**CRUDE OIL PREDICTION STATE(MACHINE LEARNING – ADVANCED – Sample PROJECT -1)**

**# \*\*1. Introduction\*\***

**### Company Introduction - Agip Oil Company Limited (AOCL)**

***Your client for this project is a is a Oil company - Agip Oil Company Limited (AOCL)***

- 1) They are one of the largest Oil Company. By 2017 they had daily production of 25.451 million BOE(barrels of oil equivalent).

- 2) In 2017, this was approximately 13% of world production, which is less than several of the largest state-owned petroleum companies.

- 3) They want to increase their production in order to compete with larger oil companies.

- 4) They want to automate the process of keeping track of the state and trend of production using previous Crude Oil Data.

- 5) Their research and development teams are trying to understand properties of the previous years Crude Oil data so that they can use it to increase their production.

***### Current Scenario:***

- Determining state of production as growth or decay, based upon historical crude oil data.

- The traditional crude oil production state prediction methods are based on statistical analysis, extracting effective information from historical data and making reasonable judgements for classification.

- Statistics based models use a sample of data to perform classification, so these cannot provide accurate results.

- However, designing a computer program to do this turns out to be a bit trickier.

**# \*\*2. Problem Statement\*\***

- This section is emphasised on providing some generic introduction to the problem that most companies confront.

### The company suffers from the following problems:

- The current process is a manual classification of production state using statistical methods.

- This is very tedious and time-consuming as it needs to be repeated for every new customer.

The company has hired you as data science consultants.

They want to automate the process of predicting the production state using properties of the crude oil trend rather than doing this manual work.

**### Your Role**

- You are given previous year Crude Oil Production data.

- Your task is to build a classification model using the dataset.

- Because there was no machine learning model for this problem in the company, you don’t have a quantifiable win condition. You need to build the best possible model.

**### Project Deliverable**

- Deliverable:\*\*CrudeOil Production will grow or not.\*\*

- Machine Learning Task: \*\*Classification\*\*

- Target Variable: \*\*activity\*\*

- Win Condition: N/A (best possible model)

**### Evaluation Metric:**

- The model evaluation will be based on the \*\*F1 Score\*\*.

***4. Data Description***

- This database was created to predict the Crude Oil production’s trend based on the previous year Crude Oil data.

- The column Target is also present in the dataset which has two values 1 and 0 corresponding to whether production will grow or not.

- This is the data that we have to \*\*predict for future samples.\*\*

#### The dataset is divided into two parts: Train and Test sets.

#### Training Set:

- The train set contains \*\*7619 rows\*\* and \*\*124 columns\*\*

- The last column \*\*activity\*\* is the \*\*target variable\*\*.

#### Testing Set:

- The test set contains \*\*2540 rows\*\* and \*\*123 columns\*\*.

- The test set \*\*doesn’t contain\*\* the \*\*Target\*\* column.

- It needs to be predicted for the test set.

\*\*Train Set:\*\*

| Records |Features|Target Variable|

|\*\*7619\*\*|\*\*124\*\*|\*\*Target\*\*|

\*\*Test Set:\*\*

|Records|Features|Predicted Variable|

|\*\*2540\*\*|\*\*123\*\*|\*\*Target\*\*|

|\*\*\*ID\*\*\*|\*\*\*\*Feature\*\*\*\*|\*\*\*\*Description\*\*\*\*|

|:--|:--|:--|

|01| Id | Unique Id |

|02| month | Selected Months|

|03| Country | Countries(76 in Total)|

|04| 1\_diffClosing stocks(kmt) | Closing Stocks for one Month|

|05| 1\_diffExports(kmt) | Exports for one Month|

|06| 1\_diffImports(kmt) | Imports for one Month|

|07| 1\_diffRefinery intake | Refinery Intake for one Month|

|08| 1\_diffWTI | West texas Intermediate Price for one Month|

|09| 1\_diffSumClosing stocks(kmt) | Sum Closing Stocks for one Month|

|10| 1\_diffSumExports(kmt) | Sum Exports for one Month|

|11| 1\_diffSumImports(kmt) | Sum Import for one Month|

|12| 1\_diffSumImports(kmt) | Sum Production one Month|

|13| 1\_diffSumImports(kmt) | Sum Refinery Intake|

|114| 12\_diffClosing stocks(kmt) | Closing Stocks for 12 Months|

|123|12\_diffSumImports(kmt) | Sum Refinery Intake| | Sum Refinery Intake|

|124| Target | The label for the crude oil data (whether production will grow or not)|