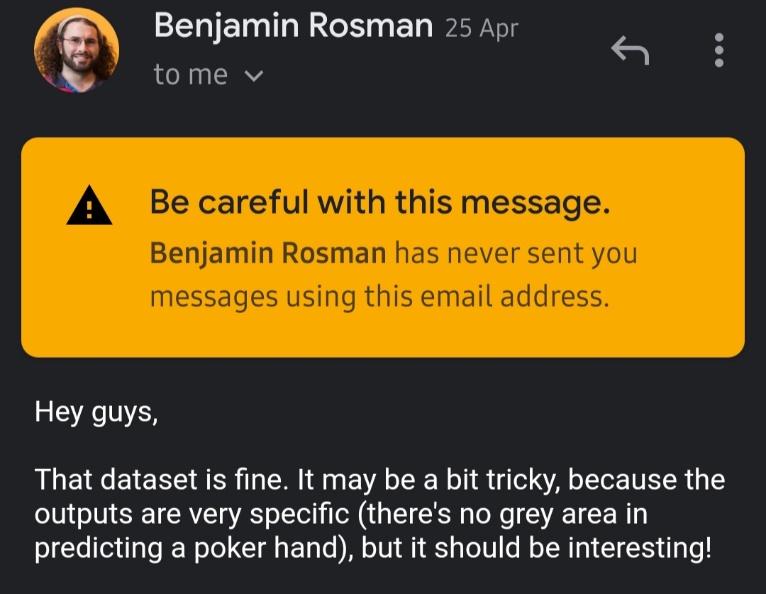
Poker dataset Decision tree report

1. Boilerplate dataset description, with all the attributes and the class of the poker hand… note that selection was called interesting in an email by Benjamin Rosman (email screenshot should be attached)
2. For the decision tree after the standard processing which included reading the data from the text files (using data reading helper classes) then splitting it along the lines of attributes and results. The dataset came with a prepared separation of data into testing and training sets, which were 25010 and 1 000 000 respectively, we then further split the 1 000 000 testing into 500 000 validation and 500 000 blind testing sets. Given that the data was given with raw attributes about suit and rank it was decided that more explanatory attributes would be needed in order to classify the given poker hands quicker. So the training data set was put through an algorithm which determined if the following could be found in the collection of ten attributes: consecutive values (if all the cards were consecutive regardless of order, for detecting straights), high ranks (if all the cards were royal in rank to better determine royal flushes), any number of duplicates (for finding pairs, two pairs, three of a kinds, and four of a kinds), and if the cards had the same suit (in order to find flushes). These new attributes were then used for the building of the tree and the inference of new data points.
3. Naïve Bayes, Decision Tree, and Neural Network

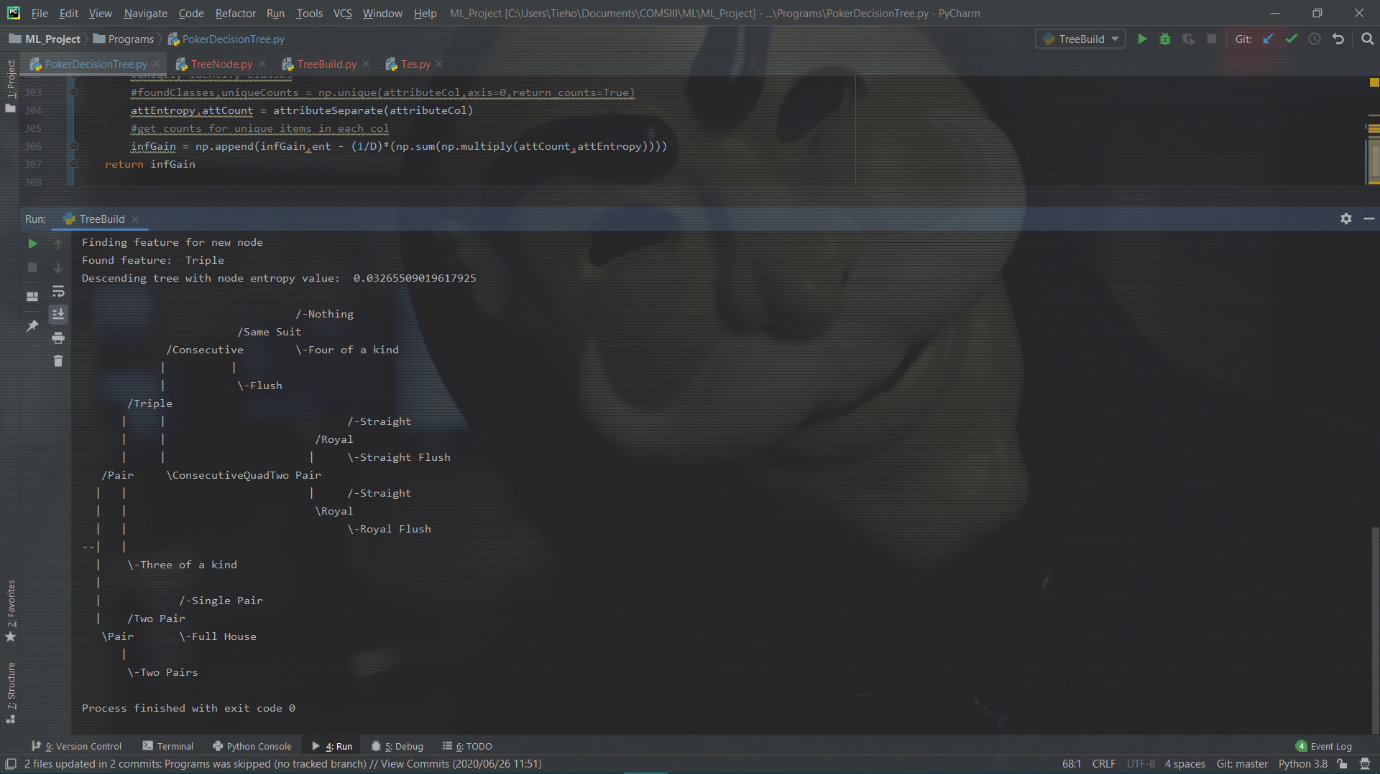


Figure : Tree drawn by algorithm

Decision Tree – Have yet to test the proper test set, still stuck on 0% for the validation set because my inference is trash. I don’t want to die; I don’t want that at all. Given that the decision tree was purely classification hyperparameters were not required. And this is the section where you want the confusion matrices. I see that, I see you and I curse Corona with my every breath. Also if I get a confusion matrix in the next… 10 minutes I’ll throw in some other equations.

1. Since each group member handled a single methodology no one member can speak to how well they compare. This section will be about the decision trees. The decision tree presents as the natural solution for classifying a poker hand as one can simply check if they have what is necessary to complete a hand and move down the ranking as they do not. The implementation, however, proved difficult. The treenode structure which simplified the drawing of the tree made passing of data rather tedious and led to certain machine precision errors that lead to higher confusion from the algorithm. The highest recorded scores are coming from the neural network on some 49% tip.