After I have imported the data set, I checked for if any missing values exist. As there are no missing values, I proceeded with dropping irrelevant features related to our decision variable (preferred position) such as name, nationality etc.

As I have only used Random Forest and Light GBM, I didn’t standardize or normalized the features. And then I used encoding as it is required.

I decided to use GridSearchCV (grid search cross validation) to find the accuracy values of Random Forest and Light GBM without feature selection.

I continued with finding the feature importance values acquired. While I could find the standardized feature values of Random Forest without any extra computation, I had to calculate the standardized feature values of Light GBM with extra calculations. I used RFE (Recursive Feature elimination) with number of 15 features both with Random Forest and Light GBM.

Finally, I have shown all of accuracy scores of each model with different feature selection methods with the original model. I realized that while feature elimination was beneficial for Random Forest model, it has a negative impact on the prediction capability of Light GBM.