

# Aprendizaje estadístico en la ciencia actuarial

## 1. PROPUESTA DE TESIS

**Asesor** Dr. Alfredo Garbuno Iñigo.

**Estudiante** Luis Eduardo Ferrari.

**Clave única** preguntar.

**email-contacto** preguntar

## 2. CONTEXTO DEL PROYECTO

El área de inteligencia artificial, en particular el aprendizaje automático a través de modelado predictivo, ha permeado muchas áreas en la ciencia y se ha utilizado para distintas aplicaciones de la vida cotidiana. Se espera que ayude a mejorar la productividad, el bienestar social, y que permita informar las decisiones que hagan frente a los problemas mundiales que tenemos en puerta. Es así, que el uso de modelos predictivos para la toma de decisiones ha levantado preguntas sobre su transparencia, sesgos y corresponsabilidad, por mencionar algunas cuestiones. Esto ha llevado a organismos internacionales como la OECD y la Unión Europea a generar manuales y guías de uso para su correcta implementación en temas sociales.

En el ramo de la actuaría se han utilizado modelos estadísticos con el objetivo de analizar datos que permitan a las aseguradoras tomar mejores decisiones como monto de reservas y tarificación. Es natural pensar en el uso de modelos predictivos para estas tareas. Sin embargo, poco se encuentra en la literatura.

## 3. OBJETIVOS

El estudiante extenderá sus conocimientos adquiridos en los cursos de Aprendizaje Estadístico e Inferencia Matemática para atacar un problema predictivo en el área de seguros. En particular, estudiará el concepto de regresión cuantílica para poder construir modelos que predigan intervalos. La construcción y ajuste del modelo se basa en la minimización de una función de error calibrada a cierto nivel de pérdida que corresponde a los cuantiles en los datos observados. Esto permite tener predicciones por intervalo en lugar de tener predicciones puntuales. La tesis busca 1) realizar predicciones para un conjunto de pruebas de regresión con distintas estructuras probabilísticas en los residuales y 2) generar un estudio de cobertura de los intervalos que pondrá a prueba si dicha cobertura es nominal bajo distintos supuestos en las colas de la distribución. Esto permitirá evaluar y entender sus limitaciones en seguros, donde métricas de incertidumbre son fundamentales para realizar tareas predictivas.

#### 4. REFERENCIAS

- Guelman, L., Guillén, M., & Pérez-Marín, A. M. (2012, May). Random forests for uplift modeling: an insurance customer retention case. In International conference on modeling and simulation in engineering, economics and management (pp. 123-133). Springer, Berlin, Heidelberg.
- Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of machine learning research*, 7(6).
- Smith, K. A., Willis, R. J., & Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: A case study. *Journal of the operational research society*, 51(5), 532-541.
- Yeo, A. C., Smith, K. A., Willis, R. J., & Brooks, M. (2001, May). Modeling the effect of premium changes on motor insurance customer retention rates using neural networks. In International Conference on Computational Science (pp. 390-399). Springer, Berlin, Heidelberg.