

Analytics of Lending

Shraddha Shrivastava¹, Harsh Gupta², Garwit Choudhary³

^{1,2,3}Department of Computer Science and Engineering, Inderprastha Engineering College, Ghaziabad

Abstract: Lending Club (LC) is a Peer-to-Peer lending company acting as a loan originator and a web platform between borrowers and investors. Marketplace lending relies on large-scale loan screening and information production by investors, a major deviation from the traditional banking paradigm. Theoretically, the participation of sophisticated investors in marketplace lending increases volumes and improves screening outcomes, but also creates adverse selection to less sophisticated investors. In this project we will analyze the factors in relation with lending of various types of loans and reach conclusions based upon that.

Keywords: Exploratory Data Analysis, Python, Jupyter, Numpy and Pandas

INTRODUCTION

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

In this case study, we used EDA to understand how consumer attributes and loan attributes influence the tendency of default.

It is not only interesting to know factors explaining P2P loan default, but also to accurately predict loan defaults. Usually, students tend to take student loans for their college studies and this was the reason we got interested in the idea of analyzing the various loan to further improve our awareness and knowledge.

LITERATURE REVIEW:

S.No	Name	Year	Author	Methods
1	Lending Club – P2P Lending Impact Of Loan Description On Loan Performance	2013-2014	Pierre-Yves FESTOC	Excel
2	A study on predicting loan default based on the random forest algorithm	2019	Lin Zhua , Dafeng Qiua , Daji Ergua , Cai Yinga , Kuiyi Liub	Rain Forest Algorithm
3	P2P Network Lending, Loss Given Default and Credit Risks	2018	Guangyou Zhou, Yijia Zhang and Sumei Luo	Theoretical Analysis and Research Hypothesis
4	Determinants of Default in P2P Lending	2015	Carlos Serrano-Cinca, Begoña Gutiérrez-Nieto, Luz López-Palacios	Empirical Study
5	Performance Analysis of Credit Scoring Models on Lending Club Data	2016-2017	Bc. Michal Polena	ML and AI (Every Algorithm)
6	Credit Risk Analysis in Peer to Peer Lending Data set: Lending Club	2019	Mohammad Mubasil Bokhari	Decision Tree, Random Forest
7	A Theoretical Analysis of Peer-to-Peer Lending	2014	Matthew Courchene	Basic Theoretical Framework
8	Project: Lending Club Data Analysis	2019	Mohammad Rafiqul Islam, Tabitha Kemboi	Jupyter
9	Online Peer-to-Peer Lending	2011	Alexander Bachmann, Alexander Becker, Michel Hilker	Naïve Bayes and Decision Tree



10	VALUATION OF A FINTECH COMPANY: LENDING CLUB	2018	JIAYU YAO	By using different accounting methodologies
11	Fintech and the Future of Financial Service: A Literature Review and Research Agenda	2020	Haitian Lu, Bingzhong Wang, Qing Wu, and Jing Ye	Blockchain, Cryptocurrencies, Artificial Intelligence, Big Data
12	A Principal Decision: The Case of Lending Club	2015	Jonathan Ricardo Guzman	Principal Component Analysis, Linear Regression
13	Credit risk analysis with machine learning techniques in peer-to-peer lending market	2018	Qionglin Shan & Mikael Nilsson	Machine Learning, SVM, MLP, Logistic Regression, Decision Tree
14	Marketplace Lending: A New Banking Paradigm?	2018	Boris Vallee and Yao Zeng	Empirical Analysis
15	Credit risk prediction in an imbalanced social lending environment	2017	Anahita Namvar, Mohammad Siami, Fethi Rabhi, Mohsen Naderpour	Feature Engineering, Logistic Regression, Linear Discriminate Analysis
16	AN OVERVIEW STUDY ON P2P LENDING	2020	Dr. Rajeshwari M. Shettar	Banking

PROPOSED METHADODOLOGY

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

- In each step of our methodology, we employed the use of Python version 3. We choose Python because of the availability of extensive machine learning and data analysis libraries. We used Scikit-Learn, Pandas, NumPy, Seaborn, Jupyter-Notebook and Matplotlib in our study.
- Data Source: We used Lending Club's data for this analysis. The data set is for the period from 2007 to 2011. There are more than 42000 observations and more than 100 variables. It is often very hard to run analysis with all the variables and observations. So, we cleaned this data.
- Data Understanding: After sourcing the data we went through the data dictionary and the actual data to understand the data we are dealing with and to draw rough estimates of what we can achieve from this.
- Data Cleaning: Data cleaning is very important to draw conclusive results and to reduce redundancy of data. We have to remove missing columns and set the data according to the data types we want.
- Data Preparation: To prepare the date we have to first categorize each column into two types that is numerical and categorical and then we proceed to form pair of these columns for further steps.
- Exploratory Data Analysis: Now comes the main work for our data in which we have to determine which type of analysis can be performed in our dataset in our case we mainly used univariate analysis and few times bivariate analysis. In this section we plot various graphs and determine the conclusions depicted upon them.

ACKNOWLEDGMENT

We take this opportunity to thank our teachers and friends who helped us throughout the project. We would also like to thank **Dr. Vijai Singh (HOD, Computer Science Department)** for his constant support during the development of the project.

REFERENCES

- [1] Jean-Francois Darre, (2015). Analysis of Lending Club's data.
- [2] J.F. darre, September 30, (2015). Project 1: Lending Club's data.
- [3] Credit Revolving Balance: <https://www.creditcards.com/credit-card-news/glossary/termrevolving-balance.php>
- [4] Lending Club: https://en.wikipedia.org/wiki/Lending_Club
- [5] Dr. Rajeshwari M. Shettar. (2018). AN OVERVIEW STUDY ON P2P LENDING.
- [6] Anahita Namvar, Mohammad Siami, Fethi Rabhi, Mohsen Naderpour. (2017). Credit risk prediction in an imbalanced social lending environment.
- [7] Boris Vallee and Yao Zeng. (2018). Marketplace Lending: A New Banking Paradigm?.
- [8] Qionglin Shan & Mikael Nilsson. (2018). Credit risk analysis with machine learning techniques in peer-to-peer lending market.
- [9] Jonathan Ricardo Guzman. (2015). A Principal Decision: The Case of Lending Club.



- [10] Carlos Serrano-Cinca, Begoña Gutiérrez-Nieto, Luz López-Palacios. (2015). Determinants of Default in P2P Lending.
- [11] Bc. Michal Polena. (2016-17). Performance Analysis of Credit Scoring Models on Lending Club Data.
- [12] Mohammad Mubasil Bokhari. (2019). Credit Risk Analysis in Peer to Peer Lending Data set: Lending Club.
- [13] Matthew Courchene. (2014). A Theoretical Analysis of Peer-to-Peer Lending.
- [14] Mohammad Rafiqul Islam, Tabitha Kemboi. (2019). Project: Lending Club Data Analysis.
- [15] Alexander Bachmann, Alexander Becker, Michel Hilker. (2011). Online Peer-to-Peer Lending.
- [16] Pierre-Yves FESTOC. (2013-2014). Lending Club – P2P Lending Impact Of Loan Description On Loan Performance.
- [17] Lin Zhua , Dafeng Qiu , Daji Ergua , Cai Yinga , Kuiyi Liub. (2019). A study on predicting loan default based on the random forest algorithm.
- [18] Guangyou Zhou, Yijia Zhang and Sumei Luo. (2018). P2P Network Lending, Loss Given Default and Credit Risks.
- [19] JIAYU YAO. (2018). VALUATION OF A FINTECH COMPANY: LENDING CLUB.
- Haitian Lu, Bingzhong Wang, Qing Wu, and Jing Ye. (2020). Fintech and the Future of Financial Service: A Literature Review and Research Agenda.