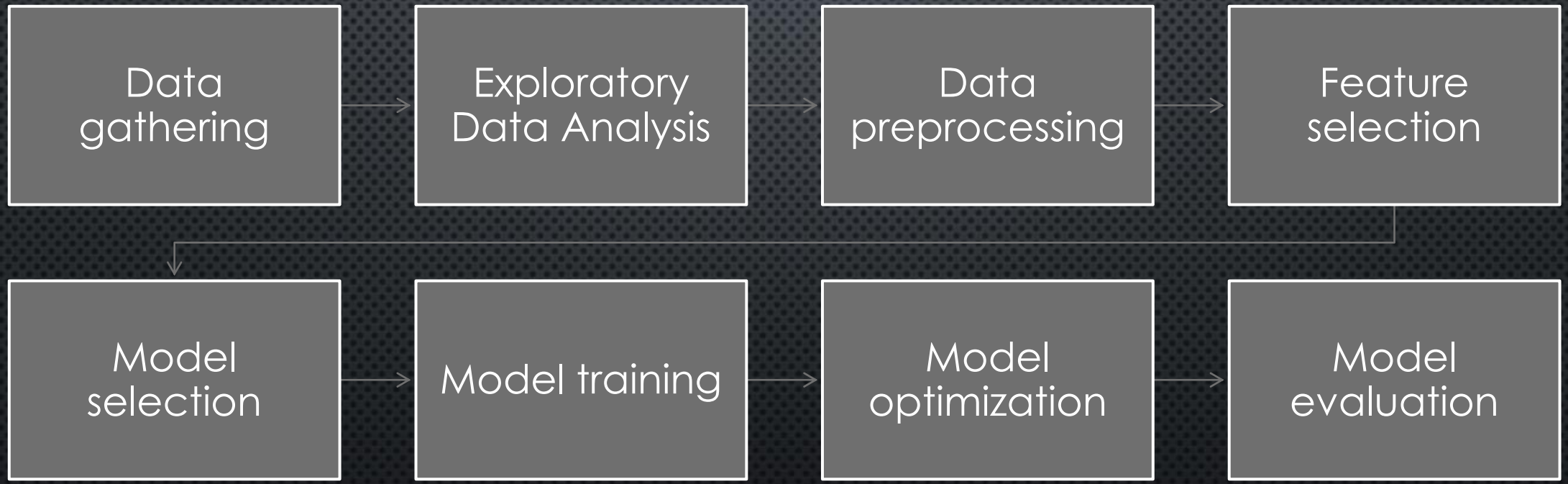
EDA PRESENTATION AND PROPOSED MODELLING TECHNIQUE

GROUP NAME: DATALEX
NAME: OVBUDE UANKHEHI
EMAIL: OVATEDOM@GMAIL.COM
COUNTRY: NIGERIA
SPECIALIZATION: DATA SCIENCE

Problem Description

- One of the challenges for all pharmaceutical companies is to understand the persistency of drug as per the physician prescription. To solve this problem ABC pharma company approached an analytics company to automate this process of identification.

PROJECT LIFECYCLE



EXPLORATORY DATA ANALYSIS

```
[5]: # Dimension of the dataset
     data.shape

[5]: (3424, 68)
```

THIS MODEL WAS BUILT WITH DATA CONTAINING 3428 OBSERVATIONS.

THERE IS NO MISSING VALUES

```
# Total number of missing values
data.isnull().sum().sum()

0
```

Note: 1 = Persistent, 0 = Non-Persistent

```
data["Persistency_Flag"].unique()

array([1, 0])
```

Balance Dataset

```
data["Persistency_Flag"].value_counts()

0    2070
1    1206
Name: Persistency_Flag, dtype: int64
```

Note: The Non-Persistent observations are almost double the Persistent observations.

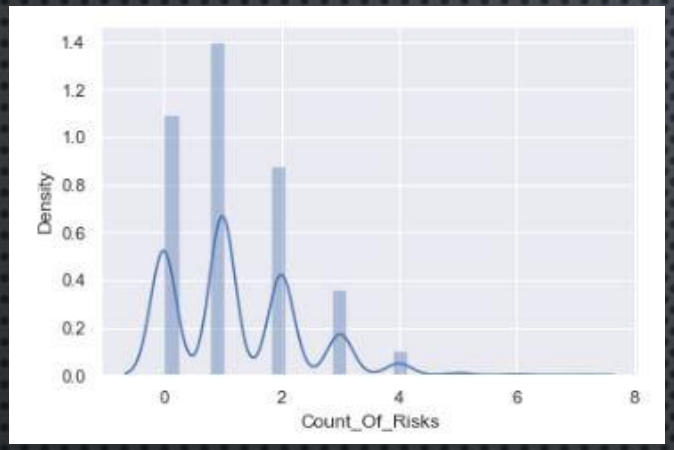
Next step is to balance the data using either the Random-Undersampling or Random_Oversampling method.

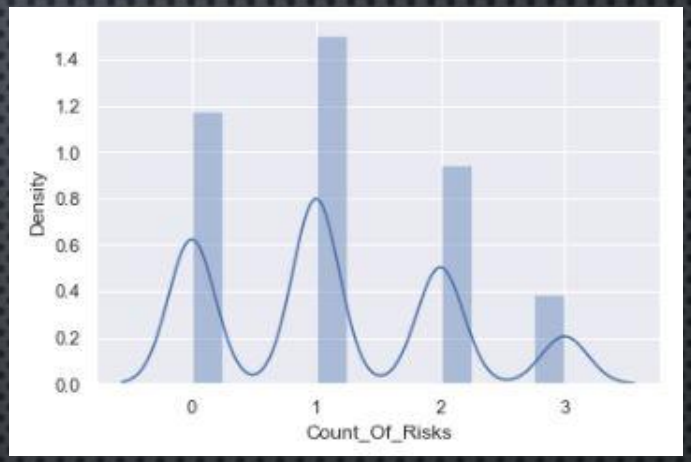
- THE DATA SET IS IMBALANCED, AS THERE IS UNEQUAL DISTRIBUTION OF CLASSES.
- PROBLEM RESOLVED WITH RANDOM OVERSAMPLING.

```
data["Persistency_Flag"].value_counts()

1    2070
0    2070
Name: Persistency_Flag, dtype: int64
```

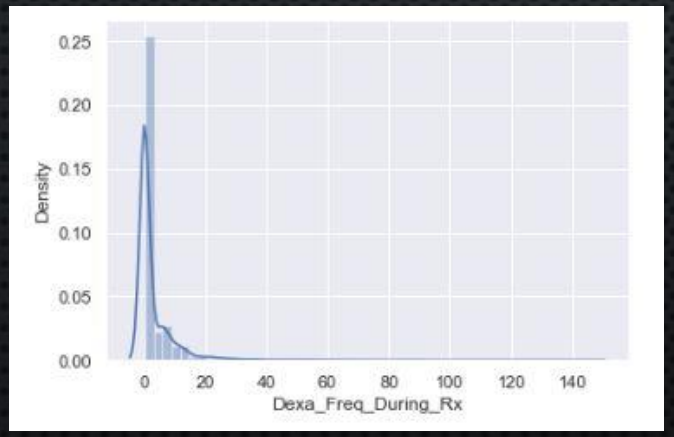
EDA – HANDLING OUTLIERS

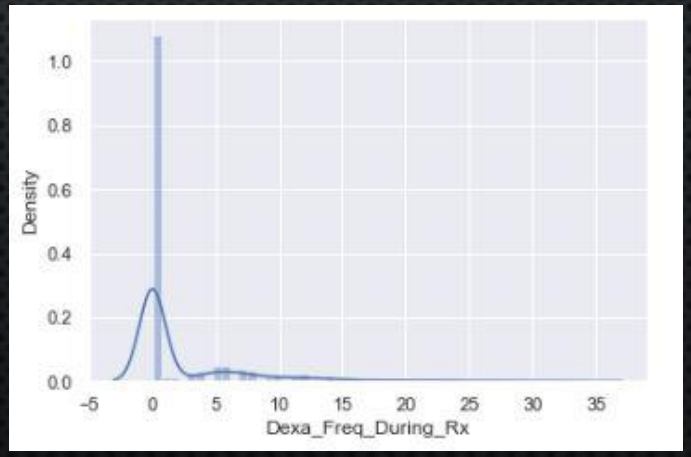




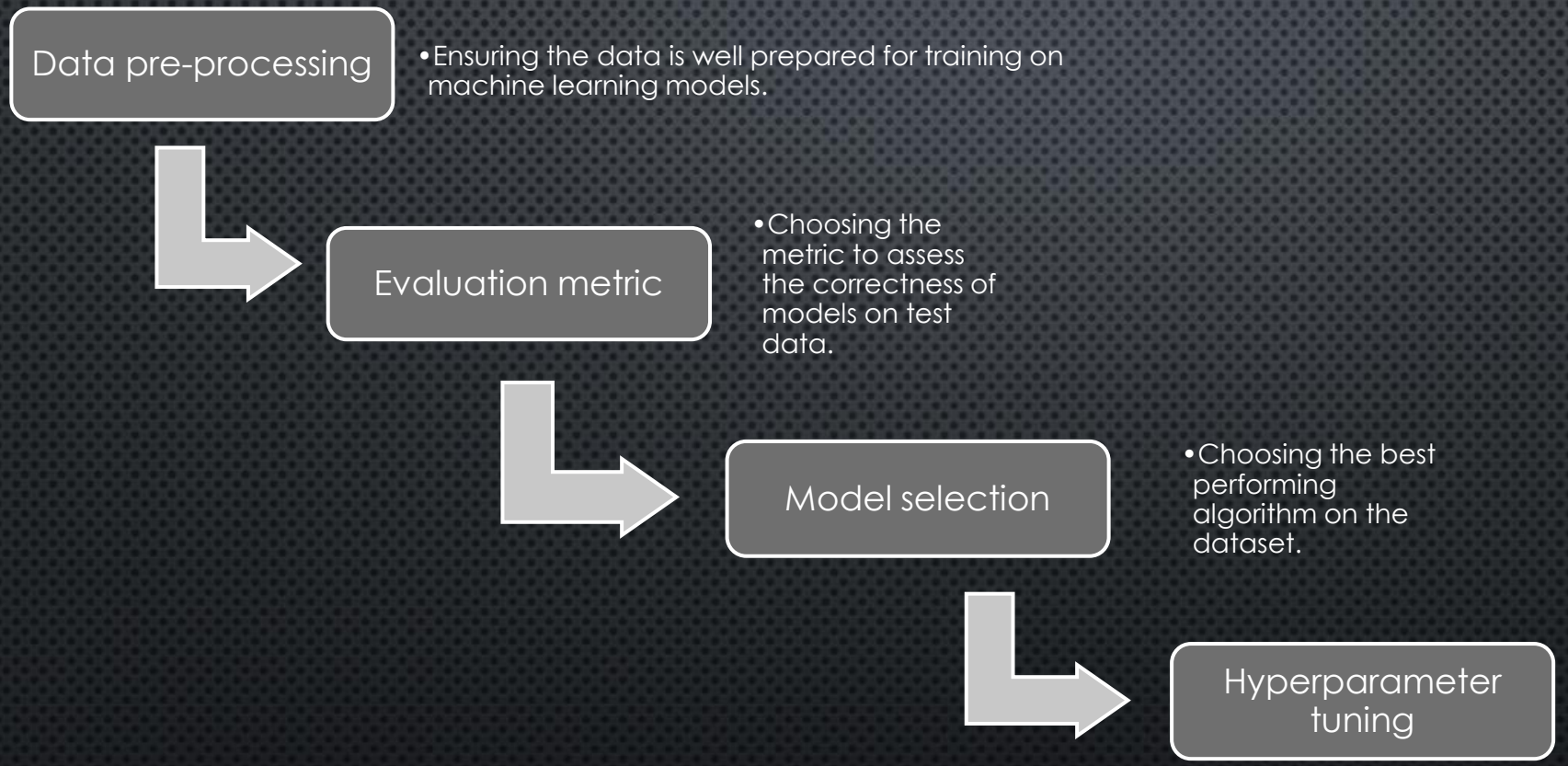
BEFORE

AFTER





MODELLING TECHNIQUE



GITHUB REPOSITORY LINK:

- HTTPS://GITHUB.COM/OVATED/PERSISTENCY-OF-A-DRUG-PREDICTION