**Configurations and parameters of the Modules**

**Step 1: Adding Dataset to Azureblobstorage**

The status column originally has 3 classes / labels. A new column was created “status\_binary” that contains two categorical values i.e., (a)operating and (b) non-operating. ‘Operating‘ values are the rows for which the original status was either “Operating” or “acquired”.

The file path for the dataset added to the designer is given below.

Graphical user interface

Description automatically generated

**Step2: Select columns in the dataset(row\_count = 19958, col\_count = 28)**

Graphical user interface

Description automatically generated

Step 3: Clean Missing Data**(row\_count = 19958, col\_count = 28)**

Graphical user interface

Description automatically generated

Step 4: Remove Duplicate Rows**(row\_count = 19907, col\_count = 28)**

Graphical user interface

Description automatically generated

Step 5: Edit Metadata**(row\_count = 19907, col\_count = 28)**

Graphical user interface

Description automatically generated

**Step 5: Clean Missing Data( row\_count = 19907, col\_count = 28)**

Usually should be put after “Clip Values” module for removing outliers in “SUM\_funding\_total\_usd” column.

Graphical user interface

Description automatically generated

Step 6: Normalize Data**( row\_count = 19907, col\_count = 28)**

Graphical user interface

Description automatically generated

Transformed Dataset

Graphical user interface, application

Description automatically generated

Step 7: SMOTE for handling imbalanced classes**( row\_count = 35667, col\_count = 28)**

Graphical user interface

Description automatically generated

Step 8: Split Data

(Result Dataset1: **( row\_count = 24967, col\_count = 28),**

Result Dataset2: **( row\_count = 10700, col\_count = 28)**

Graphical user interface

Description automatically generated

Step 9: Train Model

Graphical user interface

Description automatically generated

Step 10: Score Model & Evaluate Model **(row\_count = 10700, col\_count = 30)**

**Not-operating : 6141, Operating: 4559**

**Screenshot of the complete Pipeline Execution**

**Diagram

Description automatically generated with medium confidence**

Results

Scored Label Statistics (Distribution of the two-classes)

Graphical user interface, application

Description automatically generated

Evaluation Metrics

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated