Analyse customer eligibility for home loan based on various factors

**Steps in Data Preprocessing in Machine Learning**

1. Acquire the dataset
2. Import all the crucial libraries
3. Import the dataset
4. Identifying and handling the missing values
5. Encoding the categorical data
6. Splitting the dataset
7. Feature scaling

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.

**Predictive Modeling​**

Predictive modeling is a commonly used statistical technique to predict future behavior. Predictive modeling solutions are a form of data-mining technology that works by analyzing historical and current data and generating a model to help predict future outcomes.​

**INSIGHTS:**

* The proportion of married applicants is higher for approved loans.
* It seems people with good credit history are more likely to get their loans approved.
* Proportion of approved loans is higher for low and average loan amount as compared to that of high loan amount
* The proportion of loans getting approved in the semi-urban area is higher as compared to that in rural or urban areas.

**General parameters that are important while approving a home loan:**

Each customer corresponds to his/her account number. Based on the following parameters/data sets, decisions are made regarding approval of a home loan.

* **CIBIL score:** It is a three-digit numerical summary of one’s credit history. It gathers data from a credit information report that’s directly linked to one’s PAN number. This score gives information directly relating to the account holders prospective to repay the loan sanctioned. Hence it is of great importance to banks. The score ranges from 300 to 900.
* **Income/Salary:** It is the net income that a person generates or gains per financial year. It could be from various sources. This is the basis on which a loan is sanctioned. It also determines the maximum loan value.
* **Employment type:** It is the type of occupation ones pursuing. It could range from IT professional to selfemployed. This plays a key role in sanctioning as it decides the reimbursement probability of the loan by the individual.
* **Property validation/valuation**: The property in concern to be purchased is assessed for encumbrance and worth in monetary terms so as to decide the maximum loan amount.
* Debts: The current liability of an individual determines his ability to pay the monthly EMI premium.
* **Age:** Age of an individual is a key factor that determines how much premium is given to an individual and also the repayment period. The general age must be above 18 years of age and less than 60

We created and tested several models

* **Logistic Regression Model**
  + Accuracy Score = 84.32
* **Neural network**
  + Accuracy Score = 83.78
* **XGBoost classifier**
  + Accuracy Score = 74.069
* **Using Stacking method (Ensemble learning methods)**
  + Accuracy Score = 80.05
* **Stochastic Gradient descent classifier**
  + Accuracy Score = 82.70
* **Bagging method (Ensembled machine learning)**
  + Accuracy Score = 80.54