COS40007 Artificial Intelligence for Engineering Portfolio Assessment-3: “Let’s develop AI model by your own decision

Student name: Vivek Saini

Student ID: 103828056

Studio 1-3

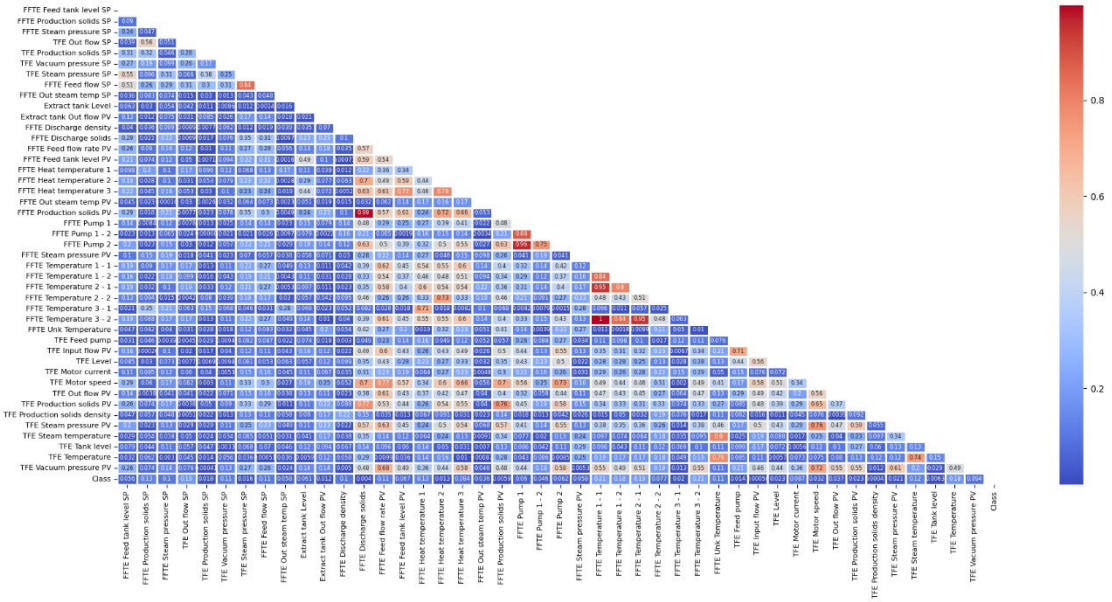
Introduction:

In this class, we have learned about different types of ML models that we have to develop by exploring data and comparing different models with the help of hyperparameter tuning and feature selection. I have applied most of the concepts from the previous 3 weeks portfolio and also I have done studio 4 to gather some more information to top it off to complete this assignment. I have used the given vegemite dataset to conduct the activities in this portfolio.

Step 1 : Data Preparation

1) The dataset contains 2 columns with constant values and they have been removed accordingly.
2) The “Class” column has few integer value but it is already in categorical value. So, I don’t have to make any changes to the file.
3) The class does not have a balanced distribution, to make some changes I have used the SMOTE function from the studio to oversample the minority classes and adjust the weights accordingly.

4) In this step, I have had to take the example from week 2 studio and use the correlation heatmap to find any covariance between the features that are already given. In the picture I have chosen all the columns with 0.85 mark and multiplied to make some composite features and get rid of any additional features. Here is the screenshot of the heatmap.
I had selected these features because there was either a lot of covariance or none at all. That’s why I put a mark above 0.85 so that we can make composite features of those affecting it the most.



5) Finally, I have 50 features in total for the dataset at the moment.

Step 2: Feature selection, Model training and Evaluation.

6) Like we did in one of the previous studios, we can apply feature selection to remove the original features of those which were used to make the composite feature, so in may case the composite features are 'FFTE Production solids PV * FFTE Discharge solids', 'FFTE Pump 2 * FFTE Pump 1', 'FFTE Temperature 2 - 1 * FFTE Temperature 1 - 1', 'FFTE Temperature 3 - 2 + FFTE Temperature 1 - 1' and 'FFTE Temperature 3 - 2 + FFTE Temperature 2 - 1', and 'FFTE Pump 1 - 2 * FFTE Pump 1'. We will add these and remove the individual columns. Again, I used this criteria to take columns above 0.85 because these correlate the most.
7) I had to train all the models that we learned in the past week to get the results and those ML models are : 'Decision Tree', 'Random Forest', 'SVM', 'MLP Classifier' and 'SGD'.

Model	Accuracy	Precision (Avg)	Recall (Avg)	F1-Score (Avg)	Confusion Matrix Deductions
Decision Tree	0.97	0.97	0.97	0.97	Very high accuracy
MLP Classifier	0.60	0.60	0.46	0.37	Average accuracy
Random Forest	0.99	0.99	0.99	0.99	Almost perfect with high accuracy
SVM	0.52	0.52	0.47	0.43	Not accurate
SGD	0.55	0.55	0.44	0.36	Not accurate

8) The table above is the deductions from the classification reports and confusion matrix.

10) I have selected Random Forest as the best model because it has given me the highest accuracy.

11) With the use of pickle library , I have successfully saved the model.

Step 3 : ML to AI

In this step I have basically followed the steps to perform the tasks and complete the specified requirements.

16) I converted the original 1000 values that were aside and converted the features to match the features with the modified csv file and run the model and not surprised I get a 100% accurate results by using those data points. With the classification report I deducted that all the accuracy comes to 1.00 which means 100% accuracy.

17) Here is the comparison table with the other models as well

Model	Accuracy	Precision (Avg)	Recall (Avg)	F1-Score (Avg)	Confusion Matrix Deductions
Decision Tree	0.99	0.99	0.99	0.99	Very high accuracy
MLP Classifier	0.66	0.66	0.47	0.37	Average accuracy
Random Forest	1.00	1.00	1.00	1.00	perfect with high accuracy
SVM	0.53	0.53	0.49	0.44	Not accurate
SGD	0.62	0.62	0.47	0.38	Not accurate

I would say I have observed the same result for these models with the big dataset but we get to see more variation in the SVM and SGD models in the smaller dataset.

Step 4 :

I have followed all the steps and made the decision tree. The file seems to be very big so I will only print the first bit of it and attach a separate file in my github repository. You can also have a look at that its called ‘decisontreeprint.txt’

Decision Tree Rules based on SP features:

|--- FFTE Feed flow SP <= 10199.77

| |--- FFTE Steam pressure SP <= 122.24

| | |--- FFTE Feed flow SP <= 9300.51

| | | |--- TFE Out flow SP <= 2177.59

| | | | |--- FFTE Steam pressure SP <= 100.96

| | | | | |--- FFTE Steam pressure SP <= 76.21

| | | | | | |--- class: 2

| | | | | |--- FFTE Steam pressure SP > 76.21

| | | | | | |--- TFE Production solids SP <= 67.50

| | | | | | | |--- TFE Out flow SP <= 2165.16

| | | | | | | | |--- FFTE Feed flow SP <= 8965.00

| | | | | | | | | |--- FFTE Steam pressure SP <= 85.00

| | | | | | | | | | |--- class: 1

| | | | | | | | | |--- FFTE Steam pressure SP > 85.00

| | | | | | | | | | |--- class: 2

| | | | | | | | |--- FFTE Feed flow SP > 8965.00

| | | | | | | | | |--- TFE Out flow SP <= 1990.72

| | | | | | | | | | |--- TFE Vacuum pressure SP <= -40.68

| | | | | | | | | | | |--- truncated branch of depth 6

| | | | | | | | | |--- TFE Vacuum pressure SP > -40.68

| | | | | | | | | | |--- class: 2

| | | | | | | | |--- TFE Out flow SP > 1990.72

| | | | | | | | | |--- class: 1

| | | | | | | |--- TFE Out flow SP > 2165.16

| | | | | | | | |--- class: 2

| | | | | | |--- TFE Production solids SP > 67.50

| | | | | | | |--- TFE Production solids SP <= 69.28

| | | | | | | | |--- FFTE Steam pressure SP <= 98.45

| | | | | | | | | |--- class: 0

| | | | | | | | |--- FFTE Steam pressure SP > 98.45

| | | | | | | | | |--- class: 1

| | | | | | | |--- TFE Production solids SP > 69.28

| | | | | | | | |--- class: 2

| | | | |--- FFTE Steam pressure SP > 100.96

| | | | | |--- TFE Out flow SP <= 2119.80

| | | | | | |--- FFTE Feed flow SP <= 9200.56

| | | | | | | |--- FFTE Production solids SP <= 40.67

| | | | | | | | |--- TFE Steam pressure SP <= 122.50

| | | | | | | | | |--- class: 2

| | | | | | | | |--- TFE Steam pressure SP > 122.50

| | | | | | | | | |--- class: 1

| | | | | | | |--- FFTE Production solids SP > 40.67

| | | | | | | | |--- TFE Out flow SP <= 1902.35

| | | | | | | | | |--- TFE Production solids SP <= 64.00

| | | | | | | | | | |--- class: 2

| | | | | | | | | |--- TFE Production solids SP > 64.00

| | | | | | | | | | |--- TFE Production solids SP <= 68.00

| | | | | | | | | | | |--- truncated branch of depth 12

| | | | | | | | | | |--- TFE Production solids SP > 68.00

| | | | | | | | | | | |--- truncated branch of depth 7

| | | | | | | | |--- TFE Out flow SP > 1902.35

| | | | | | | | | |--- FFTE Steam pressure SP <= 101.51

| | | | | | | | | | |--- class: 0

| | | | | | | | | |--- FFTE Steam pressure SP > 101.51

| | | | | | | | | | |--- FFTE Steam pressure SP <= 109.76

| | | | | | | | | | | |--- class: 1

| | | | | | | | | | |--- FFTE Steam pressure SP > 109.76

| | | | | | | | | | | |--- truncated branch of depth 11

| | | | | | |--- FFTE Feed flow SP > 9200.56

| | | | | | | |--- TFE Out flow SP <= 2067.64

| | | | | | | | |--- TFE Steam pressure SP <= 120.01

| | | | | | | | | |--- TFE Out flow SP <= 1749.54

| | | | | | | | | | |--- class: 1

| | | | | | | | | |--- TFE Out flow SP > 1749.54

| | | | | | | | | | |--- TFE Production solids SP <= 66.86

| | | | | | | | | | | |--- truncated branch of depth 14

| | | | | | | | | | |--- TFE Production solids SP > 66.86

| | | | | | | | | | | |--- truncated branch of depth 7

| | | | | | | | |--- TFE Steam pressure SP > 120.01

| | | | | | | | | |--- TFE Production solids SP <= 65.96

| | | | | | | | | | |--- TFE Vacuum pressure SP <= -79.95

| | | | | | | | | | | |--- truncated branch of depth 4

| | | | | | | | | | |--- TFE Vacuum pressure SP > -79.95

| | | | | | | | | | | |--- truncated branch of depth 3

| | | | | | | | | |--- TFE Production solids SP > 65.96

| | | | | | | | | | |--- TFE Production solids SP <= 70.54

| | | | | | | | | | | |--- truncated branch of depth 2

| | | | | | | | | | |--- TFE Production solids SP > 70.54

| | | | | | | | | | | |--- truncated branch of depth 10

| | | | | | | |--- TFE Out flow SP > 2067.64

| | | | | | | | |--- TFE Production solids SP <= 49.25

| | | | | | | | | |--- FFTE Feed tank level SP <= 42.95

| | | | | | | | | | |--- class: 1

| | | | | | | | | |--- FFTE Feed tank level SP > 42.95

| | | | | | | | | | |--- FFTE Steam pressure SP <= 105.50

| | | | | | | | | | | |--- class: 1

| | | | | | | | | | |--- FFTE Steam pressure SP > 105.50

| | | | | | | | | | | |--- class: 0

| | | | | | | | |--- TFE Production solids SP > 49.25

| | | | | | | | | |--- FFTE Steam pressure SP <= 110.04

| | | | | | | | | | |--- TFE Vacuum pressure SP <= -51.44

| | | | | | | | | | | |--- truncated branch of depth 8

| | | | | | | | | | |--- TFE Vacuum pressure SP > -51.44

| | | | | | | | | | | |--- truncated branch of depth 2

| | | | | | | | | |--- FFTE Steam pressure SP > 110.04

| | | | | | | | | | |--- TFE Out flow SP <= 2111.03

| | | | | | | | | | | |--- class: 1

| | | | | | | | | | |--- TFE Out flow SP > 2111.03

| | | | | | | | | | | |--- truncated branch of depth 3

| | | | | |--- TFE Out flow SP > 2119.80

| | | | | | |--- FFTE Out steam temp SP <= 50.02

| | | | | | | |--- TFE Out flow SP <= 2125.70

| | | | | | | | |--- TFE Out flow SP <= 2124.81

| | | | | | | | | |--- class: 1

| | | | | | | | |--- TFE Out flow SP > 2124.81

| | | | | | | | | |--- class: 0

Appendix:

Link to my Github: https://github.com/Vikksaini/AI_for_engineering_portfolio