SIMPOSIO DESCENTRALIZADO DE MATEMÁTICA Y ESTADÍSTICA
(14 de Setiembre de 2023)

Facultad de Ciencias Químicas, Físicas y Matemáticas

Machine Learning Modeling for Probability of Default

Walter Quispe Vargas Ph.D.
Departamento Académico de Matemática y Estadística
Universidad Nacional de San Antonio Abad del Cusco
Av. de la Cultura, Nro. 733, Cusco - Perú

31 de julio de 2023

Resumen: Estimar la probabilidad de impago (*Probability of Default*) es fundamental para medir el riesgo de crédito de toda entidad financiera. Tradicionalmente la probabilidad de impago se mide mediante métodos estadísticos. Los avances recientes en modelar el riesgo de crédito se basan en el aprendizaje automático (*Machine Learning*), siguiendo el ciclo de vida de los proyectos de ciencia de datos. En esta presentación, se muestra la implementación (*R+Python*) y uso del algoritmo "*Extreme Gradient Boosting*" (*XGBoost*) de aprendizaje automático para construir un modelo que prediga la probabilidad de impago en función de algunos indicadores de riesgo del portafolio de clientes de una institución financiera. Específicamente, proponemos un proceso de selección de variables robusto, usando métodos estadísticos y algoritmos de aprendizaje automático. Exploramos el uso de modelos estadíticos y algoritmos de aprendizaje automático, evaluamos sus desempeños para el proceso de selección de modelo. Discutimos los desafíos de clasificación supervisada desbalanceada y explicabilidad de los modelos de aprendizaje automático. Se muestra que el algoritmo *XGBoost* es más efectivo para nuestro conjunto de datos, y supera a resto de los métodos y algoritmos considerados, en términos de la métrica de evaluación area bajo la curva (AUC).

Palabras clave: Probabilidad de Impago, XGBoost, Clases Imbalanceadas.

Referencias

[1] Shi, S., Tse, R., Luo, W. et al. (2022). "Machine learning-driven credit risk: a systemic review." *Neural Computing and Applications* 34, 14327-14339. https://doi.org/10.1007/s00521-022-07472-2

[2] Rogojan, L. ,Croicu, A. and Iancu, L. (2023). "Modern Approaches in Credit Risk Modeling: A Literature Review." *Proceedings of the International Conference on Business Excellence* ,17(1) 1617-1627. https://doi.org/10.2478/picbe-2023-0145

[3] Aurélien, G. (2019). "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems." *O'Reilly Media, Inc.*

[4] Kuhn, M. and Silge, J. (2022). "Tidy Modeling with R: A Framework for Modeling in the Tidyverse." *O'Reilly Media, Inc.*

[5] Bellini, T. (2019). "IFRS 9 and CECL Credit Risk Modelling and Validation: A Practical Guide with Examples Worked in R and SAS." *United Kindom, Elsevier Science.*