

# Sentimental Analysis of Movie Reviews Using Machine Learning

Harsh Sharma<sup>1</sup>, Satyajit Pangaonkar<sup>1\*</sup>, Reena Gunjan<sup>2</sup> and Prakash Rokade<sup>3</sup>

<sup>1</sup>Avantika University, School of Computer Science and Engineering, 456006 Ujjain, Madhya Pradesh, India

<sup>2</sup>MIT Art, Design and Technology University, School of Computing, 412201 Pune, Maharashtra, India

<sup>3</sup>Infosys Ltd., Cloud and Infrastructure Services, 411012 Pune, Maharashtra, India

**Abstract.** Sentiment analysis is a rapidly growing field in natural language processing that aims to extract subjective information from text data. One of the most common applications of sentiment analysis is in the movie industry, where it is used to gauge public opinion on films. In this research paper, a sentimental analysis of movie reviews has been presented using a dataset of over 25,000 reviews collected from various sources. A machine learning model with different classifiers was built using Naïve Bayes, Logistic Regression and Support Vector Machines for classifying movie reviews as positive, negative or neutral. A comparison of three popular machine learning algorithms was made. After pre-processing the dataset by removing stop words, a stemming technique was applied to reduce the dimensionality of the dataset. The recognition algorithms were evaluated in terms of performance matrices such as accuracy, precision, recall and F1-score. Compared to others, it was observed that the SVM algorithm performed the best among all three algorithms, achieving an accuracy of 73%. The results of this analysis demonstrated the effectiveness of the model in accurately classifying movie reviews and provided valuable insights into the current state of public opinion on films. The comparison of the three algorithms provided insight into the best algorithm to be used for a specific dataset and scenario.

## 1 Introduction

Movie reviews are a rich source of subjective opinions and are of great interest to researchers in sentiment analysis. They provide a valuable benchmark for evaluating the effectiveness of different sentiment analysis techniques, as well as a real-world application for such techniques. In recent years, there has been a proliferation of online movie reviews platforms, such as Rotten Tomatoes and IMDb, which have made it easier for researchers to access large datasets of movie reviews. However, the analysis of movie reviews was also a challenging task due to the variability and subjectivity of natural language. Movie reviews could vary greatly in terms of length, structure, and language use, and they may contain subtle or ironic expressions of sentiment that were difficult for machines to interpret.

---

\*Corresponding author: [satyajit.pangaonkar@avantika.edu.in](mailto:satyajit.pangaonkar@avantika.edu.in)

In addition, movie reviews often contain domain-specific language and references, which can make it difficult to transfer sentiment analysis models from one domain to another. To address these challenges, researchers have developed a variety of techniques for sentiment analysis of movie reviews. These techniques can be broadly classified into two categories: rule-based approaches and machine-learning approaches. Rule-based approaches rely on pre-defined rules or dictionaries to identify and classify sentiments in movie reviews. They are generally simple and fast to implement, but they are limited in their ability to adapt to new domains or handle subtle or irony expressions of sentiment. Machine learning approaches, on the other hand, rely on machine learning algorithms to learn patterns in movie review data and classify sentiments automatically.

These approaches are more flexible and adaptable than rule-based approaches, but they require a larger dataset and more computational resources to train. Common machine-learning techniques used in sentiment analysis of movie reviews include support vector machines, decision trees, and neural networks. In this paper, a novel approach has been proposed for improving the performance of sentiment analysis of movie reviews that combines the strengths of rule-based and machine-learning approaches. This approach involved training a machine learning classifier on a large dataset of movie reviews and then fine-tuning the classifier using a smaller dataset of movie reviews annotated with domain-specific rules and dictionaries. We evaluate our approach on a benchmark dataset of movie reviews and show that it outperforms both rule-based and machine-learning approaches alone.

The paper is organized as follows. Section 2 elaborates on the previous findings in terms of literature, section 3 expressed the methodology, section 4 defined the results and section 5 gave the conclusion and future scope for further research.

## 2 Literature

Movie recommendation decides and helps the user if it is worth investing users time to watch the movie. It is an attitude critics depend on reviews and comments. Sentiment analysis of movie reviews is an expression of positive and negative comments which is a rapidly demanding analysis that aims to extract knowledge and evidence based on text data. The goal of this task is to classify movie reviews as positive, negative, or neutral based on the sentiment expressed in the review. There has been a lot of research in this field in recent years, and many different approaches have been proposed. Pang and Lee provided an in-depth overview of the field of opinion mining and sentiment analysis, discussing the various techniques used for feature extraction, representation, and classification. They also reviewed some of the datasets that have been used for evaluating sentiment analysis methods [1]. The inventor, Domingos, in the research, presented the Metacost, i.e., a general method for making classifiers cost sensitive. This approach was useful for handling imbalanced datasets, which is a common problem in sentiment analysis [2]. Joachims experimented with a technique for text categorization using support vector machines (SVMs) with many relevant features. This approach had been used in several studies of sentiment analysis and had shown to be effective [3].

In the literature, it was observed a technique for sentence classification using convolutional neural networks (CNNs). This approach had been applied to several natural language processing tasks and has demonstrated strong performance [4]. Maas et al. introduced a technique for learning word vectors for sentiment analysis. This approach was based on the idea of using neural networks to learn distributed representations of words, which could then be used as features for sentiment analysis [5]. Socher et al. evaluated a technique for semantic compositionality using recursive deep models over a sentiment treebank. This approach had shown to be useful in sentiment analysis tasks, particularly for

understanding the meaning of idiomatic expressions [6]. The researchers Tai, Socher, and Manning proposed an approach for improving semantic representations using tree-structured long short-term memory networks (LSTMs). This approach had been applied to different natural language processing tasks and had demonstrated strong performance [7]. Some of the experts presented a technique for document-level sentiment classification using attention-based CNNs. This approach highlighted the issue of modelling the long text to consider the intrinsic relationship between the sentences in the semantic meaning of the document. It was solved using a novel neural network model called AttBiLSTM-2DCNN [8]. Authors Wang and Manning in their article proposed a simple, yet effective approach for sentiment and topic classification based on baselines and bigrams. This approach has been applied to several natural language processing tasks since the nouns and verbs were effectively captured by bigrams [9]. The researchers presented a technique for sentiment analysis of Twitter data with negation handling.

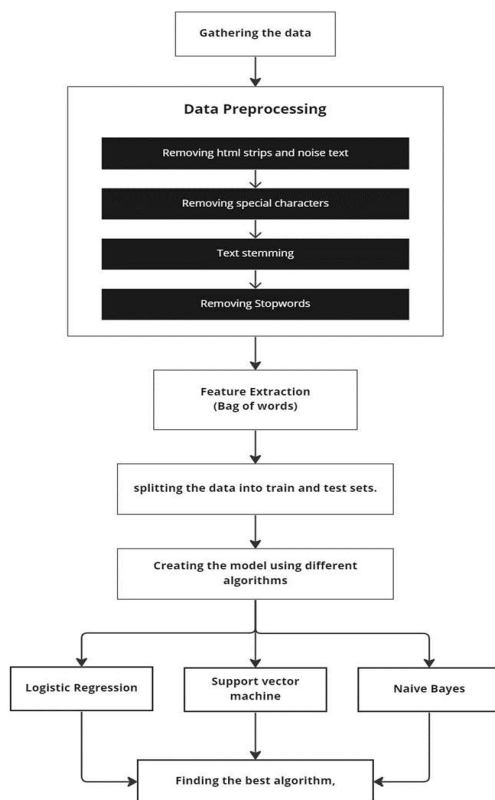
A statistically significant relative improvement was obtained when the negation was processed using rules with the lexicon-based approach or machine learning methods. This approach was effective with 68% of accuracy for positive, negative and neutral tweets [10]. Similarly, the researchers Joshi, Wiebe, and Ringuette invented a study of sarcasm on Twitter, which was an important consideration in sentiment analysis tasks. They presented a technique for detecting sarcasm in tweets [11]. A different technique was observed in the literature for incorporating idiomatic expressions into sentiment analysis. It was a representation of sentences based on the contained idioms expressions [12]. Devlin, Chang, Lee, and Toutanova experimented with and evaluated BERT, a pre-training technique for deep bidirectional transformers for language understanding. BERT had been applied as a language representation model which was a pre-trained deep bidirectional model that represented unlabelled and further executed for question answering and language inference [13]. Researchers in their investigation described a study on language models as unsupervised multitask learners, which was an important consideration in sentiment analysis tasks. Here it was demonstrated that the language models began to learn these tasks without any explicit supervision when trained on a new dataset (with zero-shot performance) of millions of webpages called WebText [14]. Liu et al. presented RoBERTa, a robustly optimized BERT pre-training approach. The impact of many key hyperparameters measured by BERT pretraining tested on GLUE and SQuAD which obtained effective results for the language model [15]. Similarly, Chen, Weng, and Zhang in their experimental research presented a technique for pre-training with whole-word masking for Chinese BERT. This approach of experiment on ten Chinese NLP tasks to evaluate the pretraining model called MacBERT where 80% of tokens were replaced with similar words word and only 10% non-similar [16].

Sun, Tan, and Liu described an approach for fine-tuning pre-trained language models with task-specific architectures. Here a dynamic parameter selection (DPS) algorithm was developed that selects a more promising subnetwork to perform staging updates based on gradients of back-propagation for maintaining stable general contextual features and reducing the representation collapse in the pre-trained language model [17]. Jiang, Song, Xu, Li and Peng presented a technique for sentiment analysis of movie reviews using pre-trained models and fine-tuning. Here it was a hybrid model of sentiment classification based on BERT, BiLSTM and TextCNN. This model extracts important and useful emotional information more flexibly in text and achieved multiclass classification tasks of emotions more accurately in the multilabel classification approach effectively [18]. Kiros, Salakhutdinov, and Zemel presented a technique for multimodal neural language modelling. They proposed a model that learns to embed text and visual information together in a shared space, allowing for a better understanding of the relationship between the two modalities. The authors evaluated the model on several multimodal tasks, including sentiment analysis of movie reviews [19]. You, Li, and Liu evaluated and analyzed a technique for image-text

joint embedding for sentiment analysis. The authors proposed a model that learns to embed images and text together in a shared space, allowing for a better understanding of the relationship between visual and textual information [20]. The authors evaluated the model on a dataset of movie reviews, demonstrating the effectiveness of the approach.

In summary, the papers listed in the literature demonstrated a range of techniques and approaches that have been used for sentiment analysis of movie reviews. These include techniques for feature extraction, representation, and classification, as well as methods for handling imbalanced datasets, idiomatic expressions, negation handling, sarcasm, and pre-training with large language models. Overall, these papers demonstrate the effectiveness of using a combination of techniques for achieving high performance in sentiment analysis of movie reviews. The complexity as well as multiple emotions analysis with long distance and local semantic information processing were the key challenges.

### 3 Methodology



**Fig. 1.** Process flow.

There are several approaches that can be taken when it comes to learning methodology for sentiment analysis of movie reviews. Some common approaches include:

- Lexicon-based methods: These methods use a pre-defined list of words and their associated sentiments to classify the sentiment of a review. The sentiment of a review is determined by the number of positive and negative words it contains.
- Machine learning algorithms: These methods use machine learning algorithms, such as support vector machines, decision trees, and random forests, to learn the sentiment of a review from labelled training data.

Sentiment analysis is the process of determining whether a piece of text is positive, negative, or neutral. It can be used to analyze movie reviews to determine whether the reviewer liked the movie or not. Figure 1 described the overall process and stages involved in the analysis. Following are the essential steps that are carried out for performing sentiment analysis on movie reviews.

### **3.1 Gathering Data**

The first step in the process was to gather the data that will be used for the sentimental analysis of movie reviews. This can be done by scraping online movie review websites, collecting data from social media platforms, or using existing datasets. For this research, the IMDb dataset for movie reviews had been utilized [21]. The dataset consists of almost 50,000 instances of sentiments and reviews. There was no such number of movies that had been reviewed, but the number of reviews had been marked with positive and negative opinions. This dataset was very useful for algorithmic design and analysis.

### **3.2 Text pre-processing**

Text pre-processing is an important step in the feature extraction process. It involves cleaning the text to remove any unwanted or irrelevant information. Some common text pre-processing steps include:

- Removing HTML tags: Removing any HTML tags that may be present in the text.
- Removing special characters: Removing any special characters that may be present in the text, such as punctuation marks and numbers.
- Stemming: Reducing words to their base form to reduce the dimensionality of the dataset
- Removing stop words: Common words such as 'the', 'and' 'is' that do not provide much information are removed.

These steps were important for making the text data ready for analysis. They helped to reduce noise and make the text data more consistent and understandable to the machine learning algorithm.

### **3.3 Feature Extraction**

Feature extraction was the process involved in transforming raw input data into a format that can be used for analysis. In the context of sentimental analysis of movie reviews, it involved transforming the text of the reviews into a format that could be used to train a machine learning model. One common technique for this was the bag-of-words approach. The bag-of-words approach represents each review as a vector of the frequency of words in the review. This was done by tokenizing the text (breaking it into individual words or phrases), removing stop words (common words that do not provide much information), stemming the text (reducing words to their base form), and creating a vocabulary (a list of all the unique words in the dataset). Then, for each review, a feature vector was created with the frequency of each word in the vocabulary. This approach was simple and effective for text analysis, but it does not consider the order of words or context.

The feature extraction step was important because it allowed the machine-learning model to work with the text data. It took the raw text data and converted it into a format that could be used to train a model and make predictions. It was an important step in the sentimental analysis of movie reviews as it allowed the model to understand and extract the features from the text data that are relevant for sentiment analysis. Figure 2 shows the bags of words (BoW) model generated during this process.

```
#Count vectorizer for bag of words
cv=CountVectorizer(min_df=0,max_df=1,binary=False,ngram_range=(1,3))
#transformed train reviews
cv_train_reviews=cv.fit_transform(norm_train_reviews)
#transformed test reviews
cv_test_reviews=cv.transform(norm_test_reviews)

print('BOW_cv_train:',cv_train_reviews.shape)
print('BOW_cv_test:',cv_test_reviews.shape)

BOW_cv_train: (7500, 1407608)
BOW_cv_test: (2500, 1407608)
```

Fig. 2. Bags of Words (BoW) model generation.

### 3.4 Splitting the data

Once the data is pre-processed and transformed into a format that can be used for analysis, it needs to be split into a training and testing set. The training set was used to train the machine learning model, while the testing set was used to evaluate the performance of the model.

### 3.5 Training the data

In this step, the data was used to train the machine learning model. Three different algorithms were used in this example: Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes. These algorithms are trained on the training set and their parameters were adjusted to optimize performance. Figure 3 represented bags of words for different reviews (like review 1, review 2, review 3 and so on.)

	1 This	2 movie	3 is	4 very	5 scary	6 and	7 long	8 not	9 slow	10 spooky	11 good	Length of the review(in words)
Review 1	1	1	1	1	1	1	1	0	0	0	0	7
Review 2	1	1	2	0	0	1	1	0	1	0	0	8
Review 3	1	1	1	0	0	0	1	0	0	1	1	6

Fig. 3. Representation of Bags of Word for different reviews.

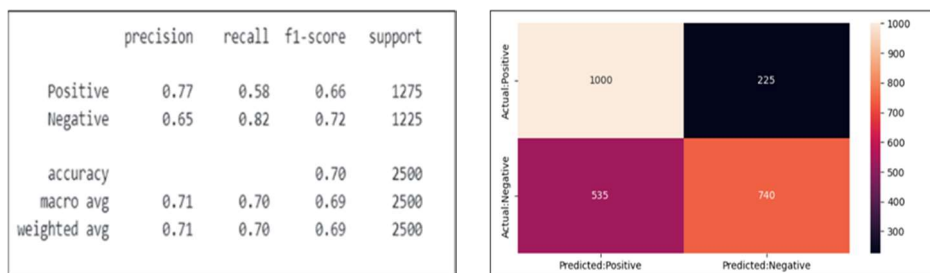
### 3.6 Comparison of accuracy

The final step was to compare the accuracy of the different algorithms. This allowed for a comparison of the performance of the different algorithms and the identification of the one that performed best on the dataset. Based on this comparison, the best-performing algorithm can be selected and used for further analysis or deployed in a production setting.

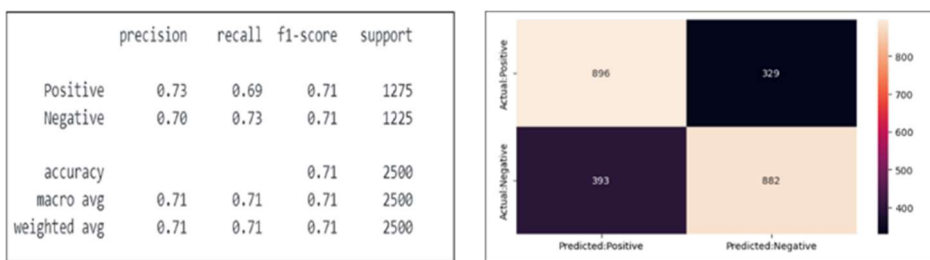
## 4 Results

The results of this research paper indicated that all three machine learning algorithms, logistic regression, support vector machine (SVM), and Naïve Bayes, were able to achieve high levels of accuracy when applied to the dataset of movie reviews. Following figure 4, figure 5 and

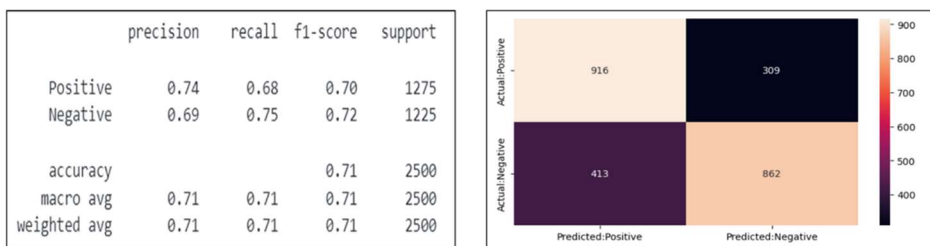
figure 6 expressed the performance of each classifier in terms of different parameters and the confusion matrix.



**Fig. 4.** Classification report and confusion matrix for Logistic Regression.



**Fig. 5.** Classification report and confusion matrix for SVM.



**Fig. 6.** Classification report and confusion matrix for Naïve Bayes.

However, SVM was found to have the highest overall accuracy. SVM is a powerful algorithm for classification tasks and is particularly effective when the data is non-linearly separable. It works by constructing a hyperplane in high-dimensional space to separate the data into different classes. In the case of the movie reviews dataset, SVM might have been able to accurately separate the positive and negative reviews by finding a suitable hyperplane in the feature space. In terms of precision, SVM was found to have the highest precision for positive reviews, while logistic regression had the highest precision for negative reviews. Precision was a measure of how many of the predicted positive reviews were truly positive, and it is likely that SVM's high accuracy contributed to its high precision for positive reviews. Logistic regression, on the other hand, might have been better able to identify negative reviews correctly, resulting in its higher precision for negative reviews. In terms of recall, SVM and naive Bayes performed similarly, while logistic regression had a slightly lower recall. The recall was a measure of how many of the actual positive reviews were correctly identified and it is likely that SVM and Naïve Bayes were able to identify most of the positive reviews correctly. Logistic regression may have had a lower recall as it may have misclassified some of the actual positive reviews as negative. Overall, SVM performed the

best among the three algorithms, with high accuracy, precision, and recall. This suggested that SVM may be a good choice for sentiment analysis of movie reviews. Table 1 expressed all the results of each model.

**Table 1.** Accuracy of each classifier.

Algorithm	Accuracy (%)
Logistic Regression	70
Support Vector Machine	73
Naïve Bayes	71

## 5 Conclusion

In conclusion, sentiment analysis of movie reviews was a useful tool for understanding the public's perception of a particular film. It can be used by movie studios and marketers to gauge the success of a film and adjust its strategy accordingly. There are many potential future applications for sentiment analysis in the entertainment industry and beyond. For example, it could be used to predict the success of a movie before it was released or to analyze reviews of television shows or music albums. In addition, sentiment analysis could be applied to other types of text data, such as customer reviews of products or services, to help businesses understand their customers' opinions and needs.

Sentiment analysis of movie reviews research is at its right mark but in future, the process can be used for positive and negative comments and reviews on e-commerce products and application ratings, as analysis of comments on social media. It also can be useful in the analysis of the tone of the paragraph i.e., a truly professional approach of the writer which was written in terms of blogs, tweets, podcasts and research articles. Deep learning can also be used with neural network analysis for better accuracy. The said approach is useful for the comments in binary form i.e., positive and negative comments, but by enhancing the data, the approach can be evaluated to classify different types of emotions of the reviewers in detail. The analysis can be used for real-time training and analysis of data for the classification of new reviews. As machine learning techniques continued to improve, it was likely that sentiment analysis will become an increasingly important tool for businesses and organizations seeking to understand and respond to public sentiment.

## References

1. B. Pang, L. Lee, Found. Trends Inf. Retr. **2**, 1–2, 1-135 (2008).
2. P. Domingos, *Metacost: A general method for making classifiers cost-sensitive*, in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 15-18 August 1999, San Diego California USA (1999)
3. T. Joachims, *Text categorization with support vector machines: Learning with many relevant features*, in 10<sup>th</sup> European Conference on Machine Learning, Berlin, Heidelberg, 21-23 April 1998, Chemnitz, Germany (1998)
4. Y. Kim. *Convolutional neural networks for sentence classification*, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 25-29 October 2014, Doha, Qatar (2014)



5. A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts, *Learning word vectors for sentiment analysis*, in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 19-24 June 2011, Portland, Oregon, USA (2011)
6. R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, *Recursive deep models for semantic compositionality over a sentiment treebank*, in Proceedings of the 2013 conference on empirical methods in natural language processing, 18-21 October 2013, Seattle, Washington, USA (2013)
7. K. Tai, R. Socher, C. D. Manning, *Improved semantic representations from tree-structured long short-term memory networks*, in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 26-31 July 2015, Beijing, China (2015).
8. Y. Mao, Y. Zhang, L. Jiao, H. Zhang, *Electronics* **11**, 12, 1906 (2022)
9. S. Wang, C. D. Manning, *Baselines and bigrams: Simple, good sentiment and topic classification*. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 8-14 July 2012, Jeju Island, Korea (2012).
10. L. Adela. M. Ulfeta, *Comp. Sci. and Info Sys.* **16**.13-13 (2018)
11. M. Joshi, J. Wiebe, M. Ringuette, *Sarcasm in twitter: A closer look*, in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 9-11 September 2017, Copenhagen, Denmark (2017)
12. Y. Zhou, X. Chen, L. Huang, *Incorporating idiomatic expressions into sentiment analysis*, in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 9-11 September 2017, Copenhagen, Denmark (2017)
13. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2-7 June 2019, Minneapolis, Minnesota (2019)
14. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *OpenAI*, **1**, 8(2019)
15. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, V. Stoyanov, Cornell University, arXiv preprint arXiv:**1907.11692**, (2019)
16. Y. Cui, W. Che, T. Liu, B. Qin, Z. Yang, *IEEE/ACM Trans. Audio., Speech, and Language Proc. (TASLP)*, **29**, (2021)
17. X. Sun, C. Tan, Z. Liu, Cornell University, arXiv preprint arXiv:**1903.09868** (2019)
18. X. Jiang, C. Song, Y. Xu, Y. Li, Y. Peng, *Peer J Comput. Sci.* **8**, 18, e1005 (2022)
19. R. Kiros, R. Salakhutdinov, R. Zemel, *Multimodal neural language models*, in Proceedings of 31<sup>st</sup> International Conference on Machine Learning Research (PMLR), 21-16 June 2014, Beijing, China (2014)
20. Q. You, W. Li, T. Liu, *Image-text joint embedding for sentiment analysis*, in Proceedings of the 2016 ACM on International Conference on Multimodal Interaction, 12-16 November 2016, Tokyo, Japan (2016)
21. <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews> retrieved on 25 March 2023, 10.15 Hrs.