



Línea de Trabajo Fin de Máster

Máster en Estadística. CURSO ACADÉMICO 2023-2024	
Título	Técnicas de corrección de errores en datos categóricos en problemas de clasificación
Tipo	INVESTIGACIÓN <input checked="" type="checkbox"/> ORIENTACIÓN PRÁCTICA <input type="checkbox"/>
Número de alumnos	1
Profesor(es)/ email	José Antonio Sáez Muñoz (joseasaezm@ugr.es)
Descripción	La adquisición y el procesamiento de datos están sujetos a imperfecciones. Este hecho puede dar lugar a la presencia de errores o ruido, los cuales implican un deterioro de los modelos de clasificación creados a partir de los datos. En la literatura especializada, existen distintas técnicas de correcciones de errores para datos categóricos en problemas de clasificación, que pretenden paliar las consecuencias negativas de los errores. Este trabajo pretende estudiar las técnicas existentes en este ámbito en la literatura, su implementación mediante un paquete software y el análisis de su eficacia sobre conjuntos de datos reales de distintas características.
Objetivos particulares	<ul style="list-style-type: none"> - Desarrollar tareas de investigación en un entorno multidisciplinar. - Entender algunos de los principales retos en la investigación en el campo. - Aplicar técnicas de preprocesamiento de datos y clasificación.
Prerrequisitos y recomendaciones	<ul style="list-style-type: none"> • Conocimientos avanzados de lenguaje R. • Conocimiento de técnicas de clasificación.
Plan de trabajo	<ul style="list-style-type: none"> - Revisión bibliográfica - Análisis comparativo de las diferentes metodologías y paquetes de R - Implementación de funciones en R para corrección de errores - Aplicación a conjuntos de datos reales - Interpretación de resultados
Competencias generales y específicas	CB: 6, 7, 8, 9, 10 CG: 1, 2, 3, 6, 9 CE: 10, 13, 15, 16, 17, 18, 19, 20, 22, 23, 26, 28, 29
Bibliografía	<ul style="list-style-type: none"> • C. Teng, Correcting noisy data, in: Proceedings of the Sixteenth International Conference on Machine Learning, Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999, pp. 239-248. • Y. Yang, X. Wu, X. Zhu, Dealing with predictive-but-unpredictable attributes in noisy data sources, in: Proceedings of the Eighth European Conference on Principles and Practice of Knowledge Discovery in Databases, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 471-483. • C. E. Brodley and M. A. Friedl, "Identifying mislabeled training data", J. Artif. Intell. Res., vol. 11, pp. 131–167, Aug. 1999.