**SALES EFFECTIVENESS REPORT**

(Lead Category Analysis)

**PROJECT SUMMARY:**

1. I am using XG Boost Algorithm to predict the lead Category (High potential, low potential).

2. **XG Boost algorithm can use gradient boosting framework and used for both classification and the regression kind of problems.** XG Boost is an implementation of gradient boosted decision trees designed for speed and performance.

3. In this Algorithm Learning from previous model called ADA boost.

4. ItMispredicted data is predicted of previous model.

5.The Combined having efficiency greater than and might be (80%-85%) called Boosting.

Ex: - It can be used for quality Assessment.

**XG Boost algorithm features:-**

* Tree pruning using depth-first Approach.
* Parallelised tree building.
* Effiecient handling of missing data.
* Regulization of Avoiding over-fitting.

**REQUIREMENT:**

>> Data of Sales Effectiveness.

* Software:-
* Anacondas with Jupyter Notebook.
* Hardware:-
  + Intel processor with 3 GB or higher RAM.
  + 500 GB HDD.

**DATA PROCESSING:**

* Data processing occurs when data is collected and translated into usable information.
* Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.)
* In the given dataset of Employee Performance, I checked its shape, data types, duplicates\_ values, null\_values, full information, etc.
* For achieving better result from the applied model the format of data has to be in a proper manner so the data processing is needed.
* The steps of Data Processing are- Data collection, Data Preparation, Data input, Processing, Data output/Interpretation, Data Storage.

**Modelling:**

Data divides into two parts:

Train model and

Test model

Train model has already trained and based on training model we predict test model.

I have taken 30% Test data and 70% Train data.

**Predicting:**

Here i predict Accuracy score from XG Boost Algorithm, Classification report and confusion matrix.Here we predicting the Lead Category (high potential, low potential.

**Matplotlib:**

Matplotlib and seaborn is used for visualization, plotting graphs, bar charts ,pie charts,heatmap for missing values.

**References:-**

* Counter (To count test data and predict data )
* Correlation (used to find relationship b/w different things)
* XG Boost (Algorithm to predict of Lead Category and Sales effectiveness).
* Scaling (To find standardization and normalization)

Amit Singh

(amitsingh.hw@gmail.com)

From 27th May CDS Batch