



Reto Tappx



| BARCELONA  
Fundación Telefónica



Reto Grupal

*Resumen: Este proyecto representa una inmersión en el mundo de machine learning, natural language processing y python con el tratamiento de artículos y videos*

*Versión: 1*

# Índice general

<b>I.</b>	<b>Introducción</b>	<b>2</b>
<b>II.</b>	<b>Prefacio</b>	<b>3</b>
<b>III.</b>	<b>Instrucciones generales</b>	<b>4</b>
III.1.	Reto . . . . .	5
III.2.	Entrega . . . . .	6

# Capítulo I

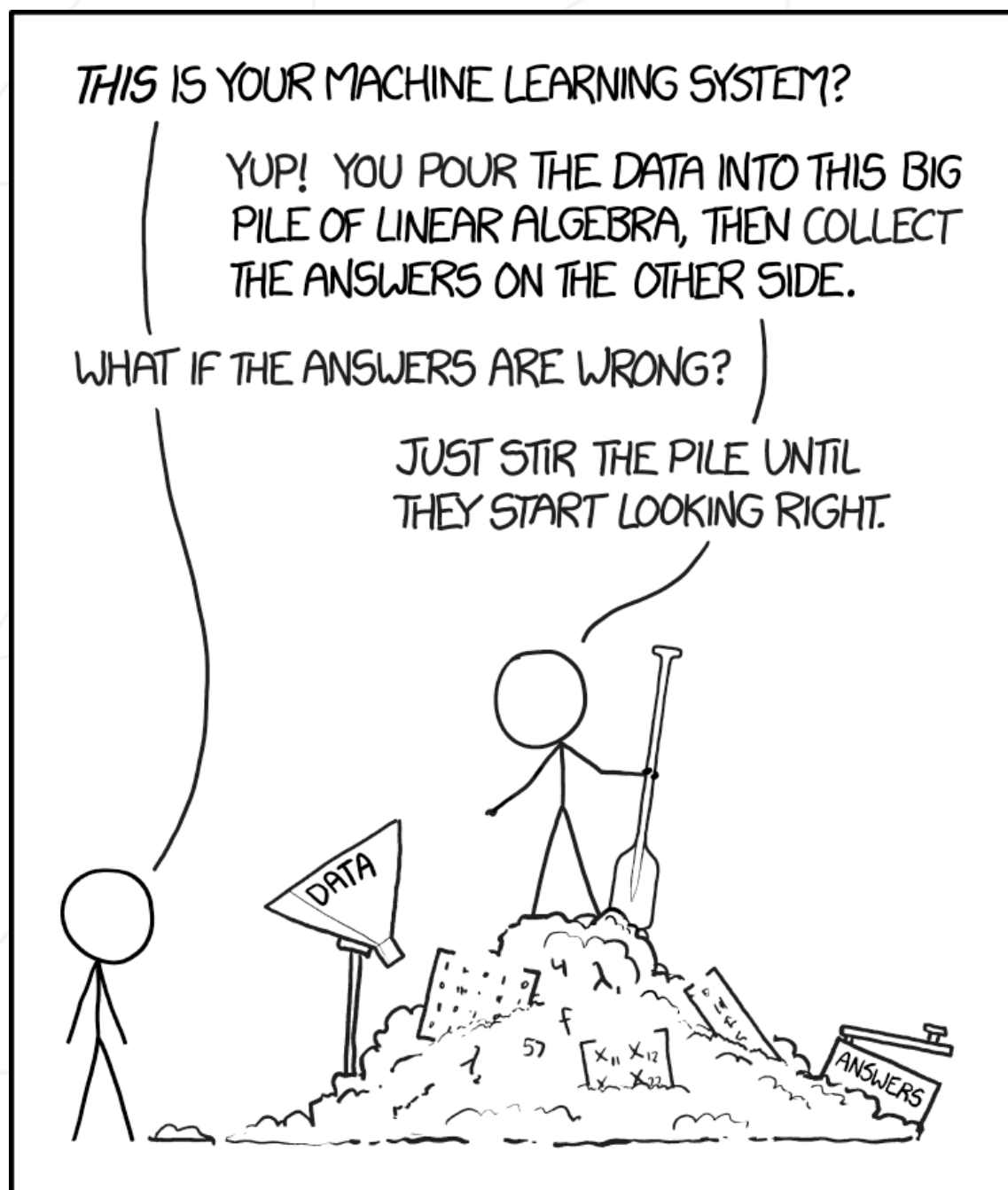
## Introducción

El Machine Learning es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo).

El procesamiento de lenguaje natural (NLP) es una tecnología de machine learning que brinda a las computadoras la capacidad de interpretar, manipular y comprender el lenguaje humano. Hoy en día, las organizaciones tienen grandes volúmenes de datos de voz y texto de varios canales de comunicación, como correos electrónicos, mensajes de texto, fuentes de noticias en redes sociales, vídeo, audio y más. Utilizan software de NLP para procesar automáticamente estos datos, analizar la intención o el sentimiento del mensaje y responder en tiempo real a la comunicación humana.

## Capítulo II

### Prefacio



# Capítulo III

## Instrucciones generales

- El código tiene que estar hecho en Python al 100 %.
- No hay limitaciones en cuanto al enfoque del problema. Cualquier algoritmo es aceptado mientras el programa final cumpla con lo pedido.
- En caso de usar inteligencia artificial, está permitido usar cualquier base de datos Online o modelo pre-entrenado.
- El reto tiene que entregarse en un archivo comprimido con los siguientes elementos:
  - Todo el código y dependencias.
  - Archivo de requerimientos.
  - Un archivo json con las características mencionadas en el apartado **entrega**.
  - Un archivo **README.txt** con la explicación del funcionamiento del programa, la definición del **score** y cualquier otra información que se considere relevante.

### III.1. Reto

Este reto consiste en encontrar relación de contexto, no necesariamente semántica, entre dos textos distintos. Disponéis de dos archivos Json:

- **articulos.json**, que contiene la transcripción del artículo.
- **videos.json**, que contiene la transcripción del vídeo.

El objetivo es asignar un conjunto de vídeos de los que se encuentran en **videos.json** a cada artículo de los que se encuentra en **articulos.json**. Los vídeos que se asignan a cada artículo deben tener relación de contexto con este.

Así pues, si tenemos un artículo que habla sobre **la victoria de Argentina en el mundial**, unos vídeos que hablen de **Messi**, o de **fútbol**, o de la **polémica de que el mundial se haya celebrado en Catar**, son buenos ejemplos para ir relacionados con el artículo.

A cada artículo se le tiene que atribuir por lo menos 2 vídeos y sin un límite máximo de número de vídeos.

La relación entre un determinado vídeo y un artículo tiene que estar representada por un **score**. Este **score** tiene que estar bien explicado (cual es su rango, qué valor representa una mayor relación, etc.)

## III.2. Entrega

Hay que entregar un **json** con las asignaciones de vídeos a cada artículo. El formato tiene que ser el siguiente:

```
{
  "id-del-articulo1": {
    "id_del_video_relacionado1": {
      "score": "score-de-la-relación"
    },
    "id_del_video_relacionado2": {
      "score": "score-de-la-relación"
    },
    .
    .
    .
  },
  "id-del-articulo2": {
    "id_del_video_relacionado1": {
      "score": "score-de-la-relacion"
    },
    .
    .
    .
  }
  .
  .
  .
}
```

Un archivo **README.txt** con la explicación del funcionamiento del programa, la definición del **score** y cualquier otra información que se considere relevante.