1. Linear Regression

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Corr | Root MSE |  |
| LinearRegM5 | 0.82 | 106.3743 |  |
| Linear Reg Greedy | 0.83 | 106.3382 |  |
| LinearRegFeatures | 0.79 | 114.5845 |  |

1. Nearest Neighbor

|  |  |  |
| --- | --- | --- |
| Model | Correlation | Root Mean |
| 1-nearest | 0.7934 | 86.3745 |
| 3-nearest | 0.7941 | 76.0591 |
| 8-nearest neighbor | 0.7654 | 72.8537 |
| FEATURE SELECTED |  |  |
| 1-nearest | 0.7033 | 133.578 |
| 3-nearest | 0.74096 | 119.5527 |
| 8-nearest neighbor | 0.72066 | 124.1453 |

\**investigate why 8 nearest neighbors is less than 3 nearest neighbors*

1. Linear Reg (LOGGED target)

|  |  |  |
| --- | --- | --- |
| Model | Corr | MSE |
| M5 | 0.8451 | 0.4079 |

\**talk about how logged has a higher correlation due to some sort of normalization*

1. Decision Trees

|  |  |  |
| --- | --- | --- |
| Model | Accuracy | Root MSE |
| Decision Tree (J48 with Pruning) | 87.0234% | 0.3193 |
| Zero R | 49.1272 | 0.5008 |
| Random Forest | 84.302 | 0.3200 |
| Logistic Regression | 84.2893 | 0.3254 |
| Decision Tree (Select Features) | 82.663 | 0.3605 |
| Random Forest (Select) | 80.6337 | 0.3665 |
| Logtistic Reg (select) | 84.917 | 0.3412 |

For select features, I used same features for regression tasks + #number of followers because that was additionally important when I ran the initial decision tree.

\*give possible reasons to why random forest was lower than decision tree (possibly pruning)