

PROPUESTA DE TRABAJO FIN DE MÁSTER (Modalidad B)
Curso 2022-2023
MÁSTER EN TÉCNICAS ESTADÍSTICAS

Título	Distancia de Levenshtein como clasificador de textos
Nombre de la Empresa Tutor/a en la empresa Tutor/a académico/a	Trucksters Lucía Tajuelo López
Descripción del trabajo a realizar	<p>Las empresas que trabajan con bases de datos grandes pueden tener duplicados en sus bases de datos. Detectar estos duplicados es una tarea que requiere tiempo, especialmente si los datos son palabras en texto plano.</p> <p>El objetivo de este trabajo es desarrollar un algoritmo que permita detectar duplicados en nuestra base de datos. Este algoritmo, basado en la distancia de Levenshtein, se implementará para clasificar nombres de localizaciones postales y nombres de clientes.</p>
Recomendaciones	El algoritmo se implementará en Python. Lectura introductoria https://blog.paperspace.com/measuring-text-similarity-using-levenshtein-distance/
Fechas de las practicas	
Lugar de trabajo y horario	Remoto y flexible
Convenio	
Otras observaciones	