

Lending club loan defaulters Prediction

Data Science Engineering Methods and Tools

August 12, 2022

Bhagyashree NAIR

SHWETA WAKALE

ANUPRITA DHAMALE

Contents

[Topic: 2](#_Toc111222134)

[Course: 2](#_Toc111222135)

[Professor: 2](#_Toc111222136)

[Group Members: 2](#_Toc111222137)

[Introduction: 3](#_Toc111222138)

[Objective: 3](#_Toc111222139)

[Methodology: 3](#_Toc111222140)

[Dataset Key Specifications: 4](#_Toc111222141)

[Working: 4](#_Toc111222142)

# Topic:

* Lending Club Data Prediction and Analysis

# Course:

* Data Science Engineering Methods and Tools

# Professor:

* Ram Hariharan

# Group Members:

* Bhagyashree Nair (002139969)
* Shweta Wakale (001518127)
* Anuprita Dhamale (001529811)

# Introduction:

In the ten years between 2005 and 2015, Lending Club, an American peer-to-peer lender, assisted over 2.5 million customers through personal, auto refinancing, business, and medical finance loans. Through the internet platform, investors give money directly to borrowers. Loans up to $40,000 are available through Lending Club over 2 terms (36 and 60 months). Loans are given out in accordance with different grades, which correlate to various interest rates.

In this project, we attempt to use our analysis to identify various factors that influence the loan market and to answer the following questions:

* Determine whether or not the Lending Club loan will be charged off.
* Analyze the factors that influence the interest rate, credit grade, and predict the defaulters over a period of years

# Objective:

LendingClub is the largest online loan marketplace, offering personal loans, business loans, and medical procedure financing. Borrowers can easily obtain lower interest rate loans by using a quick online interface.

Lending to 'risky' applicants is the most common source of financial loss for most lending companies (called credit loss). The amount of money lost by the lender when the borrower refuses to pay or flees with the money owed is referred to as the credit loss. In other words, defaulting borrowers cause the most damage to lenders. Customers who have been charged off are the 'defaulters' in this case.

If these risky loan applicants can be identified, such loans can be reduced, lowering the amount of credit loss. The goal of this case study is to identify such applicants using EDA and machine learning.

In other words, the organization seeks to understand the characteristics that are reliable predictors of loan default, also known as the driving factors (or driver variables) behind loan default. This information can be used by the business in portfolio management and risk analysis.

You should conduct some independent research on risk analytics to improve your understanding of the domain (understanding the types of variables and their significance should be enough).

# Methodology:

An effective model can be created using machine learning techniques, allowing the business to determine whether LC loans will be charged off as well as detect probable loan defaulters. We also intend to use the information to forecast future interest rates and investigate the variables that affect credit scores and interest rates.

We intend to use to the following supervised ML algorithms to train the model for predictive analysis.

* XGBoost Classifier
* Random Forest Classifier
* Deep Neural Networks (DNN)
* Linear Regression

# Dataset Key Specifications:

Files Included:

* Lending\_club\_loan\_two.csv

Number of Columns: 27

Number of Rows: 396030

Dataset Size – 100MB

Dataset Link –

<https://www.kaggle.com/code/faressayah/lending-club-loan-defaulters-prediction/data>

# Working:

* Data Description and Visualization
* Exploratory Data Analysis
* Data Pre-Processing
* Train Test Split
* Models Building: -

1. XGBoost Classifier
2. Random Forest Classifier
3. Deep Neural Networks (DNN)
4. Linear Regression