### Predict Credit Consumption for Customers for a Leading Bank

1. All the Datasets Regarding this Case-Study has been Imported.
2. All the Datasets have been merged and Missing Values is Treated.
3. Outlier Capping doesn’t impact the data as outliers don’t get capped.
4. Variables which have 0% of Co-efficient of Variance are removed.
5. Dummies are created for Binary Categorical Variables.
6. Train Dataset has been separated from Test Dataset where Train dataset contains 15,000 rows and Test dataset contains 5,000 rows.
7. Train Dataset’s Info and Description is Analyzed.
8. This is a Complex Dataset where R-Square is less than 0.2% X-Variables cannot explain Y-Variable
9. It is Required to transform all the chosen X-variables and Y-variable, else the prediction cannot be accurate.
10. Log Transformation Gives a Normal Distribution for the Y Variable.
11. Power Transformer method(box-cox) also gives a Normal Distribution.
12. RFE and Select K Best are used for Feature Reduction.
13. Segmentation is performed on Log Transformation.
14. Segmentation cannot be performed after Power Transformation as 1 cluster has around 70% of population.

Algorithms chosen for this Case-Study are:

1. OLS Regression
2. KNN Regressor
3. XG Boost Regressor

**RMSPE and MAPE Percentage For each Model:**

* **OLS Regression Model:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **OLS Models** | MAPE for Train Data | RMSPE for Train Data | MAPE for Test Data | RMSPE for Test Data |
| A)1st Segmentation Model  B)2nd Segmentation Model  C)3rd Segmentation Model  D)Without Segmentation Model | 12%  12%  12%  12% | 28%  20%  23%  22% | 12%  12%  12%  12% | 19%  20%  20%  27% |

* **KNN Regressor Model:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **OLS Models** | MAPE for Train Data | RMSPE for Train Data | MAPE for Test Data | RMSPE for Test Data |
| A)1st Segmentation Model  B)2nd Segmentation Model  C)3rd Segmentation Model  D)Without Segmentation Model | 11%  11%  11%  11% | 17%  15%  16%  18% | 16%  15%  15%  15% | 31%  21%  28%  23% |

* **XG Boost Regressor Model:**

Segmentation not done in this Algorithm as we can see segmentation is not helping in getting a stable Model.

Without Segmentation:

MAPE for Train & Test: 12% Respectively.

RMSPE for Train and Test: 22% and 21% Respectively.

**Models after Power Transformation:**

As mentioned above no Segmentation Performed after Power Transformation

* **OLS Regression Model:**

**MAPE for Train & Test: 11% Respectively**

**RMSPE for Train & Test: 19% and 25% Respectively**

* **KNN Regressor Model:**

**MAPE for Train & Test: 7% and 10% Respectively**

**RMSPE for Train & Test: 11% and 15% Respectively**

* **XG Boost Regressor Model:**

**MAPE for Train & Test: 11% Respectively**

**RMSPE for Train & Test: 20% and 19% Respectively**

**To Conclude this Report:**

**XG Boost after Log transformation and Power Transformation gives a Stable Model and has less RMSPE compared to OLS Regression and KNN Regressor. And if we need to go with one Model XG Boost after Power Transformation has 1% less MAPE and RMSPE compared to XG Boost with Log Transformation and thus Model Chosen for Predicting the Values for Test Dataset with Power Transformation but XG Boost with Log Transformation also used to Predict Values for Test Dataset!**