e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

SENTIMENT ANALYSIS OF MOVIE REVIEWS USING DIFFERENT ALGORITHMS: A SURVEY

Ayushi Jain[*1], Ankit Kapoor[*2], Medhavi Pandey[*3]

[*1,2]Department of Computer Science and Engineering, Delhi Technical Campus, Gr. Noida, UP, India

[*3] Assistant Professor, Department of Computer Science and Engineering, Delhi Technical Campus, Gr. Noida, UP, India

ABSTRACT

Sentiment analysis is a sub-domain of opinion mining within which we will focus on the extraction of emotions and opinions of the buyer on a specific product. It is a structured knowledge set or unstructured knowledge set. In this review paper, we tend to specialize in sentiment analysis on reviews of the IMDB review dataset. We'll work on the info set and check out to classify the polarity of the movie review on the dimensions of — to + (i.e. extremely negative to extremely positive). Supported that we will try to give the user or consumer a simpler and possibly feasible method based on the scores that are on the graph provided on an internet site. In addition to it, we tend to also are evaluating completely different classification approaches that may suit our project domain. We tend to review several analysis papers and located out there and found completely different approaches within the extraction of text options like a bag of words models, employing a giant motion picture review dataset, limiting adjectives and adverbs, handling negations, bonding word frequencies by the brink. We tend to conjointly target on finding the aspect-based motion picture review that conjointly affects the polarity of the review. This can be achieved by victimization bound driving strategies. As our project is consumer-based we are not dealing with accuracies of these methods, and solely work on a bag of words models employing a giant review dataset and can give visual representations of our findings the style of graphs on our website.

Keywords: Sentiment Analysis, Naïve Bayes, ML Approaches, K-Means, IMDB, Movie Reviews, Data Polarity, Bag of Words, Graphs

I. INTRODUCTION

In this growing era of digital information, the amount of user-generated data is also growing exponentially over the internet. According to a study it was observed that out of every five, four users use social media in some or other form. We can now say that the web has been totally transformed into a more creative and participating web. Each and every person present on the web can give their contribution.

One such way to participate in this contribution is by giving reviews. The majority of the sites on the web provide us a platform to share our views, opinions, and experience of a product, experience, or service used by us. One such platform is movie sites that ask us to rate the movies watched by us and give our feedback regarding it. Have you ever thought about what the purpose of asking this feedback is? Let's take an example to understand this. Imagine a scenario that this weekend you made a plan to visit for a movie but can't decide which one to go for. To get rid of such a situation, the first solution that comes to our mind is to go to a review site and check out the movie ratings. In a similar way, we can also check for our peer's feedback but as we are aware that there are 1000+ reviews, so the task becomes tedious.

Here sentiment analysis comes into play. Through this, we use many computer-based algorithms to make our task simplified. Machine learning algorithms read all the reviews giving by the users, and then find out the relevant information to generate useful results according to the user inputs. Most general algorithms that are used for this purpose are Naïve Bayes, ML approaches, SVM, SentiWordNet, K-Mean.

II. LITERATURE SURVEY

Throughout the years multiple contributions have been made in the field of Sentiment Analysis. Although few of these approaches might be obsolete and are no longer being used but they give us an insight of how the technology has evolved and why it might be better and efficient to use certain algorithms and their benefits. We have reviewed the major technology changes over the years in this paper.

Paper [1] -by Kuat Yessonov and Sasa Misailovic in 2009

Their method of sentiment analysis is based upon machine learning. They believe that in order to perform machine learning, it is necessary to extract clues from the text that may lead to correct classification. Ideally, the properties of the original text that are relevant for the sentiment analysis task are to be selected. Unfortunately, the exact algorithm for finding the best features does not exist. It is thus required to rely on our intuition, the domain knowledge and experimentation for choosing a good set of features.

Bag-of-words Model

Bag-of-words may be a model that takes individual words during a sentence as features, assuming their conditional independence. The text is arranged as an unordered collection of words. This is effectively an unigram model, where each word is independent of the others. All the words in the feature vector constitute the dictionary. The challenge with this approach is that the choice of words that are appropriate to become features. Using this model the sentence *"This is a great movie"* may be represented by the subsequent feature vector:

F0=['a':1, 'event':1, 'great':1, 'is':1, 'this': 1]

Identifying of Synonyms

Bag-of-words model doesn't capture the relations between the words. For example, it will consider the two sentences *I saw a fantastic movie* and *I saw an excellent film* as two quite different sentences. The similarity between words fantastic and great, as well as the words movie and film is obvious to the human reader. It is apparent that each of these pair of synonym words may be represented by a single feature. We modified the model so that the feature vector has only one word representing a synonym cluster. The features then become semantic similarity rather than exact word match.

Handling Negation

Negation plays an important role in polarity analysis. One of the example that we can use to justify our case is sentences from our collection "*This is not a good movie"* has the opposite polarity from the sentence "*This is a good movie",* although the features of the original model would show that they are not opposite. Words that are influenced by the negation, especially adjectives and adverbs should be treated differently. A simple, yet effective way for the support of the negation is to include an additional feature [word]-NOT, for each adjective and adverb.

Classification

Classification algorithm foretell the label for a specified input sentence.

There are two approaches which can be used for classification: Supervised and Unsupervised.

They consider a three supervised classification approach –

Naive Bayes(N),

Maximum Entropy(M),

Decision Trees(D),

one unsupervised classification approach –

K-Means clustering(K).

For feature extractors, they used –

plain bag-of-words (B),

bag-of-words using densities from the movie reviews collection (C),

bag-of-words using only adjectives and adverbs and accounting for negation (G).

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

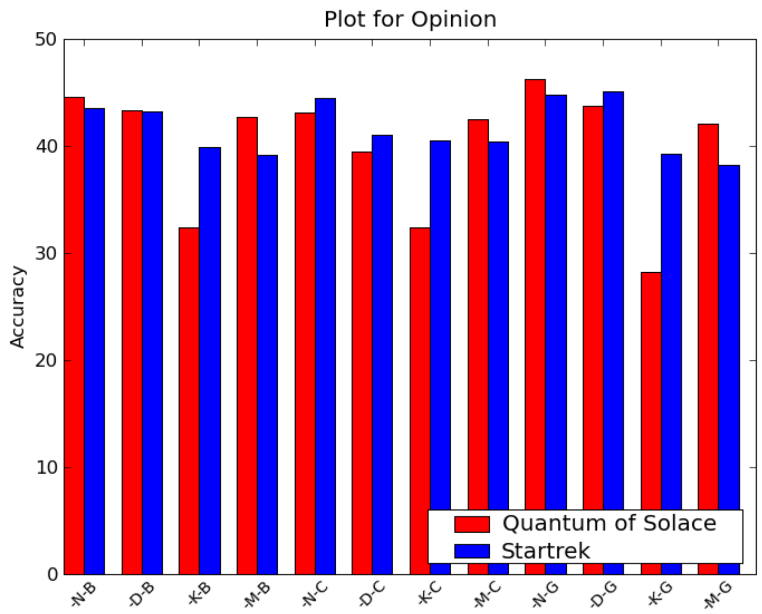

Figure 1. Result of experiment done

Paper [2] -by V.K. Singh, R. Piryani, A. Uddin & P. Walia in 2012

This paper deals with how the SentiwordNet approach can be used and how its accuracy can be achieved as much as close to other existing ML approaches used for the sentiment analysis process. They have implemented the SentiWordNet approach with different interpretation of linguistic features or classifications, scoring schemes, and aggregation thresholds and used two pre-existing large datasets of Movie Reviews and two Blog posts. The paper also presents a comparison of the accuracy of the SentiWordNet approach with two popular machine learning approaches (Naïve Bayes and SVM for sentiment classification). Increment in Accuracy of Sentiwordnet approach by using Linguistic Features.

1. Using only Adjective terms for obtaining the positive or negative polarity of the extracted content.

In this approach they have explored different linguistic features and scoring schemes. Computational Linguists features suggest that adjectives are good markers of opinions and sentiments. For example, if a review sentence says "The movie was excellent", then the use of the adjective 'excellent' tells us that the movie was liked by the reviewer and possibly he had a wonderful experience watching it. Also extracted. We computed their SentiWordNet score and then checked for the presence of 'Not' immediately before it. If a 'Not' is present the Senti_Score is negated.

2. Using Adverbs + Adjectives terms for obtaining the positive or negative polarity or the extracted content.

In the other variant, they extracted an Adverb+Adjective' combination rather than only using Adjectives had a wonderful experience watching it. Sometimes, Adverbs further alter the opinion or the sentiment expressed in review sentences. For example, the sentence "The movie was extremely good" expresses a more positive opinion and definite sentiment about the movie than the sentence "the movie was good". We have therefore explored with two linguistic classification selection schemes.

We gave equal weightage to Adjectives and Adverb scores obtained from the SentiWordNet. It would, however, be relevant to understand the point that in the English language, adjectives are largely used in an opinionated tone and adverbs are usually used as complements or modifiers.

3. By giving Adverbs relatively less weightage than Adjectives then combining them to provide the result from the dataset.

Afte+r implementing the other tests they tried to achieve more accuracy for the approach. They decided to explore this possibility by assigning different weightage to adjective and adverb SentiWordNet scores and evaluate its performance.

This results in strong adjectives being strengthened in polarity by a lesser value whereas weaker adjectives are strengthened in polarity by a relatively higher value".

In our last variation we decided to fix the scaling factor so as to give a fixed weight to adjective priority. They considered the scaling factor sf = 0.35. This means to provide only 35% weight to adverb scores. Now the modifications in adjective scores are in a fixed proportion to adverb scores. Since the value of scaling factor sf = 0.35, the adjective scores will get a higher priority in the integrated scores.

Method		Performance measure	Value
SWN-1	Dataset1	Classified Positive	513/700
		Classified Negative	387/700
	Dataset2	Classified Positive	759/1000
		Classified Negative	553/1000
	Dataset3	Classified Positive	293/435
		Classified Negative	653/1051
	Dataset4	Classified Positive	232/318
		Classified Negative	288/489
SWN-2	Dataset1	Classified Positive	526/700
		Classified Negative	359/700
	Dataset2	Classified Positive	754/1000
		Classified Negative	510/1000
	Dataset3	Classified Positive	285/435
		Classified Negative	631/1051
	Dataset4	Classified Positive	222/318
		Classified Negative	279/489
SWN (VS)	Dataset1	Classified Positive	499/700
		Classified Negative	410/700
	Dataset2	Classified Positive	725/1000
		Classified Negative	578/1000
	Dataset3	Classified Positive	288/435
		Classified Negative	649/1051
	Dataset4	Classified Positive	232/318
		Classified Negative	291/489
SWN (APS)	Dataset1	Classified Positive	508/700
		Classified Negative	406/700
	Dataset2	Classified Positive	738/1000
		Classified Negative	580/1000
	Dataset3	Classified Positive	290/435
		Classified Negative	643/1051
	Dataset4	Classified Positive	235/318
		Classified Negative	290/489

Figure 2. Result of Four different variants of SentiWordNet Implementations

Performance and Accuracy Comparison.

In order to judge the precision and efficiency of various SentiWordNet based approaches, they calculated the standard performance metrics of Accuracy in percentage, whereas F-measure and Entropy metric values range from 0 to 1. The best accuracy value is 100%, best F-measure value is 1, and of Entropy the best value is 0.

Method	Performance measure	Value
SWN-1	Accuracy	65.6%
	F-measure	0.6523113
	Entropy	0.27848914
SWN-2	Accuracy	63.2%
	F-measure	0.62643987
	Entropy	0.28467706
SWN (VS)	Accuracy	65.2%
	F-measure	0.64960706
	Entropy	0.2802905
SWN (APS)	Accuracy	65.9%
	F-measure	0.65685844
	Entropy	0.27805692
NB	Accuracy	82.9%
	F-measure	0.829
SVM	Accuracy	77.95%
	F-measure	0.779

Figure 3. Output (Accuracy, F-measure, Entropy)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

Paper [3]-by Viraj Parkhe & Bhaskar Biswas in 2015

As we know the sentiment of a consumer toward a specific product can be understood by sentiment analysis of the product review. In our case, as we are working on a movie review dataset for sentiment analysis. In this paper, they tried to include the different aspects of the movie review to direct the polarity of the review with more accuracy. They used machine learning approaches like Naive Bayes and SVM for the process.

For this they suggested an improved Naive Bayes algorithm that reduced the precision gap. It deals with each individual aspect of the movie. A movie has many aspects such as Direction, screenplay, acting, story, etc. and the reviewer may tend to provide his/her opinion or sentiment based on these aspects.

Many researchers have worked on aspect-based sentiment analysis. They et al. (2010) proposed a method for fine-grained analysis of sentiment orientation and sentiment strength of the reviewer towards the various aspects of the movie. It uses domain-specific and generic opinion lexicons to score the words and with the help of dependency tree, it identifies various inter word dependencies and helps in propagating the word score over the entire document.

For comparison and score analysis they took the help of SentiWordNet approach 2015) with feature selection comprising adjective, adverbs, verbs and n-gram features.

They used a different approach and used a randomized approach to assign values to the driving factors. Also, we choose those driving factors that give the maximum accuracy as the best driving factors. In previous studies many researchers used an algorithm in which they formulated the aspect value distribution via a Multivariate Gaussian Distribution.

Proposed Method

- The first step in the method is preprocessing of the data set in which the data is collected from different sources and preprocessed to make it suitable for use in the method.
- The next step included separating the review text into two different aspects and for that the used aspect-based text separator known as ABTS(Figure 4).
- The different aspects that they used or considered are screenplay acting music movie plot and direction of the movie.
- In the next step the separated reviews based on aspects forward to the aspect of specific classifiers. For this purpose, a naive Bayes classifier is used. As it is a machine learning process, so it needs some testing and training.
- Outputs were either -1 or 1 noting that the input text was positively negatively or neutrally oriented(Figure 6).
- Based on the weightage of the driving factors of the movie the aspect-based output is multiplied with the respective driving factor for the result. The higher the value of the driving factor of an aspect the more is its importance in the review and result is obtained(Figure 5).

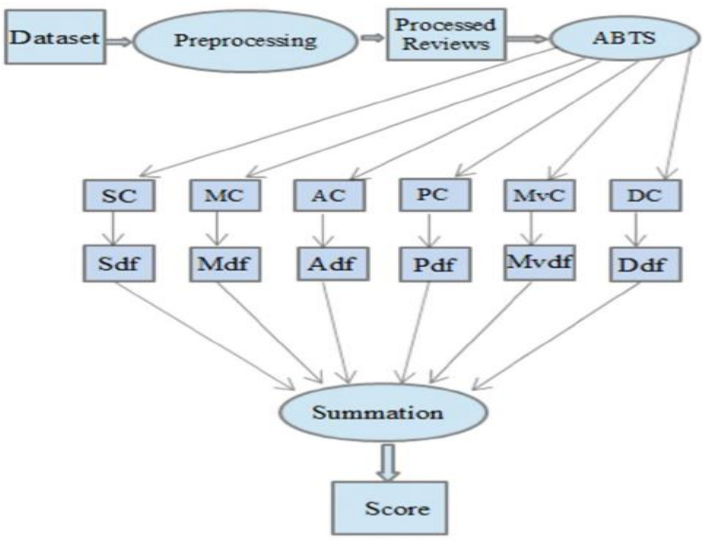

Figure 4. Flow of process.

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

Figure 5. Performance measures.

$$\sum \alpha_i = 1,$$

where $\sum \alpha_i$ is the (ith) driving factor. The net output obtained is the sum of all the classifier outputs obtained multiplied with their respective driving factors. The output is

$$\omega(\mathrm{d}) = \sum \alpha_i X_i \quad X_i \subseteq [-1, 1],$$

where α_i is the driving factor of (ith) aspect and X^i is the output of (ith) classifier and (d) is the document under consideration. Now if

$$\omega(\mathrm{d}) \leq 0 \rightarrow \text{negative classification of review} \quad \mathrm{d}$$
$$\omega(\mathrm{d}) > 0 \rightarrow \text{positive classification of review} \quad \mathrm{d}$$

Figure 6. Calculation of Different aspects.

Dataset, Experimental Results and Performance

The dataset consists of 25,000 positive and 25,000 negative reviews and was collected from IMDB. Though there is no particular time span for review collection from IMDB, but it was made sure that no more than 30 reviews from a single movie get included in the final dataset. Because of even number of positive and negative reviews, the minimum accuracy that we can obtain from the experiment is 50%.[3] . The authors of the dataset included a negative review only if it scored 4 out of 10 and included a positive review if it scored 7 out of 10 on a benchmark set by them (Maas et al. 2011). Neutral reviews were omitted. It was seen that ABTS separated thereview into various aspects having unequal text distribution.

The experiment conducted gave results as depicted in Table 1. The results in Table 1 depict the relationship between accuracy and driving factors used. The highest accuracy obtained was 0.79372, i.e. 79.372%. The corresponding factors are Screenplay—0.07877, Music—0.11756, Acting—0.28147, Plot—0.16390, Movie—

0.31225 and Direction—0.108133. Thus by using the mentioned driving factors, we get an accuracy of 79.372%. This is the highest accuracy they managed to obtain using this method.

Aspect	Aspect words
Screenplay	Scene, scenery, animation, violence screenplay, action, etc
Music	Music, score, lyric, sound, audio, musical title track, etc
Acting	Acting, role playing, act, actress, actor role, portray, character, villain performance, etc
Plot	Plot, story, storyline, tale, romance dialog, script, storyteller ending, storytelling revenge,betrayal,writing, etc
Movie	Movie, film, picture, moving picture motion picture, show, picture show, pic flick, romantic comedy, etc
Direction	Directing, direct, direction, director, filmed filming, film making, filmmaker, cinematic edition, cinematography, etc

Figure 7. Aspect words

Accuracy	Screenplay (DF)	Music (DF)	Acting (DF)	Plot (DF)	Movie (DF)	Direction (DF)
0.79372	0.07877495	0.1175615	0.218479825	0.16390358	0.31225607	0.10813343
0.78956	0.047883925	0.004667212	0.2291204	0.20019746	0.30320117	0.2136368
0.78268	0.165	0.165	0.165	0.165	0.165	0.165
0.77996	0.004460782	0.013207557	0.43918112	0.17664995	0.3274761	0.038970188
0.76912	0.07476745	0.5480689	0.09583495	0.1124063	0.1001353	0.02502478
0.7598	4.44E−05	0.001835718	0.4903199	8.41E−05	0.042956525	0.463557
0.74692	0.002677993	1.86E−04	0.9551522	0.025326777	0.010514759	0.006046253
0.7358	0.001611796	0.49006925	0.01817123	0.003027879	0.010863352	0.473160775
0.7254	0.003411332	0.000922476	0.001115599	0.49038495	3.75E−04	0.5032535
0.71812	0.24075805	0.198029235	0.00240013	0.003753847	2.46E−04	0.55194675

Figure 8. Result and output for Aspects

Paper [4] -by Tirath Prasad Sahu & Sanjeev Ahuja in 2016

They stated that the basic methodology to determine polarity is the one with a lexical approach, where we look at the words comprising the document and apply some algorithms to compute words with some sentiment score and determine the mutual polarity. They have focused on 2 areas:

1) Feature/trait Selection and Ranking.
2) Classification using Machine Learning techniques.

Their proposed methodology is –

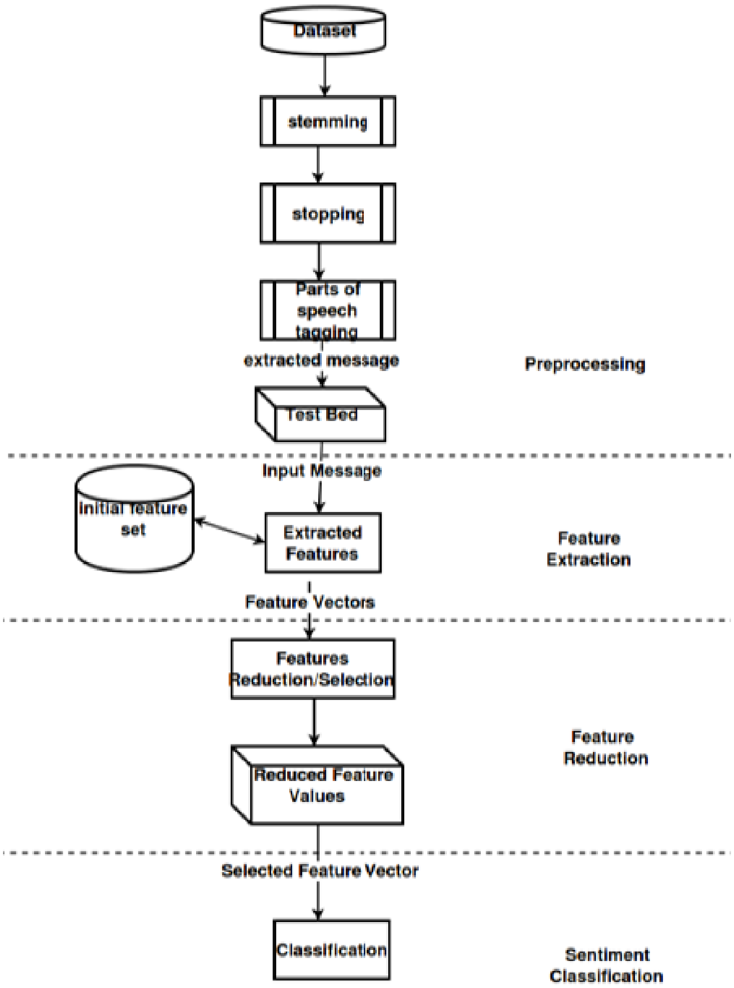

Figure 9. Proposed Methodology

After first three basic steps, sentiment classification is done according to the proposed algorithm.

They have used well known classifiers namely Bagging, Random Forest, Decision Tree, Naive Bayes, K-Nearest Neighbor, Classification via via Regression. The classification is done with the aim to predict the class level for a machine to predict the class of a movie review whenever it arrives.

Classification Technique	Precision	Recall	F-Measure	AUC	Accuracy
Random Forest	**0.892**	**0.89**	**0.89**	**0.983**	**88.95%**
Decision Tree	0.879	0.875	0.876	0.975	87.53%
COCR	0.824	0.825	0.824	0.958	82.53%
Bagging	0.888	0.886	0.886	0.966	88.57%
KNN	0.891	0.889	0.889	0.98	88.86%
Naive Bayes	0.538	0.548	0.541	0.834	54.77%

Figure 10. Evaluation Measures

III. CONCLUSION

As per the work that had been already done on this topic we can conclude that there is no one specific method to do sentiment analysis, as the data changes, we have more than one method and algorithms that can be used for this purpose. Each of the algorithms gives different accuracy according to the data provided.

For example if we talk about our 1st paper we can say that they have analyzed the sentiment of social network comments. They evaluated the fitness of different feature selection and learning algorithms (supervised and unsupervised) on the classification of comments according to their subjectivity and their polarity (positive/negative). The results show that simple bag-of-words model can perform relatively good, and it can be further refined by the choice of features. They also observed that existing corpus from apparently similar corpus which contains sentences from movie reviews. Their results show that such corpus, although contains similar polarity of the words, as well as the common topic, may not perform classification well.

Similarly in 2nd paper from test and results they observed that for movie reviews any of the SentiWordNet based approaches are not able to meet the performance of machine learning approaches like Naïve bayes and SVM. Also if we want to use SentiWordNet approach then they observed that SentiWordNet (Adjective Priority Scoring approach which provides less weightage to Adverbs) obtains the best results. They also conclude that Adjectives are no doubt the most important linguistic characteristic to make use for sentiment analysis, but adverbs which modify the adjectives, improve the performance levels further if they are combined with adjective scores in an appropriate weightage.

In 3rd paper the experiment was conducted to find which movie expects responsible for changing the orientation of the review using driving factors. It concluded that aspects like movie acting and plot adds up and results in an accuracy of 79.372% for the data set taken into consideration. The method used for classifying the text is naive Bayes classifier which uses a method of a bag of words approach. But also the accuracy can change for different genera of the movies based on their movie aspects, so we can conclude that it is not good for the classification of mixed reviews. Also, the bag of word approach does not consider the inter-word meaning dependencies and also the context in which the word was used. For that purpose, we need to develop a specific lexicon method. Thus, by combining the improvement method with the use of general specific driving factors we can get more refined scores for the movie reviews.

It was seen that if classification techniques are used for this purpose then, the highest accuracy was given by Random Forest with an accuracy of 88.95%.

IV. REFERENCES

[1] http://people.csail.mit.edu/kuat/courses/6.863/report.pdf - Sentiment Analysis of Movies Review Comments.

[2] https://ieeexplore.ieee.org/abstract/document/6514345 - Sentiment Analysis of Movie Reviews and Blog Posts Evaluating SentiWordNet with different Linguistic Features and Scoring Schemes

[3] https://link.springer.com/article/10.1007/s00500-015-1779-1 - Sentiment analysis of movie reviews: finding most important movie aspects using driving factors

[4] https://ieeexplore.ieee.org/abstract/document/7522583 - Sentiment Analysis of Movies Reviews: A Study on Feature Selection & Classification Algorithms.

[5] Donalek, Ciro. "Supervised and Unsupervised Learning." (n.d.): n. pag. Apr.2011. Web. 20 June 2016. <http://www.astro.caltech.edu/~george/aybi199/Donalek_Classif.pdf>.

[6] Ganesan, Kavita. "Opinion Mining Tutorial (Sentiment Analysis)." *Opinion Mining Tutorial (Sentiment Analysis)*. SlideShare, 7 Apr. 2011. Web. 20 June 2016. <http://www.slideshare.net/KavitaGanesan/opinion-mining-kavitahyunduk00>.

[7] Infinit Datum. "Sentiment Analysis: Everything You Need to Know." *Infinit Datum*. Infinit Datum, 26 Mar. 2015. Web. 20 June 2016. <http://www.infinitdatum.com/blog/sentiment-analysis-everything-you-need-toknow/>.

[8] Ismael, Chris. "List of 20+ Sentiment Analysis APIs." *Mashape Blog*. Mashape, 24 Apr. 2013. Web. 20 June 2016. <http://blog.mashape.com/listof-20-sentiment-analysis- apis/>.

[9] JAIN, KUNAL. "Machine Learning Basics for a Newbie." *Analytics Vidhya*. Analytics Vidhya, 11 June 2015. Web. 20 June 2016. <http://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/>. 23

[10] Kabadjov, M., Balahur, A. And Boldrini, E. 2009. Sentiment Intensity: Is It a Good Summary Indicator?. Proceedings of the 4th Language Technology Conference LTC, pp.380-384. Poznan, Poland, 6-8.11.2009.

[11] Kiser, Matt. "The Algorithmia Guide to Sentiment Analysis." *Algorithmia Developer Center Atom*. Algorithmia, 28 Jan. 2016. Web. 20 June 2016. <http://developers.algorithmia.com/guides/sentiment-analysis/>.

[12] https://gist.github.com/mkulakowski2

[13] Mora, Michaela. "The State of Sentiment Analysis." *The State of Sentiment Analysis*. Relevant Insights, 3 May 2015. Web. 20 June 2016. <http://www.relevantinsights.com/the-state-of-sentimentanalysis#sthash.pNOCPkrX.CRcKaILs.dpbs>. Narr, Sascha, Michael Hülfenhaus, and Ing. Sahin Albayrak, Dr."Language-Independent Twitter Sentiment Analysis." *Language-Independent Twitter Sentiment Analysis* (2012): n. pag. *CC IRML*. Impressum, 20 Sept. 2012. Web. 20 June 2016. <http://irml.dailab.de/wp-content/uploads/2012/09/narr-LWA2012-TwitterSentiment_Poster.pdf>.

[14] Shankhdhar, Gaurav. "Sentiment Analysis Methodology." *Edureka Blog RSS*. Edureka, 23 Jan. 2014. Web. 20 June 2016. <http://www.edureka.co/blog/sentiment-analysismethodology/>.