**REPORT**

* **Model(s) Chosen**

Firstly, considering the choice of a categorical values as the dependent variable to be predicted, it only makes sense to use classifier models instead of regression algorithms (either linear or logistic). Secondly, K-nearest neighbors is also avoided due to the fact the values for the dependent variable are not binary but rather seven-fold. (Seven videogame titles: Overwatch, Diablo, Hearthstone, The Lost Vikings, Warcraft, StarCraft, and Heroes of the Storm.)

Thusly, we are left with either going with Support Vector Machine or Random Forest models to go with. We shall report the performance of both and select which one is the better model in terms of prediction in the section below.

* **Performance: SVM vs. Random Forest**

We get the following categorical values for games by using *pandas.factorize*:

Our parameter for gauging model performance will be the Accuracy Scores generated from the confusion matrices:

Thus, the greater the score, the more our model is correctly predicting the values for game for the observations within the ‘training’ sample. From the results we can see that the Random Forest is the better model. That may be due to the inability of SVM to properly divide such a sample that’s heavily overlapped, as it possible through see through the scatter plot below. (Analysing ‘personality1’ and ‘personality2’ with regards to the ‘game\_categorical’.)

Chart, scatter chart

Description automatically generated

Meanwhile, Random Forest, sets Decision Trees to more accurately predict the favorite games of the observations.