Predicting Profits for Movie Releases

Charlie Reiney and
Noah Johnson

Data Collection

Sources

- The Movie Database
- IMDb

Query Criteria

- Profit
- Time-adjusted Popularity
  - 1991-2010: Pop. score of at least 5
  - 2011-2020: Pop. score of at least 2

Pre-Processing

- Binarize movie collection data
- One-Hot Encode genres
- Parse release date (m/d/y)
- Remove text attributes (Sentiment Analysis ☹)
- Discretize revenue (Classification)
- Normalization/Scaling (SVR, SVM)

Feature Exploration

Models Overview

Regression

- Linear
- KNN
- Support Vector
- Reg. Tree
- XGBoost

Classification

- KNN
- Naïve Bayes
- Decision Tree
- Random Forest
- XGBoost
- Neural Network

Regression Results

XGBoost

- **R^2 = 0.77850**
- MAE = 32283606.03

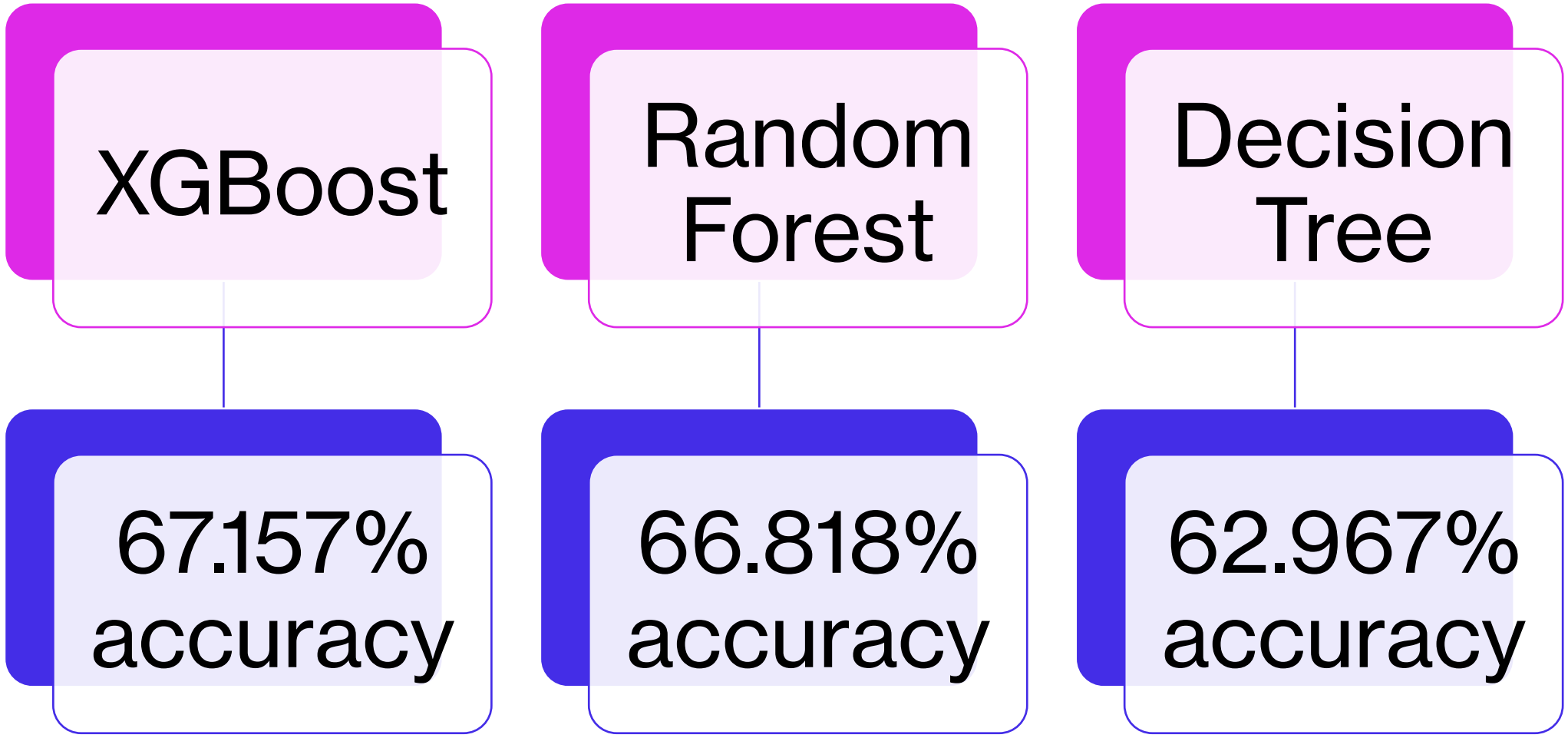
Regression Tree

- R^2 = 0.73081
- **MAE = 31457472.86**

Linear Regression

- R^2 = 0.74132
- MAE = 41071663.34

Classification Results



Real Predictions

- Movies: Venom, Dune, The Eternals, Clifford, Belfast, King Richard, Ghostbusters Afterlife, C’mon C’mon, Encanto, House of Gucci, and Resident Evil
- KNN Regression (k=2)
  - R^2 = .93495
  - MAE = 34,724,816.05
- Naïve Bayes
  - Accuracy = 81.8181%

Discussion

1

Improving Models

- Edit Neural Network
- Perform sentiment analysis
- Find a new ensemble regression model

2

Collecting More Data

- YouTube views on official trailer
- Top-billed cast members
- Social media interaction