

# **Scalable Methodologies for Big Data Analysis: Integrating Advanced Statistical Models and Optimal Designs**

Min Yang

Department of Mathematics, Statistics, and Computer Science, University of Illinois-  
Chicago, USA

## Abstract

The formidable challenge presented by the analysis of big data stems not just from its sheer volume, but also from the diversity, complexity, and the rapid pace at which it needs to be processed or delivered. A compelling approach is to analyze a sample of the data, while still preserving the comprehensive information contained in the full dataset. Although there is a considerable amount of research on this subject, the majority of it relies on classical statistical models, such as linear models and generalized linear models, etc. These models serve as powerful tools when the relationships between input and output variables are uniform. However, they may not be suitable when applied to complex datasets, as they tend to yield suboptimal results in the face of inherent complexity or heterogeneity. In this presentation, we will introduce a broadly applicable and scalable methodology designed to overcome these challenges. This is achieved through an in-depth exploration and integration of cutting-edge statistical methods, drawing particularly from neural network models and, more specifically, Mixture-of-Experts (ME) models, along with optimal designs.