Welcome to ExxonMobil DataWorks Challenge

About the challenge

Topic: Classification Model for Well Slugging

Submission: 4th Nov 2022 (Friday) at 9.00am (GMT +8)

Folders:

1. Common - essential code that can be shared between all .ipynb files
2. Dataset - source of dataset for this project
3. ModelArchive - location to store the best model

Files:

1. `analytic_workbook.ipynb` - to put all related analytics results
2. `model_testing_workbook.ipynb` - to put best model testing results
3. `model_training_workbook.ipynb` - to put model training results
4. `Data Definitions.xlsx` - definitions about the data

DataSet:

1. `complete_well_data.csv` - both training and testing data, suitable for analytics
2. `test_well_data.csv` - test data, suitable for machine learning
3. `train_well_data.csv` - training data, suitable for machine learning

Recommended Library to be installed for workbook

1. `pip install pandas`
2. `pip install seaborn`

NOTE:

1. Teams can focus on either data exploration and/or machine learning modeling using their own tools and not limited to the folders/files here.
2. Teams are required to clean their own code and ensure readability before submission.
3. Please put related files and create a manual on

    1. How the results/findings is obtained
    2. Tools used

3. Special instruction needed to run your project
4. Etc

4. Slugging Class is defined by the following pseudocode where `fluctuationTreshold` is a tunable parameter

```
if (the value of "WHP Fluctuation (%)") <= fluctuationTreshold:
    then, SluggingClass = 'Non-Slugging'
else,
     SluggingClass = 'Slugging'
```

7. In `Preprocessing.py` there is a function `getSluggingClass(df, WHPFluctuationTreshold)`, the `WHPFluctuationTreshold` is a tunable parameter that should be optimised by the teams.
8. The column `"WHP Fluctuation (%)"` should not be used for the training or testing because it is used to derive the Slugging Class.
9. Please put your final `WHPFluctuationTreshold` inside `Common/Config.py` and your manual.
10. Do not clear your workbook results.

Submission:

1. 7 minutes video (.mp4) about your findings and/or modeling
2. Documentation (.pdf) on the following

    1. Setup instruction,
    2. Development process,
    3. Findings/visualization

3. Machine Learning model file (if any)
4. Provide any file that is used for development, analytics and testing for this challenge