, if the dataset were balanced,γ5,i would be set to 1 for every i. Consequently, every sentiment score should fall into the interval of [1,5]. For positive word tokens, we expect that the median of their sentiment scores should exceed 3, which is the point of being. For negative word tokens, it is to expect that the median should be less than 2.

Random forest

The random forest classifier was chosen due to its superior performance over a single decision tree with respect to accuracy. It is essentially an ensemble method based on bagging. The classifier works as follows: Given D, the classifier firstly creates k bootstrap samples of D, with each of the samples denoting as Di. A Di has the same number of tuples as D that are sampled with replacement from D. By sampling with replacement, it means

that some of the original tuples of D may not be included in Di, whereas others may occur more than once. The classifier then constructs a decision tree based on each Di. As a result, a “forest" that consists of k decision trees is formed. To classify an unknown tuple,X, each tree returns its class prediction counting as one vote. The final decision of X’s class is assigned to the one that has the most votes.

The decision tree algorithm implemented in scikit-learn is CART (Classification and Regression Trees). CART uses Gini index for its tree induction. For D, the Gini index is computed as:



where pi is the probability that a tuple in D belongs to class Ci. The Gini index measures

the impurity of D. The lower the index value is, the better D was partitioned. For the

detailed descriptions of CART

**Evaluation of the proposed method :**

This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed.

**Evaluation methods (1.5)**

Sentiment Analysis is a classification problem as the queries; which can be either tweets or reviews or any other data, which would be classified into different sentiment labels like positive sentiment and negative sentiment. Thus, being a classification problem, the most common evaluation methods that are used are :

1. Accuracy :

The easiest performance metric to understand is accuracy, which is just the proportion of properly predicted observations to all observations.

1. Precision :

In terms of positive observations, precision is the proportion of accurately anticipated observations to all predicted positive observations. High precision signifies a low false positive rate. Thus we can evaluate how exact the positive predictions are.

1. Recall :

Recall is the percentage of accurately predicted positive observations to all of the observations in the actual class. It measures how well the model can recall the positive class that is the recall gauges how well the model can identify positive samples. The more positive samples that are identified, the larger the recall.

1. F1 score :

The weighted average of Precision and Recall is the F1 Score. Therefore, both false positives and false negatives are considered while calculating this score. Although F1 is generally more beneficial than accuracy, especially if you have an uneven class distribution, it is not intuitively as simple to understand as accuracy. The relationship between the complementary measures of precision and recall is inverse. The F1 score can be used to integrate precision and recall into a single statistic if both are relevant to us.

There have been numerous analyses and labeled sentiment datasets produced, particularly for Twitter messages and Amazon product reviews that have been used to support the above given Evaluations methodologies.

The most well-known and prevalent datasets that support the Evaluations methodologies are :

1. Stanford Twitter Sentiment :

This is also known as the sentiment140 dataset. 1,600,000 tweets that were extracted using the twitter api are included. The tweets can be used to gauge sentiment because they have been marked (0 = negative, 2 = neutral, and 4 = positive). For instance, a tweet is labeled as positive if it contains:),:-),:),:D, or =) and is labeled as negative if it contains:(,:-(, or: (.

There are 6 fields that are included in the dataset:

1. target: the polarity of the tweet (0 stands for negative, 2 stands for neutral, 4 stands for positive)
2. ids: The id of the tweet
3. date: the date of the tweet
4. flag: The query. Note : If there is no query, then this value is NO\_QUERY.
5. user: the user that tweeted
6. text: the text of the tweet
7. Sentiment Strength Twitter Dataset :

4,242 tweets that have been manually identified as having a strong positive or negative sentiment make up this dataset. A number between -1 (not negative) and -5, for instance, denotes a weak point (extremely negative). A positive strength is a number between 1 and 5, just like a negative strength (extremely positive).For Classification purposes, tweets in this dataset are re-annotated with sentiment labels (negative, positive, neutral) rather than sentiment strengths. Each tweet receives a single sentiment label based on the two principles adopted from SentiStrength's operation:

1. If a tweet's negative to positive strength ratio equals one in absolute terms, it is regarded as neutral.
2. A tweet is positive if the strength of the positive sentiment is 1.5 times that of the negative, and it is negative otherwise.

The final dataset contains 1,037 negative tweets, 1,252 positive tweets and 1,953 neutral tweets.

1. Large Movie Review Dataset :

This dataset has 50K movie reviews for natural language processing or Text analytics. This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. It is a set of 25,000 highly polar movie reviews for training and 25,000 for testing.

There is additional unlabeled data for use as well. Raw text and already processed bag of words formats are provided.

There are 2 fields that are included in the dataset:

1. Review: highly polar movie reviews for training
2. Sentiment: two unique values (positive or negative\_
3. Amazon Reviews for Sentiment Analysis :

For the purpose of training fastText for sentiment analysis, this dataset comprises a few million Amazon customer reviews (input text) and star ratings (output labels).

The concept is to train a dataset on a basic laptop in a matter of minutes using real business data that isn't just a toy.

Data is in the following format to be used with the fastText supervised learning tutorial:

Labels X and Y are \_\_label\_\_. <Text>

where the class names X and Y are. No quotations; all on one line.

In this instance, there is only one class per row, and the classes are \_\_label 1 and \_\_label 2.

The 1-star labels and 2-star labels are \_\_label 1 while the 4- and 5-star labels are \_\_label 2.

(3-star reviews, which excludes reviews with neutral emotion from the original),

The review titles are prefixed to the text with ":" and a space.

There are a few reviews in other languages, such as Spanish, but the most of them are in English.

1. Sanders Corpus :

The 5513 manually classified tweets on four distinct themes make up the Sanders Analytics-created Twitter Sentiment Corpus (Apple, Google, Microsoft, Twitter). One annotator manually classified each tweet as either favorable, negative, neutral, or irrelevant in relation to the subject. 654 negative, 2,503 neutral, 570 positive, and 1,786 irrelevant tweets were produced as a consequence of the annotation procedure. A hash file with the password for a twitter account is also there.

1. SemEval (Semantic Evaluation) dataset :

SemEval is a collection of international natural language processing (NLP) research workshops with the goal of advancing the state-of-the-art in semantic analysis and assisting in the production of high-quality annotated datasets for a variety of increasingly difficult problems in natural language semantics. Every year, the workshop includes a set of collaborative challenges where computational semantic analysis systems created by various teams are displayed and contrasted.

20K tweets were used to create the original SemEval dataset, which was divided into training, development, and test sets. Five Amazon Mechanical Turk employees manually assigned labels to each tweet that were neutral, positive, or negative. Additionally, the turkers were asked to mark if some tweets included subjective or objective language. Using a list of the dataset’s tweet ids provided by, 13,975 tweets were retrieved with 2,186 negative, 6,440 neutrals and 5,349 positive tweets.

**Comparison of base papers**

**Table - 1**

|  |  |  |
| --- | --- | --- |
| Title | Sentiment Analysis on Movie Review Data Using Machine Learning Approach | A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis |
| Problem Answered | Evaluation of opinions based on movie review data using Machine Learning Algorithms | Achieving a higher accuracy in Sentiment Analysis of movie review data using a proposed CNN-LSTM model |
| Scope | Movie review data was collected as well as used five kinds of machine learning classifiers to analyze this data. Hence, the considered classifiers are Bernoulli Naïve Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), as well as Multinomial Naïve Bayes (MNB). The analysis outlines that MNB achieves better accuracy, precision and F-score while SVM shows higher recall compared to others. | Long Short-Term Memory (LSTM) model and Convolutional Neural Network (CNN) model have been applied to different Natural Language Processing (NLP) tasks with remarkable and effective results. In this study, a hybrid model using LSTM and a very deep CNN model named as Hybrid CNN-LSTM Model is proposed to overcome the sentiment analysis problem with a higher accuracy. |
| Dataset Used | Two thousand movie review posts were collected where there exists an equal number of positive and negative reviews.  Movie Review 2000 dataset :  https://github.com/riyadatik/Sentiment-Analysis-on-Movie-Review-Data/blob/master/Data%20set.xlsx | IMDB Movie Reviews Dataset : This dataset contains a total of 50,000 binary labeled movie reviews which were taken from the rotten tomatoes website.  https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews  Amazon Movie Review Dataset : This dataset consists of a few million Amazon customer reviews (input text) and star ratings (output labels)  https://www.kaggle.com/datasets/bittlingmayer/amazonreviews |
| Algorithms | Naïve Bayes Classifier  Support Vector Machine -  Maximum Entropy Classifier  Decision Tree Classifier | Word2Vc model  Convolution Neural Network (CNN)  Long Short-Term Memory (LSTM) |
| Evaluation Metrics | Accuracy, Precision, Recall and F1 Score were taken into consideration for evaluation | Accuracy, Precision, Recall and F1 Score were taken into consideration for evaluation |
| Performance and Results | According to the research conducted, MNB outperforms the competition in terms of accuracy, precision, and F-score, while SVM displays stronger recall. | In terms of accuracy, the proposed hybrid CNN-LSTM model outperformed the single CNN and LSTM models on two benchmark datasets of movie reviews. Comparing the proposed hybrid CNN-LSTM model to more established machine learning and deep learning models, it achieved 91% accuracy. |

**Paper : Sentiment Analysis on Movie Review Data Using Machine Learning Approach**

**Discussions and Findings :**

Multinomial Naïve Bayesian Classifier is better than others and thus is the most accurate among all the classifiers in case of Sentiment Analysis. In case of smaller datasets, the above Multinomial Naïve Bayesian Classifier shows worse results, thus in such cases Support Vector Machine can act as a more accurate classifier for Sentiment Analysis. The Maximum Entropy classifier performs poorly compared to other classifiers. It is less precise and has a lower f-score than the others. But compared to Multinomial NB and Decision Tree, the precision is better.

**Drawbacks:**

Although the result shows that Multinomial Naïve Bayes classifier is better than Support Vector Machine, this is only true for selected parameters because multinomial NB shows the worst results when the training dataset is small.

**Paper : A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis**

**Discussions and Findings :**

The Proposed Hybrid CNN-LSTM model performed very well on two benchmark movie reviews datasets as compared to single CNN and LSTM models in terms of accuracy. The Proposed Hybrid CNN-LSTM model achieved 91% accuracy as compared to traditional machine learning and deep learning models and thus outperforms them too. The proposed approach also used less parameters which consume less memory and are efficient in terms of convolution layers.

**Drawbacks:**

The proposed hybrid CNN-LSTM approach uses Word2Vec which cannot handle out-of-vocabulary words well. It assigns a random vector representation for Out-Of-Vocabulary words which can be suboptimal. Word2Vec relies only on local information of language words. The semantic representation of a word relies only on its neighbors & can prove suboptimal

**Table - 2**

|  |  |  |
| --- | --- | --- |
| Title | Sentiment Analysis of Twitter Data: A Survey of  Techniques | Sentiment analysis using product review data |
| Problem Answered | In this paper, a survey is provided about comparative analyses of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning  algorithms like Naive Bayes, Max Entropy, and Support Vector  Machine, for research on twitter data streams.discuss general challenges and applications of  Sentiment Analysis on Twitter. | A general process for  sentiment polarity categorization is proposed with detailed process descriptions. Data  used in this study are online product reviews collected from Amazon.com. Experiments  for both sentence-level categorization and review-level categorization are performed  with promising outcomes. |
| Dataset Used |  | Data used in this paper is a set of product reviews collected from amazon.com. From  February to April 2014, in total, over 5.1 millions of product reviews in which the products belong to 4 major categories: beauty, book, electronic, and home. Those online reviews were posted by over 3.2 millions of reviewers (customers) towards 20,062 products. Each review includes the following information: |
| Formulae | Naive Bayes:    Maximum Entropy:  The model is represented by the following:    c is the class,d is the tweet and λi  is the weight  vector. The weight vectors decide the importance of a feature in  classification. | Sentiment score computation for sentiment tokens    Random forest  pi is the probability that a tuple in D belongs to class Ci. The Gini index measures  the impurity of D. The lower the index value is, the better D was partitioned. For the  detailed descriptions of CART |
| Evaluation Metrics | Where, TP, TN, FP, and FN respectively denote  the true positives, true negatives, false positives, and false  negatives in the predicted labels. | Performance of each classification model is estimated base on its averaged F1-score    where Pi is the precision of the ith class, Ri is the recall of the ith class, and n is the number of classes. Pi and Ri are evaluated using 10-fold cross validation. |
| Application | 1.Applications that use Reviews  from Websites:  2. Applications in Business Intelligence  3. Applications in Smart Homes | This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. |

**Paper : Sentiment Analysis of Twitter Data: A Survey of Techniques**

**Discussions and Findings :**

Machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases, which require little effort in human-labeled documents .The effects of various features on classifiers.

It can be concluded that with cleaner data, more accurate results can be obtained. Use of bigram model provides better sentiment International Journal of Computer Applications. Focus on the study of combining machine learning methods into opinion lexicon methods in order to improve the accuracy of sentiment classification and adaptive capacity to a variety of domains and different languages.

**Drawbacks:**

Identifying subjective parts of text: Subjective parts represent sentiment-bearing content. The same word can be treated as subjective in one case, or an objective in some other. This makes it difficult to identify the subjective portions of text.

Domain dependence: The same sentence or phrase can have different meanings in different domains. For Example, the word „unpredictable‟ is positive in the domain of movies, dramas ,etc., but if the same word is used in the context of a vehicle's steering, then it has a negative opinion.

**Paper : Sentiment analysis using product review data**

**Discussions and Findings :**

This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed

**Drawbacks:**

Even though there are papers talking about spam on Amazon.com, it is still contended that it is a relatively spam-free website in terms of reviews because of the enforcement of its review inspection process. The Number of companies chosen was very less as such small sample can’t give accurate measures about the population

**Table - 3**

|  |  |  |
| --- | --- | --- |
| Title | Sentiment Analysis and Classification Based On Textual Reviews | Sentiment Analysis of Movie Reviews using Machine Learning Techniques |
| Problem Answered | Determines whether an opinion document (movie review) is positive or negative or neutral sentiment. | Finding the best algorithm to identify the polarity of reviews. |
| Dataset Used | The widely used document level sentiment classification dataset is Cornell movie-review corpora. | Data was collected from 2000 user-created movie reviews archived on the IMDb web portal at http://reviews.imdb.com/Reviews and is known as “Sentiment Polarity Dataset version 2.0” 1000 positive and 1000 negative processed reviews. This data is then converted into arff format. The data was in the txt\_token file which had 2 Subfolders for +ve and -ve. The converted data was imported into the WEKA tool using the Text Directory Loader. After that text pre-processing was done on the WEKA tool. |
| Formulae | SVM uses g(x) as the discriminant function, | Random forest  When using the Random Forest Algorithm to solve regression problems, you are using the mean squared error (MSE) to show how your data branches from each node.    Where N is the number of data points, fi is the value returned by the model and yi is the actual value for data point i |
| Existing Algorithms | Naive Bayes and Support Vector Machine | Naive Bayes, K nearest neighbor and  Random forest |
| Evaluation Metrics | Accuracy, Precision and F-Measure | Accuracy, Mean absolute error and Root mean squared error |
| Application | It determines whether an opinion document (movie review) is positive or negative or neutral. | Finding the review polarity may be helpful in a variety of areas. Without the need for the user to read through individual reviews, intelligent systems can be developed that can give users comprehensive reviews of movies, products, services, etc. |

**Paper :Sentiment Analysis and Classification Based On Textual Reviews**

**Discussions and Findings :**

It is challenging for a person to forecast the movie review in sentiment analysis. The current approach uses document-level sentiment categorization to address this. It establishes the positivity or negativity of an opinion document (movie review), as well as its neural emotion. It is possible to roughly categorize the sentiment using the Bag of Words. Parts of speech can be utilized to ensure that the categorization is accurate. On the benchmark dataset of movie reviews, a novel approach dubbed Sentiment Fuzzy Classification is suggested to increase classification accuracy.

**Drawbacks:**

Sentiment Fuzzy Classification on review data are often much less accurate when applied to data from other domains such as news or social media because of the differences in how people express themselves in these domains

**Paper :Sentiment Analysis and Classification Based On Textual Reviews**

**Discussion and Findings :**

The movie reviews were examined in this research using a variety of methods, including Naive Bayes, K-Nearest Neighbor, and Random Forest. The Naive Bayes classifier produced the best results. The accuracy of the Naive Bayes classifier was 81.45%, that of the Random Forest classifier was 78.65%, and that of the K-Nearest Neighbor classifier was 55.30%. Without the need for the user to read through individual reviews, intelligent systems can be developed that can give users comprehensive reviews of movies, products, services, etc. The user can then make decisions based on the results that the intelligent systems provide without having to read individual reviews.

**Drawbacks :**

Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life.

KNN does not work well with large datasets as calculating distances between each data instance would be very costly.

Random forest limitation is that many trees can make the algorithm too slow and ineffective for real-time predictions.

**Conclusion and Future Directions**

In the paper Sentiment Analysis of Twitter Data: A Survey of Techniques comparative analyses of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics was studied. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, research on twitter data streams was provided. General challenges and applications of Sentiment Analysis on Twitter have also been discussed.

Running Naïve Bayes using Trigrams, bigrams and unigrams together gave an accuracy of 75.41 percent which is less than the accuracy obtained when Bigrams were used as a feature. Also, this feature combination bloats up the feature space exponentially and the execution becomes extremely slow. Hence for further analysis, the trigrams are not considered as they do not have a noticeable impact on the accuracy.

Research results showed that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases, which require few effort in human-labeled document. The effects of various features on classifier were also studied. It can be concluded that with cleaner data, more accurate results can be obtained. The use of bigram model provides better sentiment accuracy as compared to other models. We can focus on the study of combining machine learning method into opinion lexicon method in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages.

Smart homes are supposed to be the technology of the future. In future entire homes would be networked and people would be able to control any part of the home using a tablet device. Recently there has been lot of research going on Internet of Things(IoT). Sentiment Analysis would also find its way in IoT. Like for example, based on the current sentiment or emotion of the user, the home could alter its ambiance to create a soothing and peaceful environment. Sentiment Analysis can also be used in trend prediction. By tracking public views, important data regarding sales trends and customer satisfaction can be extracted.

In the paper Sentiment Analysis and Classification Based On Textual Reviews a document level sentiment classification is used in the existing system. A new algorithm called Sentiment Fuzzy Classification algorithm is proposed to improve classification accuracy on the benchmark dataset of Movies reviews. Several steps were used to reach the final results such as text preprocessing, tokenization, removal of stop words, Text transformation, feature selection and at the end sentiment fuzzy classification is used. It establishes the positivity or negativity of an opinion document (movie review), as well as its neural emotion. It is possible to roughly categorize the sentiment using the Bag of Words. Parts of speech can be utilized to ensure that the categorization is accurate. On the benchmark dataset of movie reviews, a novel approach dubbed Sentiment Fuzzy Classification is suggested to increase classification accuracy.

In future fuzzy approach for polarity classification will be used for identifying the relationships between classes, through looking at how the membership degrees of different classes are correlated. They will also extend our fuzzy approach for the classification of categories of emotions, an emerging subarea of sentiment analysis.

In the paper Deep learning for sentiment analysis of movie reviews, study of wide range of NLP classification models take place. Investigations consisted of two main parts. In the first part, a dataset provided by Kaggle was used and applied to the bag of words, and skip gram word2vec models to represent words numerically. Then used several classifiers, including random forest, SVM, and logistic regression to perform the binary classification task. Vector averaging, and clustering was used to produce the aggregated feature vectors. However, these suffer from losing the order of words in sentences. This motivated the second part of our work.

In the second part, implemented the recursive neural tensor networks to train a multi-class sentiment analyzer. The training of standard RNTN was computationally very expensive. So low-rank RNTN was introduced. Then showed that the low-rank RNTN can achieve comparable accuracies to that of standard RNTN much faster. This better training performance of the low-rank RNTN enables it to train several different models, and use them for ensemble averaging. A 1.5% accuracy improvement was achieved by ensemble-averaging.

In the Paper : A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis, A hybrid CNN-LSTM model was suggested for sentiment analysis. In terms of accuracy, the proposed hybrid CNN-LSTM model outperformed the single CNN and LSTM models on two benchmark datasets of movie reviews. Comparing the proposed hybrid CNN-LSTM model to more established machine learning and deep learning models, it achieved 91% accuracy.

Sentiment Analysis of Movie Reviews using Machine Learning Techniques research paper helps us to determine the polarity of the tweets and for this a variety of methodologies were employed. Naive Bayes Algorithm, K-Nearest Neighbor Algorithm and Random Forest Algorithms were performed on the Data set. The Naive Bayes classifier produced the best results. The accuracy of the Naive Bayes classifier was 81.45%, that of the Random Forest classifier was 78.65%, and that of the K-Nearest Neighbor classifier was 55.30%.

Finding the review polarity may be helpful in a variety of areas. Without the need for the user to read through individual reviews, intelligent systems can be developed that can give users comprehensive reviews of movies, products, services, etc. The user can then make decisions based on the results that the intelligent systems provide without having to read individual reviews.

In the paper Sentiment Analysis of Movie Reviews, A new Feature-based Heuristic for Aspect-level Sentiment Classification makes two important contributions. First, it explores the use of ‘Adverb+Verb’ combined with ‘Adverb+Adjective’ combined for document-level sentiment classification of a review. Second, it proposes a new feature-based heuristic scheme for aspect-level sentiment classification of a movie. The aspect level sentiment classification produces an accurate and easy to understand sentiment profile of a movie on various aspects of interest. Interestingly, the aspect-level sentiment profile result is congruent to the document level sentiment classification of reviews of a movie. Though, the aspect-level sentiment profile produces a more focused and accurate sentiment summary of a particular movie and is more useful for the users.

The aspect-level sentiment analysis algorithmic formulation designed in this research paper is a novel and unique way of obtaining a complete sentiment profile of a movie from multiple reviews on different aspects of evaluation. The resultant sentiment profile is informative, easy to understand, and extremely useful for users. Moreover, the algorithmic formulation used for aspect-level sentiment profile is very simple, quick to implement, fast in producing results and does not require any previous training. It can be used on the run and produces very useful and detailed sentiment profiles of a movie on different aspects of interest. This part of the implementation can also be used as an add-on step in movie recommendation systems that use content-filtering, collaborative-filtering or hybrid approaches. This aspect level sentiment profiling is a valuable form of sentiment analysis and subsequent exploitation of information expressed by a large number of users about a particular movie.

In the paper Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory Long Short-Term Memory (LSTM) classifier is used for analyzing sentiments of the IMDb movie reviews. It is based on the Recurrent Neural Network (RNN) algorithm. The data is effectively preprocessed and partitioned to enhance the post classification performance. The classification performance is studied in terms of accuracy. It confirms the potential of integrating the designed solution in modern text-based sentiment analyzers. In total 10k reviews are considered, 5k for positive and 5k for negative sentiments. Results show a best classification accuracy of 89.9%.

In future superior accuracy can be attained by using further data preconditioning techniques. Furthermore, higher classification accuracy can be achieved by employing the ensemble classifiers or deep learning approaches. Exploring these opportunities is another prospect.

The paper Sentiment analysis using product review data tackles a fundamental problem of sentiment analysis, namely sentiment polarity categorization . Work done in the paper is as follows: An algorithm is proposed and implemented for negation phrases identification; A mathematical approach is proposed for sentiment score computation; A feature vector generation method is presented for sentiment polarity categorization. Two sentiment polarity categorization experiments are respectively performed based on sentence level and review level; Performance of three classification models is evaluated and compared based on their experimental results.

Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process was proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization were performed.

In the Paper : Sentiment Analysis on Movie Review Data Using Machine Learning Approach, five different kinds of machine learning classifiers have been used to collect and analyze data on movie reviews. Therefore, the classifiers that are considered include Bernoulli Naive Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), and Multinomial Naive Bayes (MNB). The dataset is divided into a train set and a test set using this technique. Initially, data is collected from the movie review website. Utilize an NLP tool to preprocess the data after that. The data set is then trained using the five ML classifiers following the creation of a feature vector. The results show that Multinomial Naive Bayes classifier outperforms SVM, but only for a selection of parameters since Multinomial NB performs poorly when the training dataset is small. The aforementioned outcome exemplifies the high quality of the features vector utilized for the movie review data.

**References**

1. Mouthami, K., Devi, K.N. and Bhaskaran, V.M., 2013, February. Sentiment analysis and classification based on textual reviews. In *2013 international conference on Information communication and embedded systems (ICICES)* (pp. 271-276). IEEE.
2. Pouransari, H. and Ghili, S., 2014. Deep learning for sentiment analysis of movie reviews. *CS224N Proj*, pp.1-8.
3. Baid, P., Gupta, A. and Chaplot, N., 2017. Sentiment analysis of movie reviews using machine learning techniques. *International Journal of Computer Applications*, *179*(7), pp.45-49.
4. Singh, V.K., Piryani, R., Uddin, A. and Waila, P., 2013, March. Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. In *2013 International mutli-conference on automation, computing, communication, control and compressed sensing (imac4s)* (pp. 712-717). IEEE.
5. Rahman, A. and Hossen, M.S., 2019, September. Sentiment analysis on movie review data using machine learning approach. In *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4). IEEE.
6. Kharde, V. and Sonawane, P., 2016. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.
7. Gómez-Rodríguez, C., Alonso-Alonso, I. and Vilares, D., 2019. How important is syntactic parsing accuracy? An empirical evaluation on rule-based sentiment analysis. *Artificial Intelligence Review*, *52*(3), pp.2081-2097.
8. Qaisar, S.M., 2020, October. Sentiment analysis of IMDb movie reviews using long short-term memory. In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-4). IEEE.
9. Rehman, A.U., Malik, A.K., Raza, B. and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications*, *78*(18), pp.26597-26613.
10. Kharde, V. and Sonawane, P., 2016. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.
11. Mäntylä, M.V., Graziotin, D. and Kuutila, M., 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review, 27, pp.16-32.
12. Märkle-Huß, J., Feuerriegel, S. and Prendinger, H., 2017. Improving sentiment analysis with document-level semantic relationships from rhetoric discourse structures. In *Proceedings of the 50th Hawaii international conference on system sciences* (pp. 1142-1151). HICSS.
13. Fan, X., Li, X., Du, F., Li, X. and Wei, M., 2016, November. Apply word vectors for sentiment analysis of APP reviews. In 2016 3rd International Conference on Systems and Informatics (ICSAI) (pp. 1062-1066). IEEE.
14. Mullen, T. and Collier, N., 2004, July. Sentiment analysis using support vector machines with diverse information sources. In Proceedings of the 2004 conference on empirical methods in natural language processing (pp. 412-418).
15. Naz, S., Sharan, A. and Malik, N., 2018, December. Sentiment classification on twitter data using support vector machine. In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (pp. 676-679). IEEE.
16. Hu, M. and Liu, B., 2004, August. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177).
17. Gamon, M., 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In COLING 2004: Proceedings of the 20th international conference on computational linguistics (pp. 841-847).
18. B. Pang, L. Lee, and S. Vaithyanathan “Thumbs up?: sentiment classification using machine learning techniques,” In Proc. of ICEMNLP, pp. 79-86, 2002.
19. Woldemariam, Y., 2016, March. Sentiment analysis in a cross-media analysis framework. In 2016 IEEE international conference on big data analysis (ICBDA) (pp. 1-5). IEEE.
20. Ding, J., Sun, H., Wang, X. and Liu, X., 2018, June. Entity-level sentiment analysis of issue comments. In Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering (pp. 7-13).
21. E. M. Alshari, A. Azman, S. Doraisamy, N. Mustapha, M. Alkeshr, “Effective method for sentiment lexical dictionary enrichment based on word2vec for sentiment analysis,” In Proc. of ICIRKM, Malaysia, 2018.
22. Vanaja, S. and Belwal, M., 2018, July. Aspect-level sentiment analysis on e-commerce data. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1275-1279). IEEE.
23. Porntrakoon, P. and Moemeng, C., 2018, July. Thai sentiment analysis for consumer’s review in multiple dimensions using sentiment compensation technique (SenSecomp). In 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 25-28). IEEE.
24. Yazdavar, A.H., Ebrahimi, M. and Salim, N., 2017. Fuzzy based implicit sentiment analysis on quantitative sentences. *arXiv preprint arXiv:1701.00798*.
25. Murthy, D., 2015. Twitter and elections: are tweets, predictive, reactive, or a form of buzz?. *Information, Communication & Society*, *18*(7), pp.816-831.
26. Shaziya, H., Kavitha, G. and Zaheer, R., 2015. Text categorization of movie reviews for sentiment analysis. *International Journal of Innovative Research in Science, Engineering and Technology*, *4*(11), pp.11255-11262.
27. Dave, K., Lawrence, S. and Pennock, D.M., 2003, May. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528).
28. Kim, S.M. and Hovy, E., 2004. Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics* (pp. 1367-1373).
29. Waila, P., Singh, V.K. and Singh, M.K., 2012, December. Evaluating machine learning and unsupervised semantic orientation approaches for sentiment analysis of textual reviews. In *2012 IEEE International Conference on Computational Intelligence and Computing Research* (pp. 1-6). IEEE.
30. Benamara, F., Cesarano, C., Picariello, A., Recupero, D.R. and Subrahmanian, V.S., 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *ICWSM*, *7*, pp.203-206.
31. Maas, A., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C., 2011, June. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 142-150).
32. Pang, B. and Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, *2*(1–2), pp.1-135.
33. Pang, B. and Lee, L., 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.
34. Devi, K.N. and Bhaskarn, V.M., 2012. Text sentiments for forums hotspot detection. *International Journal of Information Sciences and Techniques (IJIST) Vol*, *2*.
35. Yi, J., Nasukawa, T., Bunescu, R. and Niblack, W., 2003, November. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Third IEEE international conference on data mining* (pp. 427-434). IEEE.
36. Pak, A. and Paroubek, P., 2010, May. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
37. Parikh, R. and Movassate, M., 2009. Sentiment analysis of user-generated twitter updates using various classification techniques. *CS224N final report*, *118*.
38. Parikh, R. and Movassate, M., 2009. Sentiment analysis of user-generated twitter updates using various classification techniques. *CS224N final report*, *118*.
39. Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R.J., 2011, June. Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)* (pp. 30-38).
40. Noulas, A., Mascolo, C. and Frias-Martinez, E., 2013, June. Exploiting foursquare and cellular data to infer user activity in urban environments. In *2013 IEEE 14th international conference on mobile data management* (Vol. 1, pp. 167-176). IEEE.
41. Noulas, A., Mascolo, C. and Frias-Martinez, E., 2013, June. Exploiting foursquare and cellular data to infer user activity in urban environments. In *2013 IEEE 14th international conference on mobile data management* (Vol. 1, pp. 167-176). IEEE.
42. Noulas, A., Mascolo, C. and Frias-Martinez, E., 2013, June. Exploiting foursquare and cellular data to infer user activity in urban environments. In *2013 IEEE 14th international conference on mobile data management* (Vol. 1, pp. 167-176). IEEE.
43. Neethu, M.S. and Rajasree, R., 2013, July. Sentiment analysis in twitter using machine learning techniques. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)* (pp. 1-5). IEEE.
44. ZhunchenLuo, M.O., 2013. TingWang, An effective approach to tweets opinion retrieval. *Springer Journal onWorldWideWeb*.
45. Liu, S., Li, F., Li, F., Cheng, X. and Shen, H., 2013, October. Adaptive co-training SVM for sentiment classification on tweets. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 2079-2088).
46. Liu, S., Li, F., Li, F., Cheng, X. and Shen, H., 2013, October. Adaptive co-training SVM for sentiment classification on tweets. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 2079-2088).
47. Pang, B. and Lee, L., 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.
48. Sailunaz, K., Dhaliwal, M., Rokne, J. and Alhajj, R., 2018. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, *8*(1), pp.1-26.
49. Bandhakavi, A., Wiratunga, N., Padmanabhan, D. and Massie, S., 2017. Lexicon based feature extraction for emotion text classification. *Pattern recognition letters*, *93*, pp.133-142.
50. Maas, A., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C., 2011, June. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 142-150).
51. Hoque, M.T., Islam, A., Ahmed, E., Mamun, K.A. and Huda, M.N., 2019, February. Analyzing performance of different machine learning approaches with doc2vec for classifying sentiment of bengali natural language. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-5). IEEE.
52. Jadav, B.M. and Vaghela, V.B., 2016. Sentiment analysis using support vector machine based on feature selection and semantic analysis. *International Journal of Computer Applications*, *146*(13).
53. Kim, S.M. and Hovy, E., 2004. Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics* (pp. 1367-1373).
54. Go, A., Bhayani, R. and Huang, L., 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12), p.2009.
55. Jiang, L., Yu, M., Zhou, M., Liu, X. and Zhao, T., 2011, June. Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies* (pp. 151-160).
56. Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), pp.1-167.
57. Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M. and Li, P., 2011, August. User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1397-1405).
58. Seref, B. and Bostanci, E., 2018, October. Sentiment analysis using naive Bayes and complement naive Bayes classifier algorithms on Hadoop framework. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-7). IEEE.
59. Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), pp.1093-1113.
60. Sahu, T.P. and Ahuja, S., 2016, January. Sentiment analysis of movie reviews: A study on feature selection & classification algorithms. In *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)* (pp. 1-6). Ieee.
61. Mahtab, S.A., Islam, N. and Rahaman, M.M., 2018, September. Sentiment analysis on bangladesh cricket with support vector machine. In *2018 international conference on Bangla speech and language processing (ICBSLP)* (pp. 1-4). IEEE.
62. Wang, Y., 2017, December. Advanced naïve bayes algorithm design with part-of-speech tagger on sentiment analysis. In *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)* (pp. 1382-1385). IEEE.
63. Seref, B. and Bostanci, E., 2018, October. Sentiment analysis using naive Bayes and complement naive Bayes classifier algorithms on Hadoop framework. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-7). IEEE.