



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 7, July 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Automatic Sentiment Analysis of Movie Reviews

Sakshi Vinod Birpan¹, Rushikesh Pravin Patil¹, Rutuja Suresh Rajput¹, Nikita Premsing Patil¹

B.E. Students, Department of Computer Engineering, SSBT's COET, Jalgaon, Maharashtra, India¹

ABSTRACT:- Entertainment is crucial part of human life entertainments like songs, music, drama and movies etc. So for watching good movies most of the peoples generally prefer theatre. If movie is not good then we feel nervous and we think wasted our money and time for watching bad movie so people prefer to go to movie by reading reviews and rating for that movie on various apps like IMDb, flixter and voice etc. But by reading one or two reviews we cannot say movie is good or bad because different peoples have different opinions some peoples like action or some like thrill or romance so people gives reviews based on their area of interest. So we cannot predict movie for that we proposed movie reviews based on sentiment analysis and classification algorithms such as Naïve Bayes and Gradient Boosting Classifier generally utilized to identify the sentiment of huge amount of text. We compare Naïve Bayes and Gradient Boosting Classifier techniques for measuring negative, positive and neutral reviews.

KEYWORDS: Sentiment analysis, Naïve Bayes algorithm, Gradient Boosting Classifier

I. INTRODUCTION

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes (Liu, 2012). In applications such as recommender systems, "what other people think" has always been an important piece of information during the decision process (Pang, Lee, et al., 2008), like recommendation of clothes in Amazon based on the sentiment analysis of user reviews. Application of sentiment analysis has been revealed in multiple areas. In social media such as Facebook and Snapchat, there are goldmines of consumer stories and opinion data (Lexalytics, 2019), which can be used for advertising, marketing, recommending new friend relationship, etc. But social posts are full of complex abbreviations, acronyms, and emoticons, which are non-trivial problems for the feature space in machine learning models. The sheer volume is a problem, too. Hopefully, a successful model can save the human beings some valuable hours parsing mountains of social data by hand (Lexalytics, 2019). Such things can also be applied to business intelligence models. For example, the sentiment analysis statistics can be used to estimate the retention rate of customers in a new product, adjusting the current marketing situation and trying to satisfy the customers in a better way (Gupta, 2018). Such a monitoring approach has enabled the companies to adapt their business plan in real time, which can lead to possible reduction of cost on marketing. For the area of analysis of movie review, sentiment analysis means finding the mood of the public about how do they judge a specific movie (Python for Engineers, 2019). For details, the documents, from the user reviews, are classified based on the mood they are expressing. Some sentiment analysis tasks do a binary classification like positive and negative, and others may do multi-level classification like positive, somehow positive, neutral, somehow negative and negative. The actual number of classes depends on the aims of specific tasks. Previous papers have talked about the traditional approaches for sentiment analysis on movie review datasets. In this paper, I am trying to extend the feature space of movie review dataset, which is rarely seen in previous work, to achieve an increase of classification performance. Parameter tuning is also included in this process to achieve a better performance.

II. RELATED WORK

Sentiment Classification has been done for Movie Review by using Supervised Machine Learning. [1] Author used two machine learning techniques for analyzing sentiments that is Naïve Bayes and Random Forest. Also utilized the Memory Comparison of naïve bayes and random forest algorithm. Naïve bayes require more memory than random forest. Based on A Literature Review on Sentiment Analysis, Stavros Karamaniolos, Geoff [2] did classification and document processing. They perform Tokenization and Stopword Filtering and Stemming. They also done classification using naïve bayes algorithm. A Study on Feature Selection and Classification Algorithm [3] in this we analyze and study the features that affect the sentiment score of the movie reviews. Also we see the state of the art classification

algorithm for the evaluation of performance and accuracy of the approach. Author RiddhiDagli, TapasDesai, PratikWayangankar, RashmiChawla [4] proposed system on movie reviews is developed conducting sentiment analysis on movie reviews. Based on Movies Reviews Sentiment Analysis and Classification, MaisYasen, SaraTedmori [5] studied different machine learning classifiers like naïve bayes, decision tree, support vector classifier, bayes network, K-nearest neighbors, ripper rule learning, random forest, stochastic gradient descent. We studied this comparatively and select the naïve bayes classifier for sentiment analysis classification. Author Pooja Mehta, Dr. Sharnil Pandya [6] compares various approaches and methods. They show basics of sentiment/opinion mining and its levels. There are various approaches and methods to identify sentiment from content. In this paper they examine machine learning procedures. And after studying all methods they examine naïve bayes shows highest accuracy. Based on this we select naïve bayes. A Survey of Sentiment Analysis from Social Media Data [7] we give sentiment analysis techniques and opinion mining. Also they show the accuracy of classifiers like naïve bayes classifier at an accuracy 75%. Gradient Boosting Machine Learning Classifier to learn general sentiment using domain free dataset. [8] Their result shows that the gradient boosting classifier is the winner in terms of accuracy and Fscore, among the other well known classifier for sentiment analysis: logistic regression, SVM, random forest, multinomial naïve bayes, deep learning.

III. MACHINE LEARNING ALGORITHM

A. NAÏVE BAYES:

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. This model is easy to build and particularly useful for very large datasets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

$$P(C|X) = P(X|C) * P(C)/P(X) \quad (1)$$

$P(C|X)$ is posterior probability of class C

$P(C)$ is prior probability of class C

$P(X|C)$ is probability of predictor given the class.

$P(X)$ is prior probability of predictor

B. GRADIENT BOOSTING ALGORITHM

Gradient Boosting is a popular boosting algorithm. In gradient boosting, each predictor corrects its predecessor's error. In contrast to Adaboost, the weights of the training instances are not tweaked, instead, each predictor is trained using the residual errors of predecessor as labels. There is a technique called **the Gradient Boosted Trees** whose base learner is CART (Classification and Regression Trees)

IV. EXPERIMENTAL RESULTS

Gaussian Naive Bayes Algorithm, Gradient Boost Classifier Algorithms were performed on the Data set. Results are presented in the form of figures and tables as shown below. Best accuracy was given by the Gradient Boost Classifier.

CLASSIFIER	ACCURACY
Gaussian Naïve Bayes	80.98
Gradient Boosted	89.78

V. CONCLUSION

The Sentiment Analysis System for Movie Reviews implements the method of Naive Bayes Classifier in order to obtain the highest posterior probability value of the two review classes of sentiment. The posterior probability value is obtained from the total sum of log prior probability and log likelihood of each term in each review of the training data for each sentiment class. Studies in SA approaches existed for more than a decade and now are exploited by enterprises as an important tool for strategic marketing planning and manoeuvring. This move is also due to the advancement in data storage, access and analytics enabled through big data frameworks.



REFERENCES

- 1) Tejaswini M. Untwale , Prof. G. Chaudhari /"Implementation of Sentiment Classification of Movie Reviews by Supervised Machine Learning Approaches, in IEEE xplore part number CFP19K25-ART,ISBN:978-5387-7808-4/ICCMC 2019
- 2) Stavros Karamaniolos , Geoff Skinner/ "A Literature Review on Sentiment Analysis and its Fundamental Technologies"/ IEEE 4th International Conference on Computer and Communication Systems 2019
- 3) Tirath Prasad Sahu,SanjeevAhuja/ "Sentiment Analysis of Movie Reviews: A Study on Feature Selection and Classification Algorithm"
- 4) RiddhiDagli, Tapas Desai,Pratik Wayangankar,RashmiChawla/Web BLogMinig using Internet Movie Database(IMDB) for Movie Review and Sunnarization "/www.irjet.com/ April2016
- 5) Asha S. Manek , P. DeepaShenoy, M. Chandra Mohan, Venugopal K. R. /"Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier"/December 2015
- 6) MaisYasen,Sara Tedmori /"Movies Reviews Sentiment Analysis and Classification"/MaisYasen, Sara Tedmori/IEEE 2019
- 7) Pooja Mehta, Dr.SharnilPandya/ "A Review On Sentiment Analysis Methodologies, Practices And Applications"/ February 2020
- 8) KoyelChakraborty, SiddhartaBhattacharyya,Rajib Bag /"A Survey of Sentiment Analysis from Social Media Data"/IEEE



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details