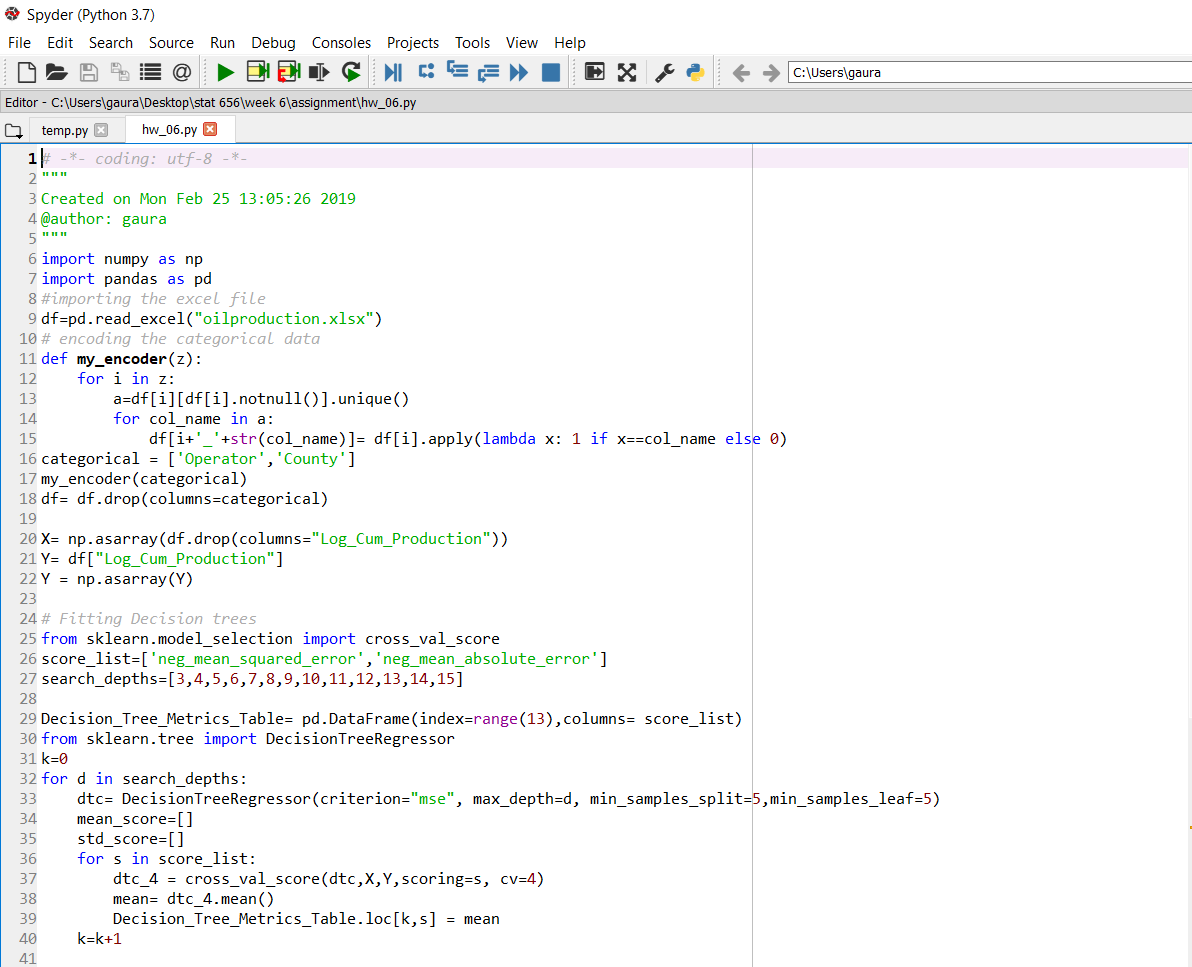
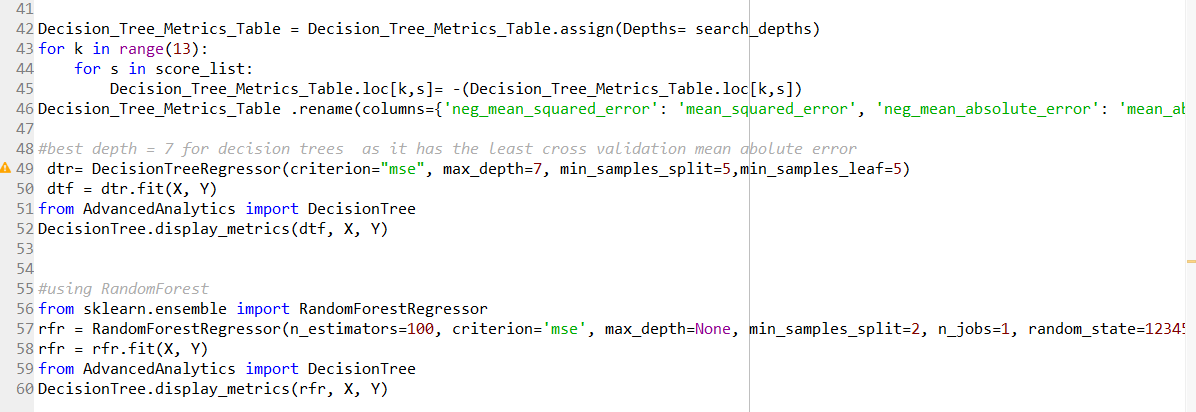
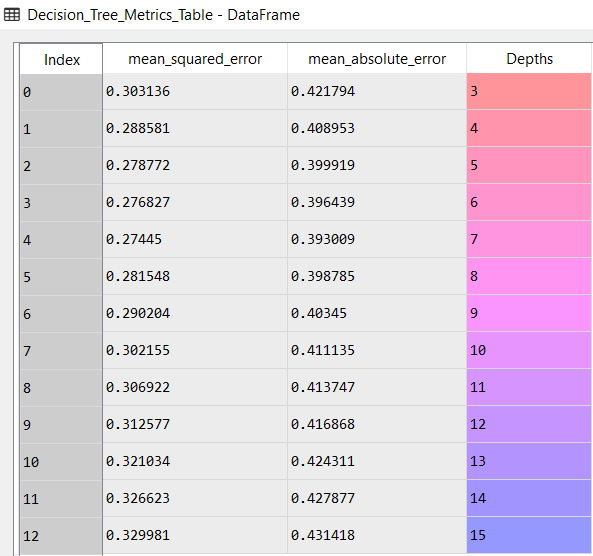
**Oil Production Prediction**

1. Python code



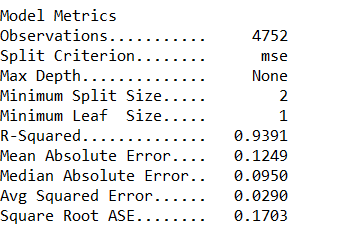
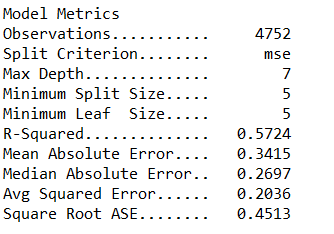


1. Table for metrics(Mean Squared Error and Mean Absolute Error for all depths for 4 fold cross validation.



1. Depth =7 was selected as the best depth as it has the lowest cross validation MSE.
2. Comparison between the model metrics for decision tree with depth =7 vs the random forest.

**Decision Tree(depth=7)** **Random Forest**

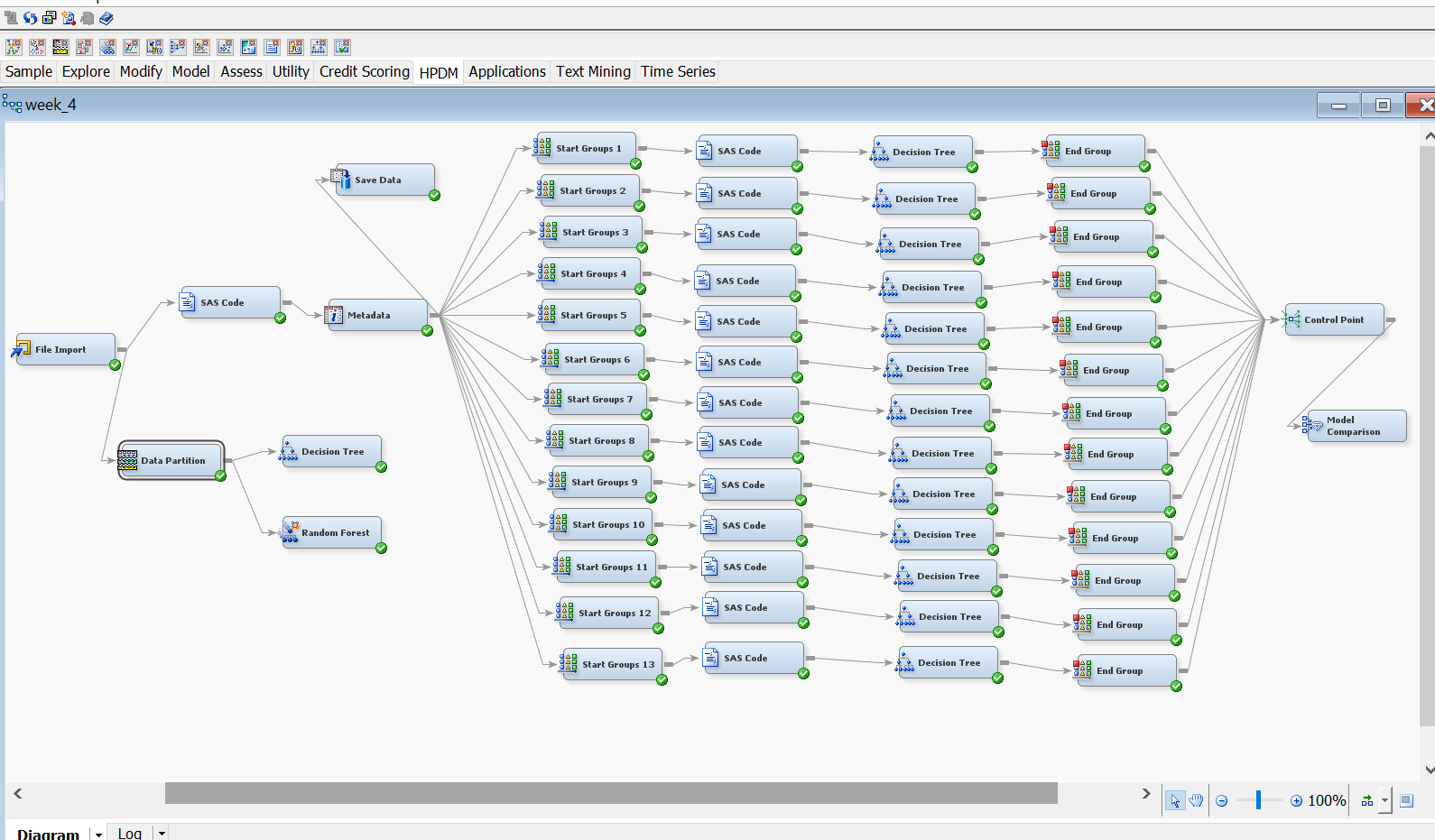


Random Forest is performing much better than a decision tree selected using cross validation.

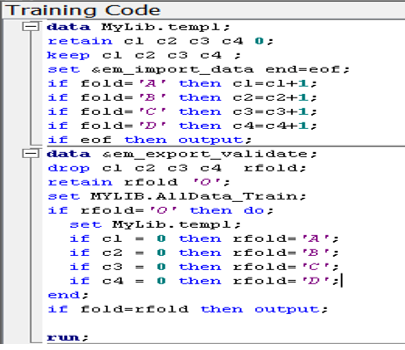
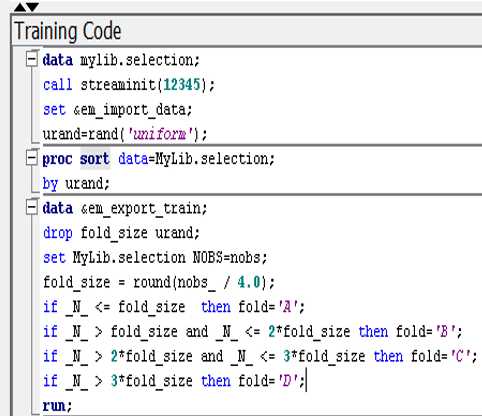
Random Forest’s Mean Squared error is almost 1/10 th of that given by the Decision Tree

**SAS**

1. Project Diagram



1. SAS Codes for Cross Validation and Folds Creation



1. Cross Validation on Decision Trees for different depths

Depth = 3

A screenshot of a cell phone

Description automatically generated

Depth = 4

A screenshot of text

Description automatically generated

Depth = 5

A screenshot of a cell phone

Description automatically generated

Depth = 6

A screenshot of a cell phone

Description automatically generated

Depth = 7

A screenshot of a cell phone

Description automatically generated

Depth = 8

A screenshot of a cell phone

Description automatically generated

Depth = 9

A screenshot of text

Description automatically generated

Depth = 10

A screenshot of a cell phone screen with text

Description automatically generated

Depth = 11

A screenshot of a cell phone

Description automatically generated

Depth = 12

A screenshot of a cell phone

Description automatically generated

Depth = 13

A screenshot of text

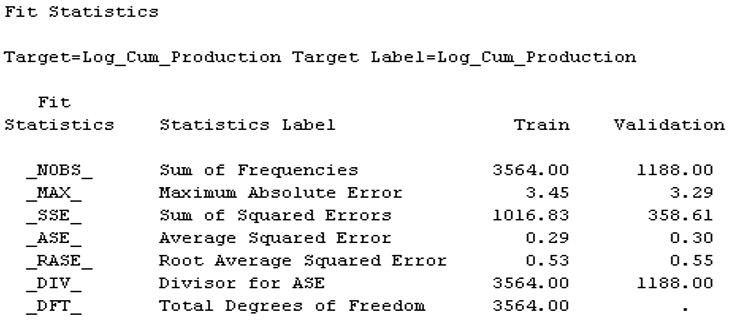
Description automatically generated

Depth = 14

A screenshot of text

Description automatically generated

Depth = 15



1. Best Depth = 8 for the decision tree as given by cross validation
2. Metrics for best Decision tree

A screenshot of text

Description automatically generated

1. Metrics for Random Forest

A screenshot of a cell phone

Description automatically generated

Random Forest is giving a lower Mean Squared error as compared to the MSE given by decision tree with depth chosen through cross validation.