**Solution Sheet**

1. Which model have you used for Total IPL 2020 Runs prediction for each player? Explain your model.

* After doing analysis on the data like OLS and RFE.
* The final two models which gave the best metrics are Random Forest and XGBoost with the r2\_score of 99.55 and 99.54 , mean\_absolute\_error of 6.74 and 7.68 ,and mean\_squared\_error of 115.11 and 118.25
* we decided to ensemble the two models, on the other hand simple models like Linear regression also gave 97.71 on the dev\_data.
* We believe the model can also be improved with more feature engineering techniques like using ratio of inns/Match and cross features.
* Additionally Data augmentation could also be done to improve the results.

Thus we have used an ensemble of two models (Random forest and XGBoost) .