**DSA210 - FINAL REPORT**

**Sinan Altıntuğ**

**31954**

**Analyzing Factors Influencing IMDb, Rotten Tomatoes and MetaCritic Ratings of Netflix Content**

**Supervised by Berke Odacı**

**SABANCI UNIVERSITY**

**Introduction**

The aim of this project was to analyze and compare the impact of various factors on the IMDb, Rotten Tomatoes, and Metacritic scores of Netflix films and TV shows. By integrating data from three large datasets—including unique features such as budget, gross, and multiple score sources—we sought to understand how language, content genre, type, director, watch time, budget, and production year (among others) influence audience and critic ratings. This investigation bridges media analytics and data science, offering valuable insights for both content creators and streaming platforms.

**Dataset Description**

**Data Sources & Structure:**  
The dataset was manually compiled by selecting around 200 Netflix films and TV shows, enriched with information from external sources such as the IMDb API, Rotten Tomatoes API, and Kaggle datasets. This combined data includes over 20,000 entries with features such as:

* Title
* Genre
* Languages
* Series or Movie
* IMDb Score
* Rotten Tomatoes Score
* Metacritic Score
* Hidden Gem Score
* Country Availability
* Runtime
* Director
* Writer
* Actors
* View Rating
* Production House
* IMDb Votes
* Budget
* Boxoffice
* Release Date
* Production Year

**Feature Engineering:**

Runtime was standardized to minutes.

Main genres were extracted and one-hot encoded.

‘Series or Movie’ was encoded as binary.

Missing values were filled from external APIs or dropped if unresolvable.

Outliers were removed for variables like budget and score.

Additional features, such as “Hidden Gem Score,” were incorporated for richer modeling.

**Exploratory Data Analysis (EDA)**

**Distribution Analysis:**  
Histograms and boxplots were used to visualize the distribution of IMDb, Rotten Tomatoes, and Metacritic scores, as well as runtimes and budgets.

**Correlation Analysis:**  
Explored relationships between scores and numeric variables (runtime, production year, IMDb votes, etc.).

**Category Comparisons:**  
Boxplots and bar charts compared average ratings by genre, language, and content type (Series vs Movie).

**Yearly Trends:**  
Visualized trends in ratings over production years.

**Key Observations:**

- Certain genres and languages consistently receive higher ratings.

- Newer productions sometimes achieve higher scores, but not always.

- Longer runtimes and higher budgets can correlate with higher ratings.

- Original Netflix content may have different rating patterns compared to licensed titles.

**Hypothesis Testing**

**Director vs Rating**

* **IMDb:** There is **NO significant difference** in average IMDb scores between directors (ANOVA F=0.0444, p=0.853). (Fail to reject H₀)
* **Rotten Tomatoes:** There **IS a significant difference** between directors (ANOVA F=4.10, p<0.001). (Reject H₀)

**Country vs Rating**

* **IMDb:** There **IS a significant difference** in average IMDb scores across countries (ANOVA F=5.80, p=0.0008). (Reject H₀)
* **Rotten Tomatoes:** There **IS a significant difference** in Rotten Tomatoes scores across countries (ANOVA F=3.94, p<0.001). (Reject H₀)

**Budget vs Rating**

* **IMDb:** There **IS a significant difference** in average IMDb scores between box office budget categories (ANOVA F=38.43, p<0.001).
  + *Low budget:* 6.67
  + *Medium budget:* 6.41
  + *High budget:* 6.73
* **Rotten Tomatoes:** There **IS a significant difference** across budget groups (ANOVA F=109.94, p<0.001).
  + *Low:* 67.6, *Medium:* 51.2, *High:* 60.9
* **Metacritic:** There **IS a significant difference** across budget groups (ANOVA F=74.84, p<0.001).
  + *Low:* 61.0, *Medium:* 52.2, *High:* 57.8

**Production Year vs Rating**

* **IMDb:** There **IS a significant difference** in average IMDb scores between 5-year periods (ANOVA F=4.90, p=0.0044). (Reject H₀)
* **Rotten Tomatoes:** There **IS a significant difference** between 5-year periods (ANOVA F=14.00, p<0.001). (Reject H₀)

**Type (Series or Movie) vs Rating**

* **IMDb Score:**  
  ANOVA analysis revealed a **statistically significant difference** in IMDb scores between movies and series (F = 24.96, p = 0.000001).  
  **Series tend to have higher average IMDb scores** compared to movies.
* **Rotten Tomatoes Score:**  
  No statistically significant difference was found between movies and series for Rotten Tomatoes scores (F = 0.45, p = 0.50).
* **Metacritic Score:**  
  No statistically significant difference was found between movies and series for Metacritic scores (F = 0.0039, p = 0.95).

**Model Building and Evaluation**

Three regression algorithms were implemented to predict IMDb, Rotten Tomatoes, and Metacritic scores:

* **Linear Regression**
* **Random Forest Regressor**
* **Gradient Boosting Regressor**

**Model Features:**

* Numeric: Runtime (minutes), IMDb Votes, Hidden Gem Score, etc.
* Categorical: Main Genre (one-hot encoded), Series or Movie

**Performance Metric:**  
R² (coefficient of determination) was used to compare model accuracy.

**Results Overview:**

* For all three scores, Gradient Boosting and Random Forest generally provided higher R² values than Linear Regression.
* Example (for Rotten Tomatoes Score):  
  Gradient Boosting R² Score: 0.95  
  Linear Regression R² Score: 0.29  
  Random Forest R² Score: 0.95
* The best predictions were achieved with ensemble methods, while linear regression, though more interpretable, underperformed.

**Prediction Examples**

For each model and each score, predictions were demonstrated on both real test set samples and two custom-made hypothetical content examples (e.g., a Comedy Movie and an Action Series). This showcased how changes in genre, runtime, content type, and votes influence the predicted score.

**Conclusion**

The project successfully demonstrated that both content features (like genre, language, runtime) and external factors (director, production house, etc.) significantly influence audience and critic ratings on Netflix. Ensemble machine learning models provided the most accurate predictions, while Linear Regression was valuable for interpretability. The findings suggest:

* Certain genres and production languages are consistently associated with higher scores.
* Director reputation and budget play important roles in predicting ratings.
* Model selection matters: ensemble models outperform simple linear models in this context.

**Challenges and Future Work**

**Challenges Faced:**

* Missing values and inconsistencies between datasets required extensive cleaning.
* Not all attributes (like streaming hours or regional viewership) were available.

**Potential Improvements:**

* Expand the dataset size for better generalizability.
* Incorporate new features, such as detailed viewer engagement statistics.

**References**

* IMDb data via Kaggle
* Rotten Tomatoes data via kaggle
* Netflix data via Kaggle
* scikit-learn, pandas, matplotlib, seaborn (Python libraries)