**CAPSTONE PROJECT**

**LEAD SCORING USING ML ALGORITHMS**

**Problem Statement**: In this capstone project, you are going to analyze the given data set and build a machine learning algorithm to predict the propensity of buying the course.

**Dataset**: Purchase\_Fraud\_Data.csv

**Dataset Description**: Every row in this dataset contains information about the lead. Along with lead details, we also have other details like number of calls made by salesperson, how long it has been since the lead has been generated etc.

**opportunity\_id:** Lead unique id

**stagename:** The stage at which the lead is

**subsource:** The subsource from which the lead has been obtained

**unanswered\_call\_counter:** Number of calls that went unanswered

**unserviced age** - No. of days since no action taken on that lead

**ageing** - Days the lead has aged since created

**source:** The source from which the lead has been obtained

**Opportunity\_created\_date:** Date on which lead was generated

**time\_taken\_for\_allocation:** Time taken to allocate the lead to a salesperson

**no\_of\_calls:** Number of calls

**calldurationinseconds:** Duration the call from salesperson to the lead lasted

**City**: City the lead belongs to

**City Type:** Whether the city belong to Tier-1, Tier-2 or Tier-3

**Workexp-** The work experience the lead has

**Laststagechangedate-** The date when the stage of the lead changes

**Sales\_end\_date-**Date at which sales end

**Sales\_start\_date-** Date at which sales start

**Program\_number:** The number of the program

**Program Code -Renewed:** The code of the program

**total\_program\_fee.1:** The fees of the program

**Category:** Category the program belongs to

**Is\_serviced\_c:** Whether action is taken on the lead

**Business\_vertical:** The business vertical that the program belongs to

**Budget:** Budget allocated for the specific program

**Lastactivitydate:** Last date of activity with the lead

**Institute:** The institute which is offering the program

**Year:** The year in which the lead was generated

**Month:** The month in which the lead was generated

**Month\_name:** The name of the month in which the lead was generated

**Week of Year:** The week of the year in which the lead was generated

**Week of Month:** The week of the month in which the lead was generated

**Day of Year:** The day of the year the lead was generated

**Day of Week:** The day of the week the lead was generated

**Day Name:** The name of the day the lead was generated

**Age:** The age range of the lead

**is\_session\_Working\_day:** Does the session fall on a working day

**workex\_reqd:** work exp required to take the course

**Session\_Start\_day\_inweek:** The day which session starts

**Weekly\_sessionsdays\_count:** How many sessions in a week

**sales\_Working\_day:** Whether it’s a working day for sales

**Time:** The time at which the lead is generated

**Hour:** The hour at which the lead is generated

**IscreatedinWorking\_hour:** Whether the lead is created in working hour of sales

**Funnel\_category:** created from stagename to decrease the number of levels in target column (Target)

Project outline :

* **Exploratory Data Analysis**

1. Summarize numerical, categorical and date columns separately and list down your inferences
2. Identify and perform missing value treatment using appropriate techniques
3. Univariate analysis: For each column perform appropriate univariate analysis. (i.e. perform distribution analysis on numerical columns and frequency analysis on categorical columns)
4. Multivariate analysis: Take combinations of multiple columns and identify the relationship between them.
5. Categorical vs numerical columns – bar charts, boxplots
6. Numerical vs numerical columns – scatter plot
7. Categorical vs multiple numerical columns – scatter plot
8. Correlation matrix
9. …

* **Feature Engineering**

* **Model Building**