Models:

Random Forest/Decision Tress

For file names prefix is rf/dt for random forest and decision trees

Three approaches:

1. Grid search

   For hyper parameters values for which I have done grid search, see dt_grid_search.py file. I performed the calculations in another HPC cluster.

   Using the saved model, I performed the evaluation on test data

   See note book dt_best_params_from_grid_search

   Metric figures ars are there in each folder.

2. DT/RF classifier with reduced feature space: I used Recursive Feature Elimination (RFE) feature selection method.

   Only used RFE feature elimintions

   Based on XGBbooster, feature selection, I used 15 features, dropping 4 features.

features_dropped: {'timeSincePP', 'friendlyPlayersOnIce', 'rebound', 'opposingPlayersOnIce'}

   Retrained the model with the best parameters from saved model from grid search

   See notebook for more details and figures dt_feature_selection_rfe. Figures are saved in t he foler.

3. DT/RF classifier after PCA Feature reduction
   Performed PCA analysis to see n_component variable in PCA. Found 10 principal compoanents account for 95% variance. See notebook pca_principal_compenent_analysis for details and figures for this step

   DT/RF training on train data with reduced feature spaces after pca dimenationaly reduction with 10 components
   See notebook dt_pca_feature_reduction for details. Figures are saved.

Gaussian Naïve Bayes Classifier

Model on Gaussian Navie bayes classifier. See notebook Gaussian Naive Bayes classifier
No time No different approaches !!!!!

Also, dt and gaussian NB were registered in comet.

Due to large pickle size file (~1.8 GB), RF model could not registered in the comet.

All Comet entries are listed in the panels. Non important ones or test experiments moved to archives.