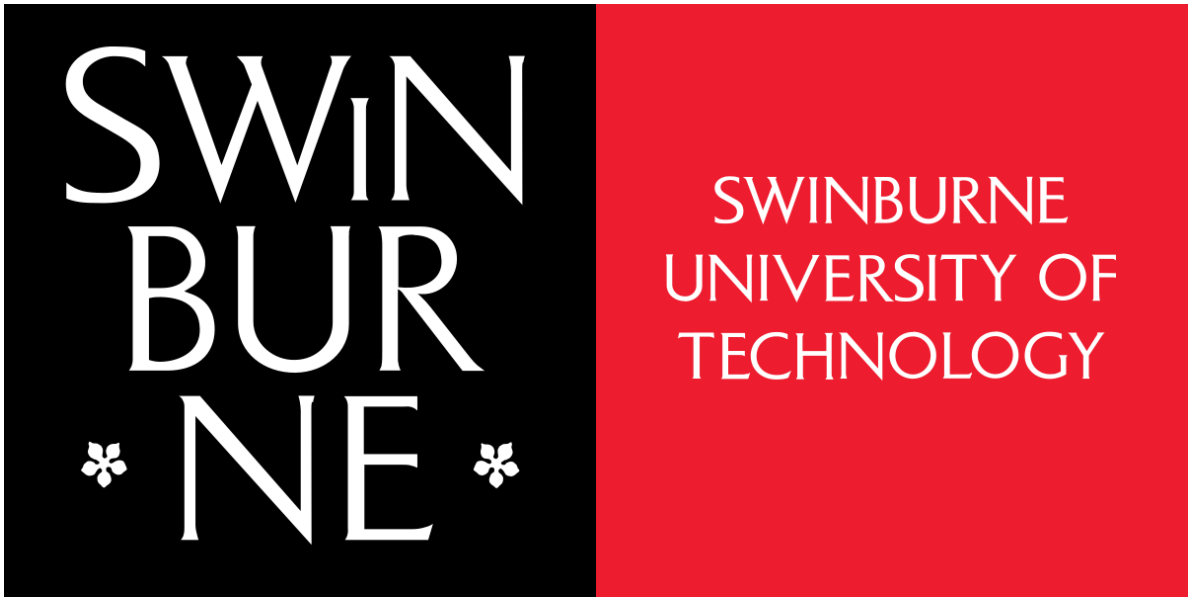
Course Name: Artificial Intelligence for Engineering (COS40007)

Studio Session: Studio 1 - 7

Studio Tutor: Irfan Mirza



Title: Week 4: Portfolio Assessment 3 - "Develop an AI model by your own decision"

Name: Ashraf Shahzad Toor

Student ID: 104586656

GitHub Code Link: https://github.com/AshrafToor/COS40007_AIE/tree/main/Assessment3

Submission Date: 06-04-2025

1) Does the dataset have any constant value column? If yes, then remove them

Two constant columns were removed: TFE Steam temperature SP TFE Product out temperature

2) Does the dataset have any column with few integer values? If yes, then convert them to a categorial feature.

Few-Value Columns: Columns with low unique values were marked as categorical:

FFTE Feed tank level SP, FFTE Pump 1, FFTE Pump 1 - 2, FFTE Pump 2, TFE Motor speed

3) Does the class have a balanced distribution? If not, then perform necessary undersampling and oversampling or adjust class weights.

Class Imbalance: Class distribution was imbalanced. Oversampling was done using RandomOverSampler to balance the classes.

4) Do you find any composite features through exploration? If so, then add some composite features to the dataset.

FFTE_Temperature_avg: Average of 3 heat temperature columns

TFE_Pressure_diff: Difference between steam and vacuum pressure in TFE

5) Finally, how many features do you have in your final dataset?

After removing constant columns and adding composites, the dataset had 46 features.

6) Does the training process need all features? If not, can you apply some feature selection techniques to remove some features? Justify your reason for feature selection.

A Random Forest was used to rank features. While not all features were essential, all 46 were retained to maintain maximum information.

7) Train multiple ML models (at least five, including DecisionTreeClassifier) with your selected features.

Five models were trained: Decision Tree Logistic Regression Random Forest Gradient Boosting Support Vector Machine (SVM)

8) Evaluate each model with a classification report and confusion matrix.

All models were evaluated using classification reports and confusion matrices.

9) Compare all the models across different evaluation measures and generate a comparison table.

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	1.00	1.00	1.00	1.00
Decision Tree	0.998	0.998	0.998	0.998
Gradient Boosting	0.938	0.935	0.935	0.935
SVM	0.570	0.498	0.424	0.498
Logistic Regression	0.505	0.511	0.504	0.511

10) Now select your best- performing model to use as AI. Justify the reason for your selection.

Random Forest was selected due to perfect evaluation metrics and generalization.

11) Now save your selected model

The selected model was saved using joblib.dump() to random_forest_model.pkl.

16) Measure the performance of your best model for 1000 unseen data points.

The model achieved perfect metrics:

Precision: 1.000

Recall: 1.000

F1-Score: 1.000

Accuracy: 1.000

17) Now, measure the performance of other models using these 1000 data points. Have you observed the same result of model selection that you identified through evaluation?

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	1.00	1.00	1.00	1.00
Decision Tree	0.998	0.998	0.998	0.998
Gradient Boosting	0.938	0.935	0.935	0.935
SVM	0.570	0.498	0.424	0.498
Logistic Regression	0.505	0.511	0.504	0.511

Outcome of Decision Tree: The decision tree learned interpretable rules based on SP features such as FFTE Feed flow SP, TFE Out flow SP, and FFTE Steam pressure SP. These were the most influential in predicting the class.

Sample Rules Derived from the Tree:

Class 1:

If FFTE Feed flow SP <= 10165

AND TFE Out flow SP <= 2249.11

AND FFTE Steam pressure SP <= 119.98

AND TFE Out flow SP <= 2154.76

AND FFTE Feed flow SP <= 9395

THEN: Class = 1

Class 2:

If FFTE Feed flow SP > 10165

AND TFE Production solids SP > 84

AND FFTE Feed tank level SP <= 37.5

AND FFTE Steam pressure SP <= 120.87

THEN: Class = 2

Class 0:

If TFE Out flow SP <= 2249.11

AND FFTE Steam pressure SP <= 119.98

AND TFE Out flow SP > 2154.76

THEN: Class = 0