



# Predicción del género de libros

---

Miguel Gutiérrez Pérez  
Mario Lozano Cortés  
Alba Reinders Sánchez  
Alejandro Valverde Mahou

---

# Índice

- **Introducción. ¿Qué pretendemos?**
- **Preprocesado de datos**
- **Conceptos teóricos**
- **Clasificación de 1 género**
- **Clasificación multigénero**
- **Conclusiones ¿Qué hemos aprendido?**

# ¿Qué problema vamos a solucionar?

*Área de aplicación = Text Mining*

- Las librerías y bibliotecas necesitan catalogar una gran cantidad de escritos
  - Necesidad de métodos de clasificación automática
  - Usaremos resúmenes y metadatos de los libros
-

# ¿Qué datos necesitamos?

**Objetivo:** Obtener el género con los mínimos datos posibles. Inicialmente se pretende utilizar solo título y resumen.

**Dataset:** Goodreads' Best Books Ever



Se deben tratar algunas cuestiones previas...

Título	Resumen	Géneros
The Hunger Games	Winning will make you famous. Losing means certain death.The nation of Panem, formed from a post-apocalyptic North America, is a country that consists of a wealthy Capitol region surrounded by 12 poorer districts....	Young Adult   Fiction   Dystopia   Fantasy   Science Fiction
Animal Farm	مزرعة الحيوانات هي رائعة جورج أورويل الخالدة.. اختيرت دائماً ضمن أفضل الأعمال الأدبية في القرن العشرين .. تُرجمت لأكثر من سبعين لغة.. وقرأها الملايين في كل أنحاء العالم طُبه...	Classics   Fiction   Dystopia   Fantasy   Literature   Academic   School   Politics   Science Fiction   Novels   Academic  Read For School

Original: 54302 columns\*12 rows



# TEXT MINING

PALABRAS



NÚMEROS



# “HACE UN ESPLÉNDIDO DÍA”

Diferentes formas de vectorizarlo

---

---

# Codificación con valor único

HACE 1

UN 2

ESPLÉNDIDO 3

DÍA 4

HACE UN ESPLÉNDIDO DÍA = [1 2 3 4]

---

---

## One-hot encoding

HACE [1 0 0 0 0 0]

UN [0 1 0 0 0 0]

ESPLÉNDIDO [0 0 1 0 0 0]

DÍA [0 0 0 1 0 0]

HACE UN ESPLÉNDIDO DÍA = [1 1 1 1 0 0]

---



---

# Embedding

HACE [2.32 7.25]

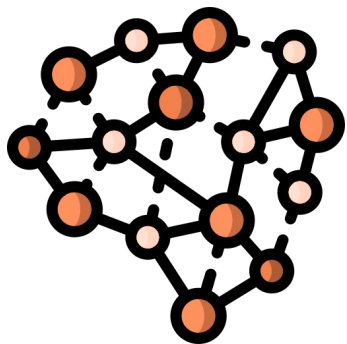
UN [0.89 4.23]

ESPLÉNDIDO [0.75 8.65]

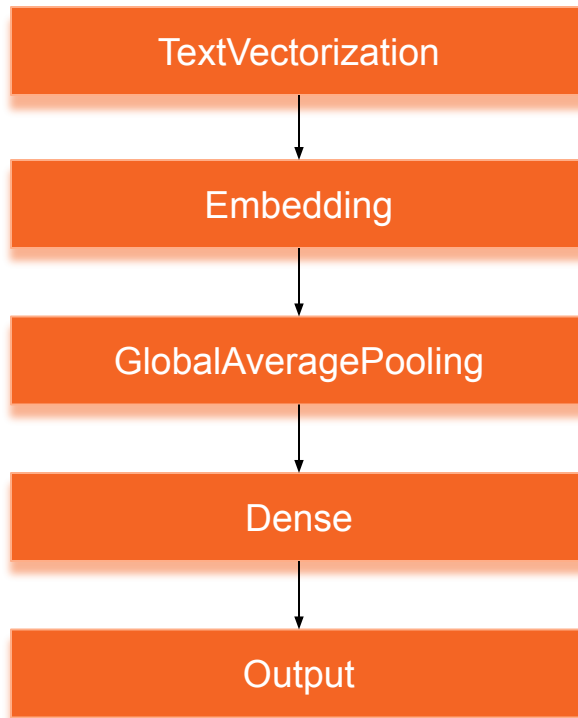
DÍA [2.65 3.00]

HACE UN ESPLÉNDIDO DÍA =  $[[2.32 \ 7.25] \ [0.89 \ 4.23] \ [0.75 \ 8.65] \ [2.65 \ 3.00]]$

---



# Arquitectura del modelo



# Clasificación de 1 género

194 géneros distintos

Ej. Ficción, Aventura, Fantasía

Aventura, Miedo, Romance, Ficción, Fantasía,...

[0, 0, 0, 1, 0, ...]



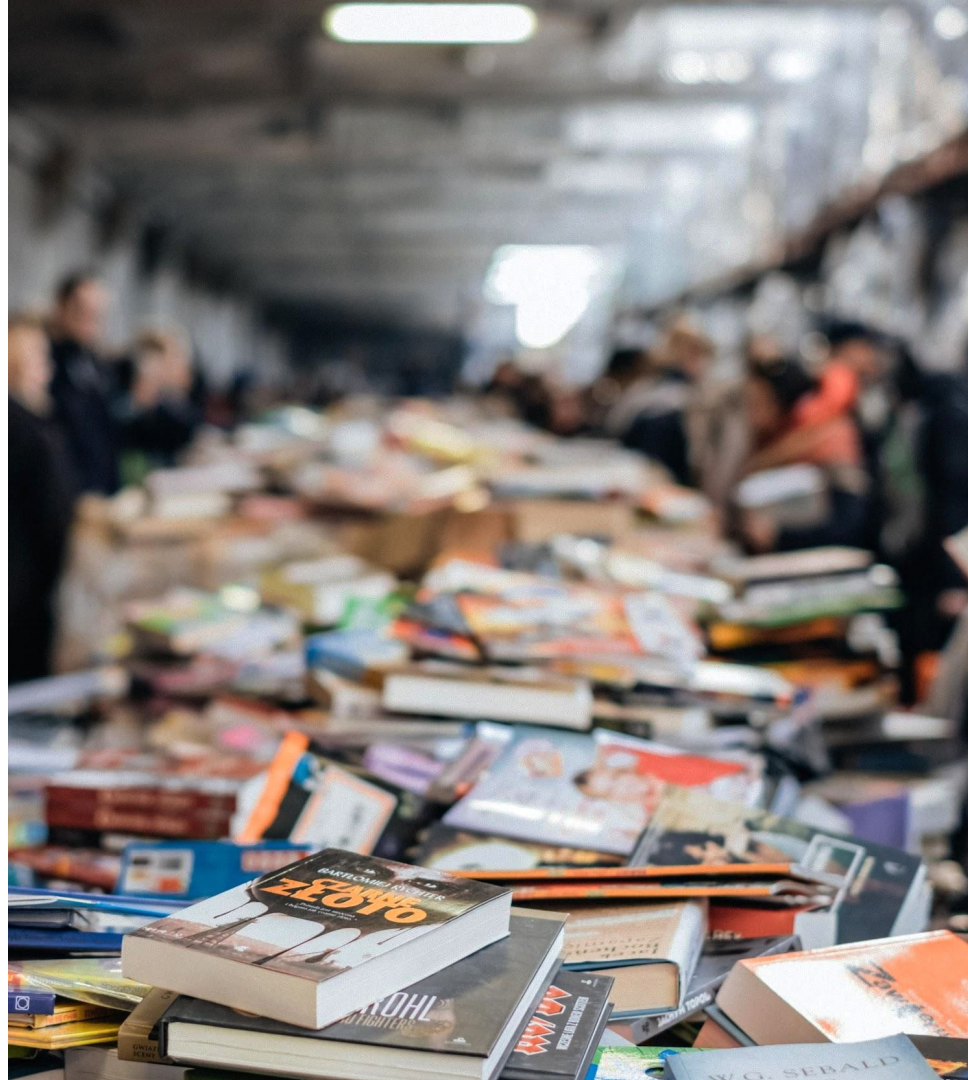
# Clasificación multigénero

625 géneros distintos

Ej. Ficción, Aventura, Fantasía

Aventura, Miedo, Romance, Ficción, Fantasía,...

[1, 0, 0, 1, 1, ...]







# RESULTADOS

---

# 44%

Clasificación sobre 1 género con:

## 194 Géneros

---

---

# 14%

Clasificación multigénero sobre:

## 625 Géneros

---





# ¿Qué nos queda?

Trabajos futuros

- Mejora de la arquitectura propuesta
- Probar distintos parámetros

En definitiva: **AFINAR LA RED**



# ¿Qué hemos aprendido? ¿Qué nos queda?

Lecciones y conclusiones

- Un buen dataset tiene una importancia capital en tareas de IA
- Existen diferentes formas de enfocar una tarea de text mining
- Solucionamos problemas reales

Gracias por escucharnos ❤️

**Estamos encantados de responder a vuestras preguntas**