**IPL 2020 Players Runs prediction.**

After observing the given data, I found that some of the values where of String Datatype. So I converted it into numerical Values.

Next, I found that Average Runs column was empty for some values, I found that It cannot be Imputed by using any statistical analysis as, In Cricket Avg Runs = Total Run / (Number of times OUT) => inn - number of times not out i.e. Zero. Which means not defined. So, it has to be filled with 0 only.

Next, plotted Heatmap for correlation between every features.

Next, I trained the dataset using multiple Regressions Model like: Linear Regression, K-Neighbors Regressor, Decision Tree Regressor, Random Forest Regressor and SVR. And found Random Forest Regressor with n\_estimators = 70 performing the best among all.

Random Forest Regressor:

Random forest is a **Supervised Learning algorithm** which uses ensemble learning method for **classification and regression**.

It operates by constructing a multitude of decision trees at training time and outputting the class that is the **mode** of the **classes (classification)** or **mean prediction (regression)** of the individual trees.

A random forest is a meta-estimator (i.e. it combines the result of multiple predictions) which **aggregates many decision trees**, with some helpful modifications:

The number of features that can be split on at each node is limited to some percentage of the total (which is known as the **hyperparameter**). This ensures that the ensemble model **does not rely too heavily on any individual feature**, and makes **fair use of all potentially predictive features**.

Each tree draws a random sample from the original data set when generating its splits, adding a further element of randomness that prevents **overfitting**.

**Feature and Advantages of Random Forest:**

* It is one of the most accurate learning algorithms available. For many data sets, it produces a **highly accurate classifier**.
* It runs efficiently on large databases.
* It can **handle thousands of input variables** without variable deletion.
* It gives estimates of what variables that are important in the classification.
* It generates an internal **unbiased estimate of the generalization error** as the forest building progresses.