

# Introducción a la Ciencia de Datos

## Edición 2024

### Informe de Laboratorio 2

v1.0.0

19 de junio 2024

Grupo 15  
Matias Rolando  
Emiliano Viotti

# Índice

<b>Índice.....</b>	<b>1</b>
<b>Introducción.....</b>	<b>2</b>
<b>1. Primera Parte: Dataset y Representación Numérica de Texto.....</b>	<b>3</b>
Preparación de los Datos.....	3
1.1. Train/Test Split.....	3
1.2. Balance de Párrafos por Personaje.....	3
1.3. Representación Bag-of-Words.....	3
1.4. Representación TF-IDF.....	3
1.5. Reducción Dimensionalidad: PCA.....	3
<b>2. Segunda Parte: Entrenamiento y Evaluación de Modelos.....</b>	<b>3</b>
2.1. Clasificador Multinomial Naïve-Bayes.....	3
2.2. Validación Cruzada (Cross-Validation).....	3
2.3. Selección Modelo.....	4
2.4. Modelando con Otras Arquitecturas.....	4
2.5. Experimentando con Personajes Adicionales.....	4
2.6. Experimentando con Representaciones de Texto.....	4
<b>3. Opcionales: Modelo Fasttext.....</b>	<b>4</b>
<b>4. Conclusiones.....</b>	<b>5</b>
<b>Referencias.....</b>	<b>6</b>



# Introducción

To Do

# 1. Primera Parte: Dataset y Representación Numérica de Texto



## Preparación de los Datos

ToDo

### 1.1. Train/Test Split

ToDo

### 1.2. Balance de Párrafos por Personaje

ToDo

### 1.3. Representación Bag-of-Words

ToDo

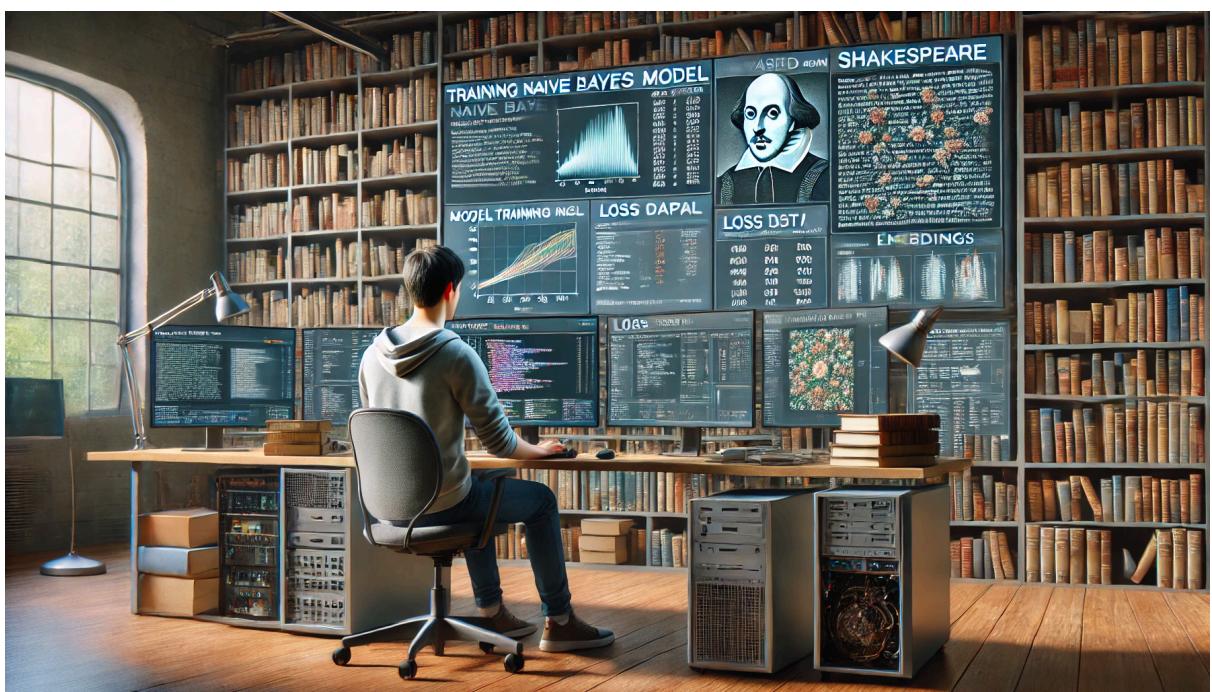
### 1.4. Representación TF-IDF

ToDo

## 1.5. Reducción Dimensionalidad: PCA

ToDo

## 2. Segunda Parte: Entrenamiento y Evaluación de Modelos



ToDo

### 2.1. Clasificador Multinomial Naive-Bayes

ToDo

### 2.2. Validación Cruzada (Cross-Validation)

ToDo

### 2.3. Selección Modelo

ToDo

### 2.4. Modelando con Otras Arquitecturas

ToDo

## 2.5. Experimentando con Personajes Adicionales

ToDo

## 2.6. Experimentando con Representaciones de Texto

ToDo

## 3. Opcionales: Modelo Fasttext





## 4. Conclusiones

A partir de los análisis realizados al conjunto de datos con obras de William Shakespeare, podemos formular las siguientes conclusiones a partir de sus obras allí presente y al conjunto de datos en sí mismo.

-



## Referencias

1. Base de Datos pública de la obra de Shakespeare con datos para este laboratorio  
<https://relational-data.org/dataset/Shakespeare>
2. Repositorio oficial del curso Intro-CD con código de referencia para Laboratorio 1  
<https://gitlab.fing.edu.uy/maestria-cdaa/intro-cd/>
3. Repositorio del grupo con código fuente y todos los entregables de esta tarea laboratorio  
<https://github.com/efviodo/mcdaa-intro-cd>