

# Loan Rangers

Nikko Gajowniczek, Conrad Barron, Oon Jie Rui

## Predicting Home Credit Default Risk

We are attempting to create a ML model that analyzes a Loan Applicant's available data to determine how likely they are to repay their debt responsibly. This problem is based on a past Kaggle competition, found [here](#). We settled on solving this problem based on a number of factors and constraints.

We didn't want to do a classification problem, and were interested in doing either NLP or image analysis. It turns out the majority of NLP/image problems are classification problems, but we happened to find a problem with a large dataset and continuous target value. We are curious to see if we can apply modern, cutting edge ML techniques to an older problem released 6 years ago.

This model will be able to combine various unrelated features and synthesize a valuable risk analysis. The output of the model will be a float between 0 and 1, a continuous problem. The dataset provided has a vast quantity of features to examine/analyze. We will most likely simplify the problem space initially, and expand if possible. Standard application data is provided; Applicant background, credit history, payment history. Our Training dataset contains 300k entries labeled with a target score, which we will split into training and validation.

Initially, we plan on training our model with basic applicant background. We expect this to be useful for obvious rejections. Beyond that, it will be important to analyze historical data of the applicant for better results.