



Línea de Trabajo fin de Máster 2024-2025

(Fecha última actualización:04/10/2024)

Máster Universitario en Estadística Aplicada.	
Título	Evaluación y Comparación de Técnicas de Preprocesamiento y Modelos de Clasificación en Datos No Balanceados
Tipo	INVESTIGACIÓN <input checked="" type="checkbox"/> ORIENTACIÓN PRÁCTICA <input type="checkbox"/>
Número de alumnos admitidos	1
Profesor(es)/ email	José Antonio Sáez (joseasaezm@ugr.es), Pedro Carmona Sáez (pcarmona@ugr.es),
Descripción	<p>El problema de clasificación con datos no balanceados se da cuando el número de muestras de cada clase es muy desigual y es común en diversas áreas como, por ejemplo, el diagnóstico de enfermedades raras. En estos casos, la distribución desigual entre clases puede afectar negativamente el rendimiento de los clasificadores tradicionales, que suelen estar optimizados para conjuntos de datos balanceados. Esta propuesta de Trabajo Fin de Máster (TFM) tiene como objetivo la evaluación de diferentes técnicas de clasificación en conjuntos de datos no balanceados, incluyendo el uso de estrategias de preprocesamiento para mejorar la precisión de los modelos.</p> <p>Se trabajará con datos reales para evaluar diferentes modelos de clasificación como SVM, Random Forest, y Redes Neuronales y técnicas de preprocesamiento, con el objetivo de identificar qué combinación de preprocesamiento y clasificador es más efectiva en contextos no balanceados de datos biomédicos.</p>
Objetivos particulares	<p>Desarrollar tareas de investigación en un entorno multidisciplinar</p> <p>Conocer algunas de las principales técnicas estadísticas de preprocesamiento y clasificación</p> <p>Implementar metodologías en R para análisis de datos reales</p>
Prerrequisitos y recomendaciones	<ul style="list-style-type: none"> Conocimientos avanzado de lenguaje R y Python
Plan de trabajo	<ul style="list-style-type: none"> - Revisión bibliográfica y selección de conjuntos de datos. - Implementación de técnicas de preprocesamiento y entrenamientos iniciales de modelos. - Comparación de resultados y ajustes en los análisis con datos reales. - Redacción del memoria de TFM y presentación.
Competencias generales y específicas	<p>CB: 6, 7, 8, 9, 10</p> <p>CG: 1, 2, 3, 6, 9</p> <p>CE: 10, 13, 15, 16, 17, 18, 19, 20, 22, 23, 26, 28, 29</p>
Bibliografía	<ul style="list-style-type: none"> - H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. - Chawla, N. V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research.

**Máster Oficial en
Estadística Aplicada.**

Dpto. Estadística e Investigación Operativa.
Universidad de Granada.



**UNIVERSIDAD
DE GRANADA**