Minh Quang Vu

**Honor Thesis Idea**

I am currently double majoring in Economics and Computer Science so I was hoping I can have my honor thesis in a topic that relates to both majors. I have always been interested in the banking industry. More specifically, I want to know more about loan, default rate and how can we apply data in this industry. I had an internship in this field during freshman summer and I did some literature review on the subject. I am attaching to this file my work so you can get a sense of my interest.

In this project, we will try to build a credit risk model to predict how likely someone will default his/her loan. First, I will have to do more literature review on the subject. Then we will use this dataset from Home Credit to build the model. The main train dataset (300,000 by 127) has more than 300,000 observations, each represents one loan application. Each obs has more than 100 features, including useful information about the customer to predict their default likelihood. These features range from specific information of the application such as income, gender, number of house/children, etc. to general information such as details of the building they live, did they submit this form, etc. Clearly, some features will be more important than the other, and my job will be finding out those features to build the best model possible. There will be a target column with only 2 possible values: 0 and 1. 1 indicates the person has some difficulty in paying back their loan, 0 indicates otherwise.

This project will work as both my honor thesis for Econ and capstone for CS so I plan on using Python and Data Science skills to tackle this problem. After doing EDA (Exploratory Data Analysis), I will apply many Machine Learning algorithms to come up with a best model. I will also use Stata and Econometrics knowledge to gain some insights.

Beside the main train dataset, there are many more datasets containing information of previous loan application of the same set of customers. The whole project is posted on Kaggle as a competition years ago so you can visit this link to learn more:

<https://www.kaggle.com/c/home-credit-default-risk>

If you find my project to be interesting, please send me an email at [mvu@clarku.edu](mailto:mvu@clarku.edu).