# **GENERAL GUIDE:**

# **DATA DICTIONARIES:**

**Datasets:**

***Raw data:*** “INX\_Future\_Inc\_Employee\_Performance\_CDS\_Project2\_Data\_V1.8.xls”

***Data after outliers handling and encoding:*** “data\_encoded.csv”

***Data after feature selection*:** “data\_fs.csv”

***Data after all preprocessing techniques are done and ready for modelling:*** *”*processed\_data.csv”

***Data after splitting into train data and test data with (X\_test, y\_test) samples for predictive model:*** “test\_data.csv”

**Variables used:**

**Minmax scaled data:** “mm\_scaled\_data”

**Training samples of splitted X and y after applying SMOTE() function:** “X\_train\_sm”, y\_train\_sm”

# **FILES CREATED:**

**Trained models:**

“trained\_model\_base\_decision\_tree.pkl”

“trained\_model\_base\_random\_forest.pkl”

“trained\_model\_base\_svm.pkl”

“trained\_model\_tuned\_decision\_tree.pkl”

“trained\_model\_tuned\_random\_forest.pkl”

“trained\_model\_tuned\_svm.pkl”

“trained\_model\_tuned\_xgboost.pkl”

**Predictive models:**

“Predict\_model\_base\_decision\_tree.csv”

“Predict\_model\_base\_random\_forest.csv”

“Predict\_model\_base\_svm.csv”

“Predict\_model\_tuned\_decision\_tree.csv”

“Predict\_model\_tuned\_random\_forest.csv”

“Predict\_model\_tuned\_svm.csv”

“Predict\_model\_tuned\_xgboost.csv”

**Univariate analysis report:**

“SWEETVIZ\_REPORT.html”