**SENTIMENT ANALYSIS ON MOVIE REVIEWS USING NATURAL LANGUAGE PROCESSING**

**Submitted By: Himanshi Baherwani**  
**Roll No.: T029**  
**B.Tech – Computer Engineering**  
**NMIMS , Indore**

**ACKNOWLEDGMENT**

I am grateful to NMIMS University, Indore, for providing the resources and environment to complete this project. I would like to thank my peers and classmates for their continuous support, insights, and feedback throughout this journey.

Their cooperation encouraged me to explore Natural Language Processing (NLP) and Machine Learning, which helped develop this Sentiment Analysis model.

**ABSTRACT**

Sentiment Analysis is a Natural Language Processing technique used to determine the emotional tone behind text data. This project focuses on classifying IMDb movie reviews as either **positive** or **negative** using machine learning techniques.

The model uses text preprocessing, TF-IDF vectorization, and supervised learning classification to understand the polarity of reviews. A trained model can assist film production houses, review platforms, and recommendation systems by identifying audience emotions automatically.

The dataset used contains a large number of movie reviews from IMDb. This project demonstrates the workflow of transforming raw text into numerical vectors, training a machine learning model, evaluating performance, and making predictions.

**TABLE OF CONTENTS**

1. Introduction
2. Literature Survey
3. Problem Statement
4. Objectives
5. Methodology
6. System Architecture
7. Implementation
8. Results & Output
9. Advantages
10. Limitations
11. Future Scope
12. Conclusion
13. References

**1. INTRODUCTION**

With exponential growth in social media and review platforms, analyzing public opinion has become essential. Manual interpretation is time-consuming, subjective, and inefficient.

Sentiment Analysis provides:

* Automated sentiment detection
* Faster decision-making
* Scalable opinion mining

In this project, Natural Language Processing (NLP) techniques are used to classify movie reviews as *positive* or *negative* using machine learning. By leveraging the IMDb dataset, the model learns commonly associated words and their emotional polarity.

**2. LITERATURE SURVEY**

Researchers worldwide are working on text analytics using machine learning.

* **Pang & Lee (2002):** Demonstrated effectiveness of NLP for movie review classification.
* **Socher et al. (2013):** Introduced Recursive Neural Networks for sentiment composition.
* **Kim (2014):** Used CNNs for sentence-level classification.

Traditional approaches include:

* Bag of Words (BoW)
* TF-IDF
* Naive Bayes classifiers

Modern advancements leverage:

* Word Embeddings (Word2Vec, GloVe)
* LSTM networks
* Transformer architectures

This project combines conventional text representation with an efficient supervised classifier technique.

**3. PROBLEM STATEMENT**

To design and implement a machine learning model that automatically classifies IMDb movie reviews into *positive* or *negative* sentiment using Natural Language Processing techniques.

**4. OBJECTIVES**

* To clean and preprocess raw text data.
* To convert text into numerical feature vectors.
* To train a classification model.
* To evaluate sentiment accuracy.
* To test the model on custom, user-given reviews.

**5. METHODOLOGY**

The workflow includes:

**Data Collection**

* IMDb dataset (built-in Keras dataset)
* Contains predefined labeled reviews

**Data Preprocessing**

* Tokenization
* Removal of stop words
* Lowercasing
* Padding/truncation

**Feature Extraction**

* TF-IDF / Word Embedding vectorization

**Model Training**

Algorithms typically used:

* Logistic Regression
* Naive Bayes
* Neural Networks

**Evaluation**

* Accuracy Score
* Precision
* Recall
* Confusion Matrix

**6. SYSTEM ARCHITECTURE**

**Input Review → Preprocessing → Feature Extraction → Model Training → Prediction → Output Sentiment**

**7. IMPLEMENTATION**

Technologies Used:

* Python
* Google Colab
* TensorFlow / Keras
* Scikit-Learn
* NumPy / Pandas

Steps:

1. Import IMDb dataset
2. Tokenize text data
3. Convert to padded sequences
4. Build neural network model
5. Train on training data
6. Evaluate on test data
7. Predict sentiment of custom reviews

**8. RESULTS & OUTPUT**

After training, the model achieved a good accuracy score on test data.  
It was able to correctly classify new user-provided movie reviews.

Example Predictions:

* “This movie was fantastic!” → Positive
* “Worst movie ever made.” → Negative

**9. ADVANTAGES**

* Fast processing of large text data
* No human bias
* Useful for recommendation engines
* Real-time sentiment understanding

**10. LIMITATIONS**

* Sarcasm is difficult to detect
* Dataset domain restricted to movies
* Requires continuous retraining
* Context understanding limited

**11. FUTURE SCOPE**

* Multilingual sentiment detection
* Transformer-based models (BERT)
* Fine-grained sentiment categories
* Emoji sentiment analysis
* Integration with social media analytics

**12. CONCLUSION**

This project successfully demonstrates how Natural Language Processing can classify movie reviews based on sentiment. By transforming raw text data into meaningful numerical representation, we trained a model capable of detecting positive and negative reviews.

The implementation highlights the importance of preprocessing, vectorization, and model evaluation in sentiment analysis applications.

**13. REFERENCES**

1. Bing Liu, Sentiment Analysis: Mining Opinions, 2015.
2. Research papers on IMDb sentiment classification.
3. Keras, TensorFlow official documentation.
4. Wikipedia – Sentiment Analysis
5. Medium / Analytics Vidhya NLP articles