|  |  |  |  |
| --- | --- | --- | --- |
| Model | Correlation | Mean Sq Error | Datasets |
| LinearReg M5 | 0.7023 | 76.9082 | Cross Valid 10 Train |
| LinearReg Greedy | 0.8224 | - | Cross Valid 10 Train |
| LinearRegFeatures | 0.8027 | 80.5633 | Cross Valid 10 Train |
|  |  |  |  |
| LinearReg M5/Greedy | 0.838  (no diff) | 76.3631 | Training Set |
|  |  |  |  |

1. Linear Regression
2. Nearest Neighbor

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Correlation | Root Mean |  |
| 1-nearest | 0.7929 | 86.3745 | Cross Validation 10 |
| 3-nearest | 0.8249 | 76.3174 | Cross Validation 10 |
| 8-nearest neighbor | 0.8319 | 74.9065 | Cross Validation 10 |
| FEATURE SELECTED |  |  |  |
| 1-nearest | 0.7743 | 89.3561 | CV 10 |
| 3-nearest | 0.812 | 78.9068 |  |
| 8-nearest neighbor | 0.80 | 81.2838 | CV10 |

1. Linear Reg (LOGGED target)

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Correlation | Mean Sq Error | Datasets |
| LinearReg M5 | 0.7023 | 76.9082 | Cross Valid 10 Train |
| LinearReg Greedy | 0.8224 | - | Cross Valid 10 Train |
| LinearRegFeatures | 0.8027 | 80.5633 | Cross Valid 10 Train |
|  |  |  |  |
| LinearReg M5/Greedy | 0.838  (no diff) | 76.3631 | Training Set |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Corr | MSE | Data |
| M5 | 0.8241 |  | CV10 |
| Features |  |  |  |
| M5 | 0.6718 |  | CV10 |
| Greedy | 0.6719 | SAD |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Correlation | Mean Sq Error | Datasets |
| c | 0.7023 | 76.9082 | Cross Valid 10 Train |
| LinearReg Greedy | 0.8224 | - | Cross Valid 10 Train |
| LinearRegFeatures | 0.8027 | 80.5633 | Cross Valid 10 Train |
|  |  |  |  |
| LinearReg M5/Greedy | 0.838  (no diff) | 76.3631 | Training Set |
|  |  |  |  |

1. Decision Trees

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Accuracy | Root MSE |  |
| Decision Tree (J48 with Pruning) | 85% | 0.3486 | CV10 |
| Zero R | 52% | 0.500 | CV10 |
| Random Forest | 85.439% | 0.324 | CV10 |
| Logistic Regression | 85.7949 | 0.3208 | CV10 |
| Decision Tree (Select Features) | 84.1206 | 0.3593 | CV10 |
| Random Forest (Select) | 84.87 | 0.331 | CV10 |
| Logtistic Reg (select) | 85.07 | 0.3388 | CV10 |

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.