# **GENERAL GUIDE**

# **DATA DICTIONARIES**

Datasets:

**Raw data:** “inx\_dataset.xls”

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

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

# **FILES CREATED:**

|  |  |
| --- | --- |
| **Model Name** | **Model saved File name** |
| Decision Tree | decission\_tree\_base\_model.pkl |
| Tunned Decision Tree | decission\_tree\_trained\_and\_tunned\_model.pkl |
| Random Forest | random\_forest\_base\_trained\_model.pkl |
| Tunned Random Forest | random\_forest\_trained\_and\_tunned\_model.pkl |
| Support Vector Machine | trained\_model\_base\_svm.pkl |
| Tunned Support Vector Machine | trained\_model\_tuned\_svm.pkl |
| Tunned XG Boosting | xgboost\_trained\_and\_tunned\_model.pkl |