# 中央研究院統計科學研究所
# 學 術 演 講

講 題：Advancing Credit Scoring Accuracy through Explainable AI: A
　　　　Rescaled Cluster-then-Predict Approach

演 講 人：Prof. Hung-Yin Chen（ 陳虹吟 助理教授 ）
　　　　　Department of Accounting, Tamkang University

時 間： 2024-09-02(Mon.) 10:30-12:00

地 點：Auditorium, B1F, Institute of Statistical Science；The tea
　　　　reception will be held at 10:10.

備 註： Online live streaming through Cisco Webex will be available.

## Abstract

Accurately predicting default risk is paramount for financial institutions in credit scoring. Traditional methods often face significant challenges when dealing with imbalanced datasets, resulting in biased models and unreliable predictions. This paper extends Teng et al. (2024) to enhance the accuracy and interpretability of credit scoring models through a novel rescaled cluster-then-predict approach. Initially, the data undergoes preprocessing, resulting in a data matrix of  features. A key aspect of our methodology is rescaling the feature set with a p-variate weight vector $w = (w(1),....,w(p))$, leading to the rescaled data set where the i-th feature is multiplied by $w(i)$ for $i = 1, 2,...,p$ . We define the total entropy as the w-weighted average of individual entropies at each cluster. Then, the optimal w is determined by minimizing the total entropy for the rescaled data matrix, ensuring more homogeneous clusters regarding default risk. In the training phase, data is clustered, and separate models are trained for each cluster. In the testing phase, the appropriate cluster for each new data point is identified, and the corresponding model is used for prediction, ensuring accurate and reliable risk assessment. This proposed rescaled cluster-then-predict approach is validated through both in-sample and out-of-sample performance metrics, specifically focusing on the area under the curve (AUC) for model evaluation, and progressively improves model robustness and interpretability.

中 央 研 究 院

統計科學研究所