CinemaScore Predictor: An AI Model for IMDb Rating Prediction

Purpose:

The purpose of this project is to develop a supervised machine learning model that can predict the IMDb score of a movie based on key attributes such as the director, budget, lead actor, and genre. This tool aims to assist film studios, investors, and enthusiasts in predicting the potential reception of a movie before its release, thereby aiding in strategic decision-making for production, marketing, and investment.

Dataset:

- Source: We will compile data from the IMDB movie database like IMDb. This dataset will include:
    - Director: Names of directors.
    - Budget: The total budget in USD.
    - Actor: Names of lead actors.
    - Genre: Categorized genres of the movie.
    - IMDB Score: The actual IMDB score of the movie.

- Size: 5,000 movies to ensure the model has enough data to learn from, considering the variety in directors, actors, and genres.
- Cleaning: Data cleaning will involve handling missing values, normalizing budget figures, and encoding categorical variables like director, actor, and genre.

Initial Analysis:

- Exploratory Data Analysis (EDA):
    - Visualize the distribution of IMDB scores.
    - Analyze the correlation between budget and IMDB scores.
    - Explore how different genres affect scores.

- Feature Engineering:
    - Create interaction features between director-genre, actor-genre, etc., to capture combined effects.
    - Use techniques like one-hot encoding or embedding for categorical variables.

- Data Insights:
    - Look for patterns or anomalies, such as if certain directors consistently perform better in specific genres or if high-budget films correlate with higher or lower scores.

Goals:

- Model Development:
    - Regression Models: Develop and evaluate regression models for training
    - Classification Models: Develop and evaluate classification models for training

- Performance Metrics:
    - Use Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared to evaluate model performance.
    - Implement cross-validation to ensure model robustness.

- Deployment:
    - Develop a simple web interface or an API where users can input movie details and receive a predicted IMDB score.
    - Ensure the model can be updated periodically with new movie data to keep predictions relevant.

- Future Enhancements:
    - Integrate more variables like release date, production company, or even sentiment analysis from movie trailers or reviews to improve prediction accuracy.
    - Consider real-time updates or predictions based on initial audience reactions or festival screenings.
    - Integrate other databases including: Box Office Mojo, and The Numbers for better predictive analysis.

This proposal outlines a structured approach to developing a predictive model for movie success, potentially revolutionizing how films are evaluated before they hit the screens.