

Title: Chunking: A practical approach to manage consent-based data sources in credit scoring

Abstract: The increasing number of data sources available to credit bureaus and the growing control individuals have over their data in credit scoring applications necessitate a tractable approach to integrating available data in the modelling process. The challenge posed by consent-based data sources is that they may lead to 2^n distinct subpopulations depending on which of the n data sources individuals grant permission for. We introduce "chunking," a concept related to "dependent staging" (Anderson, 2007) that allows for sequential training on the expanding number of data sources while ensuring transparency, fairness, and consistency in their integration. The chunking process begins with training a model on a base set of variables and then subsequently adding new data sources by fixing the effects of the base variables. We will discuss how this method mitigates the challenges of integrating consent-based data sources, handling scarcity in data sources, and managing drift in subpopulations.

In our talk, we will explore the pros and cons of chunking, including its ability to control adverse effects that skewed data sources might have on the overall model and its usefulness in streamlining the modelling process for the possible combinations of subpopulations which reduces the overhead of scorecard implementation and maintenance. Chunking may result in a loss of the models' statistical power due to the possibility of data sources being statistically dependent. We will discuss how to quantify this loss of statistical power and determine when it is an acceptable trade-off. Furthermore, we will examine situations in which differences between subpopulations are significant enough to warrant entirely separate scorecards. By employing the chunking technique, credit scoring models can better adapt to the evolving landscape of data sources while upholding the principles of consistency, fairness, and transparency.