

Feature-Based Opinion Mining and Managed Machine Learning with Sentiment Classification Models



Jagdish Chandra Patni, Shubham Billus, Shubhita Garg, Shivam Billus, Romika

Abstract: Sentiment Analysis is individuals' opinions and feedbacks study towards a substance, which can be items, services, movies, people or events. The opinions are mostly expressed as remarks or reviews. With the social network, gatherings and websites, these reviews rose as a significant factor for the client's decision to buy anything or not. These days, a vast scalable computing environment provides us with very sophisticated way of carrying out various data-intensive natural language processing (NLP) and machine-learning tasks to examine these reviews. One such example is text classification, a compelling method for predicting the clients' sentiment. In this paper, we attempt to center our work of sentiment analysis on movie review database. We look at the sentiment expression to order the extremity of the movie reviews on a size of 0 (highly disliked) to 4 (highly preferred) and perform feature extraction and ranking and utilize these features to prepare our multilabel classifier to group the movie review into its right rating. This paper incorporates sentiment analysis utilizing feature-based opinion mining and managed machine learning. The principle center is to decide the extremity of reviews utilizing nouns, verbs, and adjectives as opinion words. In addition, a comparative study on different classification approaches has been performed to determine the most appropriate classifier to suit our concern problem space. In our study, we utilized six distinctive machine learning algorithms – Naïve Bayes, Logistic Regression, SVM (Support Vector Machine), RF (Random Forest) KNN (K nearest neighbors) and SoftMax Regression.

Keywords: Sentiment Analysis, Opinion Mining, Movie Review, Machine learning, Classification Algorithms

I. INTRODUCTION

Today with the huge help of innovation, the web has turned into an exceedingly profitable spot in which thoughts has been traded effectively, internet learning, surveys for an administration or item or motion pictures.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Dr. Jagdish Chandra Patni, Department of Virtualization, University of Petroleum & Energy Studies, Dehradun (Uttarakhand) India.

E-mail: JCPATNI@ddn.upes.ac.in

Shubhita Garg, School of Computer Science, University of Petroleum & Energy Studies, Dehradun (Uttarakhand) India.

E-mail: gargshubhita1@gmail.com

Romika, School of Computer Science, University of Petroleum & Energy Studies, Dehradun (Uttarakhand) India.

E-mail: romikasingh13@gmail.com

Shubham Billus, School of Computer Science, University of Petroleum & Energy Studies, Dehradun (Uttarakhand) India.

E-mail: shubham.billus24@gmail.com

Shivam Billus, School of Computer Science, University of Petroleum & Energy Studies, Dehradun (Uttarakhand) India.

E-mail: billusshivam24@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

It makes hard to comprehend and record the feelings of the client since surveys on the web are accessible for millions for an administrations or items. Sentiments are feelings of the clients seeing substances, for example, items, occasions, issues and administrations that might be great, phenomenal or awful. Analysis of individuals' feelings, responses dependent on input from web is known as sentiment analysis. Sentiment analysis is additionally called as opinion mining. For research, Sentiment analysis is a developing region to gather the abstract data from source material by applying Linguistics and content examination, Natural Language processing, Computational and arranged the extremity of the sentiment or assessment. Sentiment analysis is language preparing task which utilizes computational way to deal with the assessment of client and arrange as positive, negative or impartial. The principle point of analysis of sentiment is to distinguish the opinion of an author or a speaker concerning some theme. The two associations and clients can use opinion mining and sentiment analysis. At the point when a watcher needs to know whether a movie is worth viewing or not, he/she can get to extensive number of clients audits however perusing and dissecting most of the surveys could be a protracted and baffling procedure.

A. Features Based Opinion Mining

In Features based Opinion Mining, important sentences are chosen from a colossal measure of information gathered from studies, remarks, and surveys. After extraction of helpful information from a vast piece of information, keywords identified with item includes are removed. There are ventures for highlight based conclusion mining.

B. Model of Sentiment Analysis

Opinions are commonly communicated for anything. For Example, an administration, an item, an individual, a point, or an association. The substance under perception has distinctive segments and may likewise have sub-parts. In this manner, the element is called an item for sentiment analysis. Highlight based sentiment analysis utilizes the progressive model since articles are various leveled in nature. The article may have sub-parts and characteristics. Thus, it is troublesome for general individuals to comprehend these specialized terms (characteristic or parts). Along these lines, a basic word "Feature" is utilized for opining mining, which is featured-based. Feeling can be communicated in one sentence or in different sentences as a passage. Conclusion word introduction decides the introduction of opinion. One single sentence can have at least one-opinion words.



II. RELATED WORK

From the past years, numerous articles and books have been composed on sentimental analysis.

In the meantime, few scientists center more around topics like finding the subjectivity expression, subjectivity pieces of information, abstract sentence, points, and sentiments of words and separating views source, while others target is on allocating sentiments to entire document. All the analyzers of sentiment analysis have designed a few strategies to naturally predict expression, sentiments of the words or of a whole document. The informational index for sentimental analysis considered are films, item reviews or data from social media through internet. They use methodologies, Natural Language Processing (NLP) and AI techniques.

Cagatay CATAL “unpublished” [1] target of the paper is examine the potential advantage of numerous classifier frameworks idea on Turkish sentiment classification issue and suggest a novel characterization procedure.

Vote calculation had been utilized related to three classifiers, to be specific Naive Bayes, Support Vector Machine (SVM), and Bagging. Parameters of the SVM have been improved when it was utilized as an distinct classifier. Test results demonstrated that numerous classifier frameworks increment the execution of individual classifiers on Turkish sentiment classification datasets and Meta classifiers add to the intensity of these different classifier frameworks. The proposed methodology accomplished better execution over Naive Bayes, which was accounted for the best individual classifier for these datasets, and Support Vector Machines. Various classifier frameworks are a decent methodology for sentiment classification, and optimization of parameters of individual classifiers must be considered while creating MCS-based forecast frameworks.

Rajesh Piryani “unpublished”. exhibited an exploratory work on perspective dimension of sentiment analysis of movie reviews[2]. It fundamentally contains client feeling for different viewpoints, for example, heading, acting, movement, cinematography, and so on. They had planned a linguistic rule based methodology which perceive the perspectives from film audits, finds supposition about that angle and list the sentiment polarity of that conclusion utilizing linguistic approaches. The framework produces a viewpoint level feeling summary. The exploratory plan is assessed on datasets of two movies. The outcomes accomplished great exactness and shows guarantee for sending in a sentiment profiling framework. Asha S Manek “unpublished” [3] actualized sentiment analysis for film audits utilizing different component determination techniques with Naive Bayes and Support Vector Machine (SVM). The proposed work utilizes number of steps, for example, accumulation of movie audits datasets, pre-handling, include choice, order strategies. Result demonstrates that list technique gives better execution with SVM for arrangement for vast amount of dataset and Correlation based component choice with SVM for the little amount of dataset[15]. Bogdon Batrinca “unpublished” proposed a plan of programming device for internet-based life, web journals, talks, newsfeeds and so on and how to utilize them for scratching, purifying and analyzing [4]. For scratching the web-based social networking data it proposes

the difficulties, for example, Data purifying, Data assurance, Data analysis and Visualization and examination Dashboard. This paper displays an overview on strategy of social media, information, suppliers and investigation strategies, for example, stream preparing, sentimental analysis. An outline of various devices required for social analysis design is likewise exhibited. There has been simple accessibility of APIs given by Twitter, Facebook and News administrations which prompted blast of information administrations to scrape and sentiment analysis.

Mrs. R.Nithyaet al. [5] had represented the Sentiment analysis which chiefly on emotional and polarity identification. A planned work includes:

- Feature Extract-Commonly, Sentiment analysis utilizes AI calculations and a strategy to extricate highlights from writings and after that train the classifier.
- Preprocessing-stemming alludes diminishing words to their underlying foundations. Porter is stemming calculation utilized for expelling stop words. For the most part, descriptive word words have sentiment.
- Product angles Text detail is an unreservedly accessible that can be utilized for separating design.
- Find extremity of stubborn sentence-here SentiStrength lexicon-based classifier used to recognize sentiment quality.

Here, 575 surveys have been taken from shopping locales. Tanagra1.4 instrument utilized for information mining. Naive Bayes order done through this device dependent on every individual component, for example, show, extras, battery life and cost. Results demonstrates that 'battery life' have best esteem so it advances marking and 'cost' have low positive esteem that show dealer to focus extra on notoriety and item quality. Minhoe Hur “unpublished” [6] proposed a framework to foresee Box-office accumulation dependent on Sentiments of film survey. They have utilized Viewer conclusions, which are utilized as input factors notwithstanding indicators and three AI based calculations (SVM, Regression tree, AI neural network) were utilized to get non-direct 12789 connection between the movies and its gathered predictors. Aurangzeb Khan, [7] proposed a standard based strategy in which SentiWordNet is utilized to get more precision than a lexicon-based system in its pure form for sentiment analysis for viewer audits and software surveys. The proposed framework has 91% of exactness at the report level and 86% of precision at the sentence level. Mudinas and Zhang, [8] proposed a half and half strategy which gives preferable execution over the lexicon and nearly performs like learning based method. Techniques which are hybrid are steady as lexicon system and execution as AI based methods. The framework has a general exactness of 82.3%. Lei Zhang “unpublished” 2010 [9] proposed a positioning and separating feature of product in opinion records calculation. At first, they have audits of clients and it was hard to decide by the machine to separate between positive surveys and negative audits. They utilized the related rule mining technique for extricating features item.

Seven Rill “unpublished” 2014 [10] an application is proposed which demonstrates Early discovery of developing political themes on Twitter and the effect on idea level sentiment analysis.

In this Paper twitter, hashtags are utilized to decide the election results in the USA even before "Google Trends". Twitter API is utilized to gather information and Analyze utilizing sentiment dissecting a calculation[16,17].

Monu Kumar and Dr. Manju Bala, 2016 [11] recommended that it is hard to break down the colossal measure of unstructured information nowadays assembled from different social networking sites. Accordingly, they utilized cloud administration and used Hadoop for canny analysis and capacity of huge information. Sentimental Analysis of Twitter is finished utilizing the cloud.

Martin Wöllmer “unpublished” 2013 [12] proposed a system to break down sentiments in Audio – Video setting of a YouTube Movie. They utilized Metacritic database to get client surveys as info. They assessed the learning-based methodology, applying information-based methodology in an in-area setting just as in a cross-space setting.

Giuseppe Di Fabrizio “unpublished” 2013 [13] proposed perspective rating appropriations and language displaying which utilized for condensing on the web item and administration reviews. They utilized a novel methodology for separating multi-archive outline for textual information that considers viewpoint rating circulations and language displaying as synopsis highlights.

Chirag Sangani 2013 [14] proposed a technique for investigating client sentiments towards applications through their audit remarks and appraisals can be financially beneficial to application engineers. They propose a framework that gives a rundown of audits to every subject that speaks to client feelings towards that point and a many-to-numerous connections depicting from surveys to themes of intrigue.

The machine learning approaches we are using are:

A. Naïve Bayes

Naïve Bayes is a model of conditional probability. Despite it is very simple and make use of strong assumptions, naïve Bayes classifier worked satisfactorily in many different domains. Bayesian classification gives practical knowledge of learning algorithms and prior information and data observed can be consolidated. In Naïve Bayes strategy, the essential thought is to discover the probabilities of classes given a content record by utilizing the joint probabilities of the words and classifications. It depends on the suspicion of word independence. The Bayes theorem, states that for a given information point x and class C :

$$P(C/x) = P(x/C)/P(x) \quad (1)$$

Besides, by making the supposition that for an information point $x = \{x_1, x_2, \dots, x_j\}$, the probability of every one of its attributes occurring in the given class is independent and probability of x can be estimated as:

$$P(C/x) = P(C). \prod P(x_i/C) \quad (2)$$

B. Support vector machine

Support vector machine is a machine learning strategy for data of linear or nonlinear type. Support vector machine (SVMs) have been appeared to be very successful at conventional text classification. In SVM the information is

mapped to high dimension. SVM scans for hyperplane with the biggest edge, that is, the maximum marginal hyper-plane. The related margin gives the biggest separation between the classes. The fundamental SVM takes a lot of information and predicts, for each given info, which of two potential classes frames the output. In the two classification case, the essential thought behind the preparation methodology is to locate a maximum margin hyper plane, which is represented by the vector, that not just separates vectors in a single class from those in the other, but for the one for which the partition, or margin, is as substantial as could be expected under the circumstances.

C. Logistic Regression

Logistic Regression is a kind of algorithm for classification including a linear discriminant. In contrast to actual regression, logistic regression does not attempt to predict the estimation of a numeric variable given a input data set. Rather, the output is given as probability for provided inputs has a place with a specific class. For example, how about we expect that we have just two classes and the probability being referred to is $P_{+} >$ the probability that a specific information point has a place with the ‘+’ class and obviously $P_{-} = 1 - P_{+}$. The output of the Logistic Regression will be in between $[0, 1]$. The main reason of Logistic Regression is the supposition that your information space can be segregated into two 'areas', one for each class, by straight boundary. So, for two dimensions, the linear boundary is a straight line and for three dimensions, it's a plane. This line will obviously be chosen by the input information and the learning calculation. In any case, for this to bode well, the information must be segregated into the two areas by a direct limit. In the provided input data do fulfill this requirement, they are said to be separable linearly.

D. Softmax Regression

Softmax regression is general form of logistic regression used when we want multiple classes to be handled. Logistic regression has assumption that labels are binary: $y(i) \in \{0,1\}$. Whereas Softmax regression lets us to handle $y(i) \in \{1, \dots, N\}$ where N is the number of classes. Through logistic regression, we can only classify the information into two classes. Whereas in softmax regression, we can classify into multiple classes (opposed to binary classification), and so label y can output on N different values, rather than two. Thus, in the training set as $\{(x(1), y(1)), \dots, (x(m), y(m))\}$ now output label y can have more than two values $y(i) \in \{1, 2, \dots, K\}$.

E. KNN

KNN calculation is one of the least complex or simplest classification approach and it is most utilized learning approach. KNN is a non-parametric and lazy approach. Its motivation is to utilize a database in which the information focuses are segregated into a few classes to predict the out of another example. When we state a strategy is non-parametric, it suggests that it doesn't make assumptions at all on the basic information. As it were, the model structure is resolved from the information.

All things considered, it's quite valuable, because in "this present reality", the vast majority of the information does not comply with the regular hypothetical assumptions made (as for example in linear regression). In this manner, KNN could and most likely ought to be one of the primary decisions for classification when there is practically zero earlier information about the distribution information.

KNN is additionally lazy which imply that KNN does nothing, which means it doesn't use the training data to do any type of generalization. At last, there is no training model, it is insignificant. This likewise implies the preparation stage is quick. Absence of generalization implies that KNN keeps all the training information. To be increasingly precise, all (or most) the training information is needed amid the testing stage.

F. Random Forest

Random Forest is an adjustable, simple to use algorithm that results in, without the hyper-parameter tuning, an extraordinary results usually. It is also a standout amongst the most used approaches, as it's very simple and the mode that it tends to be used for both regression and classification type of tasks. As clear from its name, it makes a forest and makes it some way or another random. The "forest" it constructs, is a gathering of Decision Trees, more often trained with 'bagging' technique. The general thought of the bagging technique is that a mix of learning models builds the general outcome. To state it in basic words: Random forest constructs decision choice trees and combines them to get an increasingly precise and stable prediction.

Random Forest has about a same hyperparameters as decision tree. Luckily, a decision tree need not to be combined with a bagging classifier and just classifier-class of Random Forest is needed to be used. With Random Forest, Regression tasks can be dealt by utilizing the Random Forest regressor. Random Forest adds extra randomness to model, while developing the trees.

III. METHODOLOGY

To compare various techniques to classify the sentiment of a text review, we have defined various buckets under which an algorithm works. One algorithm from a bucket is compared to other algorithms from the same bucket, and in the end, an overall comparison of all the algorithms through all the buckets is performed. A bucket is defined by two things:

- The text representation algorithm used, and
- How multi-class classification is performed.

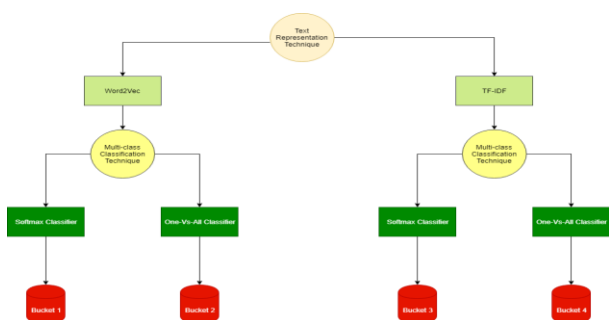


Fig 1. Buckets

We have analyzed the two most popular ways to represent text, namely TF-IDF word representation and Word2Vec word representation. Then further, we have analyzed both these word representations over both the ways that are available to perform a multi-class classification, namely, a softmax classifier and a one-vs-rest classifier.

Step-1 Input to the System

There are three inputs to the system:

- Labeled Training Dataset with (Review, Sentiment) pairing.
- Labeled Test Dataset with (Review, Sentiment) pairing.
- Pre-defined weight values for Word2Vec representation of words through Google's 300 dimensional weights (negative weight dimensions included), which is trained over a corpus size of 5 billion text sentences from Wikipedia and contains 300 million unique vocabulary words.

Step-2 Deciding on the Deep Learning Algorithms

The next step is to decide which algorithms are to be used to classify sentiment of a text. For example, when using Word2Vec representation of text, a Naïve Bayes classifier does not work as some of the dimensions in text representation have negative values. Therefore, we must set a very specific set of algorithms for each bucket to ensure:

- Comparable algorithms across all buckets.
- Valid/Usable algorithms for the specific bucket.

Step-3 Implement each selected algorithm for every bucket

After selecting upon the algorithms for each bucket, all the algorithms are implemented and trained over the training data for sentiment classification in each bucket.

The process of training a deep learning model can be complex. Special attention must be given to over-fitting and under-fitting on the training data. To do this, the training data is split into train-set and validation-set. The validation-set acts as unseen internal test-set to detect if a model is not over-fitting or under-fitting on the trained data. Once the model performance is good, enough on both, train-set and validation-set, the model is ready to be tested on the actual test set.

Step-4 Test performance of algorithms

After training each algorithm properly, the performance of the algorithms is tested on unseen training data.

The accuracy of each model in each bucket is noted for further analysis of the outcomes from the testing phase.

Step-5 Analyze the Results

In the final step, all the accuracy scores from various algorithms are analyzed to draw insights from this research experiment.

The two main tasks at hand in this final step are:

- Compare the performance of all algorithms within a bucket to know which algorithm performed the best in local conditions.



- Compare the best performers from all buckets to analyze which algorithm was the best overall and which bucket was the ideal bucket for the task. After performing both the tasks, hypotheses are drawn based on the outcomes from both these tasks. One hypothesis is drawn for why the best performing algorithm in a bucket came out to be the top performer, and why the worst performer is at the bottom and one is drawn for the overall best performer which is then used to draw insights about the dataset and sentiment analysis process itself.

IV. RESULTS

A. Data Overview

The distribution of user reviews across all the possible sentiments is as shown in Fig. 2. As one can observe, the data is not very well balanced and hence it makes the job even harder for the models to train and capture characteristics of each class properly.

B. TF-IDF Word Representaion with Multi-class Classifier

The best possible results from all the classification models in their best found hyper parameters settings is shown in Fig. 3. These results came out from the softmax classifier that was moderately deep but not as deep as softmax_1. This goes on to show again that a deep neural network is capable enough to extract patterns from sentence representations of reviews to classify them properly.

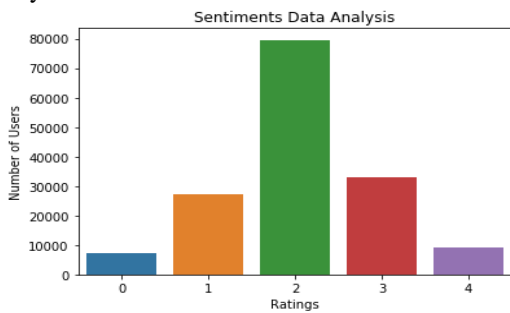


Fig 2. Data Distribution

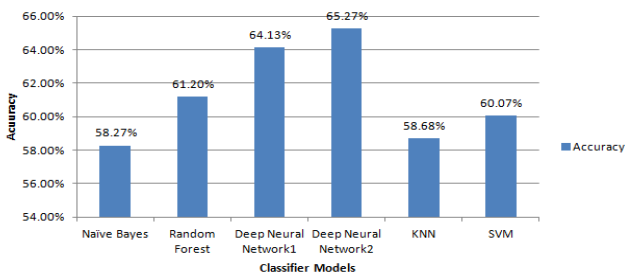


Fig 3. TF-IDF Weights with Multi-Class Classifier

Also, another observation from this experiment is that Naïve Bayes performed the worst for the dataset which is presumably due to the fundamental assumption that Naïve Bayes classifier follow that each input dimension is independent of every other input. The text data that goes as input in NLP applications is rarely independent of sequence. In fact, textual input is notorious of being difficult to work with due to high interdependence of words in a textual

sentence. Hence, Naïve Bayes performed the worst of them all.

The only decision-based model on our list of models – Random Forest Classifier, performed adequately well too. The reason for that is there being 100 trees (hyper parameter) in the classification model, the general error, in theory, tends to get mitigated by averaging all of them out.

C. TF-IDF Word Representaion with One-vs-All Classifier

The results from using tf-idf word representation with a one-vs-all classifier are as shown in Fig. 4. One important thing to note here, we were able to use a logistic regression classifier by using a one-vs-all classifier. And evidently, logistic regression came out to be on top of all the different classification models. The reason for that is the sophisticated yet simple mathematics that logistic regression classifier follows. Again, the worst performer of all came out to be Naïve Bayes classifier, again because of the assumption that the input parameters are independent of each other.

D. Word2Vec Word Vectors with Multi-class Classifier

The results from using pre-trained word2vec weight representation with a multi-class softmax classifier came out to be as in Fig. 5.

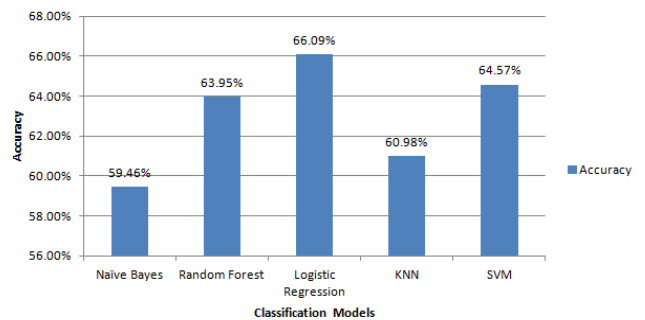


Fig 4. TF-IDF Weights with One-vs-All Classifier

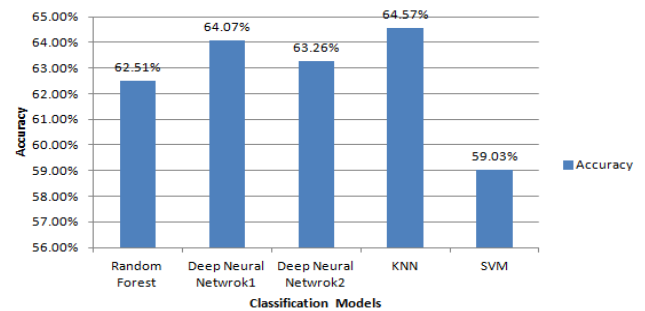


Fig 5. Word2Vec Weights with Multi-Class Classifier

KNN came out to be the best model in the case. This was expected as word2vec is a more proximity-based representation of words. So if the sentence formed through the weights follow the same property as the words that made up the sentence, then the sentences made from a complete different set of words (used to define a different sentiment class perhaps?) would be far apart to be distinguished based upon the distance of that vector in the vector space of all sentences. The worst of the performers came out to be SVM.

This may happen because of the multi-dimensional data not being as linearly separable as before (using TF-IDF weights). One important thing to note here, a Naïve Bayes model couldn't be trained using these weights as Naïve Bayes algorithm assume that the input data come from a normal data distribution. And normal data distribution cannot have negative values, which Word2Vec dimensions have but TF-IDF representations didn't (the lowest value of TF-IDF Weight is 0). Therefore, a Naïve Bayes classifier couldn't be trained on Word2Vec weights.

E. Word2Vec Word Vectors with One-vs-All Classifier

The results from using pre-trained word2vec weight representation with a one-vs-all SoftMax classifier came out to be as in Fig. 6. KNN came out to be the best model in the case. This was expected as word2vec is a more proximity-based representation of words. The continuous good performance scores of logistic classifiers testify the mathematics that it follows.

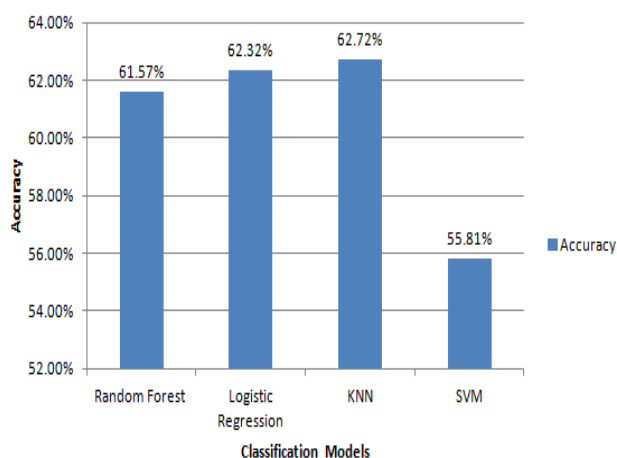


Fig 6. Word2Vec Weights with One-vs-All Classifier

The worst of the performers came out to be SVM. This may happen because of the multi-dimensional data not being as linearly separable as before (using TF-IDF weights).

V. CONCLUSIONS

The best results from all the different models came out to be Logistic Regression with TF-IDF Weights and One-vs-All Classifier with an accuracy of 66.09%. This again goes on to prove that the simple sophisticated mathematics behind the logistic regression backpropagation formula is capable enough to draw hyperplanes that very well distinguish between different binary classes. The classification technique that came on top was also expected as a standard One-vs-All Classifier theoretically should beat a softmax classifier, given more time.

The only surprising result from the experiment came out to be the TF-IDF word representation performing better than Word2Vec word representation which is pretty much a standard in this industry. This goes on to show that the sequencing information does not matter when it comes to movie reviews sentiment analysis and one can simply skip this translation for classification.

REFERENCES

1. C. Catal, M. Nangir, "A Sentiment Classification Model Based On Multiple Classifiers", Applied Soft Computing Elsevier, Vol.50, pp 135-141, 2017.
2. R. Piryani, V. Gupta, V. K. Singh and U. Ghose, "A Linguistic Rule-Based Approach for Aspect-Level Sentiment Analysis of Movie Reviews", Advances in Computer and Computational Sciences, Springer Nature Singapore Pte Ltd, Vol 1, pp 201-209, 2017.
3. A.S. Manek, P.D. Shenoy, M.C. Mohan and Venugopal K R, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier", World Wide Web Internet and Web Information Systems Springer, Volume 20, Issue 2, pp 135-154, 2016.
4. B. Batrinca, P. C. Treleaven, "Social media analytics: a survey of techniques, tools and platform", AI & Society Springer, Vol.30, Issue.1, pp 89-116, 2015.
5. Mrs. R.Nithya, Dr. D.Maheshwari, "Sentiment Analysis on Unstructured Review", 14 Proceedings of the International Conference on Intelligent Computing Applications IEEE, , pp 367-371, 2014.
6. Minhoe Hur Seoul National University "Box-office forecasting based on sentiments of movie reviews and Independent subspace method", Information Sciences, 2016.
7. A. Khan, B. Baharudin, K. Khan, "Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure" ICSECS 2011: 2nd International Conference on Software Engineering and Computer Systems, Springer, pp. 317-331, 2011.
8. A. Mudinas, D. Zhang, M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York, NY, USA, Article 5, pp. 1-8, 2012.
9. Lei Zhang University of Illinois, Chicago "Extracting and Ranking Product Features" Coling 2010: Poster Volume, pages 1462-1470, Beijing, 2010.
10. Seven Rill Goethe University Frankfurt, Germany "PoliTwi- Early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis", Elsevier, 2014.
11. Monu Kumar Thapar University, Patiala "Analyzing Twitter sentiments through big data", IEEE, 2016.
12. Martin Wöllmer Technical University of Munich, Germany "YouTube movie reviews- Sentiment analysis in an audio-visual", IEEE Computer Society, 2013 .
13. Giuseppe Di Fabbri A&T Research Labs, USA "Summarizing Online Reviews Using Aspect Rating Distributions and Language Modeling", Digital Object Identifier IEEE, 2013 Chirag Sangani Stanford University, USA "Sentiment Analysis of App Store Reviews", 2013.
14. Chirag Sangani Stanford University, USA "Sentiment Analysis of App Store Reviews", 2013
15. Parul Mittal, Piyush Mishra, Vivek Patel , Jagdish Chandra Patni, "Comparison of Runtime Performance Optimization using Template-Metaprogramming" presented and published in International Conference on Next Generation Computing Technologies (NGCT), CCIS, volume 827 ,pp. 139-147, 30-31 October 2017, DOI https://doi.org/10.1007/978-981-10-8657-1_11.
16. Piyush Mishra, Vivek Patel, Parul Mittal, Jagdish Chandra Patni "Algorithm Analysis Tool Based on Execution Time Input Instance-based Runtime Performance Benchmarking" published in International journal of computer applications, pp: 27-30, 2018, ISSN: 0975 - 8887.
17. Jagdish Chandra Patni, Ravi Tomar, Ankur Dumka, Hitesh Kumar Sharma, "A Model for Embedded Machine learning and Genetics in IoT" published in International Journal of Current Engineering and Scientific Research (IJCESR), 2017, Vol 4, Issue 10, pp 12-20, ISSN-2394-0697.



AUTHORS PROFILE



Dr. Jagdish Chandra Patni, working as Assistant Professor at School of Computer Science , University of Petroleum and Energy Studies , Dehradun. He completed Ph.D. in the area of High Performance computing in the year 2016 M. He has completed M. Tech. and B. Tech. respectively in the year 2009 and 2004.. His area of research are Database Systems, High Performance computing, Software Engineering, Machine Learning. He has published more than 50 research papers in Journals/Conferences of Repute and published 3 International books, 2 book chapters. He is Guest Editor/Reviewer of various referred International journals. He has delivered 15 Keynote/Guest speech in India and abroad. He has organized more than 5 International Conferences as Chair/Secretary and 6 National Workshops as Chair/Co-Chair.

Shubham Billus, is student B.Tech CSE at University of Petroleum and Energy Studies, Dehradun. His area of interest is Data mining, Machine learning and image processing.



Shubhita Garg, is student B.Tech CSE at University of Petroleum and Energy Studies, Dehradun. His area of interest is Data mining, Machine learning and image processing.



Shivam Billus, is student B.Tech CSE at University of Petroleum and Energy Studies, Dehradun. His area of interest is Data mining, Machine learning and image processing.



Romika, is student B.Tech CSE at University of Petroleum and Energy Studies, Dehradun. His area of interest is Data mining, Machine learning and image processing.