UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

Propuesta TFG. Curso 2026/2027

GRADO: Grado en Estadística

CÓDIGO DEL TFG: 223-023-2026/2027

1. DATOS BÁSICOS DEL TFG:

Título: Bases estadísticas de modelos de clasificación basados en boosting como XGBoost, LightGBM, o CatBoost

Descripción general (resumen y metodología):

Hoy en día el aprendizaje estadístico se ha consolidado como uno de los pilares para trabajar con problemas de predicción. Existen diversos modelos, pero uno de los más simples es el árbol de decisión. Éstos destacan sobre todo debido a su fácil interpretación, aunque cuentan con ciertas restricciones y limitaciones, como el sobreajuste y una alta varianza.

Para intentar superar estas limitaciones, surgieron las técnicas de ensamblado, que combinan predicciones de múltiples árboles para obtener resultados más robustos. A su vez existen diferentes formas de combinar los árboles, siendo una de las más destacadas la basada en boosting, una técnica muy potente con la que es posible corregir varias de las limitaciones anteriores.

Dentro de los métodos de boosting, encontramos distintas variantes, como pueden ser XGBoost, LightGBM y CatBoost. Éstas nos permiten resolver problemas de clasificación complejos, con muchas menos limitaciones que los modelos de ensamblado originales, y actualmente estos dominan en su campo debido a su capacidad de trabajar con grandes volúmenes de datos sin perder ni precisión ni eficiencia.

En este proyecto se pretende profundizar en las bases estadísticas de estos modelos, así como analizar y comparar la potencialidad y limitaciones de las distintas variantes.

Tipología: Trabajos bibliográficos sobre el estado actual de una temática relacionada con el Grado.

Objetivos planteados:

- Realizar una revisión bibliográfica del estado del arte de los modelos basados en boosting.
- Comprender las bases estadísticas de estos modelos.
- Comparar las diferentes variantes de boosting.
- Implementar un caso de uso sobre datos reales.

Bibliografía básica:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2.ª ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2.ª ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. En I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 30 (pp. 3146–3154). Curran Associates, Inc.

- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. En S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 31 (pp. 6638–6648). Curran Associates, Inc.

Recomendaciones y orientaciones para el estudiante:

Cursar las asignaturas de Técnicas Avanzadas de Estadística Multivariante y Minería de Datos.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: FRANCISCO JAVIER ARNEDO FERNÁNDEZ

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: arnedo@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: VICTOR BARRERO ROMERO

Correo electrónico: victorbarrero@correo.ugr.es