COMPARATIVE STUDY OF SENTIMENT ANALYSIS FOR MOVIE REVIEWS

STUDENT NAME: SANKET TALAVIYA

STUDENT ID: 969198

INTERIM REPORT

LIVERPOOL JOHN MOORES UNIVERSITY – MS IN ML AND AI

DECEMBER 2020

**DEDICATION**

I dedicate this work to my wife Bhumika, my little daughter Praanvi and My parents, who have been there for me throughout this research work. You all have been my biggest supporters. Without your support I couldn’t have continued my dream of doing Masters in Machine Learning and Artificial Intelligence.

**ACKNOWLEDGEMENTS**

I would like to thank Dr. Ahmed for the weekly sessions and guidance in preparing this report. I would also like to thank my thesis supervisor Anurag Maji for his valuable suggestions and feedbacks that helped preparing this research report.

# 

**TABLE OF CONTENTS**

DEDICATION ii

ACKNOWLEDGEMENTS iii

ABSTRACT vii

LIST OF TABLES xii

LIST OF FIGURES xv

LIST OF ABBREVIATIONS xvi

CHAPTER 1: INTRODUCTION .............................................................................................. 1

1. Background of the Study............................................................................................... 1
2. Problem Statement ……………………….................................................................... 2
3. Aim and Objectives ………………………………....................................................... 3
4. Scope of the Study ........................................................................................................ 4
5. Significance of the Study .............................................................................................. 5
6. Structure of the Study.................................................................................................... 5

CHAPTER 2: LITERATURE REVIEW ................................................................................... 6

1. Introduction ................................................................................................................. 10
2. Sentiment Analysis on movie reviews.......................................................................... 10
3. Sentiment Analysis on various product reviews........................................................... 10
4. Sentiment Analysis in other domains........................................................................... 10
5. Discussion.................................................................................................................... 10
6. Summary ...................................................................................................................... 21

CHAPTER 3: RESEARCH METHODOLOGY ..................................................................... 22

1. Introduction ................................................................................................................. 22
2. Methodology …………………………........................................................................ 22
   * 1. Data Selection ……………………….............................................................. 22
     2. Data Pre-Processing …………………………………………………………. 31
     3. Exploratory Data Analysis…………………………………………………… 32
3. Proposed Model ……………………………………..…............................................. 29
4. Proposed Flow Diagram .............................................................................................. 31

3.5 Evaluation Metric Used…………………………….…………………….………….. 32

3.6 Expected Outcomes………………………………………………………………...…32

3.7 Requirements/Resources…………………………………………………………...…32

3.7.1 Hardware Resources…………………………………………………………… 32

3.7.2 Software Resources….………………………………………………………… 32

3.7.3 Additional Resources……………………...…………………………………… 32

3.6 Expected Outcomes………………………………………………………………...... 32

3.7 Summary………………………………………………………………….………….. 33

REFERENCES ........................................................................................................................ 26

APPENDIX A: RESEARCH PROPOSAL ............................................................................. 27

APPENDIX B: DATASET AND CODE URLs ..................................................................... 28

# 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. One such domain is the entertainment industry, especially movies. Users have been expressing their views on a particular movie through various popular websites such as IMDB, Rotten Tomatoes, Twitter etc. These reviews help other people to decide whether they should spend their precious time in watching a particular movie or not. And hence lot of people go through movie review comments to find out whether they should watch a particular movie. Sentiment analysis is the process of classifying these review comments. If one doesn’t apply machine learning and AI techniques to find out overall sentiments of the topic going through all review 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. So far lot of experiments have been done using feature engineering and pre-processing using BOW, TF-IDF, Doc2Vec, Word2Vec embeddings and different machine learning models like Logistic Regression, SVM, Random Forest. Although good results have been achieved using these techniques there is still scope of doing better in feature engineering space through applying contextual embedding methods like BERT embeddings which has been introduced by the state-of-the-art model BERT. Context free embedding techniques like TF-IDF, Word2Vec etc. ignore the meaning of the word in different contexts as it creates only one representation of a given word in entire vocabulary. In contrast to that, contextual embedding technique like BERT learns different representations of a given word based on their context. While there are few experiments already done on IMDB movie review dataset using context free embedding techniques. In this research we have tried applying contextual BERT embeddings with Logistic Regression, SVM and Random Forest. And we have also tried implementing transformers-based model BERT and then compared performance of these models to other similar papers.

# LIST OF FIGURES

|  |  |
| --- | --- |
| Figure 1 | TF-IDF Representation |
| Figure 2 | Skip Gram Representation |
| Figure 3 | BERT embeddings Representation |
| Figure 4 | Research Methodology Flow-Chart |
| Figure 5 | AUC ROC Curve |

# 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 |
| TPR | True Positive Rate |
| FPR | False Positive Rate |
| TN | True Negative |
| FP | False Positive |
| TP | True Positive |
| FN | False Negative |

# CHAPTER 1

# INTRODUCTION

# 1.1 Background of the Study

Sentiment analysis is useful for finding out how customers think about a particular product or topic. Before computer era started, In the earlier times sentiment analysis was done primarily on non-online modes like paper or written documents as there was no internet services. In the recent times internet services have become part of the human life and hence more and more people have started expressing their views on online forums including social medias, publications, blogs, websites etc. Sentiment Analysis is a process of extracting sentiments from these raw reviews expressed over above various internet forums using NLP and statistics.

As Internet services are evolving more and more each day, Exponential growth has been witnessed in terms of information availability on the internet. This includes opinions and reviews for particular product or topics like movies. (Yousefpour et al., 2014). Because of tremendous value sentiment analysis brings to the table by eliminating manual efforts of parsing all reviews and figure out overall sentiment of the subject in question. Lot of researchers are attracted to this field. This area of research broadly involves NLP and Machine learning models. (Qaisar, 2020). Sentiment analysis in each field had different objectives for example if its e-commerce website reviews for a product. From the sentiment analysis product companies can gather information how good or product is doing in the market by deriving sentiments from the reviews. If the subject is movie reviews, It helps users to decide whether they should view this movie or not based on sentiment analysis that way it helps them save their time and movie if overall sentiments are negative.

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)

Film Industries including songs, trailers, movies etc rely heavily on sentiment analysis to figure out how people are feeling about particular song/trailer or movie. This sentiment analysis can also give pointers to other producers on what people don’t like in movies or songs etc based on that they can take actions for their next movies. And existing movie reviews help user to decide on which movie to watch or not to watch etc and film makers in a way that what particular things people like the most and disliked the most. The dataset here is taken from Kaggle and it contains IMDB movie reviews. (IMDB Dataset of 50K Movie Reviews | Kaggle, 2021)

Lot of research has already been done in this area using different context-free embedding techniques like BOW, TF-IDF, Word2Vec with different machine learning models but there is a scope of doing more research on sentiment analysis with recently introduced contextual embedding techniques like BERT embeddings with different machine learning models.

Broadly below are the two types of sentiment analysis techniques.

1. **Machine Learning based techniques**: This technique mainly falls into supervised classification category, in this technique two set of datasets are required, train dataset and test dataset, Train dataset is used to learn various patterns between variables of the subject and test document is used to check the robustness of the model on the dataset that model hasn’t seen yet. 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. The most widely used features in this method are as (1) Term 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. In this technique features of a given statements are compared with sentiment lexicons whose values are known in advance. 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 few methods to build sentiment lexicons (1) Manually (2) From Corpus (3) From Dictionary. (Vohra and Teraiya, 2013)

# 1.2 Problem Statement

This research focuses on saving users’ precious time by suggesting them accurate sentiments of a particular movie. This research also can be used by other movie makers and actors to understand how their movie is being rated by critics. That will help them taking better actions based on these sentiments in the future movies. Most of the existing researches have solved this problem with use of different embedding techniques like BOW, Word2Vec, Doc2Vec, or TF-IDF with different machine learning models like SVM, LR, Decision Tree, Random Forest, Naïve Bayes etc. So there is a scope of solving following problems.

1. Can we get more accurate model that can help predicting movie reviews better to users?
2. Can contextual embedding techniques like BERT Embeddings do better than other traditional embedding techniques like Word2Vec, TF-IDF etc.

# 1.3 Aim and Objectives

This research is aimed to propose a model to classify sentiments on IMDB movie review with high accuracy and the goal of this research is to contribute to users who want to get accurate idea of overall sentiments on a particular movie which can help them deciding if they should watch particular movie or not.

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

1. Figure out research that has been done so far in sentiment analysis domain for movie reviews.
2. Figure out the best embedding technique and feature engineering.
3. Build machine learning models and compare the performance of these models and Identify the most accurate model.
4. Provide interpretability for these models using LIME/SHAP.

# 1.4 Scope of the Study

As the time duration for this research is limited, research scope will be limited to following points:

1. The dataset used here is of IMDB movie reviews and it has been taken from Kaggle, so assuming data is already valid.
2. In this research, scope will be limited to develop and compare the performance of different machine learning classification models using TF-IDF, Word2Vec Embedding and BERT embeddings only.
3. Machine learning Model experiments will be limited to Logistic Regression as interpretable models and SVM, Random Forest 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.

# 1.5 Significance of the Study

Critics today often uses social media and movie rating websites to post their comments or rating for the movies. This data can eventually help users in decision making whether to watch that movie or not and that can really save users from wasting their precious time. Now if user starts reading all movie reviews that are provided on internet or social media sites, it will consume a lot of time as it’s a non-trivial task to go through all reviews manually and figure out overall sentiments on a particular movie. Through machine learning models we can automate the process of sentiment analysis from the pool of movie review comments and figure out what is the overall sentiment for the movie. This involves Natural language processing underneath and benefit that we get is incredible in terms of money and time saved. This sentiment analysis is basically a classification problem whether a particular comment is positive or negative or neutral and then for the whole movie how many comments are positive and how many of them are negative, this way we can achieve overall sentiment for the movie. So, this research would provide more accurate results to end users to help them taking better decisions on whether they should watch a particular movie or not based on sentiments expressed through movie reviews.

# 1.6 Structure of the Study

The first chapter contains an introduction to the issue, a background examination of the research, and issue statements. The research's goals and objectives are also given, followed by a discussion of the study's significance and scope.

The second chapter contains literature review of sentiment analysis in various domains like product reviews, movie reviews etc. It covers existing research that have been done using variety of techniques like data pre-processing, feature extraction, embeddings, machine learning models and deep learning models. Towards the end of this chapter various outcomes have been documented and conclusions have been provided.

# CHAPTER 2:

# LITERATURE REVIEW

# 2.1 Introduction

(Lewis, 1945) in his research did the comparative analysis of the performance of various models like Naïve Bayes, Maximum Entropy and Support Vector Machines using various features like unigram, bigram, combination of both, POS with unigrams etc. and the analysis showed that presence of the feature is more beneficial than frequency of the feature. When features are lesser Naïve Bayes works better but as it increases SVM starts doing better, also with increased number of features Max Entropy does better than Naïve Bayes but overfits.

(Tamara and Milićević, 2018) did the comparative analysis by applying BOW, BOW with n-grams and TF-IDF embeddings to different machine learning models like Logistic Regression, SVM and few deep learning models like LSTM and ConvNets. Glove and Word2Vec embeddings were also tested here. And paper concluded that deep learning models perform better than traditional machine learning models. LSTM yielded greater than 95% accuracy and BOW n-grams + TDF with SVM model gave around 92% accuracy. BOW with Boosting yielded around 80%+ accuracy.

(Al-Saqqa and Awajan, 2019) in their survey found the in the word embedding techniques CBOW and skip-gram, Skip-gram works better for infrequent words than the other. And CBOW works well with frequent words and is faster compared to skip-gram.

(Tripathy et al., 2016) concluded in the research that accuracy of classification model is inversely proportional to the value of n in n-gram. When tests were done using higher n-gram models where n is greater than 3, Accuracy decreased for models like Maximum Entropy, SVM, NB, Stochastic Gradient De-scent(SGD). However unigrams and bigrams results were remarkably better. In this study IMDB dataset was used.

(Sahu and Ahuja, 2016) examined IMDB dataset and proposed an approach to classify sentiment expressions from zero to four on scale and extracted the features and applied ranking and used these features for training multiclass classifier to classify 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, In this study they applied Decision tree, NB, SVM machine learning algorithms on a vector space that was created in KNIME platform. That is one analytics platform. Results they obtained were showing Decision tree was performing better than NB and SVM. And when they tried for other datasets as well and finally concluded that SVM yields better results across different datasets.

(Yasen and Tedmori, 2019) In this research, applied around 8 different machine learning models after cleaning and pre-processing the dataset and they evaluated using 5 different metrics. Results concluded that Random Forest outperformed the other classifiers and Ripper Rule Learning performed the worst on the datasets.

# CHAPTER 3:

# Research Methodology

# 3.1 Introduction

In this section sequential steps that are going to be performed in this whole research is detailed out.

This research is going to be carried out in three parts sequentially using 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 Forest 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.

# 3.2 Research Methodology

# 3.2.1 Dataset Description

This dataset consists of IMDB movie reviews taken directly from Kaggle Website. Dataset has 2 columns Review and Sentiment. Review column represents user reviews some of the are large comments and some of them are small. Sentiment column has two labels “Positive” and “Negative”, Overall, there are 50000 records in the dataset and classes of positive and negative reviews are balanced that is 25000 each.

(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

# 3.2.2 Data Clean up and Pre-Processing

There is an approach called Exploratory Data Analysis (EDA) that helps extracting the information enfolded in the data and based on that we can summarise the main attributes or characteristics of the data that is more important to study for a given problem.

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 different embedding techniques like TF-IDF, Word2Vec Embeddings and BERT Embeddings.

# 3.2.3 Exploratory Data Analysis

# 3.3 Model Creation

We have created two set of data from the dataset in the 70:30 ratio where 70% data will be considered as training data and 30% will be used for testing purpose. We are also going to perform K-Fold validation technique so that original dataset entries have equal probability of appearing in training as well as testing set. And it will help keeping models less biased. We can apply following models on it with different embeddings and compare their performances later.

1. Logistic Regression (Whitebox)
2. SVM (Blackbox)
3. XGBoost (Blackbox)
4. Random Forest (Blackbox)
5. BERT (Blackbox)

This IMDB movie reviews dataset has target values as positive or negative only hence it’s a binary classification problem. So that’s the thought process for selecting above models for classification. These all models we need to apply separately with TF-IDF, Word2Vec Embeddings and BERT Embeddings. Below is a summary around these methods.

**TF-IDF:**

Assume that we have a problem at hand where we want to figure out a which document is the most relevant for the review statement or comment that we want to search in these documents. So one naïve solution for this problem is to simply gather words of the review statement or comment and figure out how many documents do not include all the words of the review statement or comment and simply eliminate them. But since its very naïve approach intuitively we can think of that still a lot of documents will still remain relevant for the searching operation. So the next addition we can do in filtering is to count frequency of each term in the document. This is called Term Frequency.

A picture containing letter

Description automatically generated

Document frequency is similar to term frequency, It represents the count of a particular term in the set of document. In other words, It is the number of documents where the given term is present.

Text

Description automatically generated

In the term frequency, all the terms are given equal importance or weightage. But in practical scenarios some words occur too frequently like “is”, “are”, “of” etc. and that has very less to no importance, hence we need to rationalise or weigh down high frequency terms and weigh up the low frequency terms, IDF is an factor which helps achieving this by lowering weightage of high frequency terms and increasing weightage for low frequency terms.

Logo

Description automatically generated

Here N is total number of documents and DF is number of documents containing given term.

Table

Description automatically generated

Figure 1: TF-IDF Representation

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

**Word2Vec Embeddings:**

Word vector helps representing the words as vectors. So you can vectorize these words and match these vectors to find similarities, But this sounds impractical if you have millions of words. So we might need to find similarities between these encoded vectors.

Earlier N-Grams were used heavily to extract the meaning of the word with surrounded other words but it can’t capture the context.

Based on the different contexts the words are found, Word Embeddings can be defined as high-dimensional vectorised view of words. Word embeddings helps comparing the similarity of words and provide some useful info for machine learning models to consider as input.

(Introduction to Word Vectors and Word2Vec - Studytonight, 2021)

Word2Vec is very well known predictive model that strives to output a target word provided a context or the context words from the target. It’s a vector representation of words such that the cosine distance between those vectors reflects the semantic “distance” between the words they represent. For example if word1 and word2 have a similar meaning, then cos distance of word1 vector and word2 vector should be small.

Word embeddings help achieving dimensionality reduction so number of feature required to build the model is less. Also it captures the meanings of the words and semantic relationships as well as context.

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 Forest 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 comprises of 3 different embeddings.

1. Position Embeddings are used to understand the location index information of the input. This can overcome the shortcomings of RNN which doesn’t keep location information.
2. Segment Embeddings are unique embeddings for different sentences that is learnt by BERT. In these embedding different sentences are given different embeddings.
3. Token Embeddings are unique embeddings in which BERT learn specific token from the word token vocabulary.

# 3.4 Research Methodology Flow Chart

Diagram

Description automatically generated

Figure 4: Research Methodology Flow-Chart

# 3.5 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.

By calculating AUC-ROC, you can determine the goodness of the model. If the curve is sticky towards upper left side, it means that the model is very good and if it is more towards the 45-degree diagonal, it means that the model is almost completely random. So, the larger the AUC, the better will be your model. So In the best case it passes through the upper left corner of the graph. So the least area that an ROC curve can have is 0.5, and the highest area it can have is 1.

 ROC Curves shows the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR)

Chart

Description automatically generated

Figure-5: AUC-ROC Curve

(Zhu et al., 2010)

**Terms used in AUC-ROC Curve**

1. TPR can be defined as below

A picture containing diagram

Description automatically generated

1. Specificity can be defined as below.

Text

Description automatically generated with low confidence

1. FPR can be defined as below.

A picture containing text

Description automatically generated

So for the best model we can expect high TPR and low FPR and depending on the study we can decide if high precision is more important or high recall.

Text

Description automatically generated with medium confidence

# 3.6 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. LIME outputs set of interpretations that helps interpreting contribution of each variable or feature to the prediction for a given input sample. This is in form of local interpretability. Linear Regressions, Decision Trees etc. are used as interpretable models with this algorithm and these models are trained with small augmentation so to the original data to achieve good local approximation. (LIME Explained | Papers With Code, 2021)
2. **SHAP**: This is a technique with strong support from game theory that helps explaining the predictions of any Machine Learning model. It helps deriving local variable contributions to overall prediction. It presents essentially three properties 1. Local accuracy 2. Missingness 3. Consistency. (5.10 SHAP (SHapley Additive exPlanations) | Interpretable Machine Learning, 2021)

# 3.7 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.

# 3.8. Required Resources

**3.8.1 Hardware Requirements**

* Operating System: MacOS BigSur.
* Processor: 2.2 GHz 6-Core Intel Core i7
* Memory: 16 GB

**3.8.2 Software Requirements**

* Python Version: 3.3
* Jupyter Lab: 2.2.6
* Python Libraries: Pandas, NumPy, Matplotlib, Seaborn, scikit-learn.
* Some interpretation libraries available in python like LIME, SHAP etc
* Access to cloud offerings like Kaggle notebooks, Google Collab etc.
  + 1. **Additional Requirements**
* UpGrad team has provided GPU enabled Nimble box for processing needs.

# 10. 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) *Introduction to Word Vectors and Word2Vec - Studytonight*. [online] Available at: https://www.studytonight.com/post/introduction-to-word-vectors-and-word2vec [Accessed 17 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) *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].

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.

Zhu, W., Zeng, N. and Wang, N., (2010) *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS ® Implementations*.