PREDICTING BOX OFFICE SUCCESS USING PRE-RELEASE METADATA:

A DATA-DRIVEN APPROACH

BRANDON KHUU

ABSTRACT

- OBJECTIVE: PREDICT BOX OFFICE REVENUE USING PRE-RELEASE METADATA.
- IMPORTANCE: IMPROVE DECISION-MAKING IN FILM PRODUCTION AND MARKETING.
- APPROACH: USE MACHINE LEARNING MODELS ON DATASETS FROM IMDB AND TMDB.
- OUTCOME: IDENTIFY KEY FACTORS INFLUENCING MOVIE SUCCESS AND DEVELOP ACCURATE PREDICTIVE MODELS.

INTRODUCTION

- CHALLENGE: HIGH FINANCIAL RISK IN MOVIE PRODUCTION.
- NEED: DATA-DRIVEN INSIGHTS FOR BETTER RESOURCE ALLOCATION.
- FOCUS: PRE-RELEASE FACTORS LIKE BUDGET, CAST, AND GENRE.
- GOAL: DEVELOP A PREDICTIVE FRAMEWORK FOR BOX OFFICE SUCCESS.

RELATED WORKS

- **BOX OFFICE REVENUE PREDICTION USING LINEAR REGRESSION:** EARLY ATTEMPTS USING REGRESSION MODELS.
- **EARLY PREDICTIONS OF MOVIE SUCCESS:** HIGHLIGHTED IMPORTANCE OF TIMING AND BUDGET.
- **PREDICTING BOX OFFICE MARKETS:** INTRODUCED MACHINE LEARNING BUT LACKED DIVERSE DATASETS.
- **GAP:** LIMITED INTEGRATION OF MODERN MACHINE LEARNING TECHNIQUES AND EMERGING FACTORS LIKE DIGITAL MARKETING.

PROPOSED WORK

- **DATASETS:**
  - TMDB AND IMDB DATASETS FOR COMPREHENSIVE METADATA.
- **TOOLS:**
  - PYTHON LIBRARIES (PANDAS, SCIKIT-LEARN, XGBOOST).
- **TECHNIQUES:**
  - DATA CLEANING AND PREPROCESSING.
  - EXPLORATORY DATA ANALYSIS.
  - PREDICTIVE MODELING WITH LINEAR REGRESSION AND GRADIENT BOOSTING.

EVALUATION PLAN

- METRICS:
  - MEAN ABSOLUTE ERROR (MAE).
  - ROOT MEAN SQUARE ERROR (RMSE).
  - R-SQUARED.
- VALIDATION:
  - K-FOLD CROSS-VALIDATION FOR ROBUSTNESS.
  - COMPARATIVE ANALYSIS OF MULTIPLE MODELS.
- SUCCESS CRITERIA:
  - ACCURATE PREDICTIONS.
  - IDENTIFICATION OF SIGNIFICANT FEATURES.

TIMELINE

- WEEK 1: DATASET ACQUISITION AND PREPROCESSING.
- WEEK 2: EXPLORATORY DATA ANALYSIS AND FEATURE ENGINEERING.
- WEEK 3-4: MODEL DEVELOPMENT AND TUNING.
- WEEK 5: MODEL EVALUATION AND COMPARISON.
- WEEK 6: REPORTING AND INSIGHTS GENERATION.

CONCLUSION

- **EXPECTED IMPACT:**
  - REDUCE FINANCIAL RISKS IN MOVIE PRODUCTION.
  - PROVIDE ACTIONABLE INSIGHTS FOR FILMMAKERS AND MARKETERS.
- **FUTURE DIRECTIONS:**
  - INCORPORATE AUDIENCE SENTIMENT ANALYSIS.
  - ANALYZE COMPETITION FROM STREAMING PLATFORMS.

CITATIONS

- TMDB MOVIE DATASET (HTTPS://WWW.KAGGLE.COM/DATASETS/TMDB/TMDB-MOVIE-METADATA?RESOURCE=DOWNLOAD&SELECT=TMDB_5000_MOVIES.CSV)
- TMDB CREDITS DATASET (HTTPS://WWW.KAGGLE.COM/DATASETS/TMDB/TMDB-MOVIE-METADATA?RESOURCE=DOWNLOAD&SELECT=TMDB_5000_CREDITS.CSV)
- RESEARCH PAPERS:
  - *BOX OFFICE REVENUE PREDICTION USING LINEAR REGRESSION.*
  - *EARLY PREDICTIONS OF MOVIE SUCCESS.*