Telugu Movies Prediction Report

1. Dataset Description

1.1 Source: Public movie databases and entertainment sites (IMDb, Wikipedia, BookMyShow).

1.2 Columns:
• title – Movie name
• genre – Movie category (Action, Drama, Comedy, etc.)
• director – Director's name
• lead_actor – Main actor
• budget – Production budget (in crores)
• runtime – Movie duration (minutes)
• release_year – Year of release
• rating – IMDb-style rating (1–10 scale)
• collection – Box office earnings (crores)

1.3 Data Quality:
• Dataset of 500+ Telugu movies (2010–2023).
• Missing values handled through imputation and data cleaning.
• Duplicates removed; consistent numeric formatting.
• Categorical features encoded for modeling.

2. Operations Performed

2.1 Data Cleaning & Exploration:
• Checked for missing/null values and handled them appropriately.
• Standardized monetary units (budget, collections).
• Analyzed unique values for genre, director, and actor fields.

2.2 Descriptive Analytics:
• Genre distribution visualization (pie chart).
• Yearly trend of movies released and collections (line chart).
• Budget and rating distributions (histograms).
• Relationship plots between budget, rating, and collection.

2.3 Predictive Modeling:
• Model 1: Box Office Collection Prediction $\rightarrow$ Regression (Linear & Random Forest).
• Model 2: Rating Prediction $\rightarrow$ Regression/Classification.
• Train–test split: 80–20.
• Evaluation metrics: R^2 Score, RMSE, MAE.

3. Key Insights

3.1 Genre & Audience Preferences:
• Action and Drama dominate Telugu cinema (~40%).
• Comedy and Romance have moderate representation.
• Family-oriented genres show higher audience ratings.

3.2 Box Office Insights:
• Budget shows a strong positive correlation (r ≈ 0.78) with collection.
• Star-driven films often achieve higher returns, even with moderate ratings.
• Average movie collection: ~■25 crores.

3.3 Rating Insights:
• Average rating: 6.8/10.
• Movies directed by top-tier filmmakers average above 7.2/10.
• Ratings improved notably after 2018 due to better storytelling trends.

3.4 Model Insights:
• Random Forest outperformed Linear Regression in predicting collections.
• Feature importance: Budget > Genre > Lead Actor > Director.
• Predicted and actual collections closely align, confirming good model fit.

4. Recommendations

4.1 Pre-Production Strategy:
• Allocate budget strategically based on genre and lead actor popularity.
• Focus on strong scripts and proven directors to improve both ratings and collections.

4.2 Marketing & Release Strategy:
• Festival releases (e.g., Sankranti, Dasara) correlate with higher earnings.
• Increase promotional spending for low-rated but high-budget films to ensure visibility.

4.3 Data Expansion & Future Work:
• Integrate social media sentiment analysis to assess pre-release buzz.
• Include OTT performance data for post-release predictions.
• Explore clustering to identify movie "success archetypes."

4.4 Technical Enhancement:
• Deploy predictive models as an interactive dashboard for producers.
• Continuously retrain models using new movie data.