|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sl. NO** | **Author Name** | | **Year** | | **Type of Methodology** | | **Sample** | **Data Collection Approach** | | | | **Key Findings** | | | | **Researchers Remarks** | |
| 1. | Prachi Shevate,  California State University, Sacramento  prachishevate@csus.edu | | 2018 | | Bag of words and Word2Vec combined with Random Forest Classifier. | | IMDB review Dataset | The data is from IMDB, which a famous movie online  database of information related to world movies and other  entertainment sources. | | | | Data pre-processing is extremely necessary to get  better accuracy. It helps us understand the model  construction.  Word2Vec model gives an  accuracy rate of just 75% whereas using the Bag-of-words  models can provide us with an accuracy rate that is greater than 80% or above. | | | | “I plan to extend my work by implementing a  model which will improve my accuracy more than 92%. Also, to include comparisons with different machine learning models.” | |
| 2. | Ankit Goyal,  a3goyal@ucsd.edu  Amey Parulekar,  [aparulek@ucsd.edu](mailto:aparulek@ucsd.edu) | | 2015 | | Bag of Words, N-gram modelling, TF-IDF modelling coupled with classifiers like Maive Bayes, Random Forest, Logistic Regression, SGD classifier and kNN classifier. | | IMDB movie review dataset | The dataset used for this task was collected from Large Movie Review Dataset, which was used by the AI department of Stanford University for the associated publication. | | After comparing the different approaches, we can infer that for our problem statement, Logistic Regression Model  with feature set using mixture of Unigrams and Bigrams is best. Apart from this, one can also use  a Naïve Bayes’ Classifier or a SGD classifier as they also provide good accuracy percentage.  One peculiar thing to note is low accuracy with Random Forest classifier. This might be because of over-fitting of decision trees to the training data. Also, low accuracy of kNN Classifiers shows us  that people have varied writing styles and kNN Models are not suited to data with high variance. | | | | One of the major improvements that can be incorporated as we move ahead in this project is to merge words with similar meanings before training the classifiers. Another point of improvement can be to model this problem as a multi-class classification problem where we classify the sentiments of reviewer in more than binary fashion like “Happy”, “Bored”, “Afraid”, etc.  This problem can be further remodeled as a regression problem where we can predict the degree of affinity for the movie instead of complete like/dislike. | | | | |
| 3. | Palak Baid  Department of Computer Science & Engineering  Jaipur  Apoorva Gupta  Department of Computer Science & Engineering  Jaipur  Neelam Chaplot  Department of Computer Science & Engineering  Jaipur | 2017 | | Sentiment Analysis, Opinion Mining, Movies Reviews, Naive Bayes, K-Nearest Neighbour, Random Forest  It is a technique based on Bayes’ Theorem. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature | | Cornell movie review data | | It is a technique based on Bayes’ Theorem. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature | | | The Naïve Bayes classifier achieved 81.45% accuracy,  Random Forest classifier we achieved 78.65% accuracy  , K-Nearest Neighbour classifier achieved 55.30% accuracy. | | | | “I would try and work on a model which will improve all the accuracy i.e naïve bayes,random forest and k-nearest neighbour by more than 90%. Also, to include comparisons .” | |
| 4. | Gurshobit SinghBrar,  Asst.Prof.Ankit Sharma | 2018 | | Opinion mining,Sentiment analysis, natural language ,sentiment score ,sentiment lexicon | | Movie Lens Dataset | | There are two ways to give input to the movie review sentiment analyzer. One by providing a list of reviews in JSON file format,or by providing TMDB ID of movie title. | In this paper ,movie reviews are classified into positive or negative polarity.  Best thing about the system is that it is a web based API for sentiment analysis for movie review s with JSON output to display the results on any operating system. | | | | This API can be trained for other reviews like smartphones,laptops or clothes etc ,therefore would like to improve the accuracies a further more and accessibilies to a wide range of technological devices. | | | | |
| 5. | Ashok kumar pant central department of computer science&it,tu kirtipur,kathamandu , nepal  Abhimanu yadav college of applied business naksal kathmandu,nepal | | 2013 | | Their exist the approachs towards sentiment analysis meachine learning based methods, lexicon methods and linguistic analysis meachine learning methods based on training an algorithm | | Stanford sentiment treebank  (rotten tomatoes reviews) | Movie reviews centiment analysis can be done to identify user attitude and opinion toward perticular movie,in this research,you have considered two clases of gloabal subjective polarity(positive and negitive) of movie reviews texts  sentiment classification of reviews has been the focous of recent research.it has been attempted in different domains such as movie reviews | | | | The top level sentiment classification system for nepali movie reviews.it is divided into four sub system,data acquisition,feature extraction,preprocessing,and classification preprocessing the data is the process of cleaning and preparing the text for future extraction and clasification. | | | | Navi baves based meachine learning technique is used for the clasification of the sentiment.empirical results shows,clasification accuracis are,79.23% of precision,78.57% of recall and 78.90% of f-score | |
| 6. | Amit p.pimpalkar department of computer science(software system)  Shri ram institue of technology,jabalpur  Prof.mahendra kumar rai  Department of information technology,jabalpur | | 2016 | | Sentiment analysis(SA)is a  “web openion mining”where the primary objective is so clarify do openions according to a verify and range.the boundiries on the range usually corresponding to +ve or -ve feelings | | IMDB movie review dataset | The focous of the system is to analyge the sentiments for the movie reviews. The input is to be taken for the movie reviews sites and the social networking sites on which the cooments are proposal for the particular movie | | Natural language processing meachine learning,movie reviews,sentiment analysis system,web opinion mining text tekmisson,fnzzymesures, sentiwordnet | | | | Today the internet holds on continues amount of textual data which is also growing every day.peole communite through the onlineresourses , discussions,groups and blocks while facts focous on data transmission the opinion express the sentiment of there person behind currently google searchs for facts and facts can be expressed in keywords | | | | |