**Artificial Intelligence of Engineering**

**COS 40007**

**Name: Mohammed Barsat Zulkarnine**

**Student Number: 103810626**

**Studio: 1 - 5**

Step 1: Data Preparation

Answer:

Class Distribution Before Oversampling:

Class 0: 2468 samples

Class 1: 4716 samples

Class 2: 7053 samples

Class Distribution After Oversampling:

Each class has been oversampled to 7053 samples to ensure balanced class distribution.

Data Cleansing:

Constant Value Columns: Any columns with constant values were removed.

Few Integer Values Columns: Columns with a limited range of integer values were converted to categorical features.

Class Imbalance: Applied oversampling using SMOTE to ensure balanced class distribution.

Composite Features:

Added composite features identified through data exploration.

Final Features Count: The final dataset contains [number of features] features after data preparation.

Source Code and Data: [Provide a link to your source code and data in a shared folder]

Step 2: Feature Selection, Model Training, and Evaluation

Answer:

Feature Selection:

Selected Features: 20 features were selected using Recursive Feature Elimination (RFE).

Reason for Selection: RFE was chosen to reduce dimensionality while maintaining the most predictive features for the model.

Model Training and Evaluation:

Trained five models including DecisionTreeClassifier, RandomForestClassifier, GradientBoostingClassifier, SVC, and LogisticRegression.

Evaluation Metrics:

Best Performing Model: RandomForestClassifier with the highest accuracy of 0.9999.

Comparison Table: [Include the table you printed in the output]

Selected Model:

Reason: RandomForestClassifier was chosen due to its high accuracy, precision, recall, and f1-score across all classes.

Model Saving:

The best-performing model (RandomForestClassifier) was saved using joblib.

Source Code and Data: [Provide a link to your source code and data in a shared folder]

Step 3: ML to AI

Answer:

Performance on Unseen Data Points:

The performance of the best model (RandomForestClassifier) on 1000 unseen data points was evaluated. The model achieved an accuracy of [insert accuracy] and an f1-score of [insert f1-score].

Performance of Other Models:

All models were evaluated on the unseen data points, and RandomForestClassifier remained the best performing model.

Source Code and Data: [Provide a link to your source code and data in a shared folder]

Step 4: Develop Rules from ML Model

Answer:

Decision Tree Rules: The decision tree was built using only SP features, and the rules were generated as follows:

Example Rules:

For Class 2: If TFE Production solids SP > 84.99 and FFTE Feed tank level SP <= 37.50 and FFTE Steam pressure SP <= 120.87, then classify as Class 2.

For Class 0: If TFE Production solids SP > 84.99 and FFTE Feed tank level SP <= 37.50 and FFTE Steam pressure SP > 120.87, then classify as Class 0.

For Class 1: If TFE Production solids SP > 84.99 and FFTE Feed tank level SP > 37.50 and FFTE Out steam temp SP > 52.04 and TFE Steam pressure SP > 115.00 and TFE Vacuum pressure SP <= -36.40, then classify as Class 1.

Source Code and Data: [Provide a link to your source code and data in a shared folder]