**LHL Final Project Pitch**

Josh Ho

**Project Title**

Credit Scoring Model Using Supervised Learning

Topics: Machine Learning, Probabilistic Classification, Finance, Banking, Data Science, Business, Modeling

**Introduction**

* Banks are large contributors to the economy, having immense monetary influence on how countries, governments, societies, and individuals’ function and develop
* By providing credit to borrowers in search of a capital, borrowers are able to start businesses, purchase homes, pay for school tuition, go on dream vacations and fund other important life decisions
* But before approving credit, banks and lenders require understanding their clients needs and analyzing their financial background in order to decide whether offering credit will be beneficial to both parties, or harmful
* This is where creating a credit scoring model can be developed to optimize and support this process
* Outputs received from model can be used to help inform creditors and borrowers alike on whether they should opt to lend or borrow money respectively

**Goal**

* Develop a credit scoring algorithm that can predict the probability of a borrower defaulting on a line of credit within 2 years
* Deploy model to cloud for prediction use and decision making

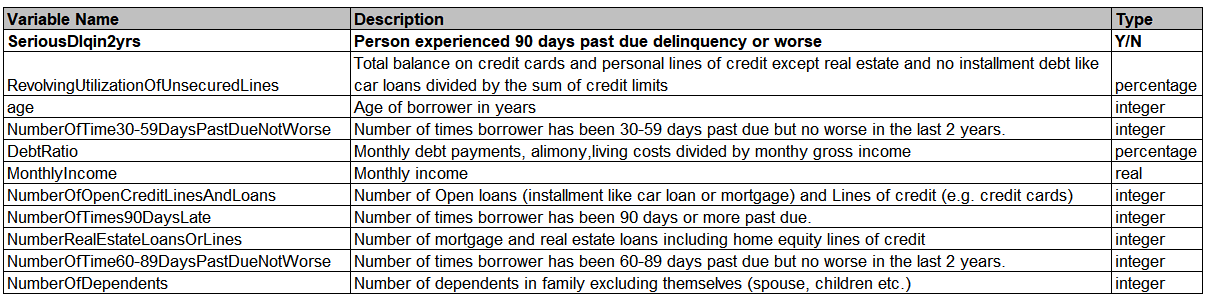
**Description**

* Classification task involving creating a model that will output the probability on whether a borrower will default on a line of credit
* Uses data on past borrowers mapped to several relevant features, as well as a target variable (whether the borrower was 90 days past due credit delinquency or worse)
* Tech Stack
* Python (Pandas, Matplotlib, Plotly, Scikit-learn, Streamlit)
* VS Code
* Docker
* AWS
* GitHub

**Dataset**

* Pulled from Kaggle “Give Me Some Credit” Competition
* 150000 rows of training data with several labeled features, and a target variable (outlined below)
* Train and test set divided into two separate datasets
* <https://www.kaggle.com/competitions/GiveMeSomeCredit/overview>

Feature Descriptions



**Project Workflow**

**Machine Learning Process**

1. Data Collection
2. Data Preprocessing
3. Exploratory Data Analysis
4. Feature Engineering
5. Model Building
6. Model Pipeline
7. Model Evaluation & Tuning
8. Model Deployment
9. Version Control

**Presentation**

1. Create Slides
2. Rehearse