Model Metrics - Video Game Classification

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Balanced Accuracy | Precision - Above Average | Precision - Equal / Below Average | Recall - Above Average | Recall -  Equal / Below Average |
| Entropy - No Max Depth | 79.57% | 72.86% | 76.43% | 80.50% | 53.68% | 92.03% |
| Gini - No Max Depth | 79.54% | 72.81% | 76.36% | 80.48% | 53.62% | 92.01% |
| Entropy - Max Depth 16 | 79.89% | 73.42% | 76.55% | 80.91% | 54.93% | 91.90% |
| Gini - Max Depth 15 | 79.84% | 73.28% | 76.68% | 80.79% | 54.54% | 92.02% |
| Random Forest - No Max Depth | 80.00% | 73.26% | 77.64% | 80.69% | 54.00% | 92.52% |
| Random Forest - Optimized Params | 80.33% | 73.61% | 78.46% | 80.88% | 54.40% | 92.81% |

In a real world situation, when predicting that a game will have above average suggestions, false positives would lead to costly investment in a game that might not be popular enough to make back the money. A false negative in this situation would mean that money would not be invested in a particular game that is likely to be successful, however the company would likely choose a different game idea that is identified as likely to succeed, and thus would still make a profit. Therefore it is important to avoid false positives when predicting success and inversely avoid false negatives when predicting failure. Using this logic, the most important metrics are precision for above average games and recall for equal or below average games. The optimized random forest model performs the best in both of these metrics and is also the most accurate model.