**Credit Risk Analysis and Prediction of Loan Defaults**

**Introduction**

Credit risk is the possibility of a loss resulting from a borrower's failure to repay a loan or meet contractual obligations. In order for lenders help to minimize the possible losses and increase the volume of credits, Credit Risk assessment is a crucial to evaluate if a loan applicant can be a defaulter at a later stage so that lenders can go ahead and grant the loan or not. This project focuses to analyze the credit risk involved in peer-to-peer (P2P) lending system of LendingClub Company based in the United States, in which investors provide funds for potential borrowers and investors earn a profit depending on the risk they take. Machine learning algorithms and preprocessing techniques will be used to explore, analyze and determine the factors, which play crucial role in predicting the credit risk involved in LendingClub. The dataset existed in the open source Kaggle website through the URL link,<https://www.kaggle.com/wendykan/lending-club-loan-data>that ranges from 2007 to 2015. It has 890 thousand records and 75 features.

**Statement of the problem**

In the financial industry, lending loans has been playing a significant role and it is quite beneficial for both the lenders and the borrowers. However, evaluating the creditworthiness of loan applicants is a common challenge in the financial industry typically where loans are unsecured. Therefore, in decision-making process whether or not to fund particular loans, it is important for the lending industry to apply risk management by providing investors with comprehensive risk assessment.

**Methodology**

Initially, I will be focusing on understanding which features are most relevant or most predictive for separating those loans likely to be paid from those loans likely to default. The level of default risk will be best predicted with predictive modeling using machine-learning techniques. I will perform Data Processing, Exploratory Data Analysis (EDA) and data Visualization using Python.

**Goal**

Using Lenders Club dataset, the goal is to provide a data product that enables investors to avoid loans likely to default. This will be achieved by conducting EDA and by developing a model using various machine-learning techniques that predict the probability of default for a potential loan to avoid loans that are predicted to default.

**Reference:**

Credit Risk by Olivia Labarre, May 2019 retrieved from <https://www.investopedia.com/terms/c/creditrisk.asp>

Default rates at Lending Club & Prosper: When loans go bad by Simon Cunningham on October 17, 2014 in [P2P Lending Basics](https://www.lendingmemo.com/category/p2p-lending-basics/) <https://www.lendingmemo.com/lending-club-prosper-default-rates/>

<https://www.roselladb.com/credit-risk-analysis.htm>