A COMPARATIVE STUDY OF SENTIMENT ANALYSIS

FOR MOVIE REVIEWS

Student Name: SANKET TALAVIYA

Student ID – 969198

Under the supervision of

ANURAG MAJI

**Research Proposal**

Liverpool John Moores University - Master’s in ML and AI.

DECEMBER 2020

# Abstract

In today’s world, Sentiment analysis has become the need of the hour for opinion mining on variety of subjects. It has various applications in marketing, e-commerce, advertising, politics and research. Day by day people expressing their sentiments over internet, different websites and social media are increasing. Sentiment analysis is the process of identifying the classifying these opinions. If one doesn’t apply machine learning and AI techniques to find out overall sentiments of the particular topic going through all reviews/comments. It would take humongous effort to label this review comments manually to figure out overall sentiments. So, If we automate sentiment analysis activity using machine learning and AI, It can help saving lot of human efforts. Also, one problem in this space is that best model depends on the characteristics of the dataset model is trained on. In this research work I am going to compare different models like Logistic Regression, SVM, XGBoost, Random Forrest Ensemble and the state-of-the-art models like BERT. All these models I am going to try after applying ﻿(1) TF-IDF (2) Word2Vec Embeddings and (3) BERT Embeddings. And then using different interpretability techniques I will derive some insights for each of the models.

**Table of Contents**

[Abstract 2](#_Toc69066180)

[1. Introduction 5](#_Toc69066181)

[2. Background 6](#_Toc69066182)

[3. Problem Statement OR Related Research OR Related Work 6](#_Toc69066183)

[4. Research Questions (If any) 6](#_Toc69066184)

[5. Aim and Objectives 6](#_Toc69066185)

[5. Significance of the Study 6](#_Toc69066186)

[6. Scope of the Study 6](#_Toc69066187)

[7. Research Methodology 6](#_Toc69066188)

[7.1 Dataset Description 7](#_Toc69066189)

[7.2 Text Pre-Processing 7](#_Toc69066190)

[7.3 Model Building 8](#_Toc69066191)

[7.4 Research Methodology Flow Chart 10](#_Toc69066192)

[7.5 Model Evaluation 11](#_Toc69066193)

[7.6 Interpretability of the models 12](#_Toc69066194)

[7.7 Results 12](#_Toc69066195)

[8. Requirements Resources 13](#_Toc69066196)

[9. Research Plan 13](#_Toc69066197)

[References 13](#_Toc69066198)

**List Of Figures**

|  |  |
| --- | --- |
| Figure 1 | TF-IDF Representation |
| Figure 2 | Skip Gram Representation |
| Figure 3 | BERT embeddings Representation |
| Figure 4 | Research Methodology Flow-Chart |

**List Of Tables**

|  |  |
| --- | --- |
| Table1 | Dataset columns overview |

**List Of Abbreviations**

|  |  |
| --- | --- |
| EDA | Exploratory Data Analysis |
| AUC | Area Under ROC Curve |
| ROC | Receiver Operating Characteristic |
| BOW | Bag Of Words |
| TF | Term Frequency |
| IDF | Inverse Document Frequency |
| BERT | Bidirectional Encoder Representations from Transformers |
| SVM | Support Vector Machine |
| XgBoost | Extreme Gradient Boosting |
| LIME | Local Interpretable Model-Agnostic Explanations |
| SHAP | Shapley Additive Explanation |
| KNIME | Konstanz Information Miner |

# 1. Introduction

With the genesis of the internet and the world wide web, we have seen an unprecedented growth of data and information on the web as well as a huge growth in digital or textual opinions, sentiments and attitudes as remarked upon in reviews. (Yousefpour et al., 2014). Sentiment analysis field has attracted researchers from different disciplines, especially from computer science as it falls under interactive computation or human computer interaction (HCI). Sentiment analysis is mainly a classification problem that merges both domains natural language processing (NLP) and machine learning (ML)(Qaisar, 2020). Sentiment analysis helps us to understand the relationship between natural text and humans emotions or judgment. It helps us to review a person’s perspective about an entity which means a great deal to the producer of the entity. For instance, in this age, nobody goes to watch a movie unless they heard some good reviews of that film in social media or from some film critics The case is also the same in buying products. So reviews are taking on the marketing world.(Haque et al., 2019)

Although algorithms have progressed pretty well in last few years few challenges in this field have not been completely eliminated yet. One of these challenges is to classify statements to positive or negative even though it doesn’t have any clear emotions and sometimes statements can have words which have different meanings based on the contexts so that leads in to ambiguity in the end result of the classification.(Qaisar, 2020)

﻿Why movie domain?

﻿ Recently studies conducted regarding online opinions, comments, writing reviews, discussion blogs. The most of the piles are used by the film industry includes songs, movie premier, trailers, television programs, and radioactivity to convey the ﻿actual profit earned by the movies. Since movies play a significant role in the entertainment market not only to entertain people but also dependent on ratings and profits on the basis of reviews across the globe. Ultimately it helps in generation of leads with the help of previous data. The movie review dataset is obtained from website Kaggle. The publicly accessible website for movie review is IMDB website. (Dholpuria et al., 2018)

In this research I am going to apply following steps on IMDB movie review dataset.

1. Retrieve IMDB dataset.
2. Pre-processing and balancing dataset.
3. Feature extraction.
4. Applying classical models like Logistic Regression, SVM, XGBoost, Random Forrest and state of the art BERT model on top of vectorization methods like TF-IDF, Word2Vec embeddings and BERT embeddings.
5. Compare performances of each model using AUC (area under ROC curve) method and confusion metrics will be shared in the results.
6. Also, I will try to extract the part of the sentence which contributed the most to derive to specific class of the sentiment.

# 2. Background

Sentiment analysis is useful for finding out how customers feel about a particular topic or product or an idea. Before computer era started, Sentiment analysis used to be done primarily on written paper documents. Today, Sentiment analysis as a field has evolved and it is used to mine information from internet including texts, blogs, social medias, news articles, publications, blogs, reviews, comments etc. This is achieved using variety of techniques like NLP, Statistics and machine learning methods. Mainly there are two types of sentiment analysis techniques. (1) Machine Learning based techniques. (2) Lexicon based techniques

1. **Machine Learning based techniques**: This approach mainly falls into supervised classification category, in this technique two set of datasets are needed, training dataset and test dataset, Training dataset is used to learn various characteristics of documents and test document is used to check how well the classifier has learnt. Logistic Regression, SVM and Naïve Bayes have achieved good performance in this area. One important decision to make in this method is to do feature selection. It can tell how documents are represented. The most commonly used features in this method are as (1) Term Presence and their frequency (2) Part of Speech Information (3) Negations (4) Opinion words and phrases.
2. **Lexicon based techniques**: It is unsupervised technique as it requires prior knowledge of sentiment lexicons. Classification is done by comparing the features of a given statement against sentiment lexicons whose sentiment values/labels are known prior to their use. These lexicons contain the word list that represents people’s subjective feelings and opinions. Once we have these lexicons handy next task is to go through document and find out positive and negative lexicons in each statement and if statement has more positives it’s a positive sentiment otherwise it’s a negative sentiment. There are three methods to construct a sentiment lexicon (1) Manual construction (2) Corpus based (3) dictionary based.

(Vohra and Teraiya, 2013)

In this research we are going to focus on Machine learning based techniques and try out different word embedding techniques with machine learning models and try finding out which works the best for movie review dataset.

# 3. Related Work

(Lewis, 1945) compared the performance of 3 classifiers Naïve Bayes, Maximum Entropy and Support Vector Machines in sentiment classification using various features like unigram, bigram, combination of unigram and bigram, combining unigrams with part of speech, considering only adjectives and combining unigrams and position information and result shown than feature presence is more important than feature frequency. And when feature space is small naïve bayes performs better than SVM and when feature space is large SVM performs better than naïve bayes, when feature space is large Maximum Entropy’s performance is better than Naïve Bayes, but it may suffer overfitting.

(Tamara and Milićević, 2018) compared ﻿traditional models such as a bag-of-words, bag-of n-grams and their TF-IDF variants combined with linear classifiers such as Logistic Regression and SVM, and deep learning models such as word-based convolutional neural networks (ConvNets) and the simple long short-term memory (LSTM) ﻿recurrent neural network. Various document representation techniques such as Paragraph Vector or using pre-trained Word2Vec and Glove word embeddings to compute the vector for each word in the document were tested, and word vectors are aggregated using the element-wise mean. It is shown that deep learning models perform better on our large dataset than traditional models. LSTM resulted with the best accuracy of 95.55%. Deep learning models generally work better than traditional models as training set size increases. In the traditional models Bag of n-grams + TF-IDF with Linear SVM as machine learning model gave the best accuracy score of 92% and Bag of n-grams with Gradient boosting as machine learning model gave the worst accuracy of 81%.

(Al-Saqqa and Awajan, 2019) did the survey and found that most studies were used the two methods of word2vec: CBOW and skip-gram and compared the results from each method. Skip gram is better for infrequent words than CBOW, however, CBOW is faster and works well with frequent words. Many of the studies in literature applied word2vec using tools such as word2vec tool and FastText.

(Tripathy et al., 2016) concluded in the research that, As the value of ‘n’ in n-gram increases the classification accuracy decreases i.e., for unigram and bigram, the result obtained using the algorithm is remarkably better; but when tri-gram, four-gram, five-gram classification are carried out, the value of accuracy decreases. Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient De- ﻿scent(SGD), and Support Vector machine (SVM) algorithms were applied using n-gram approach on IMDb dataset in this research.

(Sahu and Ahuja, 2016) examined IMDB dataset and proposed an approach to classify sentiment expressions from 0 (highly disliked) to 4 (highly liked) and performed feature extraction and ranking and use these features for training multi-label classifier to classify movie reviews to a correct label. They followed approach based on n-grams and after applying classification models they achieved accuracy of 88.95% as a best one.

(Tarımer et al., n.d.) conducted a study on IMDB and twitter dataset, they created a vector space in KNIME analytics platform and did a classification study on this vector space using Decision Tree, Naïve Bayes SVM classification algorithms. The classification results for IMDB movie comments are obtained as 94,00%, 73,20%, and 85,50% by Decision Tree, Naive Bayes and SVM algorithms. And they concluded that for both IMDB dataset and twitter dataset SVM produced the best results.

(Yasen and Tedmori, 2019) In this research, tokenization was employed to transfer the input string into a word vector, stemming was utilized to extract the root of the words, feature selection was conducted to extract the essential words, and finally classification was performed to label reviews as being either positive or negative. A model that makes use of all of the previously mentioned methods was presented. The model was evaluated and compared on eight different classifiers. The model was evaluated on a real-world dataset. In order to compare the eight different classifiers, five different evaluation metrics were utilized. The results show that Random Forest outperforms the other classifiers. Furthermore, Ripper Rule Learning performed the worst on the dataset according to the results attained from the evaluation metrics.

# 4. Research Questions

After all related research following research questions have been formulated.

1. Which vectorization technique works best with white box and black box models from TF-IDF, Word2Vec embeddings and BERT embeddings.
2. If these white box and black box models can be interpreted on this dataset to explain why they worked better/worse than other models.

# 5. Aim and Objectives

Aim of this research is to find out which vectorization technique and machine learning model works the best for analysing sentiments on movie review dataset. Here in this research we are going to apply 3 different vectorization techniques TF-IDF, Word2Vec Embeddings and BERT Embeddings to various machine learning models like Logistic Regression, SVM, Random Forrest and BERT. Also provide interpretability for these models using various interpretability techniques at the end.

Based on the aim of the study, Following objectives have been identified.

1. Build different word embeddings using TF-IDF, Word2Vec and BERT.
2. Build predictive models using all above 3 embeddings and evaluate model’s performance.
3. Compare the performance of all the models that are built and figure out which worked the best.
4. Provide interpretability for each of the models in terms of features and relevant explanations.

# 5. Significance of the Study

Movie reviews help users decide if the movie is worth their time. A summary of all reviews for a movie can help users make this decision by not wasting their time reading all reviews. Movie-rating websites are often used by critics to post comments and rate movies which help viewers decide if the movie is worth watching. Sentiment analysis can determine the attitude of critics depending on their reviews. Sentiment analysis of a movie review can rate how positive or negative a movie review is and hence the overall rating for a movie. Therefore, the process of understanding if a review is positive or negative can be automated as the machine learns through training and testing the data. This project aims to rate reviews using two classifiers and compare which gives better and more accurate results. Classification is a data mining methodology that assigns classes to a collection of data in order to help in more accurate predictions and analysis.

(Lakshmi Devi et al., 2020)

# 6. Scope of the Study

Due to the limitation of time frame, the scope of the research will be limited as below:

1. The data for research is directly taken from Kaggle and the data validation will not be part of this research.
2. The research will include the development and comparing the performance of different machine learning models using TF-IDF, Word2Vec Embedding and BERT embeddings.
3. The study will limit the use of algorithms to Logistic Regression as interpretable models and SVM, XGBoost, Random Forrest as part of black box models. Deep learning models will not be evaluated in this research or one model like BERT will be evaluated only if time permits.
4. For comparing performance of different models AUC will be used and for interpretability we are going to limit 2 methods LIME and SHAP only.

# 7. Research Methodology

The detailed flow of research will be discussed here in detail which includes sequential flow to perform this study.

This research will be carried in three parts sequentially with the help of IMDB movie review dataset. The main objective of this study is to predict whether given comments have positive or negative sentiments and possibly explain the phrase/word of the comment that contributes the most to derive to a particular classification (positive/negative) of the sentiment.

● First part entails understanding the dataset and then apply pre-processing steps, perform EDA followed by feature extraction.

● In the second part, apply create different vectors using (1) TF-IDF (2) Word2Vec Embeddings (3) BERT embeddings.

● Third part is to create models like Logistic Regression, SVM, XGBoost, Random Forrest with all these 3 embeddings and conclude the study by performing comparative study on these model’s performance and derive inferences on which model worked better with some interpretability facilitator libraries like LIME and SHAP.

# 7.1 Dataset Description

IMDB dataset having 50K movie reviews for natural language processing or Text analytics.  
This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, predict the number of positive and negative reviews using either classification or deep learning algorithms.(IMDB Dataset of 50K Movie Reviews | Kaggle, 2021)

|  |  |
| --- | --- |
| Variable | Definition |
| Review | User’s review statements are given as review column. |
| Sentiment | Classification of sentiment, i.e. positive, negative |

Table1: Dataset columns overview

# 7.2 Text Pre-Processing

Exploratory Data Analysis (EDA) is an approach which seeks to explore the most important and often hidden patterns in a data set by using a stepwise approach as below.

1. Remove BR tags
2. Remove all single characters
3. Substituting multiple spaces with single space
4. Remove links
5. Remove words containing numbers
6. Converting to lowercase
7. Lemmatization
8. Feature Extractions using TF-IDF, Word2Vec and BERT Embeddings

# 7.3 Model Building

We have divided the dataset with 80:20 ratio where 80% data will be considered as training data and 20% will be treated as test dataset (hidden).

The dataset we have chosen has target values as positive or negative only hence it’s a binary classification problem. So, we can apply following models on it with different embeddings and compare their performances later

1. Logistic Regression
2. SVM
3. XGBoost
4. Random Forrest
5. BERT

These all models we need to apply separately with TF-IDF, Word2Vec Embeddings and BERT Embeddings. Below is a brief summary around these methods.

**TF-IDF:**

This is another method which is based on the frequency method but it is different to the bag-of-words approach in the sense that it takes into account not just the occurrence of a word in a single document (or tweet) but in the entire corpus.

TF-IDF works by penalising the common words by assigning them lower weights while giving importance to words which are rare in the entire corpus but appear in good numbers in few documents.(Twitter Sentiment Analysis - word2vec, doc2vec | Kaggle, 2021)

Let’s have a look at the important terms related to TF-IDF:

* TF = (Number of times term t appears in a document)/(Number of terms in the document)
* IDF = log(N/n), where, N is the number of documents and n is the number of documents a term t has appeared in.
* TF-IDF = TF\*IDF

Table

Description automatically generated

Figure 1: TF-IDF Representation

(BoW Model and TF-IDF For Creating Feature From Text, 2021)

**Word2Vec Embeddings:**

Word embeddings are the modern way of representing words as vectors. The objective of word embeddings is to redefine the high dimensional word features into low dimensional feature vectors by preserving the contextual similarity in the corpus. They are able to achieve tasks like **King-man + woman = Queen**, which is mind-blowing.(Twitter Sentiment Analysis - word2vec, doc2vec | Kaggle, 2021)

The advantages of using word embeddings over BOW or TF-IDF are:

1. Dimensionality reduction - significant reduction in the no. of features required to build a model.
2. It capture meanings of the words, semantic relationships and the different types of contexts they are used in.

Word2Vec is not a single algorithm but a combination of two techniques – CBOW (Continuous bag of words) and Skip-gram model. Both of these are shallow neural networks which map word(s) to the target variable which is also a word(s). Both of these techniques learn weights which act as word vector representations.

CBOW tends to predict the probability of a word given a context. A context may be a single adjacent word or a group of surrounding words. The Skip-gram model works in the reverse manner, it tries to predict the context for a given word.(Twitter Sentiment Analysis - word2vec, doc2vec | Kaggle, 2021)

We will go ahead with the Skip-gram model as it has the following advantages:

* It can capture two semantics for a single word. i.e it will have two vector representations of ‘apple’. One for the company Apple and the other for the fruit.
* Skip-gram with negative sub-sampling outperforms CBOW generally.

A picture containing graphical user interface

Description automatically generated

Figure 2: Skip Gram Representation

(NLP 101: Word2Vec — Skip-gram and CBOW | by Ria Kulshrestha | Towards Data Science, 2021)

**BERT Embeddings:**

DistilBERT is a lightweight version of BERT with comparable accuracy. DistilBERT processes the sentence and passes along some information it extracted from it on to the next model that is Logistic regression/SVM/Random Forrest model from scikit learn, It will take the result of DistilBERT’s processing, and classify the sentence.

**A screenshot of a game

Description automatically generated with medium confidence**

Figure 3: BERT embeddings

(What is BERT | BERT For Text Classification, 2021)

BERT Embeddings is a combination of 3 different embeddings.

1. Position Embeddings: BERT uses positional embeddings to keep the positions of the words in a sentence. These are added to overcome shortcomings of RNN – that is not able to capture position or sequence of the information.
2. Segment Embeddings: BERT can learn unique embeddings for different sentences. As per above figure Different sentences are given different embeddings.
3. Token Embeddings: BERT can learn specific token from the word token vocabulary.

(What is BERT | BERT For Text Classification, 2021)

# 7.4 Research Methodology Flow Chart

Diagram

Description automatically generated

Figure 4: Research Methodology Flow-Chart

# 7.5 Model Evaluation

Since it’s a binary classification problem we are going to use AUC (Area Under Curve) – ROC (Receiver Operating Characteristics) as a metric to evaluate various model performances.

AUC - ROC curve is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s. By analogy, the Higher the AUC, the better the model is at distinguishing between reviews with positive sentiments and negative sentiments.

**Defining terms used in AUC and ROC Curve**

1. TPR (True Positive Rate) / Recall /Sensitivity = TP / (TP + FN)
2. Specificity = TN / (TN + FP)
3. FPR (False Positive Rate) = 1 - Specificity

Sensitivity and Specificity are inversely proportional to each other. So when we increase Sensitivity, Specificity decreases, and vice versa. When we decrease the threshold, we get more positive values thus it increases the sensitivity and decreasing the specificity.

Similarly, when we increase the threshold, we get more negative values thus we get higher specificity and lower sensitivity.

Also, as we know FPR is 1 - specificity. So, when we increase TPR, FPR also increases and vice versa.

The ROC curve is used to understand the strength of the model by evaluating the performance of the model at all the classification thresholds. The Area Under the Curve (AUC) summarizes the overall performance of the classifier.

● Best threshold would be one at which the True Positive Rate is high and False Positive Rate is low, i.e., misclassifications are low.

● Depending upon the problem statement, study will identify which measure is more important: high precision or high recall.

● F1-Score is the harmonic mean of precision and recall values for a classification problem where the requirement is of having best precision and recall at the same time. The formula for F1-Score is: 2\*(precision\*recall)/(precision + recall)

# 7.6 Interpretability of the models

1. **LIME**: LIME is an algorithm that helps explaining predictions of any classification or regression model by approximating it locally with some interpretable model. The output of LIME is a set of explanations representing the contribution of each feature to a prediction for a single sample, which is a form of local interpretability. Interpretable models in LIME can be, for instance, linear regression or decision trees, which are trained on small perturbations (e.g. adding noise, removing words, hiding parts of the image) of the original model to provide a good local approximation. (LIME Explained | Papers With Code, 2021)
2. **SHAP**: The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction. The SHAP explanation method computes Shapley values from coalitional game theory. SHAP unify methods such as LIME, LOCO, and Shapley Values to derive consistent local variable contributions to model predictions. It presents desirable three properties 1. Local accuracy 2. Missingness 3. Consistency. (5.10 SHAP (SHapley Additive exPlanations) | Interpretable Machine Learning, 2021)

# 7.7 Results

The research will be concluded based on comparative study of the performance of the above models with all embeddings mentioned(TF-IDF, Word2Vec Embeddings, BERT Embeddings). Performances of the models will be compared based on AUC-ROC primarily. I will also try to achieve interpretability of classical model using techniques like LIME and SHAP. Then I will compare the insights drawn from LIME and SHAP both.

# 8. Requirements Resources

1. Hardware Requirements
2. Operating System: MacOS BigSur
3. Processor: 2.2 GHz 6-Core Intel Core i7
4. Memory: 16 GB
5. Software Requirements
6. Python Version: 3.3
7. Jupyter Lab: 2.2.6
8. Python Libraries for machine learning: Pandas and NumPy for data processing, Matplotlib and Seaborn for data visualization, scikit-learn learn for data pre-processing, predictive modelling & model evaluation.
9. Python libraries for interpretability methods like LIME, SHAP etc
10. Access to cloud offerings like Kaggle notebooks, Google Collab etc.

# 9. Research Plan

Chart

Description automatically generated

# References

Al-Saqqa, S. and Awajan, A., (2019) The Use of Word2vec Model in Sentiment Analysis: A Survey. *ACM International Conference Proceeding Series*, June 2020, pp.39–43.

Anon (2021) *5.10 SHAP (SHapley Additive exPlanations) | Interpretable Machine Learning*. [online] Available at: https://christophm.github.io/interpretable-ml-book/shap.html#definition [Accessed 11 Apr. 2021].

Anon (2021) *BoW Model and TF-IDF For Creating Feature From Text*. [online] Available at: https://www.analyticsvidhya.com/blog/2020/02/quick-introduction-bag-of-words-bow-tf-idf/ [Accessed 11 Apr. 2021].

Anon (2021) *IMDB Dataset of 50K Movie Reviews | Kaggle*. [online] Available at: https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews [Accessed 9 Apr. 2021].

Anon (2021) *LIME Explained | Papers With Code*. [online] Available at: https://paperswithcode.com/method/lime [Accessed 11 Apr. 2021].

Anon (2021) *NLP 101: Word2Vec — Skip-gram and CBOW | by Ria Kulshrestha | Towards Data Science*. [online] Available at: https://towardsdatascience.com/nlp-101-word2vec-skip-gram-and-cbow-93512ee24314 [Accessed 11 Apr. 2021].

Anon (2021) *Twitter Sentiment Analysis - word2vec, doc2vec | Kaggle*. [online] Available at: https://www.kaggle.com/nitin194/twitter-sentiment-analysis-word2vec-doc2vec [Accessed 10 Apr. 2021].

Anon (2021) *What is BERT | BERT For Text Classification*. [online] Available at: https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/ [Accessed 11 Apr. 2021].

Dholpuria, T., Rana, Y.K. and Agrawal, C., (2018) A sentiment analysis approach through deep learning for a movie review. *Proceedings - 2018 8th International Conference on Communication Systems and Network Technologies, CSNT 2018*, pp.173–181.

Haque, M.R., Akter Lima, S. and Mishu, S.Z., (2019) Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews. *3rd International Conference on Electrical, Computer and Telecommunication Engineering, ICECTE 2019*, pp.161–164.

Lakshmi Devi, B., Varaswathi Bai, V., Ramasubbareddy, S. and Govinda, K., (2020) Sentiment analysis on movie reviews. In: *Advances in Intelligent Systems and Computing*. [online] Springer, pp.321–328. Available at: https://link.springer.com/chapter/10.1007/978-981-15-0135-7\_31 [Accessed 10 Apr. 2021].

Lewis, S.J., (1945) Thumbs up. *American Journal of Orthodontics and Oral Surgery*, 319, pp.481–482.

Qaisar, S.M., (2020) Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory. *2020 2nd International Conference on Computer and Information Sciences, ICCIS 2020*, pp.7–10.

Sahu, T.P. and Ahuja, S., (2016) Sentiment analysis of movie reviews: A study on feature selection and classification algorithms. *International Conference on Microelectronics, Computing and Communication, MicroCom 2016*, pp.0–5.

Tamara, K. and Milićević, N., (2018) Comparing Sentiment Analysis and Document Representation Methods of Amazon Reviews. *SISY 2018 - IEEE 16th International Symposium on Intelligent Systems and Informatics, Proceedings*, pp.283–288.

Tarımer, İ., Çoban, A. and Kocaman, A.E., (n.d.) *Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques*.

Tripathy, A., Agrawal, A. and Rath, S.K., (2016) Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, [online] 57March, pp.117–126. Available at: http://dx.doi.org/10.1016/j.eswa.2016.03.028.

Vohra, M.S.M. and Teraiya, P.J.B., (2013) Journal of Information, Knowledge and Research in Computer Engineering a Comparative Study of Sentiment Analysis Techniques. *Journal of Information,Knowledge and Research in Computer Engineering*, pp.313–317.

Yasen, M. and Tedmori, S., (2019) Movies reviews sentiment analysis and classification. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings*, pp.860–865.

Yousefpour, A., Ibrahim, R., Hamed, H.N.A. and Hajmohammadi, M.S., (2014) A comparative study on sentiment analysis. *Advances in Environmental Biology*, 813, pp.53–68.