**Problem Statement:** To predict the CLTV based on the user and policy data of one of the leading insurance companies in India (VahanBima).

**Given data:**

* **Input Variable**
  + Id - Unique identifier of a customer
  + Gender - Gender of the customer
  + Area - Area of the customer
  + Qualification - Highest Qualification of the customer
  + Income - Income earned in a year (in rupees)
  + marital\_status - Marital Status of the customer {0: Single, 1: Married}
  + Vintage - No. of years since the first policy date
  + claim\_amount - Total Amount Claimed by the customer (in rupees)
  + num\_policies - Total no. of policies issued by the customer
  + Policy- Active policy of the customer
  + type\_of\_policy - Type of active policy
* **Target Variable**
  + Cltv - Customer lifetime value

**Observation: *Target variable represents continuous value, so all the below approaches are***   ***based on Regression.***

**Solution:**

* **Approach-1:** 
  + **Preprocessing (EDA/Data Engineering):**
    - Checked missing values (None Found).
    - Converted the categorical data into nominal and ordinal value using MsExcel.
    - Scaled the numerical data (claim\_amount & cltv) using sklearn MinMax().
    - Checked relationships between input and target variable (Correlation)
  + **Processing (Model Creation and training/validating):**
    - Looped through Regression models (RandomForestRegressor, GradientBoostingRegressor, DecisionTreeRegressor, LinearRegression, Lasso, Ridge) to get the best r2\_score possible with the above pre-processed data – Found GradientBoostingRegressor performed well in comparision to all the models with a r2\_score of 0.1603.
* **Approach-2:**
  + Based on the above preprocessed data, a deep neural network is trained to see if it can be used to solve the problem statement.
* **Approach-3:**
  + Used CatBoostRegressor wherein preprocessing steps like conversion and scaling are not required as it is capable of giving quick results without the need of any preprocessing steps to save time.

**Note:** JOB-A-THON - January 2023 - Using CatBoostRegreesor .ipynb is used for final submission.