**Algorithms**

1. **K Nearest Neighbor Algorithm:**Chart

   Description automatically generatedChart, scatter chart

   Description automatically generated

Plotting predicted and actual avg\_vote revealed issues with our model that our evaluation metrics did not immediately make clear

**KNN Classification Model:**Chart, scatter chart

Description automatically generated

Table

Description automatically generated

Testing on Unseen Data Results:

Accuracy: 69.11%

Sensitivity: 74.66%

Precision: 72.38%

1. **Random Forest (BEST MODEL):**

Chart

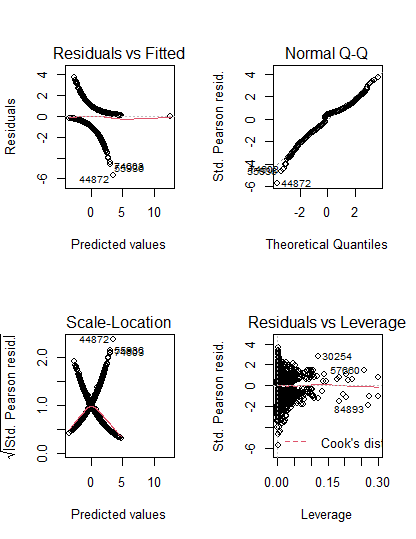
Description automatically generated

Model was built using random forest classification using duration, genre\_1, genre\_2, continent, language, oscar\_actor, oscar\_director, oscar\_writer, and top\_100 as predictors

**Chart, scatter chart

Description automatically generated**

Important variable that we found was duration

1. **Logistic Regression:**

**Chart

Description automatically generated**Model\_2 took in less predictors than model\_1, resulting in a better AIC score.

Results on Unseen Data (using probability threshold of 50%):

Accuracy: 70%

Sensitivity: 56%

Precision: 72%

1. Chart, scatter chart

   Description automatically generatedA picture containing text, receipt

   Description automatically generated**Linear Regression:**

For the most part, our residual vs fitted had no funneling shape, which lead to no indication of heteroscedasticity. Also, it had a random scattering around the 0 line, indicating a linear relationship.

Our Q-Q graph showed us a normally distributed graph because it did not have an S pattern.

**Forward Selection:**

**A picture containing text, music, piano

Description automatically generated**

Forward Selection to figure out our optimal set of predictors based on lowest RSS