**PROJECT SYNOPSIS**

**Credit Risk Analysis**

*Submitted towards the partial fulfillment of the criteria for award of Genpact Data Science Prodegree by Imarticus*

*Submitted By:*

*Member 1 (Roll #)*

*Member 2 (Roll #)*

*Member 3 (Roll #)*

*Course and Batch: DSP 16*



# Scope & Objective:

To build a data model to predict the probability of default, and choose a cut-off based on what feels is suitable. Alternatively we can also use a modelling technique which gives binary output.

# Business Problem Statement:

In this project you will have to put yourself in the shoes of a loan issuer and manage credit risk by using the past data and deciding whom to give the loan to in the future. The text files contain complete loan data for all loans issued by XYZ Corp. through 2007-2015. The data contains the indicator of default, payment information, credit history, etc.

The data should be divided into train (June 2007 - May 2015) and out-of-time test (June 2015 - Dec 2015) data. You will have use the training data to build models/analytical solution and finally apply it to test data to measure the performance and robustness of the models.

You should use the variable : 'issue\_d' to divide the data in the above time periods, the variable is in <month>-<year> format.

# Data Sources:

# Analytics Tools:

Jupiter notebook,

# Analytics Approach:

1.Understanding the data and framing the objectives

2.Exploratory data analysis

3.Data cleaning i.e treating missing values, replacing categorical variables

4.Model fitting without feature engineering using logistic regression, Random Forest and gradient boosting

5.Feature engineering

6.Model fitting with feature engineering using logistic regression, Random Forest and gradient boosting

7.Testing and cross validation

8.Results and plots

# KPIs, Timelines, Milestones

# Anything else which one might want to add