WHY I CHOSE THE MODEL?

The model I chose to solve this challenge was the AdaBoost model. There were several reasons for this.

The quality of the data is not very good. This meant that Neural Networks would not do well here. My suspicions were confirmed later what a Simple Feed Forward Neural Network attained an accuracy of about 38 to 40%. Neural Networks only do well if it’s an image/video classification problem or if the data is ideal. The analogy

I always use to explain this is , “Imagine if a NN is being used to predict the winner of the horse race, horse 1 has very high odds, horse 2 has medium odds and horse 3 has low odds. Say in the race horse 1 falls and horse 2 trips on horse 1 and falls as well and horse 3 wins the NN would try and explain why horse 3 won instead of understanding what happened in the race.”

Based on the data I had the intuitive feeling that this was going to be a tree based model. After some experimenting with Decision Trees, Random Forests and AdaBoost models, I concluded that AdaBoost was the best model.