**Predictive Model for Movie Ratings**

**1. Introduction**

This project aims to develop a machine learning model to predict movie ratings based on various attributes such as genre, director, duration, and votes. By leveraging feature engineering techniques and training a robust predictive model, we can provide insights into movie success factors. The model is evaluated using standard performance metrics like RMSE and R² scores.

**2. Dataset Description** The dataset consists of movie-related information, including:

* **Name**: Title of the movie
* **Year**: Release year
* **Duration**: Runtime in minutes
* **Genre**: Category of the movie
* **Rating**: IMDb rating
* **Votes**: Number of votes received
* **Director**: Movie director
* **Actor 1, Actor 2, Actor 3**: Leading actors

**3. Data Preprocessing**

* **Handling Missing Values**: Median imputation for numerical fields and mode imputation for categorical fields.
* **Data Type Conversion**: Extracting numeric values from text fields like Year and Duration.
* **Encoding Categorical Variables**: Label encoding for Genre and Director.

**4. Feature Engineering**

* **Director Success Rate**: The average IMDb rating of movies directed by a given director.
* **Genre Average Rating**: The average rating of movies belonging to the same genre.

**5. Model Implementation** The dataset is split into training and testing sets (80:20). A **Random Forest Regressor** with 100 trees is trained to predict ratings. This model is chosen for its ability to handle non-linearity and categorical variables effectively.

**6. Model Evaluation**

* **Root Mean Squared Error (RMSE)**: Measures the average prediction error.
* **R² Score**: Indicates how well the model explains variance in ratings.

**7. Results and Analysis**

* RMSE: Indicates the accuracy of predictions.
* R² Score: Measures model performance.
* Feature importance analysis highlights the most influential factors in predicting ratings.

**8. Conclusion and Future Scope** The model successfully predicts movie ratings using key attributes. Future improvements include integrating NLP-based sentiment analysis on reviews, refining feature engineering, and testing deep learning models for enhanced accuracy.

**9. References** Appropriate references to datasets and machine learning documentation.