Movie Success Prediction and Sentiment Analysis

Objective

- Predict movie success based on IMDB movie data and ratings.
- Analyze the sentiment of movie reviews using VADER Sentiment Analyzer.

Tools & Libraries

- Python
- Pandas, NumPy for data manipulation
- NLTK (VADER) for sentiment analysis
- Scikit-learn for machine learning models and evaluation
- Matplotlib, Seaborn for visualizations

Dataset Description

- Movie Data: Contains attributes like Title, Genre, Description, Director, Revenue, Rating, Votes, etc.
- Review Data: Contains movie review texts, user ratings, and sentiment labels (positive = 1, negative = 0).

Workflow Steps

- Step 1: Importing and Cleaning Data
  - Loaded movie and review datasets.
  - Filled missing values using median.
  - Ensured clean datasets.
- Step 2: Sentiment Analysis with VADER
  - Applied VADER SentimentIntensityAnalyzer.
  - Classified reviews as positive or negative.
  - Achieved an accuracy of 100% for sentiment classification.
- Step 3: Feature Engineering for Movie Success Prediction
  - Selected features: Rating, Votes, Revenue (Millions), Metascore.
- Step 4: Model Building for Success Prediction
  - Built Logistic Regression model.
  - Achieved 100% accuracy on test set.

- Step 5: Visualization
  - Created simple visualizations like sentiment distribution and Revenue vs Rating scatter plots.

Results Summary

- Sentiment Analysis Accuracy: 100%
- Movie Success Prediction Accuracy: 100%
- Data Handling: No missing values after cleaning
- Extra Feature Used: None

Conclusion

- VADER sentiment analysis matched human-labeled sentiments very well.
- Logistic Regression model successfully predicted movie success with high accuracy.
- Suggested to test further on larger, unseen datasets for real-world deployment.