**India ML Hiring Hackathon 2019**

**Madhuri Konnur**

Loan Delinquency Prediction

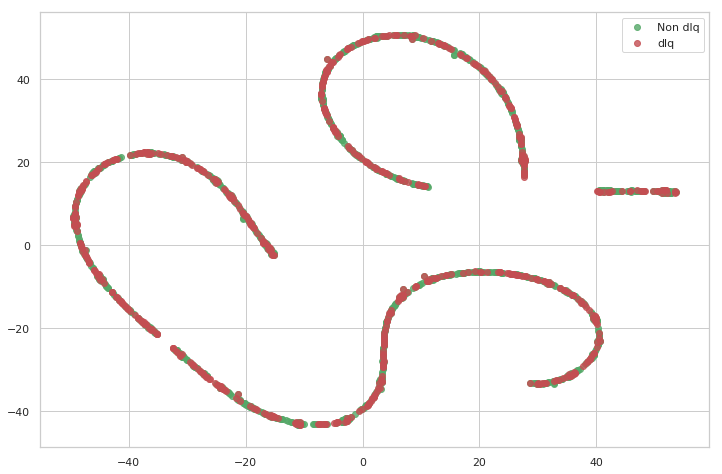
Loan default prediction is one of the most critical and crucial problem faced by financial institutions and organizations as it has a noteworthy effect on the profitability of these institutions

Given the information like mortgage details, borrowers related details and payment details, our objective is to identify the delinquency status of loans for the next month given the delinquency status for the previous 12 months (in number of months)

Brief Approach:

1. Since it is highly imbalanced data, solution approach lies around same
2. Only 0.55 % cases are Delinquency transactions ([‘’m13’’]==1)
3. Undersampling of majority class worked well here
4. This approach helped for clear visualization and thought for rescue method to separate two types of data

***Visualize Delinquency and Non-Delinquency Transactions***



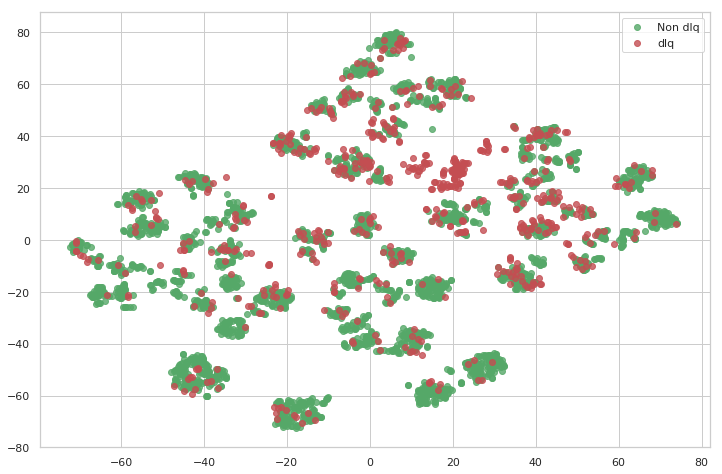
**Validation strategy:** **AutoEncoders**

**Preprocessing /Feature Engineering:**

**From the initial data analysis, there was not much scope for missing data handling, LabelEncoder was used for converting categorical var to int and feature variable dependencies were most of the time intercorrelated on each other impact on target is less decisive.**

**Here 3 layered AutoEncoder came to rescue with less loss after multiple level of tuning. This gave much better separable visualization of delinquency and non- delinquency**

***Visualize the latent representations: Delinquency Vs Non-Delinquency***



**Final Model:**

**Final model used is “XGBoost Classifier”. Model was built on Autoencoded data with undersampling and an AUC score of 0.772**