

TRABAJO FIN DE MÁSTER

Máster en Ciencia de Datos e Ingeniería de Datos en la Nube

WiDS Dathaton 2024 - Challenge 2: Modelos de regresión para estimación del periodo de diagnóstico metastático

Autor: Luna Jiménez Fernández

Tutor: Juan Carlos Alfaro Jiménez

Junio, 2025

*Dedicado a la gente que, pese a todo,
sigue persiguiendo sus sueños.
Nunca os rindáis.*

Declaración de autoría

Yo, **Luna Jiménez Fernández**, con DNI **47092045M**, declaro que soy la única autora del Trabajo Fin de Master titulado ***“WiDS Dathon 2024 - Challenge 2: Modelos de regresión para estimación del periodo de diagnóstico metastático”***, que el citado trabajo no infringe las leyes en vigor sobre propiedad intelectual, y que todo el material no original contenido en dicho trabajo está apropiadamente atribuido a sus legítimos autores.

Albacete, a ... de **Junio de 2025**

Fdo.: **Luna Jiménez Fernández**

Resumen

TODO RESUMEN AQUI

Abstract

TODO ABSTRACT HERE

Agradecimientos

En primer lugar, quiero agradecer a todos mis compañeros y amigos del grupo de **Sistemas Informáticos y Minería de Datos (SIMD)** - y, especialmente, a mi amigo y director **Juan Carlos Alfaro Jiménez** - por su apoyo, recursos y consejos durante la realización de este trabajo. Aunque ya no sea formalmente parte de este grupo, siempre me sentiré vinculada a él.

Además, quiero agradecer a mis amigos y familia del **Curso de Comic Online de la Escola Joso - Arai, Aina, Arkaitz, Clara, Irene, Martín, Pau, Rafi...** -, con los que compartí un proyecto de gran importancia personal, mi primer comic publicado, y en los que he encontrado un grupo al que pertenecer. Muchas gracias por todo.

Finalmente, quiero agradecer a **mi familia y seres queridos** - tanto los que me acompañan presencialmente como los que se encuentran a distancia. Vuestro apoyo y cariño continuo me ha ayudado a seguir adelante y acabar este trabajo a pesar de todas las dificultades.

Índice general

1	Introducción	1
1.1	Objetivos	1
1.2	Estructura de la memoria	1
2	Revisión de técnicas	3
2.1	Ciencia de datos	3
2.1.1	Búsqueda de hiperparámetros y validación cruzada	3
2.2	Modelos propuestos	3
2.2.1	Modelos lineales	3
2.2.2	Máquinas de vectores de soporte	3
2.2.3	Árboles de decisión y ensembles	3
3	Estudio del problema	5
3.1	Definición del problema	5
3.1.1	Atributos del problema	5
3.2	Análisis exploratorio de datos	5
3.2.1	Variable objetivo - distribución y comportamiento	5
3.2.2	Valores perdidos	5
3.2.3	Atributos categóricos	5
3.2.4	Atributos numéricos	5
3.2.5	Variables geográficas, sociales y económicas	5
4	Preprocesamiento del conjunto de datos	7
4.1	Selección de atributos	7
4.2	Procesamiento de los datos	7

5	Modelado y experimentación	9
5.1	Selección de modelos	9
5.2	Experimentación	9
5.2.1	<i>Ajuste de hiperparámetros y selección de subconjuntos de atributos</i>	9
5.2.2	<i>Validación y selección de modelo final</i>	9
5.3	Análisis de resultados	9
5.3.1	<i>Rendimiento de los subconjuntos de hiperparámetros</i>	9
5.3.2	<i>Rendimiento de los modelos entrenados</i>	9
5.3.3	<i>Rendimiento del modelo final</i>	9
6	Aplicación web	11
6.1	Aplicación para usuario - predicción individual	11
6.2	Aplicación <i>batch</i> - predicción en grupo	11
7	Conclusiones	13
7.1	Trabajo futuro	13
	Referencia bibliográfica	15
A	Anexo 1	17

Índice de figuras

Índice de tablas

Índice de algoritmos

Índice de listados de código

1. Introducción

1.1. Objetivos

1.2. Estructura de la memoria

2. Revisión de técnicas

2.1. Ciencia de datos

2.1.1. Búsqueda de hiperparámetros y validación cruzada

2.2. Modelos propuestos

2.2.1. Modelos lineales

2.2.2. Máquinas de vectores de soporte

2.2.3. Árboles de decisión y *ensembles*

3. Estudio del problema

3.1. Definición del problema

3.1.1. Atributos del problema

3.2. Análisis exploratorio de datos

3.2.1. Variable objetivo - distribución y comportamiento

3.2.2. Valores perdidos

3.2.3. Atributos categóricos

3.2.4. Atributos numéricos

3.2.5. Variables geográficas, sociales y económicas

4. Preprocesamiento del conjunto de datos

4.1. Selección de atributos

4.2. Procesamiento de los datos

5. Modelado y experimentación

5.1. Selección de modelos

5.2. Experimentación

5.2.1. Ajuste de hiperparámetros y selección de subconjuntos de atributos

5.2.2. Validación y selección de modelo final

5.3. Análisis de resultados

5.3.1. Rendimiento de los subconjuntos de hiperparámetros

5.3.2. Rendimiento de los modelos entrenados

5.3.3. Rendimiento del modelo final

6. Aplicación web

6.1. Aplicación para usuario - predicción individual

6.2. Aplicación *batch* - predicción en grupo

7. Conclusiones

7.1. Trabajo futuro

Referencia bibliográfica

A. Anexo 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.