

# CS236r Project Proposal

## Scoring Systems for Predicting Credit Risk

Alex Lin

Melissa Yu

## 1 Introduction

In this proposal, we apply the Supersparse Linear Integer Model (SLIM) scoring system to make interpretable predictions of loan default risk. We motivate this research by describing how the proposed mechanism can develop an accurate model that is simultaneously easy to understand for the average layman.

## 2 Background

Consumer spending is among the most important determinants of macroeconomic conditions and systemic risk, accounting for around 70% of the U.S. GDP between 2001 and 2010 [2]. As the credit industry has rapidly expanded over the last few decades, the risk associated with consumer lending has multiplied as well. Today, it is more important than ever that financial institutions leverage accurate, interpretable prediction models and algorithms to make the many lending decisions they must process on a daily basis, instead of relying on potentially biased, inaccurate, and un-scalable human discretion. These predictive algorithms leverage characteristics pulled from the consumer credit files collected by credit bureau agencies to discern the riskiness of a loan.

### 2.1 Credit scoring models

*Credit scoring models* are quantitative models widely used by financial institutions to determine the probability of delinquency or default for loan applicants [4, 7]. These models have as their goal accurately classifying loan applicants into one of two groups: “good” customers, who are likely to repay their debt, and “bad” customers, who are denied credit on the basis that they are likely to default.

In order for lenders to leverage scoring algorithms to make smart lending decisions, the model they use must satisfy several requirements [1]:

1. **Predictive power:** The model should make accurate predictions of customer type – good or bad.
2. **Fairness:** The model must comply with the Equal Credit Opportunity Act (ECOA) and Fair Credit Reporting Act (FCRA). It cannot discriminate on the basis of race, age, education, employment history, gender, marital status, or wealth.
3. **Broad coverage:** The model is adaptable enough to apply across large datasets of people.
4. **Data transparency:** The data used in the model is verifiable and correctable by consumers.

5. **Interpretability (consumer-focused):** The model’s classification process can be easily explained to and understood by consumers, and the most important predictive factors can be intuitively identified to aid loan seekers in improving their profile.

## 2.2 Scoring Systems

Scoring systems are linear classification models that only require users to add, subtract and multiply a few small numbers in order to make a prediction [5]. This makes them remarkably easy to understand for the average layman, especially when compared to other machine learning models used for prediction. When working in high-risk settings such as credit scoring, it is imperative that users feel that they can interpret the models that are available to them [3]. If they see an algorithm as nothing more than a black-box, then there is a good chance that they will inherently mistrust its capabilities.

This is exactly why scoring systems may be desirable for high-risk problems in domains such as medicine and law [5, 6] in which the cost of a misclassification can have grave consequences. Using classical optimization techniques in Mixed Integer Programming (MIP), Ustun and Rudin were able to develop an algorithm called the Supersparse Linear Integer Model (SLIM) that can generate generic scoring systems while taking into account factors such as algorithmic fairness and interpretable coefficients [5]. Although SLIM produces simple models, its models also seem to exhibit high accuracy under AUC measures. The authors further developed RiskSLIM, which can produce scoring systems that not only make a prediction, but can also incorporate risk assessment into that prediction [6].

## 3 Research Proposal

The previously described attributes of the credit risk prediction problem motivate us to suggest a scoring system-based solution. We will apply the work of Ustun and Rudin to the domain of finance and implement a SLIM-based algorithm that can produce a scoring system to determine if an individual will default on his/her loan. In creating the scoring system, we will enforce constraints that are specific to our domain of study, such as fairness requirements as detailed under ECOA and FCRA. We will assess the model’s accuracy using AUC analysis and also examine the tradeoff between the algorithm’s runtime and optimality gap.

Given enough time, we would also like to incorporate default risk into our model using principles from RiskSLIM.

### 3.1 Dataset and Implementation

There are many different datasets online for loan default prediction. We have not settled on our dataset of choice yet, but we have examined a few possibilities:

- <https://www.kaggle.com/c/GiveMeSomeCredit>
- <https://www.kaggle.com/c/loan-default-prediction>
- <http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29>
- <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

The key will involve selecting a dataset that covers a desirable feature space. We plan to do our implementation of SLIM in Python and use a library such as **PyCPX** to solve the MIPs.

### 3.2 Immediate First Steps

There are two main avenues on which we will focus our initial efforts. The first is *model-based* and will require us to develop a deep intuition for Ustun and Rudin’s papers [5, 6]. We will read these papers very closely and understand them to the point at which we would feel comfortable presenting them in a lecture.

The second is *application-based* and will require us to compile a list of reasonable credit scoring-related constraints to encode into the SLIM algorithm. For example, one such constraint could be ensuring that the scoring system does not discriminate based on race.

### 3.3 Support from the Teaching Staff

It would be great if the teaching staff could help us choose a dataset that they feel would work well for this project. Any papers related to fairness or interpretability in the financial domain, or any further work on scoring systems would also be greatly appreciated.

## References

- [1] FICO. 5 building blocks of a great scoring system, 2015.
- [2] A. E. Khandani, A. J. Kim, and A. W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787, 2010.
- [3] L. J. Kleinberg Jon and S. Mullainathan. A guide to solving social problems with machine learning. 2016.
- [4] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng. Building credit scoring models using genetic programming. *Expert Systems with Applications*, 29(1):41–47, 2005.
- [5] B. Ustun and C. Rudin. Supersparse linear integer models for optimized medical scoring systems. *Machine Learning*, 102(3):349–391, 2016.
- [6] B. Ustun and C. Rudin. Optimized risk scores. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1125–1134. ACM, 2017.
- [7] D. West. Neural network credit scoring models. *Computers & Operations Research*, 27(11-12):1131–1152, 2000.