DEVELOPING A FLIGHT DELAY MODEL USING MACHINE LEARNING

TEAM ID: PNT2022TMID27775

Splitting The Dataset Into Dependent And Independent Variables

- In machine learning, the concept of dependent variable (y) and independent variables(x) is important to understand. Here, Dependent variable is nothing but output in the dataset and the independent variable is all inputs in the dataset. We can denote with any symbol (alphabets). In our dataset we can say that class is the dependent variable and all other columns are independent. But in order to select the independent columns we will be selecting only those columns which are highly correlated and some value to our dependent column.
- With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.
- Let's create out independent and dependent variables

```
dataset = pd.get_dummies(dataset, columns=['ORIGIN', 'DEST'])
dataset.head()
```

```
: x = dataset.iloc[:, 0:8].values
  y = dataset.iloc[:, 8:9].values
```

- In the above code we are creating a DataFrame of the independent variable x with our selected columns and for dependent variable y we are only taking the class column.
- Where DataFrame is used to represent a table of data with rows and columns.