

Micro-Credit Defaulter Prediction

Submitted by:

AMIT PURI GOSWAMI

**ACKNOWLEDGMENT**

The project is done under the kind guidance of Mr. Kasif Khan. Some learnings are made from different informative websites to come out of the some difficult situations. The material which is freely available for any aspirants are used for better understanding of concepts of different techniques. Few of them to name are [www.google.com](http://www.google.com), <https://www.ibm.com> etc.

**INTRODUCTION**

* Business Problem Framing

One of our clients in Telecom Sector wants to provide its micro finance services to low-income people with a check of their repayment habits. The clients is a fixed wireless telecommunications network provider and wish to check the repayment history of low income people and want to help them with micro finance credit if the repayment history is good or the loan aspirant is not a defaulter.

With the help of data available we need to build a model which can predict that whether a customer can be defaulter or not. If it is not then the financial product can be approved to him or her.

* Conceptual Background of the Domain Problem

In Bank , Credit Agencies or Micro Finance Agencies they have some analysts called credit managers who are responsible to decide whether any loan application should be approved or not. If they approve it then the loan is issued to the customer after due approval of the same.

The task of credit managers is very important as it is very crucial to decide whether the loan should be given or not. The credit manager need a mechanism to take their decision as it may result in profit if the load is repaid or will result in financial loss to the organization.

We need to develop a model on the basis of data available to find out whether the customer can repay the loan amount or not.

* Review of Literature

It is researched that the problem is very important and how the micro finance institution work. It is found that the work of micro finance institution is very tedious. They have credit managers who decide that the loan should be given or not. We need to help them by making a model which can predict the defaulter candidate.

* Motivation for the Problem Undertaken

The Micro Finance Agency is very keen to help the low-income people to establish their business in less time with small capital at the same time wish to earn some profit mutually with the loan aspirants.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Checked for Null values and found no null. If null values are found then these are replaced with mean and median values which requires statistical calculation by the system.

Moreover correlation and describe are also mathematical calculation done by the system. Then Transformation requires many mathematical calculation.

Data Sources and their formats

We received Data File.csv file it came from the client to us and which contains the data for using in the prediction model.

* Data Preprocessing Done

Treating Null Values

Dropping less important features

Transforming data

Scaling data

Balancing the class

* Data Inputs- Logic- Output Relationships

There are certain features which are strongly correlated with the target features. Highest correlated feature is cnt\_ma\_rech30 which affects the output very much.

* State the set of assumptions (if any) related to the problem under consideration

Previous history of repayment is very important for deciding the defaulter.

* Hardware and Software Requirements and Tools Used

Basic Hardware are : i5 processor, 8GB RAM, other standard h/w

Softwares used : Jupyter, MS Excel, MS Word

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Standard process of data cleansing, feature selection, EDA, Outliers Removal, Skewness Removal, Handling Multi collinearity , dividing data in train and test, Standardization, finding best random state, model selection, finding best cv, Parameter finetuning, Finalization of the model and predicting.

* Testing of Identified Approaches (Algorithms)

Algorithms used are

1. LinearRegression
2. DecisionTreeRegressor
3. RandomForestRegressor
4. Ridge
5. SVR
6. KNeighborsRegressor

* Run and Evaluate selected models

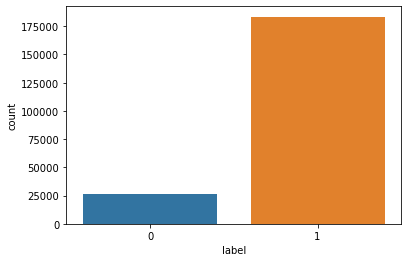
After running and evaluating RandomForestRegressor is found to be the best model.

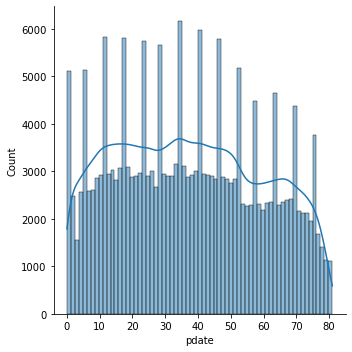
* Key Metrics for success in solving problem under consideration

Best feature selection

* Visualizations

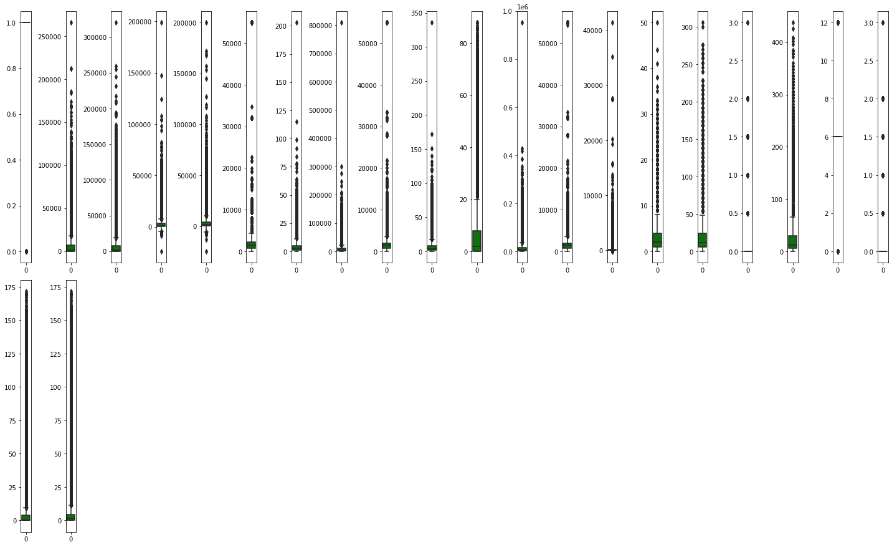
1. Scatterplot
2. CountPlot
3. Boxplot











* Interpretation of the Results

The insight we get from the model that if the credit history is not good the customer may be a defaulter and we should not approve him the loan.

**CONCLUSION**

* Key Findings and Conclusions of the Study

The insight we get from the model that if the credit history is not good the customer may be a defaulter and we should not approve him the loan.

* Learning Outcomes of the Study in respect of Data Science

Visualization helps us a lot in understanding the problem and identifying the work which will be suitable for model building.

* Limitations of this work and Scope for Future Work

No. of columns are 37 which is very high for a standard computer to handle and if we get high performing hardware little more accuracy can be obtained by getting best parameters. Though some features are dropped as they were not contributing much.