

2/4/2024

Project Proposal

CMP4011 – Big Data Course

Name	Sec	BN
Khaled Hesham	1	21
Kirollos Samy	2	13
Abdelaziz Salah		
Abdelrahman Noaman		

Problem Statement:

In this project we will be doing credit risk modelling of peer to peer lending Bondora systems, Credit risk modeling involves analyzing data to assess the likelihood that a borrower will default on a loan or fail to meet their financial obligations. Peer-to-peer lending platforms like Bondora facilitate lending directly between individuals, bypassing traditional financial institutions. In the context of Bondora's peer-to-peer lending system, credit risk modeling would likely involve analyzing borrower data, such as credit scores, income, employment history, and other relevant factors, to predict the likelihood of default for each borrower. This helps investors on the platform make informed decisions about which loans to fund and manage their risk exposure.

Dataset :

Link: <https://www.bondora.com/en/public-reports>

Rows: **372541**, Columns: **112**, Size: **303 MB**

Planned approach:

Preprocess the data handling missing or corrupted values, perform EDA on the data to extract insights, Utilize **PySpark's** capabilities to analyze and extract insights from the lending data, enabling data-driven answers to common questions such as:

- **What are the key factors contributing to loan defaults?**
- **How does the default rate vary across different borrower demographics or loan characteristics?**
- **Are there specific patterns or trends in loan repayment behavior over time?**

And more

Present findings using descriptive statistics, data visualizations, and interactive dashboards to facilitate understanding and support informed decision-making processes.