**Milestone 2 Report**

* **Graphs:**

**FeatureSelection:**

We used correlation to filter the data and get the top features that affect more than 50% on the value so it became 23 only from 91 column on the contrary of the feature selection process for linear Regression models which generated only 8 features from 91 due to the change of y feature as it was “value” and now “PlayerLevel” .

**Hyperparameter tuning:**

**Logistic Regression** : solver = “saga“🡪 saga solver gave the best accuracy for this model because it’s suitable for multinomial logistic regression and it’s suitable also for very large data .

**SVM** : kernel = “poly” 🡪 We tried to use linear, sigmoid kernel functions and poly kernel function gave the best accuracy than the other kernel functions as it helps in representing the similarity in feature space allowing learning of non linear models .

**Decision Tree** : max-depth = “10” 🡪 We tried from 1 – 11 and number 10 gave the best result as this determines the maximum depth of the tree .

**Conclusion :**

The Model with best Accuracy is SVM with kernel “Poly” but it takes long time because of huge dataset Logistic Regression and Decision tree takes much less time but with less accuracy