**Approach**

**Feature Engineering**

1. Checked Correlation values of independent variables and found out some features are highly multicollinear and dropped few multicollinear features. Eg – Player height and player
2. Checked the missing value counts. Found out few features had no records and about 20% features had more than 15% missing values.
3. Dropped Features having more than 13% missing values. As statistically it will not make sense to perform imputation.
4. Encoded two categorical features – winner and team by labelling.
5. Segregated the categorical and numerical features and performed imputations with mode and median for categorical and numerical features respectively.
6. Checked for outliers – but outliers’ values are acceptable. Therefore, no outlier treatment was performed.
7. Found out missing values in player position 1 and position 2 occurring simultaneously. Imputed them with mode.

**Feature Selection**

1. Feature Selection was based on Recursive feature Elimination method. Obtained ranking of all the features according to their importances.

**Modelling**

1. As from the kdeplot of target feature ‘rating num’, normality was not observed and the distribution was left skewed, applied log transformation which did not help. Therefore, did not go for linear models.
2. Used RandomForestRegressor as the base model. Obtained R2score as 0.16 on the test dataset. Used Cross validation with 5 folds but average score obtained was 0.17.
3. Tried different models like Support Vector Regressor, MLPRegressor, AdaBoost, Catboost Regressor, Gradient Boost Regressor, Decision Tree Regressor, XgBoost Regressor.
4. Only XgBoost regressor was performing better compared to other models. Applied Hyperparameter tuning using GridSearchCV. Used the best parameters to fit the model and obtained the best model result.
5. Used Ensemble techniques like averaging the models and stacking to improve accuracy.