# Weekly Project Meeting Minutes

*The main purpose of the document is to capture all the work that has been done by the group over the course of one week and* ***not*** *to write down what was discussed in a single meeting. You should be meeting and/or working throughout the week*.

Time group spent on project: \_\_\_\_\_10 hours\_\_\_\_\_\_\_\_\_\_

Group Number: 01

Group members present (Name, ID):

* Jasleen (0734327)
* Arvind Sharma (0730475)
* Harpreet Kaur (0733894)

Specific Activities from prior week:

* List brief description of activities carried out **by group member**

In this week we worked on Jupyter file as well as Report file. In the Jupyter file we are doing the models for getting the best accuracy and in report file we are working on literature reviews and other parts of the report.

Specific Output from prior week:

* Include brief summary of any written work, experiments, or code developed

In this we divided the Jupyter notebook to all of the three members so, everyone has their individual tasks in python file.

Arvind did the Random Forest Model and check for the accuracy and he also perform the feature selection.

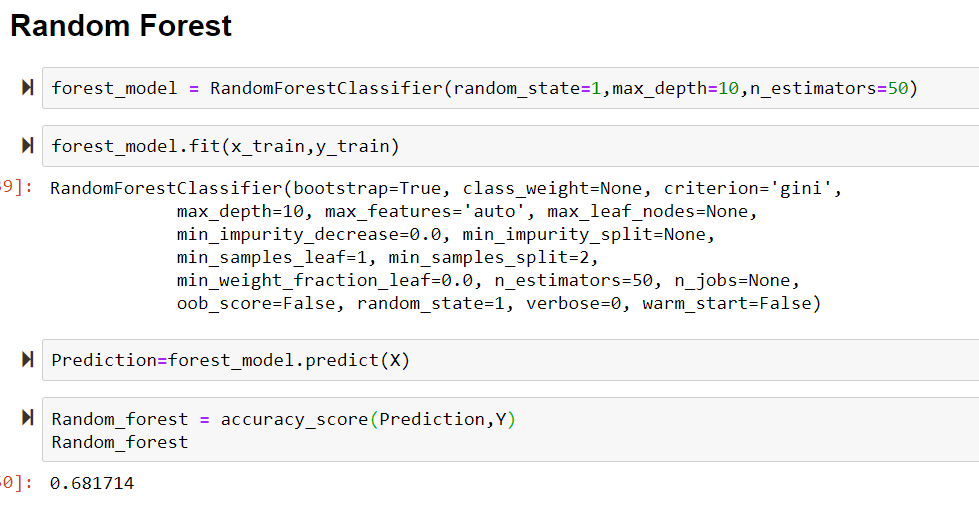
Harpreet did the Iteration part of all the three models to recheck the accuracies after the feature selection.

Jasleen did the Cat-Boosting algorithm to check the accuracy by selecting Root Mean Square Method as loss function.

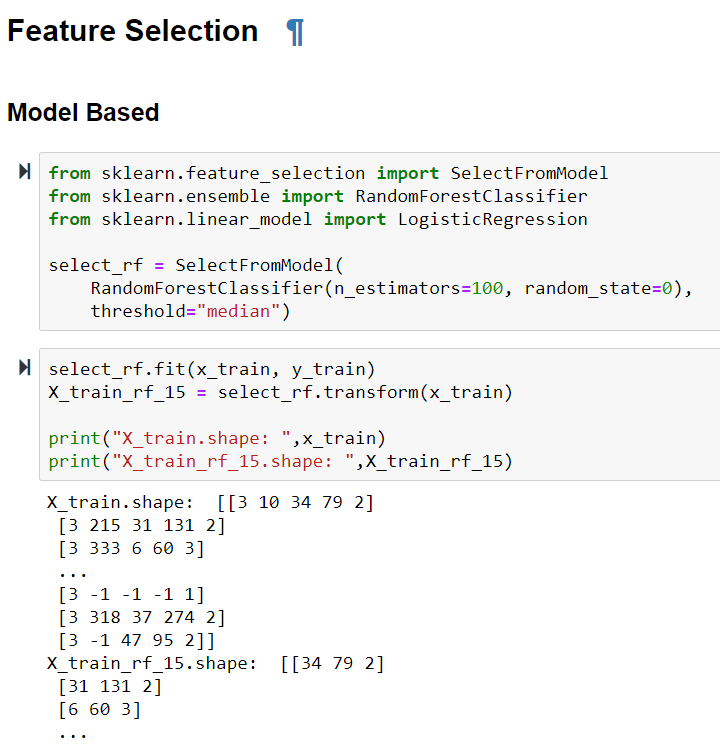
* Attach actual output as a separate file when submitting minutes; for example, export your Jupiter notebook as an html file and upload that with your minutes.

**Model Building: -**

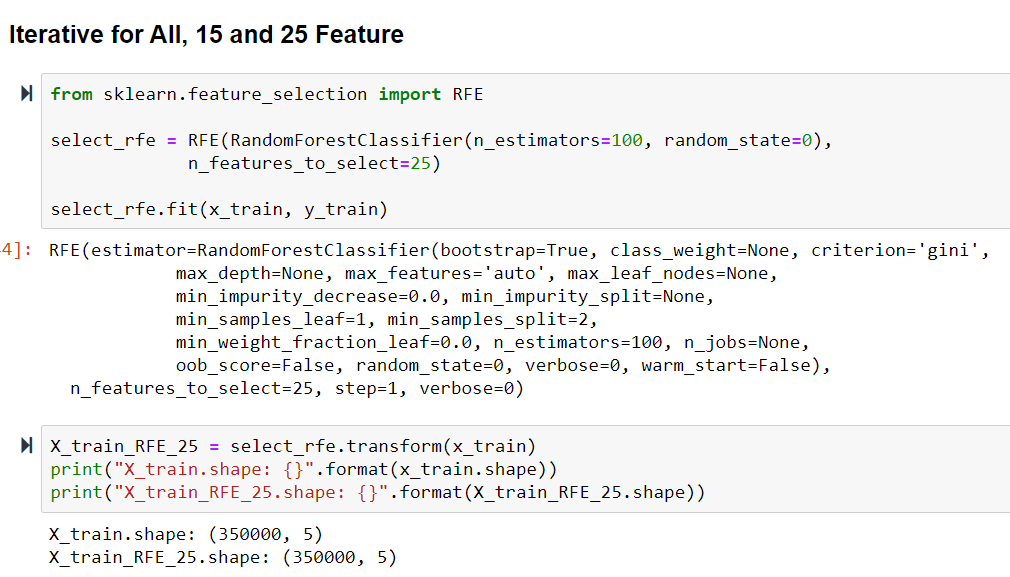
* This time we perform Random Forest Classifier which has higher accuracy from all the models that we previously performed on our dataset but there is not a big difference between the accuracies of these models as Logistic Regression with accuracy 0.66 and Decision Tree with 0.67 so the difference is minor. We choose this model because the accuracies we got from previous models is unsatisfied.



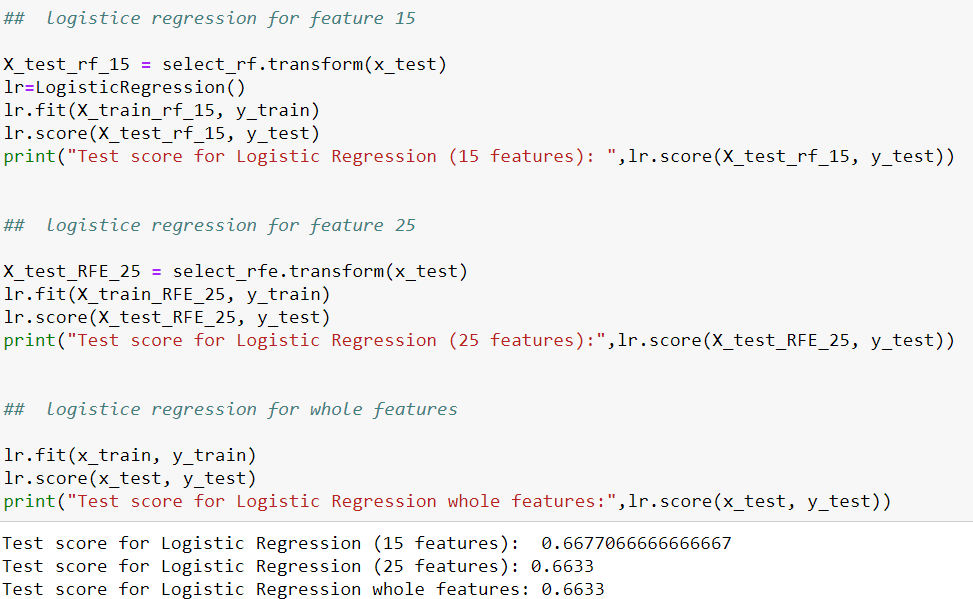
* Then we perform Feature Selection as we know feature selection play an important role in the performance of your model which selects the features which provide most contribution to your output.



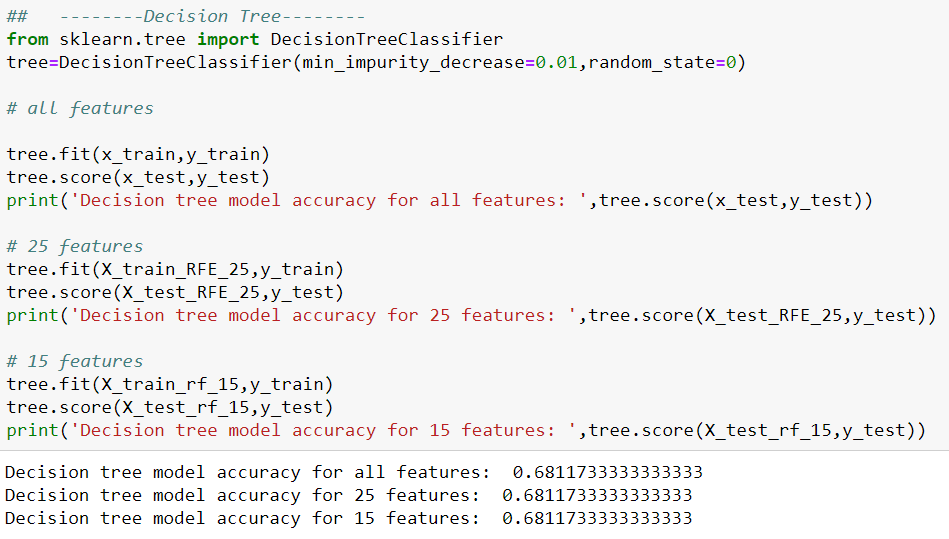
* Next we perform iterative for features in three categories that is by considering all features, 15 features and 25 features. We tried iterative because after doing feature selection we are expecting the change in the accuracies as feature selection always help in improving the output of model.



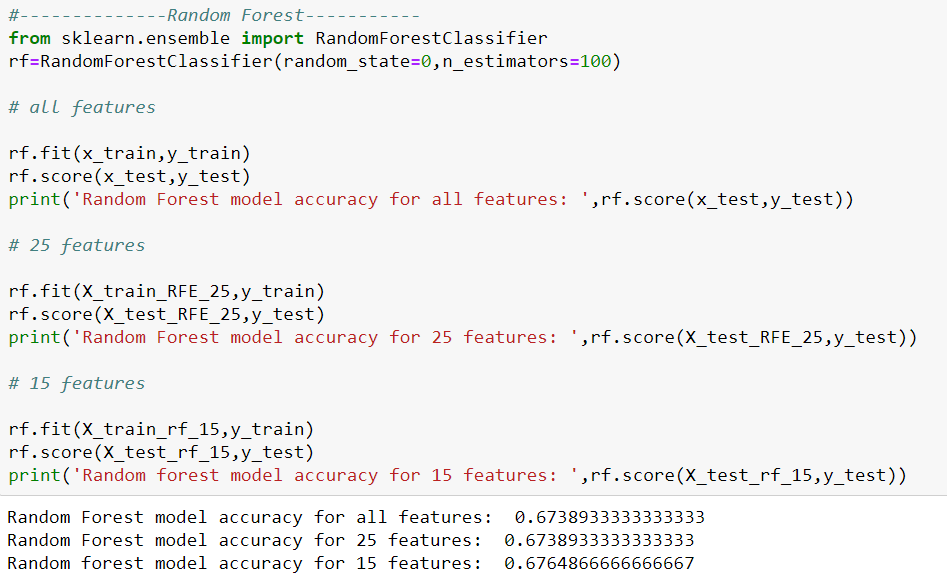
* **Iterative for Logistic Regression:**



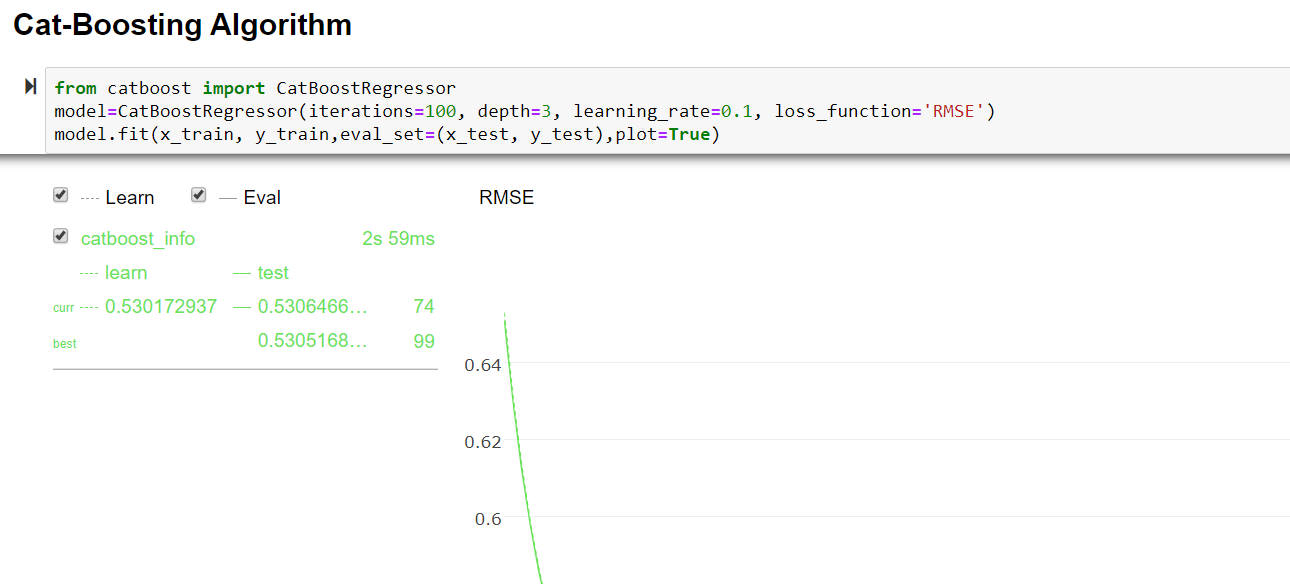
* **Iterative for Decision Tree:**

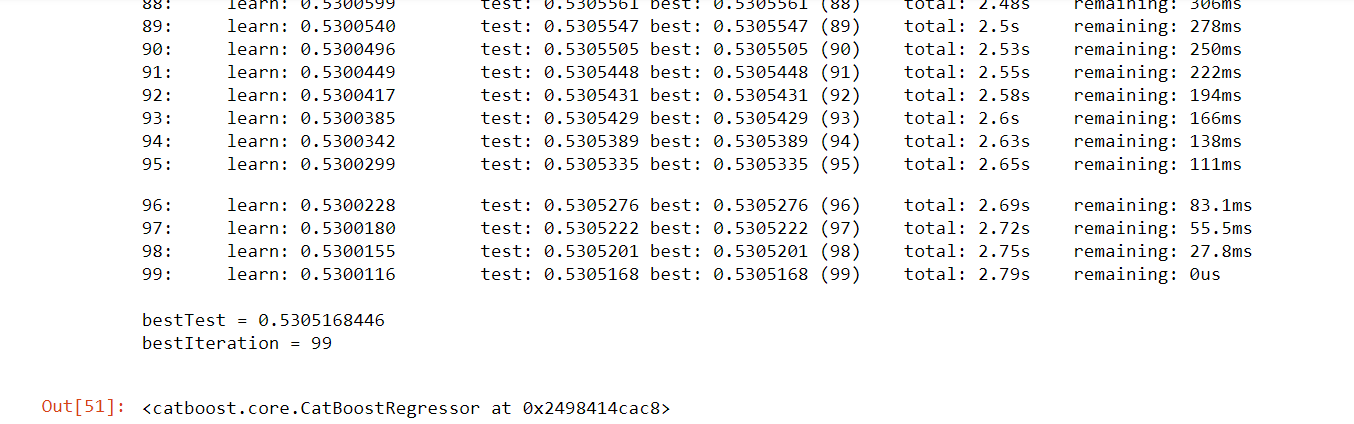


* **Iterative for Random Forest:**



* So, we can clearly see the changes in the accuracies i.e. Decision Tree has higher accuracy than the Random Forest after the Feature Selection which is 0.681 for Decision Tree and 0.676 for Random Forest.
* Then we perform Cat Boosting i.e. a gradient boosting algorithm on Decision trees. After running this model, we get 0.53 as test accuracy which is very less as compared to other models.





On Target:

* Indicate the current status of your project
  + \_\_\_\_\_ green: everything on track for completion by due date



* + \_\_\_\_\_ yellow: a small number of tasks are off track and completion by due date is at risk
  + \_\_\_\_\_ red: many tasks are off track and project will not be completed by due date

Challenges/Disagreements:

* List any particular challenges identified/discussed and possible solutions

No, there is challenge or disagreement this week.

Planned Activities for coming week:

* List brief description of activities **by group member**

For next week we finalize to perform another model to see if they are having better accuracy then these and we will try to complete the report writing.