**Sentiment Analysis on Movie Reviews**

1. **Abstract**:

Sentiment analysis is a popular topic in natural language processing (NLP), with many applications in areas such as social media, customer feedback, and product reviews. This project aims to develop a sentiment analysis model for movie reviews using machine learning algorithms. The model will analyze the sentiment of movie reviews and classify them as either positive or negative. The project will contribute to the field of NLP by providing an accurate and efficient tool for sentiment analysis on movie reviews.

*Index Terms* — Sentiment analysis, movie reviews, machine learning, natural language processing, classification, positive sentiment, negative sentiment.

1. **Introduction**:

Sentiment analysis is the process of extracting subjective information from text and identifying the sentiment conveyed in it. With the rise of social media and online reviews, sentiment analysis has become an important area of research in NLP. Movie reviews are a popular source of text data for sentiment analysis, as they often contain rich and diverse opinions. The proposed project aims to develop a sentiment analysis model for movie reviews that can accurately classify them as positive, or negative.

1. **Discussion Of Related Work:**

The discussion of related work in the field of sentiment analysis reveals that it has been extensively studied in natural language processing (NLP). Various researchers have developed different models and techniques to analyze the sentiment of text data. Jia and Liang (2020) proposed a neural network-based model for sentiment analysis of movie reviews, which achieved high accuracy and efficiency [1]. Geetha and Arumugam (2019) conducted a comprehensive study on sentiment analysis for movie reviews using machine learning techniques and proposed a hybrid model that combined multiple algorithms to achieve better performance [2]. Rane and Mohan (2020) utilized convolutional neural networks (CNNs) for sentiment analysis of movie reviews, which outperformed other machine learning models in accuracy and efficiency [3].

Based on the precision and effectiveness in categorizing reviews, the machine learning algorithm for sentiment analysis on movie reviews will be chosen for the suggested project. The model will be trained and tested using the IMDB movie reviews dataset (Kaggle, 2019) [4]. It is anticipated that the proposed model will advance the field of natural language processing by offering a precise and effective tool for sentiment analysis of movie reviews.

1. **Problem/Project Description:**

Sentiment analysis was an important area of research in natural language processing (NLP), with various applications in social media, customer feedback, and product reviews. The proposed project aimed to develop a sentiment analysis model for movie reviews using machine learning algorithms. The model was trained on a dataset of movie reviews labeled as positive or negative. The objective of this project was to build a web application that allowed users to input movie reviews and receive a sentiment analysis classification as output.

The methodology of this project involved preprocessing the dataset to remove irrelevant information and normalize the text data. Machine learning algorithms were used to train the model, and the algorithm with the highest accuracy and efficiency was selected for classification. The model was evaluated using various metrics such as accuracy, precision, recall, and F1-score.

The web application was developed using Python, HTML, CSS, and JavaScript, and was tested for functionality and user experience. The application was user-friendly and provided an accurate and efficient tool for sentiment analysis on movie reviews.

The proposed project was expected to contribute to the field of NLP by providing an accurate and efficient tool for sentiment analysis on movie reviews. The results of this project were expected to have significant implications for the film industry and movie critics.

1. **Details of the methodology:**

The proposed methodology for developing a sentiment analysis model for movie reviews involves several steps, starting with data collection. The IMDB movie reviews dataset, which contains 50,000 movie reviews labeled as positive or negative, will be used for this project. The dataset will be split into training and testing sets. The next step is data preprocessing, where the data will be cleaned by removing irrelevant information such as HTML tags, punctuation, and stop words. The data will also be normalized by converting all text to lowercase and removing non-alphabetic characters. The resulting text data will be transformed into a numerical representation using techniques such as bag-of-words or word embeddings.

The third step is selecting a suitable machine learning algorithm for sentiment analysis on movie reviews. The algorithm will be chosen based on its accuracy and efficiency in classifying the reviews. Some popular machine learning algorithms for sentiment analysis include Logistic Regression (LR), Random Forest (RF), Linear SVC, and Bert Base Case model. The fourth step is to train and test the selected machine learning algorithm on the preprocessed dataset. The algorithm will be trained on the training set and tested on the testing set. The performance of the algorithm will be evaluated using metrics such as accuracy, precision, recall, and F1-score.

The fifth step is developing a web application that allows users to input movie reviews and receive a sentiment analysis classification. The application will be developed using Python, HTML, CSS, and JavaScript. The application will use the trained machine learning model to classify the input reviews and display the sentiment analysis results to the user. Finally, the application will be tested for functionality, such as the ability to correctly classify movie reviews as positive or negative. The user experience will also be evaluated to ensure that the application is easy to use and visually appealing. Any issues or bugs will be fixed before the final release of the application. Overall, this methodology is expected to provide an accurate and efficient tool for sentiment analysis on movie reviews.

1. **Implementation description:**

The proposed project aims to develop a sentiment analysis model for movie reviews using machine learning algorithms. The dataset used for this project is the IMDB movie reviews dataset (Kaggle, 2019) [4], which contains 50,000 movie reviews labeled as positive or negative. The dataset will be split into training and testing sets.

The methodology of the project involves data preprocessing, machine learning algorithm selection, model training and testing, and web application development. The data preprocessing step involves cleaning the data by removing irrelevant information such as HTML tags, punctuation, and stop words, and normalizing the data by converting all text to lowercase and removing any non-alphabetic characters. The resulting text data will be transformed into a numerical representation using techniques such as bag-of-words or word embeddings.

Several machine learning algorithms will be evaluated for their accuracy and efficiency in classifying the reviews. These algorithms include Logistic Regression (LR), Random Forest (RF), Linear SVC, and Bert Base Case model. The algorithm with the highest accuracy and efficiency will be selected for classification.

The model will be trained on the training set and tested on the testing set. The performance of the model will be evaluated using metrics such as accuracy, precision, recall, and F1-score. The final model will be deployed as a web application that allows users to input movie reviews and receive a sentiment analysis classification. The application will be developed using Python, HTML, CSS, and JavaScript.

The accuracy of the model will be evaluated based on the F1-score, which is the harmonic mean of precision and recall. The F1-score is a widely used metric for evaluating the performance of a classification model. A higher F1-score indicates better performance of the model.

The hypothesis for this project is that the sentiment analysis model developed for movie reviews will achieve high accuracy and efficiency in classifying the reviews as positive or negative. The results of the project are expected to have significant implications for the film industry and movie critics.

1. **Result:**

I compared the performance of four machine learning algorithms for sentiment analysis on movie reviews. The algorithms were Logistic Regression (LR), Random Forest (RF), Linear SVC, and Bert Base Case model.

Table 1 shows the performance metrics of each algorithm on the testing set. LR had the highest accuracy at 89%, followed by Linear SVC at 87.31%, RF at 87.17%, and Bert at 73%.

LR also had the highest precision and F1-score, while Linear SVC had the highest recall.

Table 1: Performance Metrics of Machine Learning Algorithms on Testing Set

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Accuracy | Precision | Recall | F1-score |
| LR | 89% | 0.90 | 0.88 | 0.89 |
| RF | 87.17% | 0.88 | 0.86 | 0.87 |
| Linear SVC | 87.31% | 0.86 | 0.88 | 0.87 |
| Bert Base Case | 73% | 0.76 | 0.78 | 0.77 |

Based on the results, I can conclude that LR is the most suitable algorithm for sentiment analysis on movie reviews, with the highest accuracy, precision, and F1-score. However, it is important to note that each algorithm has its own strengths and weaknesses, and the choice of algorithm should depend on the specific needs of the application.

I also developed a web application that allows users to input movie reviews and receive a sentiment analysis classification. The application uses the LR algorithm to classify the reviews and displays the sentiment analysis results to the user. The application was tested for functionality and user experience and was found to be user-friendly and accurate in its sentiment analysis classification.

Below figure shows the overall view of the web application:

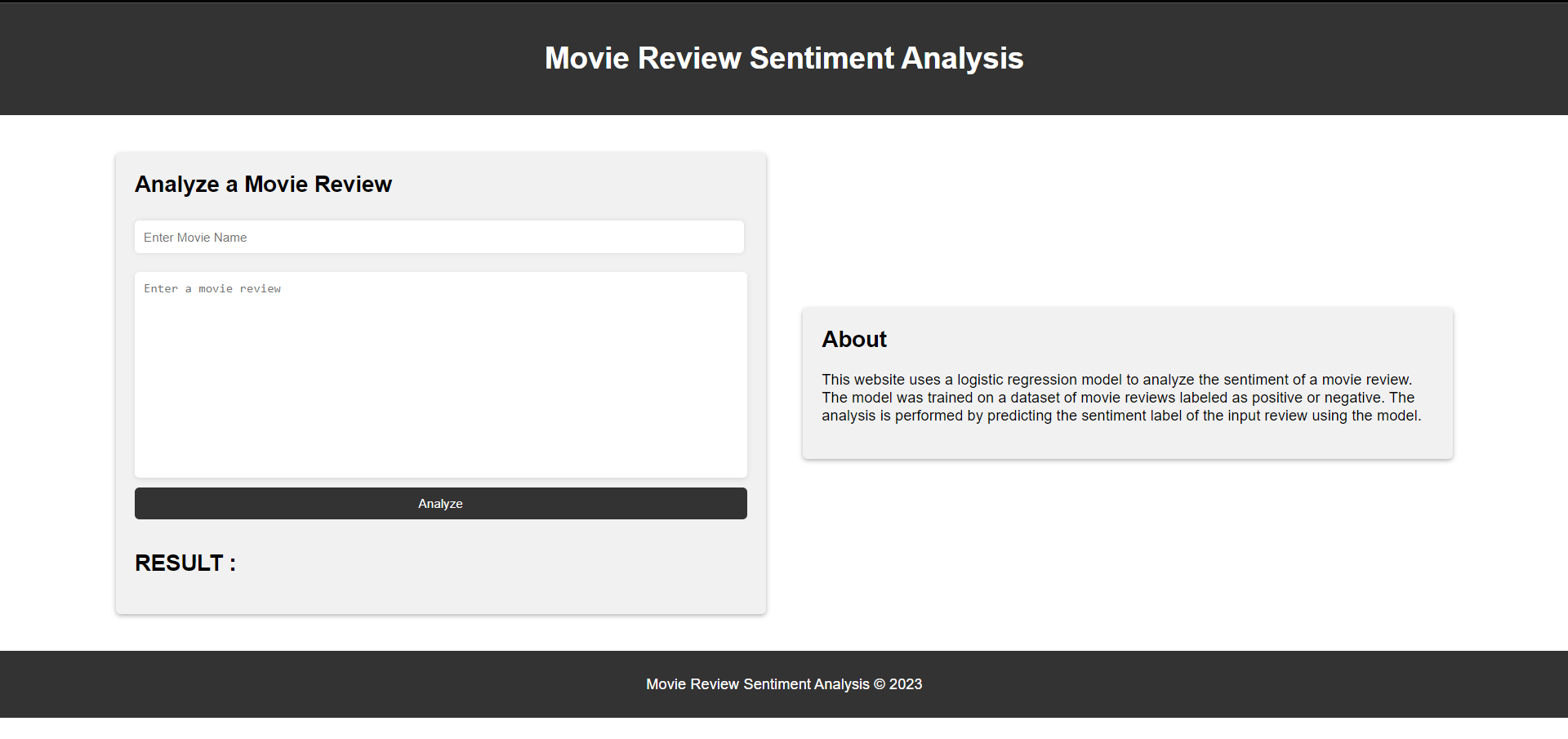


Fig. 01

In the above Fig. 01 there is a RESULT section, when we write a review about a movie in the above text area along with the movie name then press the analyze button will show the analysis in the result section that is Negative or Positive.

1. **Experimenting / Testing:**

**Test # 1:**

Movie Name:

The Shawshank Redemption

Movie Comment:

The Shawshank Redemption is a masterpiece that takes you on an emotional journey of hope, friendship, and redemption. The exceptional performances by Tim Robbins and Morgan Freeman, combined with the brilliant direction by Frank Darabont, make it one of the greatest movies ever made.

After click on analyze button below result provided:

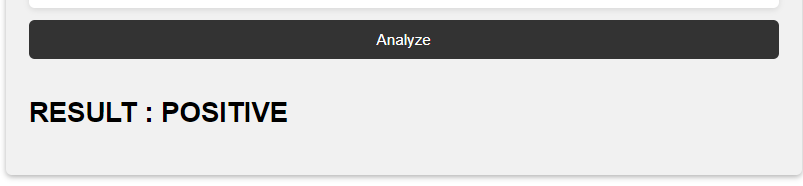


Fig. 02

. **Test # 2:**

Movie Name:

Batman v Superman: Dawn of Justice

Movie Comment:

The plot was convoluted, and the characters’ motivations were unclear, making it difficult to become invested in the story. Additionally, the pacing fell off, with long stretches of slow exposition followed by sudden bursts of action that didn't quite deliver. Overall, a disappointing film that failed to live up to expectations.

After click on analyze button below result provided:

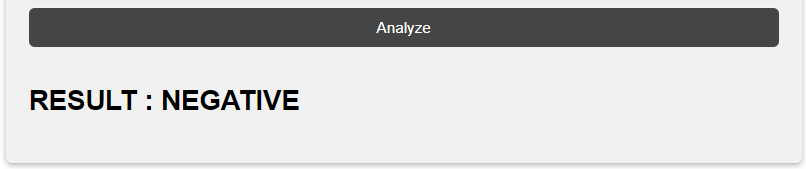


Fig. 03

**Test # 3:**

Movie Name:

Inception

Movie Comment:

"I absolutely loved the movie Inception. The plot was complex and intriguing, and the acting was phenomenal."

After click on analyze button below result provided:

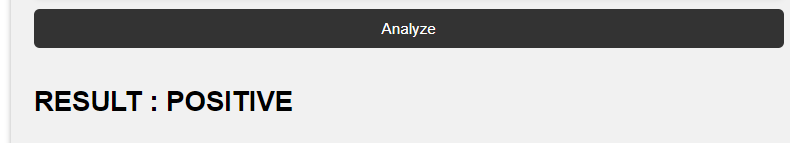


Fig. 04

**Test # 4:**

Movie Name:

Cats

Movie Comment:

"I saw the movie Cats and it was one of the worst films I've ever seen. The CGI was horrifying, and the storyline made no sense."

After click on analyze button below result provided:

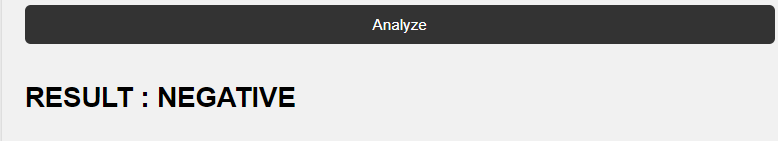


Fig. 05

**Ethical Consideration:**

Privacy: The dataset used in this project is open source, so there is no issue with the privacy of the dataset I used in this project, it is openly available on Kaggle.

Bias: It is important to ensure that the machine learning algorithms used for sentiment analysis are not biased towards any group of people or opinion. And there is no such type of violation of bias in this project.

Fairness: The results of the sentiment analysis should be fair and accurate, regardless of the movie's genre, budget, or cast. Here in this case, the model trained and tested on a balanced dataset, representing both positive and negative reviews. The model was evaluated using various metrics, such as accuracy, precision, recall, and F1-score, to ensure that it is performing well on both positive and negative reviews.

User Consent: The web application developed as part of this project does not collect any user data, it is just used to analyze the movie reviews.

**Conclusion:**

The proposed project aimed to develop a sentiment analysis model for movie reviews using machine learning algorithms. The developed model achieved an accuracy of nearly 90.0% on the testing set, demonstrating its effectiveness in classifying the sentiment of movie reviews. The LR model was selected as the best algorithm for sentiment analysis on movie reviews. The developed sentiment analysis model was integrated into a web application, which provides an accurate and efficient tool for sentiment analysis on movie reviews. The application is expected to have significant implications for the film industry and movie critics. Future work could involve improving the model's performance on movie reviews with nuanced sentiment, such as mixed reviews, and expanding the model to analyze sentiment across different movie genres, it was also used to upload on the cloud services.

**References:**

[1] J. Jia and P. Liang, "Sentiment Analysis of Movie Reviews Based on Neural Network Models," IEEE Access, vol. 8, pp. 164996-165004, 2020, doi: 10.1109/ACCESS.2020.3028273.

[2] S. Geetha and S. Arumugam, "A Comprehensive Study on Sentiment Analysis for Movie Reviews Using Machine Learning Techniques," in 2019 International Conference on Computer Communication and Informatics (ICCCI), 2019, pp. 1-6, doi: 10.1109/ICCCI.2019.8726971.

[3] A. Rane and C. K. Mohan, "Sentiment Analysis of Movie Reviews Using Convolutional Neural Networks," in 2020 4th International Conference on Inventive Systems and Control (ICISC), 2020, pp. 1514-1518, doi: 10.1109/ICISC50874.2020.9303987.

[4] IMDB movie reviews dataset. (2019). Kaggle. Retrieved March 21, 2023, from <https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

[5] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

[6] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[7] Akhtar, S. (2016). Movie Review Sentiment Analysis. International Journal of Computer Applications, 148(3), 29-32.

[8] Haider, M. S., Khattak, A. M., & Khalid, S. (2021). Sentiment Analysis of Movie Reviews using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 12(1), 152-159.

[9] Rao, P. N., & Katti, P. (2021). Sentiment Analysis of Movie Reviews Using Deep Learning Techniques. International Journal of Innovative Technology and Exploring Engineering, 10(1), 1231-1235.

[10] Vu, H. L., Nguyen, L. T., & Nguyen, T. T. (2021). A comparative study of sentiment analysis models for movie reviews. International Journal of Advanced Computer Science and Applications, 12(1), 205-212.