IEC
Institute of Electronics and Computer

Journal of the Institute of Electronics and Computer, 2022, 4, 15-33
https://iecscience.org/journals/JIEC
ISSN Online: 2643-8240

Movie Success-Rate Prediction System through Optimal Sentiment Analysis

Dewan Muhammad Qaseem[1], Nashit Ali[1], Waseem Akram[1], Aman Ullah[2], Kemal Polat[3, *]

[1]Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 61100, Pakistan
[2]School of Computer Science and Engineering, Central South University, Changsha, 410083, China
[3]Department of Electrical and Electronics Engineering, Bolu Abant Izzet Baysal University, Bolu 14280, Turkey
* Corresponding Author: Kemal Polat, Email: kpolat@ibu.edu.tr

How to cite this paper: Dewan Muhammad Qaseem, Nashit Ali, Waseem Akram, Aman Ullah, Kemal Polat (2022). Movie Success-Rate Prediction System through Optimal Sentiment Analysis. Journal of the Institute of Electronics and Computer, 4, 15-33. https://doi.org/10.33969/JIEC.2022.41002.

Received: August 10, 2022
Accepted: October 6, 2022
Published: October 18, 2022

Abstract

With the speedy growth of social media, it has become easy for people to express their feelings about anything and give their opinions. These opinions are helpful in business plans development, marketing trends, political parties' popularity. Different social media sources are used for this purpose, i.e., Facebook, Instagram, YouTube, Twitter, etc. The rapid growth of text data on social media is required to develop algorithms and techniques for recognizing people's opinions towards a specific subject. Nowadays, Twitter becomes a rapidly used social media application where people feel free to share their feelings about anything and give their opinions. Film Industry is one of the revenue-generating Industries for the economic growth of any country. People express their views about any upcoming movie by watching its trailer using social media. The practical sentiment analysis of opinions on social media such as Twitter can be helpful to predict movie ratings. This research focused on developing a technique to predict movie success rates based on viewers' tweets on movie trailers. The results provide the movie rating in the star's form (1-5). We have collected tweets about different movies after their trailer was released by applying the hashtag method (#Hash). We have used four key algorithms (Naïve Bayes, SVM, decision tree, and KNN) on NLTK Movie review corpora and train & test our models. Machine learning training data sets were not readily available for movies ratting; then, we shifted towards a lexicon-based approach. All these three dictionaries have a different word count, and each word in these dictionaries has its own polarity in the form of a score. Finally, we have also compared our results with other movie rating sites like IMDB rating, which are satisfactory.

Keywords

Optimal Sentiment Analysis, Machine Learning, Analysis, Decision Making.

1. Introduction

In our research, we have predicted the popularity of a movie through sentiment analysis of social media and proposed a movie rating system. People on social media are allowed freely to express their opinions and suggestions about any topic. Social media can be the source to collect the user's reviews on a particular topic [1].

User-generated data on social media is increasing day by day. In 2013, SINTEF published a research report that tells that 90% of data on the web or internet was produced in the past two years, and this increasing data speed continues. The World Wide Web (www) is proof of large data because it has approximately 4.55 billion pages. In this large data, it can be difficult to find our required information [2]. A critical part of collecting information about a new thing is that we should find the information that other people think about it. It is human nature to decide that they must talk and collect recommendations from other people. For example, if we want to buy a new car, we must ask our relatives or friends which car is best for us. A large amount of reviews data on social media makes it easier to find out the related opinions of those we have never met before [3]. The fastest growth of information on social media makes it easy for buyers to decide what to purchase.

And on the other hand, sellers can improve their product ads based on this information [4]. Social media plays an important role in decision making (e.g., forum, discussion, comments, blogs, Twitter, microblogs, and posting on social network websites). For example, if someone wants to buy a product, one should need some opinions. This one is not bound to ask one's family or friends for opinions about the product because there are many reviews/comments and suggestions on social network sites about the product. But there are some issues for average readers because they can feel difficulty to identify those sites and extract and summarize the useful opinions from a large number of opinions. So the automated sentiment analysis systems are used to solve this problem [5].People have used sentiment and websites to post the feelings of people in recent years. Sharing opinions or current issues/topic is their main purpose. Therefore, among Internet users, a networking tool has become called the micro-blogging service. From Alta Plana’s Text Analytics research study [6], there is a graph in which, based on 962 sources of textual information, 216 respondents were surveyed.

Nowadays, Twitter that was launched in 2006, is considered one of the most widely used platforms. Twitter is being used by people for communication, for expressing their opinions and feelings by uploading messages. For example, a student can express his emotions or feelings by uploading a positive message on Twitter when he received a dream job offer. On Twitter, 500 million tweets are generated by more than 330 million active users per day. There is a huge amount of information that is valuable is hidden behind Twitter. Various research and industrial information such as marketing, social studies etc. can use such information. Tweets are those messages that are read or sent by users on Twitter. Tweets are of a maximum of 140 words, and they are short in form. So, there is no chance to explain the full detail of one's emotions or information about events of daily life. People often make spelling mistakes; use emoticons manually created words or other characters to express special meanings. People do all these things and mistakes to contain as much information as possible. "@" symbol is used by Twitter users to refer to other users of Twitter. People automatically alert by referring them to this symbol. For making topics, hashtags are usually used by the users. To increase the views of tweets, all these things take place.

A huge number of messages have been produced because of a considerable number of user-generated tweets. Due to the huge volume of tweets or messages (500 million per day), the hidden useful information behind tweets cannot be analyzed by people through hand. Therefore, a valuable and effective technique is needed that automatically evaluates and analyzes sentiment information on tweets, called tweet sentiment analysis. There are two main implemented research tasks for tweet sentiment analysis. The first task focused on promoting people to determine new analysis methods that are based on twitter relationships and data on Twitter. The second task is concerned with using some tweet features and training various models [7]. Movie reviews, product experience, and political comments are areas where common sentiment analysis is focused. The second task has been studied extensively. Due to the short and messy tweets, some new and unique challenges were used in sentiment analysis. The primary problems for solving are abnormal structures and content length limitation. In tweet sentiment analysis, typically, the essential thing is a classifier in which analysis and classification of tweets take place. Common classes for tweet classification are positive, negative, and neutral. Generally, to build a sentiment classifier, there are several techniques such as lexicon-based classifiers, machine learning, sentiment analysis, etc. In recent years the approach that is mostly used and famous in various applications is the machine learning approach.

Our research is about success rate prediction based on reviews. For this purpose, we used Twitter as a data source. Because twitter now a day is a top-rated source for expressing people's opinions about any topic. When a movie trailer is released, Twitter becomes a platform for people to express their thoughts or opinions about the movie after watching the trailer. They can judge the movie based on cast, cast popularity, music, graphics, or story highlights in a movie trailer. After watching the trailer, they express their likes and dislikes about the movie in textual form, which is tweets. We have collected these tweets on the trailer about the particular movie before its release. This research investigates how to preprocess tweet data and apply a linguistic approach with three different dictionaries, compare their results, and check whether these results are useful for achieving our success rate prediction of a movie. Built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables, are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

2. Literature Review

Sentiment analysis in reviews investigates item surveys on the web to decide the general sentiment or feelings about an item. Reviews speak to the so-called user-generated content. This is of developing consideration, and a rich asset for showcasing groups, sociologists, analysts, and other people who may be worried about assessments see the open state of mind and general or individual states of mind [27]. The immense number of audits on the web speaks to the present type of client's input. Choosing about the slant of sentiment in a survey is a testing issue because of a few elements. One issue is the subjectivity in the writings and the need to recognize supposition bearing from non-stubborn sentences. Another issue that makes it difficult to characterize surveys is the alleged "foiled desire" which implies that the essayist composes many sentences one way which can be comprehended as positive, and after that closes with one negative sentence that turns around the importance of the whole content [19]. This is the reason there ought to be better techniques for highlight choice. Turnery [35] infers that motion picture audits are challenging to arrange because the general conclusion in the survey is not the whole of all of the conclusions said in the content. Many investigations have investigated different techniques in examining item audit assumption, some of which utilize machine learning approaches [19] [36], and some utilize lexical techniques [35] [31].

There is a vast number of online surveys, and that some of these contain just a little division of sentences that express a conclusion makes it harder for clients' and organizations to know the general feeling about an item and set up a conclusion about it. Thus, assessment condensing is proposed. Feeling condensing could be led by revising the first content and centering on the primary point. It can also utilize angle-level notion examination and select the primary elements in the writings, their angles, and the related conclusions to create an "element-based survey outline". Cases can be found in [32] [37]. In this theory, film audits are examined, utilizing a standout among sentiment analysis's most well-known information sets. The issue is moved toward utilizing the machine learning approach and the feeling holder point of view.

2.1. Machine Learning Approach for Sentiment Analysis and Opinion Mining

Machine learning could be considered a part of artificial intelligence. Supervised Learning and Unsupervised Learning [38] are the main two categories included in machine learning. Usually, on concluding information about the characteristics of sets of data, machine learning algorithms work well. For sentiment analysis, natural language processing (NLP) techniques from which sentiment features are selected and extracted and the success of machine learning relies on them.

From the machine learning approach, the majority of research work was focused on product reviews that are highly subjective English normal texts. On the other side, there has also been some multilingualism, short and messy sentiment analysis of tweets. With distant supervision sentiment of Twitter messages that are automatically classified with the help of machine learning algorithms was presented by Go et al. [39]. They used noisy labels of Twitter messages in order to show that pre-processing steps are essential. For making distant supervised learning feasible, after training three machine learning algorithms, they found using tweets with emotions. Twitter API for sentiment analysis that automatically collects twitter corpus was used by Pak and Paroubek [5]. Based on the multiracial Naïve Bayes Classifier, they built a classifier with the help of a document that could be annotated in negative, positive, and neutral sentiments. Using social relationships, the performance of user-level sentiment analysis could be improved. It is founded by Ten et al.[40] . For example, to refer other users for reasons, users of Twitter often use the symbol "@". In this manner, referring to the other users automatically alert them. According to Ten et .al that people may hold similar opinions who use "@" symbol to connect to each other. Therefore, in their experiments "@" symbol became an important feature. Their outcome demonstrated that incorporating link information from Twitter could enhance sentiment classification performance essentially in light of the Support Vector Machine. A method based on the context and syntactic relationships for target-dependent twitter sentiment classification was introduced by Dong et al. [41] called Adaptive Recursive Neutral Network (AdaRNN). AdaRNN contained various composition functions. In the experiment, an interesting target was created with the given tweets. A few researchers intended to continue Twitter sentiment analysis on another domain with a specific end goal to get more attention to tweets of different languages (not English) [42].Naïve Bayes, Support Vector Machine, and Maximum Entropy are three machine learning classifiers with them [42] did sentiment analysis of Turkish Political columns. Their inspiration was deciding positive and negative opinions from entire documents without considering subjects. From unlabeled tweets data to labeled political columns, to improve classifications performance Transfer Learning was used. In an unsupervised way extracting features from unlabeled data and accuracy of sentiment classification can be improved with the help of transfer learning method. A first ranking algorithm with bi-grams and uni-grams and second skip-grams to the speech processing are two approaches provided by Fernandez at.al [43] for sentiment analysis of Spanish tweets. Positive, strongly positive, none, strong negative and negative are five levels in which the polarity of Spanish Tweets was divided. As data set in the paper [44], movie reviews were used. In the paper, standard machine learning strategies were used which out-perform human-produced baselines. As compared to traditional topic-based categorization the three machines learning techniques do not perform efficiently. From the performance point of view, SVMs performs worst. In a paper [45] the documents are classified as positive, negative and neutral with the help of a system called Document Level Opinion Mining System.

Negation also handled by the proposed systems from IMDB movie dataset used by experimental. POS tagging used in a movie review; Document Based Sentiment Orientation System performs better than AIRC Sentiment Analyzer. Movie reviews dataset that contain 1000 positive and 1000 negative reviews are classified and in the paper [40] for this classification, three supervised machine learning algorithms such as KNN, SVM, and Naïve Bayes were compared. It is recognized in this approach that the training dataset had a larger number of reviews and a large number of reviews and as compare to Naïve Bayes the SVM approach performed well. In classifying data correctly more than 80% accuracies acquired by SVM approach, for all three algorithms about movie reviews in sentiment classification when a huge amount of training dataset containing 800 to 1000 reviews will perform better.

In this paper [46], sentiment analysis of movie reviews was proposed using a combination of natural language processing and machine learning approach. Firstly, the dataset data pre-processing was done. Secondly, to obtain the results for sentiment analysis, different feature selection schemes combined with two classifiers, Naïve Bayes and SVM, were used. Thirdly, for obtaining the results for higher-order n-grams, the model of sentiment analysis was extended. As compare to Naïve Bayes, the linear SVM classifier gives more accuracy shown in the classification of a movie review.

In this research [47], using emoticons for training data to perform distant supervised learning is an effective way. High accuracy can be achieved by using different machine learning algorithms in this approach—tweet sentiment classified by machine learning algorithms with the same performance because twitter messages have unique characteristics.

The previous study [41] performed sentiment analysis on a movie review; a new approach was used called a Combined Approach. Two separate classifiers, such as Support Vector Machine and Hidden Markov Model (HMM), were combined in this approach, and the results of these classifiers were also combined. By using the combination of these classifiers, there is a possibility to improve the results of classification. The classifier handles slang words and smiley. With higher accuracy, a good sentiment classification is achieved in this approach.

In a study [48], Random Forests Classifier by changing the values of different hyperparameters supervised learning technique called Random Forests for classification results. The comparison of some supervised learning techniques like BN, C4.5, and ID3 with Random Forests Classifiers was focused on in this paper. And this comparison is focused with respect to ROC Area and incorrectly classified instances. Random Forests outperforms all three classifiers in terms of incorrectly and correctly classified instances and ROC Area.

This research [43] investigates parametrization rule for the Forest-RI algorithm should be considered in order to focus on the radiance that is an important property. The influence of Hyperparameter on Random Forest Accuracy is presented in this paper.

By tuning of Hyperparameters in Random Forest, to perform sentiment analysis of movie review dataset on using the Random Forest was focused on the paper [49]. On the movie review dataset, Random Forest performed well based on experimental results. The result with high accuracy of 87.85% was of dataset V1.0, and the result with an accuracy of 91.0% was provided by the dataset V2.0. Most of the past work has focused on Maximum Entropy, Naïve Bayes, and Support Vector machines for the sentiment classification. But in this paper, the experiments that are carried out show that if Hyperparameters are fine-tuned, then Random Forest can give better outcomes.

2.2. Lexicon-based Approach

For the lexicon-based approach, as the name infers, the sentiment lexicon is the fundamental reason for the time spent sentiment analysis, for instance, gatherings of sentiment phrases or reference books of sentiment words. As a matter of fact, lexicon-based techniques are isolated into two sorts, which are dictionary-based and corpus-based [38]. Dictionary-based techniques ordinarily depend on lexicons that are aggregated and characterized by a number of standard sentiment terms. A decent illustration of this term is Sentiwordnet [44]. Despite the fact that it is a devoted answer for sentiment analysis, there are a few confinements when managing outstanding commitments or analyzing contexts. On the opposite side, corpus-based techniques are additionally in view of dictionaries. However, these lexicons have associated sentiment terms or phrases for specific fields. Additionally, to create dictionaries, factual and semantic procedures are every now and again utilized [38].

Although numerous scientists tend to pick the machine learning approach in sentiment analysis, the lexicon-based approach, be that as it may, in any case, plays a huge part in the sentiment analysis of Twitter. As we talked about in the next Section, features and resources, for example, automatic part-of-speech tags and notion dictionaries, have demonstrated helpful in sentiment analysis for the applications, for example, product surveys. Is it accurate to say that they are additionally valuable for tweet sentiment analysis? Kouloumpis et al. [50] concentrated on inquiring about and noting this inquiry.

In addition to this query, another issue was the colossal expansiveness of tweets' topics. Particularly for this issue, the authors proposed utilizing Twitter hash-tags that distinguished topics rapidly. At that point, they examined and analyzed four features: n-gram features, vocabulary features, grammatical form features, and microblogging features in light of polarity sources in the MPQA subjectivity lexicon that was right off the bat recorded by Wilson et al. [51]. What they found are: for tweet sentiment analysis, macro- blogging features were generally valuable; what's more, part-of-speech features were not obviously supportive. Zhang et al.[52] went for joining the lexicon-based approach with learning-based methods for Twitter sentiment analysis. For the lexicon-based methods, there are three sections: sentence type detection, coreference resolution, what's more, opinion lexicon. Declarative sentences, imperative sentences, and interrogative sentences are the three fundamental sorts of tweet sentences. Declarative sentences and imperative sentences, as a rule, contain opinion features, be that as it may, interrogative sentences don't. In this manner, sentence type recognition, a procedure of recognizing and expelling interrogative sentences from tweet information, would be required. Coreference resolution is a helpful procedure for tweet sentiment analysis. For instance, "I had a sweetheart. She is wonderful." In the second sentence, coreference determination can make sense of that "she" alludes to "sweetheart". The authors depended on opinion lexicons from others and manually included numerous valuable opinion hash-tags into the dictionary. Their tests demonstrated that the combination of the techniques is successful and promising. The semantic strategy is a piece of corpus-based approach. Saif et al. [7] worked at semantic sentiment analysis of Twitter. For sentiment analysis, their approach was to include semantic features into the preparation set. For instance, a tweet is "The new Mazda 3 is great!". They may include a semantic idea "Mazda car" as another component for the separated substance "Mazda 3", furthermore, estimated the polarity of the extra feature. Besides, they moreover thought about the semantic approach and the sentiment-bearing topic analysis approach. The outcomes demonstrated that the semantic approach could increment both F1 scores of positive sentiment and negative sentiment.

In a comparative study, the semantic approach showed signs of improvement review and F1 scores in negative sentiment characterization and better exactness in positive sentiment classification. There are various existing lexicons for sentiment analysis, such as Affective Norms for English Words (ANEW) [45], Sentiwordnet, and Sentiment Finder. In particular, with the end goal of sentiment analysis, researchers have been taking a shot at developing new word lists. Nielsen [53] made another word list for Twitter sentiment analysis which performed better than ANEW. As per tweet attributes, the author built a new dictionary including opinion words, slang, and indecent words. The message-level task for detecting message sentiment item-level task for detecting the sentiment of a term are the two tasks that are described by Mohammad et.al [54]. For the lexicon-based approach MPGA mentioned before, they used lexicons that are existing and created manually and like NRC Hashtag Sentiment Lexicon in which new lexicons are generated. In the end, an F1 score of 88.93 was obtained for the Term-level task and an F1 score of 69.02 was obtained for the Message-level task. Sanchez-Mirabel et al. [55] used sentiment 140 lexica and NRC-Hashtag sentiment and also included two new lexicons and new features. He did all of his work just based on the work and approaches of Mohammad et al. [54]. The two new lexicons that Mirabel added are NRC Emotion Lexicon 1.0 and NRC

Emotion Lexicon 2.0. The first new lexicon just noticed the values of positive and negative it was not capable of calculating the sentiment score. On the other side, the second lexicon calculated the positive values such as (sum of the values of positive, trust, surprise, joy, and anticipation) and negative score (sum of the values for negative fear, sadness, anger, and disgust) from NRC Emotion lexicon 2.0. Their system did not work as well as it was expected. According to the discovery of some researcher's tweet sentiment classification can be improved with the help of topic-sentiment analysis, just because tweets are messy messages and short. Wang et al. [46] discovered that for making keywords or phrases connected hashtags are most widely used in tweets. Rather than analyzing the polarity of tweet topics directly hash tag-level sentiment classification might be much better and beneficial.

The main thing on which their task-focused is automatically generation and identification of sentiment polarity of three types of information: sentiment polarity of hashtags, meaning, and the relationship of hashtags. Still proposed method was complicated and defective, and the experimental results were good. Also, Canneyt et al. [40] concentrated on utilizing tweets' topics to help classify sentiment, but in actuality, through analyzing their relevance, they detected tweet topics. From the above method Wang et al.[56] their technique was completely different, even though sentiment analysis in Twitter was the topic on which they all worked. For Canneyt et al. [40] approach: training for the tweets with some topic and the training for all types of tweets are the two classifiers that were implemented. With the help of experiments, they found that a topic-specific sentiment classifier could truly improve the performance of Twitter sentiment analysis.

Petros ventis et al. describe in their research about predicting movie success by using blogs. Generally, the release data of a movie is known for all users before its release. The blogs can provide a platform for discussion about the movie before and after the movie release. They selected the top 300 movies from 2008 based on revenue, and after it, they filtered the movie list by selecting movies with the word "WANTED" in the title. They created 120 features on the following criteria. Every reference to movie count on the blog is used for movie ranking; all the features have a date range limit. They making the factors for reference counts e.g. if a reference came into the movie title, it must be considered. In this research, analysis of date ranges distinct features before and after five weeks of movie release for obtaining each week buzz. For sentiment analysis to evaluate the positive and negative posts, they applied hierarchal classification by using the Ling Pipe approach. For pointing out the spam posts, they ignore short posts and considered only some large posts. They use Kullback-Leibler and Pearson'scorrelation divergence for evaluating their features by using some different outcome variables. These variables were an average rating of user's reviews, 2008 average sales, and the opening 5-week sales on box office. They found more success in predicting movie ratings than other critics and user ratings [57].

In contrast to D. Jennifer (2014) research on “Affective text-based emotion mining on social media”. In this research author proposed “Latent Dirichlet Allocation” for emotion modeling. The basic aim of this model was to mine the textual information from social media [58].

VASU JAIN (2013) discusses in his research on “Prediction of Movie Success using Sentiment Analysis of Tweets”. The author did a fundamental study on sentiment analysis for determining a movie's success on box office. The study's outcomes show that the movie success can be anticipated by sentiment analysis of movie by some matrices and good efficiency. The author observed that more than one variable might influence the movie's success at the box office. However, in this research author focuses on sentiment analysis. The author faced many challenges during sentiment analysis on Twitter.

- Having restrictions onTwitter APIs (only 1500 tweets per day). Also, have not sufficient resources for collecting data which can cause some inaccurate results.

- If the author randomly picked 200 tweets, there may be lots of noise and spam data.

- The author was not able to take all the tweets for prediction matrices.

- Author’s sentiment analyzer has a very low accuracy [59].

Kaut yessenov and Sasa Minailovic described in “sentiment analysis of movie review comments” their research based on machine learning. They use social websites for data collection e.g. Digg. Authors use dig articles for obtaining comments. Digg is a voting system where users vote for some particular topic in the format of +1 as positive or -1 as negative according to their interest. The difference between positive and negative comments defines the polarity of the movie or topic. To train the classifier, they use the existing movie review corpus. To evaluate the accuracy of the comment classification, they use learning algorithms (unsupervised and supervised) and feature selection. The result of the study shows that “Bag of words” model works well. They investigate the impact of feature vectors on the accuracy of classification. They discovered that the current corpus from the comparative corpus contains movie review sentences. Research outcomes demonstrate that such corpus has the comparative polarity of words [60].

Aina Elisabeth Thunestveit was researching “Sentiment Analysis on User-Based Reviews: Movie Recommendation Case”. In this research, they describe a new model based on analyzing and extracting adjectives from the user reviews. They describe that every adjective contains an opinion and their system is based on adjectives as opinion deciders. They use classical machine learning techniques like support vector machine and random forest to anticipate user reviews' opinion orientation. They use a large dataset for experiments which contains 50000 movie reviews. The author presents the following model for “Sentiment Analysis on User-Based Reviews” [61].

Sentiment analysis helps people to judge quality by analyzing the reviews. In this era where everyone is so busy in their routine work, it feels like a very difficult and time-consuming task to check all available studies on a product. [62]

Sentiment analysis has been an appealing topic for the researcher. Research has been done on social media blogs and other online documents. [63]

3. Research Method

The main purpose of this research is to find out the movie success rate through sentiment analysis on Twitter data. Our research is based on the lexicon approach. This chapter describes the overall methodology process. We present a lexicon-based approach to extracting sentiment from text. In this research, we have used various dictionaries to calculate the sentiment score of a particular movie data set. Each dictionary has its classifications to rate any movie. We have used three different dictionaries for this purpose AFINN, Socialsent, and Rating Warriner. AFINN has a 2478-word count, and it provides the classification result between -5 to 5. Socialsent has 6967-word count, and it provides classification result between -5 to 5 while Rating Warriner has a 13915-word count and provide the classification from 0 to 8.

We have calculated the rating of the specific movie reviews for all of these dictionaries. After calculating a rating for these dictionaries, we have rescaled those ratings from 0-5 and compare the result of all dictionaries, and analyzed the variation between the ratings of these dictionaries. The prediction of IMDb scores through data from social media has been explored before in a research study by Oghina et al. [17]. However, this study differs from previous work because the machine learning approach was used in the previous work, and now we have used a lexicon-based approach. In our approach, we used three different dictionaries with a large collection of words with their ratings, improving the accuracy of prediction results. Following diagram 3.5 shows our system flow.

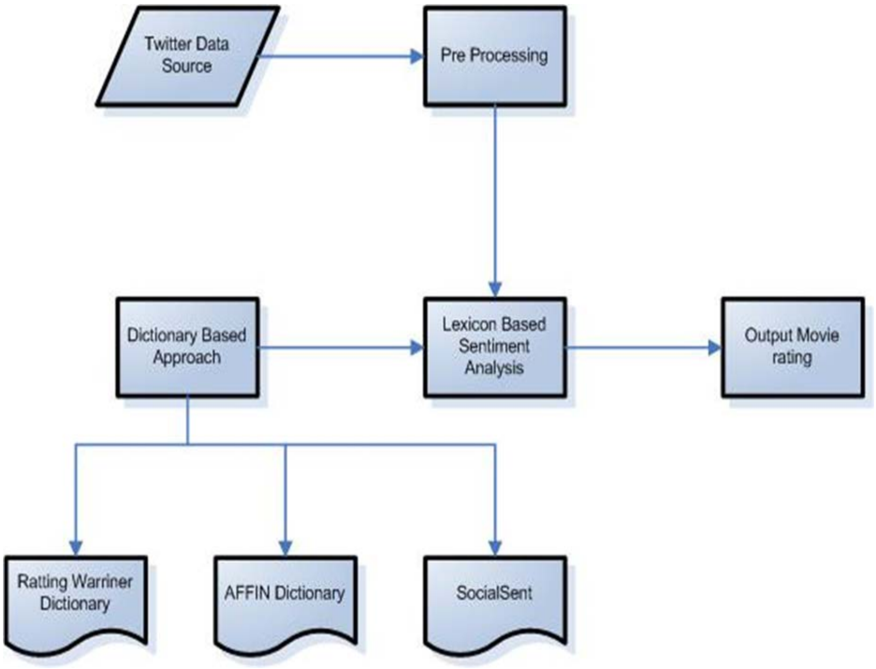

Figure 1 Research Methodology

4. Results and Comparison

First of all, we predict the success rate of using tweets. For this purpose, after the preprocessing process, we calculate each tweet sentence polarity score. And they are stored in a variable. The other variable is used for storing the total count of tweets. Finally, after calculating all the sentences rating, we calculate the mean of all these ratings.

$$Movie\ Rating = \frac{\sum_{i=0}^{n} Tr}{Tn} \quad \ldots\ldots \quad (1)$$

Here Tr of rating of all tweets and Tn is the Total number of tweets.

By using this formula, we got a rating of movie success. Then, according to this same method, we calculate each movie rating in all three dictionaries. But now the problem is that all the ratings we got have different scales. And we need only a five-star rating. For this purpose, we need to be rescaled all the values from 0 to 5 rating.

4.1. Rescaling

Rescaling is the process of normalizing the values into the desired scale.in the previous step, we have the predicted value of any movie. But these values are not from 0 to 5 rating. For this purpose, we use the rescaling standard formula, which is

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

X' = rescaled value

X = Score of each sentence

X_{min} = minimum value of the dictionary

X_{max} = maximum value of the dictionary

After applying this formula, we got rescaled values from 0 to 5. For example, “Ratings_Warriner_et_al” has 0 to 8 words ratings. But after rescaling this dictionary, results were rescaled into 0 to a 5-star rating.

Table 1 Predicted Score comparison with IMDB

Sr #	Movie Titles	Dictionaries Results			IMDB Rating
		AFFIN	Ratings_Warriner	Socialsent	
1	Star Wars	3.22	3.28	2.56	3.65
2	Thor Ragnarok	3.41	3.26	2.63	3.95
3	Bright Movie	2.82	3.42	2.29	3.2
4	Bomb City	2.63	2.69	2.80	3.5
5	Mamma Mia	3.19	3.30	2.55	3.2
6	Lady Bird	3.21	3.41	2.67	3.75
7	Daddy Home 2	3.37	3.54	2.55	3

4.2. Upcoming movies ratting prediction

Table 2 Upcoming movies ratting prediction

Movie	Trailer Releasing Date	Movie Releasing Date	No. of Tweets	Rating
Mulan	July,2019	27th March,2020	2970	2.25
Jai Mummy di	12th December,2019	16th January,2020	973	2.76
UnderWater	22nd December,2019	8th January,2020	2970	2.9
Tanhaji	19th November,2019	10th January,2020	3214	2.68

We have also used the lexical-based approach for movie success rate prediction; in our approach, we have used three different dictionaries with different word counts. We have analyzed the success rate for each movie using all three different dictionaries and identify which dictionaries provide a better result on the same dataset. The results show that ratings_ Warriner dictionary provides more accurate results than the other two dictionaries. For Accuracy purposes, we compared our results with IMDB movie ratings and analyze that results are comparative with IMDB social network.

4.3. Comparison

Table 3 Comparison with other techniques

	Precision	Recall	F1	Accuracy	
Naïve Bayes	82	82	82	82	Naïve Bayes
Decision Tree	64	64	64	63	Decision Tree
K-Nearest Neighbor	84	61	66	61	K-Nearest Neighbor
Linear SVC	84	84	84	83	Linear SVC
NUSVC	77	60	64	60	NUSVC
SGD Classifier	83	83	83	83	SGD Classifier
Multinomial NB	82	82	82	82	Multinomial NB

4.4. Comparison Graph

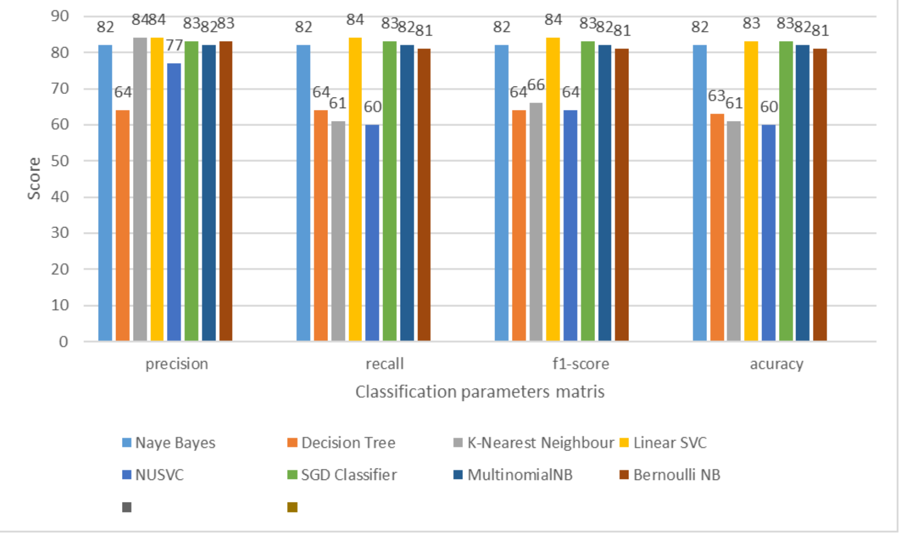

Figure 2 Comparison Graph

5. Conclusion and Future Work

In this research, we have worked on predicting movie success rates using a dictionary-based approach. For this purpose, we have used three different dictionaries. Each dictionary has another word count. We have collected data set for multiple movies from Twitter by using the hashtag method. We have analyzed the success rate for each movie using all three different dictionaries and find which dictionaries provide a better result on the same dataset. We have also used the machine learning-based approach to predict movie success rate. We have used four different algorithms (Navie Bayes, Neural Network, Decision Tree, SVM) and find SVM provides better accuracy in results than the other algorithms. We have Analyzed challenges we have faced in machine learning approaches like delivering positive and negative results. We were unable to find a suitable data set for predicting movie success rate in Star rating format. We have concluded that a dictionary-based approach can provide better results than the machine learning approach to predict movie success rate in star rating method. We have also compared our results with other social media sites like IMDB rating and analyze our results are compared with IMDB social network.

Currently, we worked on three dictionaries. In the future, we can perform this technique in multiple dictionaries and compare the variation of results. In the future, we can build a domain-specific dictionary about movies for more accurate results. We can work on creating a suitable dataset for movie star ratings for the machine learning method. In the future, we can work on the classification of movie reviews using machine learning techniques for multi-class sentiment analysis. Finally, we can extract movie reviews from other social media applications like Facebook and IMDB.

References

[1] N. Banovic, T. Buzali, F. Chevalier, J. Mankoff, and A. K. Dey, "Modeling and understanding human routine behavior," 2016, doi: 10.1145/2858036.2858557.

[2] R. Feldman, "Techniques and applications for sentiment analysis," Commun. ACM, vol. 56, no. 4, pp. 82–89, 2013.

[3] G. Di Fabbrizio, A. Aker, and R. Gaizauskas, "Starlet: multi-document summarization of service and product reviews with balanced rating distributions," in Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, 2011, pp. 67–74.

[4] K. I. Asad, T. Ahmed, and M. S. Rahman, "Movie popularity classification based on inherent movie attributes using C4. 5, PART and correlation coefficient," in Informatics, Electronics & Vision (ICIEV), 2012 International Conference on, 2012, pp. 747–752.

[5] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining." in LREc, 2010, vol. 10, no. 2010, pp. 1320–1326.

[6] S. Grimes, "Text/content analytics 2011: user perspectives on solutions and providers," Alta Plana, 2011.

[7] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in International semantic web conference, 2012, pp. 508–524.

[8] D. J. Kim, D. L. Ferrin, and H. R. Rao, "A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and

their antecedents," Decis. Support Syst., vol. 44, no. 2, pp. 544–564, 2008.

[9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[10] W. Fan, L. Wallace, S. Rich, and Z. Zhang, "Tapping the power of text mining," Commun. ACM, vol. 49, no. 9, pp. 76–82, 2006.

[11] D. Delen and M. D. Crossland, "Seeding the survey and analysis of research literature with text mining," Expert Syst. Appl., vol. 34, no. 3, pp. 1707–1720, 2008.

[12] D. S. Soper and O. Turel, "An n-gram analysis of Communications 2000--2010," Commun. ACM, vol. 55, no. 5, pp. 81–87, 2012.

[13] D. M. Blei, "Probabilistic topic models," Commun. ACM, vol. 55, no. 4, pp. 77–84, 2012.

[14] M. Rossetti, F. Stella, and M. Zanker, "Analyzing user reviews in tourism with topic models," Inf. Technol. Tour., vol. 16, no. 1, pp. 5–21, 2016.

[15] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends® Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.

[16] S. Asur and B. A. Huberman, "Predicting the future with social media," in Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, 2010, pp. 492–499.

[17] A. Oghina, M. Breuss, M. Tsagkias, and M. de Rijke, "Predicting imdb movie ratings using social media," in European Conference on Information Retrieval, 2012, pp. 503–507.

[18] S. R. Das, "News analytics: Framework, techniques and metrics," 2010.

[19] M. Daiyan, D. S. K. Tiwari, M. Kumar, and M. A. Alam, "A Literature Review on Opinion Mining and Sentiment Analysis," Int. J. Emerg. Technol. Adv. Eng., vol. 5, no. 4, pp. 262–280, 2015.

[20] [20] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, 2011, pp. 142–150.

[21] B. Liu, "Sentiment analysis and opinion mining," Synth. Lect. Hum. Lang. Technol., vol. 5, no. 1, pp. 1–167, 2012.

[22] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences," in Proceedings of the 2003 conference on Empirical methods in natural language processing, 2003, pp. 129–136.

[23] L. K.-W. Tan, J.-C. Na, Y.-L. Theng, and K. Chang, "Sentence-level sentiment polarity classification using a linguistic approach," in International Conference on Asian Digital Libraries, 2011, pp. 77–87.

[24] O. Täckström and R. McDonald, “Semi-supervised latent variable models for sentence-level sentiment analysis,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, 2011, pp. 569–574.

[25] S.-M. Kim and E. Hovy, “Automatic detection of opinion bearing words and sentences,” 2005.

[26] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, “User-level sentiment analysis incorporating social networks,” in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 1397–1405.

[27] H. Tang, S. Tan, and X. Cheng, “A survey on sentiment detection of reviews,” Expert Syst. Appl., vol. 36, no. 7, pp. 10760–10773, 2009.

[28] R. Narayanan, B. Liu, and A. Choudhary, “Sentiment analysis of conditional sentences,” in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1, 2009, pp. 180–189.

[29] C. Zirn, M. Niepert, H. Stuckenschmidt, and M. Strube, “Fine-grained sentiment analysis with structural features,” in Proceedings of 5th International Joint Conference on Natural Language Processing, 2011, pp. 336–344.

[30] P. C. Tetlock, M. Saar‐Tsechansky, and S. Macskassy, “More than words: Quantifying language to measure firms’ fundamentals,” J. Finance, vol. 63, no. 3, pp. 1437–1467, 2008.

[31] A.-M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews,” in Natural language processing and text mining, Springer, 2007, pp. 9–28.

[32] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima, “Collecting evaluative expressions for opinion extraction,” in International Conference on Natural Language Processing, 2004, pp. 596–605.

[33] P. Melville, W. Gryc, and R. D. Lawrence, “Sentiment analysis of blogs by combining lexical knowledge with text classification,” in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 1275–1284.

[34] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, “Structured models for fine-to-coarse sentiment analysis,” in Proceedings of the 45th annual meeting of the association of computational linguistics, 2007, pp. 432–439.

[35] P. D. Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” in Proceedings of the 40th annual meeting on association for computational linguistics, 2002, pp. 417–424.

[36] S.-M. Kim and E. Hovy, “Automatic identification of pro and con reasons in online reviews,” in Proceedings of the COLING/ACL on Main conference poster sessions, 2006, pp. 483–490.

[37] K. Lerman, S. Blair-Goldensohn, and R. McDonald, “Sentiment summarization: evaluating and learning user preferences,” in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009, pp. 514–522.

[38] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, “Sentiment analysis: A review and comparative analysis of web services,” Inf. Sci. (Ny)., vol. 311, pp. 18–38, 2015.

[39] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” CS224N Proj. Report, Stanford, vol. 1, no. 12, 2009.

[40] S. Van Canneyt, N. Claeys, and B. Dhoedt, “Topic-dependent sentiment classification on twitter,” in European Conference on Information Retrieval, 2015, pp. 441–446.

[41] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, “Adaptive recursive neural network for target-dependent twitter sentiment classification,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2014, vol. 2, pp. 49–54.

[42] M. Kaya, G. Fidan, and I. H. Toroslu, “Transfer learning using Twitter data for improving sentiment classification of Turkish political news,” in Information Sciences and Systems 2013, Springer, 2013, pp. 139–148.

[43] J. Fernández, Y. Gutiérrez, J. M. Gómez, P. Martínez-Barco, A. Montoyo, and R. Munoz, “Sentiment analysis of spanish tweets using a ranking algorithm and skipgrams,” 2013.

[44] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.” in Lrec, 2010, vol. 10, no. 2010, pp. 2200–2204.

[45] M. M. Bradley and P. J. Lang, “Affective norms for English words (ANEW): Instruction manual and affective ratings,” Citeseer, 1999.

[46] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, “Identifying sources of opinions with conditional random fields and extraction patterns,” in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 355–362.

[47] K. Dave, S. Lawrence, and D. M. Pennock, “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews,” in Proceedings of the 12th international conference on World Wide Web, 2003, pp. 519–528.

[48] K. Eguchi and V. Lavrenko, “Sentiment retrieval using generative models,” in Proceedings of the 2006 conference on empirical methods in natural language processing, 2006, pp. 345–354.

[49] [C. Fernandez-Lozano et al., “Improving enzyme regulatory protein classification by means of SVM-RFE feature selection,” Mol. Biosyst., vol. 10, no. 5, pp. 1063–1071, 2014.

[50] E. Kouloumpis, T. Wilson, and J. D. Moore, “Twitter sentiment analysis: The good the bad and the omg!,” Icwsm, vol. 11, no. 538–541, p. 164, 2011.

[51] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity in phrase-level sentiment analysis,” in Proceedings of the conference on human language technology and empirical methods in natural language processing, 2005, pp. 347–354.

[52] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, “Combining lexiconbased and learning-based methods for twitter sentiment analysis,” HP Lab. Tech. Rep. HPL-2011, vol. 89, 2011.

[53] F. Å. Nielsen, “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs,” arXiv Prepr. arXiv1103.2903, 2011.

[54] S. M. Mohammad, S. Kiritchenko, and X. Zhu, “NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets,” arXiv Prepr. arXiv1308.6242, 2013.

[55] P. A. Sánchez-Mirabal, Y. R. Torres, S. H. Alvarado, Y. Gutiérrez, A. Montoyo, and R. Muñoz, “UMCC_DLSI: Sentiment Analysis in Twitter using Polirity Lexicons and Tweet Similarity,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014, pp. 727–731.

[56] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, “Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach,” in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 1031–1040.

[57] E. Sadikov, A. G. Parameswaran, and P. Venetis, “Blogs as Predictors of Movie Success.” 2009.

[58] D. Jennifer, “Affective Text based Emotion Mining in Social Media,” Int. J., vol. 2, no. 3, 2014.

[59] V. Jain, “Prediction of movie success using sentiment analysis of tweets,” Int. J. Soft Comput. Softw. Eng., vol. 3, no. 3, pp. 308–313, 2013.

[60] K. Yessenov and S. Misailovic, “Sentiment analysis of movie review comments,” Methodology, vol. 17, pp. 1–7, 2009.

[61] A. E. Thunestveit, “Sentiment Analysis on User-Based Reviews: Movie Recommendation Case.” NTNU, 2016.

[62] Nashit Ali, Anum Fatima, Hureeza Shahzadi, Nasrullah Khan, and Kemal Polat. "Online Reviews & Ratings Inter-contradiction based Product’s Quality-Prediction through Hybrid Neural Network." Journal of the Institute of Electronics and Computer 3, no. 1 (2021): 24-52.

[63] Nashit Ali, Anum Fatima, Hureeza Shahzadi, Aman Ullah and Kemal Polat. “Feature Extraction aligned Email Classification based on Imperative Sentence Selection through Deep Learning.” Journal of Artificial Intelligence and Systems, 3, (2021): 93–114.