PREDICTING BOX OFFICE SUCCESS USING PRE-RELEASE METADATA

Checkpoint Presentation - Progress and Updates

ABSTRACT

- Objective: Develop a machine learning model to predict box office revenue.
- Importance: Minimize financial risks and optimize decision-making in film production.
- Approach: Leverage datasets from TMDB and IMDb with advanced ML techniques.
- Outcome: Identify key factors influencing success and deliver actionable insights.

INTRODUCTION

- Problem: Unpredictability of box office performance presents high financial risks.
- Solution: Use pre-release metadata (cast, budget, genre) to build predictive models.
- Goal: Deliver a robust framework for predicting box office success.

PROGRESS SUMMARY

- 1. Data Acquisition: TMDB and IMDb datasets processed.
- 2. Exploratory Data Analysis: Correlation analysis and data distribution visualizations.
- 3. Preprocessing: Addressed missing values, normalized numerical features, encoded categories.
- 4. Initial Model Development: Baseline models with moderate predictive accuracy.

CHALLENGES

- 1. Data Quality: Significant preprocessing to handle missing and inconsistent data.
- 2. Feature Selection: Difficulty in identifying impactful features.
- 3. Modeling: Initial models showed room for improvement in accuracy.

PROPOSED NEXT STEPS

- 1. Refine Models: Use Gradient Boosting and Random Forest for better accuracy.
- 2. Feature Engineering: Add derived features like social media sentiment.
- 3. Optimization: Apply hyperparameter tuning and cross-validation.

EVALUATION PLAN

- Metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R-squared.
- Validation: k-fold cross-validation to ensure robustness.
- Success Criteria: High predictive accuracy and identification of key features.

TIMELINE

- Week 1-2: Feature engineering and advanced model development.
- Week 3-4: Model evaluation and fine-tuning.
- Week 5: Generate insights and prepare the final report.

CONCLUSION

- The project is progressing as planned with key milestones achieved.
- Next steps focus on enhancing model accuracy and addressing challenges.
- Expected Outcome: Robust predictive framework for actionable insights.