

Predicción del Riesgo de Bancarrota Empresarial mediante Machine Learning: Un Enfoque Comparativo con Random Forest y Gradient Boosting

La presente tesis evaluará la capacidad de modelos de *machine learning* para estimar la probabilidad de riesgo de bancarrota (*financial distress*) en empresas, utilizando como insumo principal información contable proveniente de estados y razones financieras. A partir de datos históricos, se construirá un conjunto de variables que representan dimensiones clave del desempeño corporativo –liquidez, apalancamiento, rentabilidad, cobertura e indicadores de eficiencia– con el objetivo de identificar patrones que preceden episodios de deterioro financiero severo que pueden culminar en bancarrota, conforme a una definición operativa establecida en el estudio.

El componente central del trabajo será la comparación sistemática entre enfoques predictivos basados en *machine learning* y modelos tradicionales utilizados en la literatura y en la práctica financiera. En el bloque tradicional se considerarán aproximaciones clásicas como el modelo Z-score de Altman (1968). En el bloque de *machine learning*, la investigación se centrará en tres algoritmos supervisados de ensamble particularmente efectivos para este tipo de problemas: *Random Forest*, *Gradient Boosting* (potenciación del gradiente) y *XGBoost*. Estos algoritmos han demostrado capacidad superior para capturar relaciones no lineales y efectos de interacción entre variables financieras (Barboza et al., 2017; Zięba et al., 2016).

Random Forest, propuesto por Breiman (2001), construye múltiples árboles de decisión mediante *bootstrap aggregating* (bagging) y selección aleatoria de características, lo que reduce la varianza y mejora la generalización. Por su parte, *Gradient Boosting* construye árboles secuencialmente, donde cada árbol corrige los errores del anterior, optimizando una función de pérdida mediante descenso del gradiente. Implementaciones modernas como *XGBoost* (Chen & Guestrin, 2016) han mostrado resultados excepcionales en competencias de ciencia de datos y aplicaciones financieras reales.

La comparación se realizará bajo un marco metodológico homogéneo, priorizando validación fuera de muestra (*out-of-sample*), control de sobreajuste mediante validación cruzada, manejo apropiado del desbalance de clases (dado que los eventos de bancarrota son relativamente raros) y métricas de desempeño relevantes para clasificación y sistemas de alerta temprana, incluyendo precisión, *recall* y *F1-score*.

Finalmente, la investigación buscará determinar en qué condiciones los modelos de *machine learning* ofrecen ventajas sobre los enfoques tradicionales, y cuáles son los compromisos asociados en términos de interpretabilidad, estabilidad y requerimientos de datos. Con base en los resultados, se propondrán implicaciones aplicables a contextos de gestión de riesgo, monitoreo crediticio y análisis financiero corporativo, incluyendo recomendaciones sobre selección de variables, criterios de evaluación y lineamientos para implementar un esquema de predicción de riesgo de bancarrota que sea robusto, replicable y útil para la toma de decisiones empresariales y de inversión.

Referencias

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
- Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58, 93-101.

Sobre el autor:

Jorge Emiliano Ornelas Salgado

Estudiante de la Licenciatura en Banca e Inversiones en la Universidad de las Américas Puebla (UDLAP). Interesado en la aplicación de técnicas cuantitativas y de inteligencia artificial al análisis financiero y la gestión de riesgos.

Contacto: jorge.ornelasso@udlap.mx

Tutor

Dr. Jaime González Maiz Jiménez

Doctor en Ciencias Administrativas con especialidad en Finanzas por la EGADE Business School sede Monterrey. Profesor de tiempo completo en la UDLAP desde el 2015. Apasionado de la docencia e investigación. Es miembro del Sistema Nacional de Investigadores y su campo de especialidad son las finanzas corporativas, mercados financieros y la aplicación de la inteligencia artificial a las finanzas.

Contacto: jaime.gonzalezmaiz@udlap.mx