**TRƯỜNG ĐẠI HỌC THĂNG LONG**

-----------------------



**BÁO CÁO DỰ ÁN**

**MÔN HỌC DỮ LIỆU LỚN**

**SINH VIÊN THỰC HIỆN:** ĐỖ HOÀNG ANH - A44498

VƯƠNG HOÀNG GIANG - A44526

**GIẢNG VIÊN HƯỚNG DẪN:** ĐINH THUỶ TIÊN

**HÀ NỘI - 2024**

  A blue text on a black background

Description automatically generated

# HOME CREDIT DEFAULT RISK

**Overall**

|  |  |
| --- | --- |
| **Challenge Statement** | Classifying Customers Experiencing Difficulties in Debt Repayment |
| **Team Name** | Dom Dom |

**Team Members**

|  |  |  |
| --- | --- | --- |
| **Full Name** | **Email Address** | **LinkedIn Profile URL** |
| Đỗ Hoàng Anh | dohoanganh.work@gmail.com | [Linkedin](https://www.linkedin.com/in/hoang-anh-do-a64872254/) |
| Vương Hoàng Giang | gvuonghoang123@gmail.com | Linkedin |

# Content Outline

|  |  |
| --- | --- |
| **Topic** | **Page No.** |
| Overview |  |
| Explanation |  |
| Data Preparation |  |
| Solution |  |
| Application |  |

**MỤC LỤC**

**A. FrontEnd ............................................................................................................................................................ 4**

**I. Khởi tạo ứng dụng ............................................................................................................................... 4**

**1. Tạo ứng dụng ReactJS.......................................................................................................... 4**

**2. Cài đặt các library và framework ......................................................................................... 5**

**3. Cấu trúc thư mục dự án ....................................................................................................... 8**

**4. Thêm một số file cấu hình cho các library và framework vừa cài đặt ............................ 10**

**II. Xây dựng các hàm gọi API ................................................................................................................ 15**

**1. FakeAPI ................................................................................................................................. 1**

**III. Xây dựng FrontEnd các trang web dự án ....................................................................................... 63**

**1. Web User.............................................................................................................................. 63**

**2. Điều hướng trang .............................................................................................................. 257**

**I, Solution Introduction**

1, General introduction

Many people struggle to get loans due to insufficient or non-existent credit histories. And, unfortunately, this population is often taken advantage of by untrustworthy lenders.

Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience, Home Credit makes use of a variety of alternative data--including telco and transactional information--to predict their clients' repayment abilities.

2, **Customer Lifetime Value (CLV)**

CLV in this context is determined by classifying customers based on their likelihood of facing repayment difficulties. By segmenting customers into different risk categories, financial institutions can predict long-term profitability and tailor their lending strategies accordingly. Customers identified as high-risk for repayment issues can be offered customized solutions, such as adjusted loan terms, to mitigate default risk and enhance overall profitability. This approach ensures effective risk management while optimizing CLV.

3, Stakeholders

* **Executive Leadership**: Provides strategic direction, approves budgets, and makes high-level decisions essential for the project's success.
* **Relevant Departments**: This group includes Risk Management, Data Science, IT, and Compliance teams. They develop predictive models, manage risks, ensure technical and regulatory standards, and use these models to effectively address repayment issues.

4, Technical Objectives

* **Data Pipeline**: Collects and processes transaction, card, and user data to ensure a comprehensive dataset for analysis.
* **Machine Learning Model**: Segments customers to identify high-value groups, leveraging algorithms to distinguish between different customer profiles.
* **Visualization Dashboard**: Presents demographic and creditworthiness features of each segment, enabling easy interpretation and decision-making.
* **Data Standardization**: Ensures consistency in data formats and scales, which is crucial for accurate analysis and modeling.
* **VectorAssembler**: Combines multiple feature columns into a single vector, facilitating the use of machine learning algorithms by preparing data in a format suitable for model training.
* **BinaryClassificationEvaluator**: Assesses the performance of classification models by evaluating metrics such as accuracy, precision, and recall, ensuring the reliability of customer segmentation.
* **Intersection Threshold**: Defines a threshold for segment overlaps or decision boundaries, helping to fine-tune the segmentation process and improve the precision of high-value customer identification.

5, P**rocessing Environment:**

* **PySpark**: Used for scalable and efficient data processing. It supports distributed data handling and parallel computation, making it ideal for managing large datasets and executing complex machine learning tasks.

6,