

# Movie Success Prediction and Sentiment Study

## Introduction

The movie industry thrives on the anticipation of a film's success. With the advent of online reviews and social media, user sentiments have become a key indicator of a film's performance. This project aims to leverage sentiment analysis and machine learning to predict a movie's success based on user reviews.

## Abstract

This project combines Natural Language Processing and machine learning to evaluate and predict the success of movies. Using IMDB data, the study analyzes the sentiment of viewer reviews via VADER sentiment analysis and applies classification models to predict whether a review is positive or negative. A genre-wise sentiment trend was also examined to understand viewer preferences.

## Tools Used

- Python (Pandas, NLTK, Scikit-learn)
- VADER Sentiment Analyzer
- TF-IDF Vectorization
- Matplotlib & Seaborn (for visualization)
- Jupyter Notebook

## Steps Involved in Building the Project

1. Data Collection: Imported IMDB dataset containing 50,000 labeled reviews.
2. Preprocessing: Cleaned HTML tags, removed special characters, and tokenized text.
3. Sentiment Analysis: Applied VADER to compute sentiment scores for each review.
4. Genre Simulation: Randomly assigned genres to each review for sentiment grouping.
5. Feature Extraction: Used TF-IDF to convert text into numerical features.
6. Model Building: Trained Logistic Regression, Naive Bayes, Random Forest, and SVM models.
7. Evaluation: Compared accuracy, precision, and ROC-AUC across models.
8. Visualization: Created charts to represent sentiment scores and model results.

## Conclusion

The sentiment of user reviews can be a strong predictor of a movie's success. Among the tested models, Support Vector Machine (SVM) with TF-IDF features delivered the highest accuracy. Additionally, genre-wise sentiment analysis provided useful insights into audience preferences. This project demonstrates how machine learning and NLP can offer valuable tools for forecasting trends in the entertainment industry.