

# **MACHINE LEARNING**

# ESQUEMA DE LA PRESENTACION

---

## **1) SISTEMA DE RECOMENDACION**

**DISEÑO**

**FUNCIONAMIENTO**

## **2) RESUMEN INFORMATIVO**

**ENFOQUES CONSIDERADOS**

**PROBLEMAS Y SOLUCIONES**

**ARQUITECTURA**

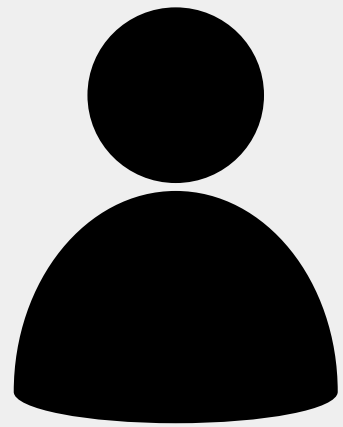
**FUNCIONAMIENTO**

## **3) CIERRE**

**ESTRUCTURA FINAL**

# SISTEMA DE RECOMENDACION

## PRIMERA PARTE FILTRO NAIVE

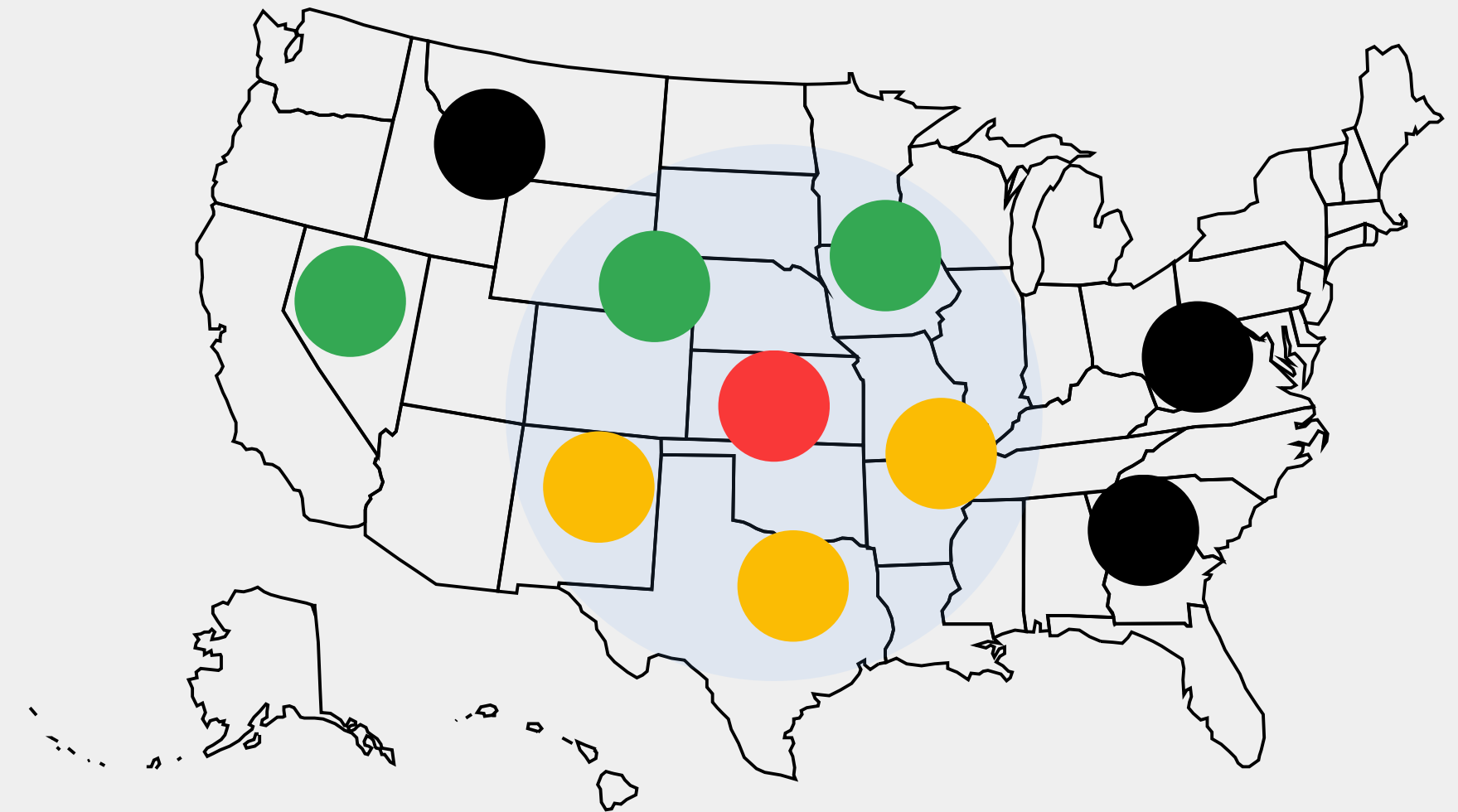


ID DEL USUARIO

LATITUD | LONGITUD

CATEGORIA

DISTANCIA



**USUARIO**



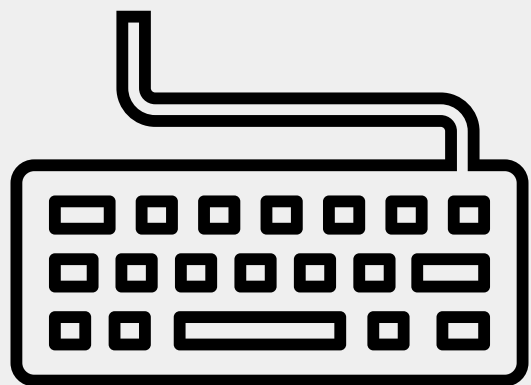
**INSTITUCIONES VISITADAS**



**INSTITUCIONES NO VISITADAS**

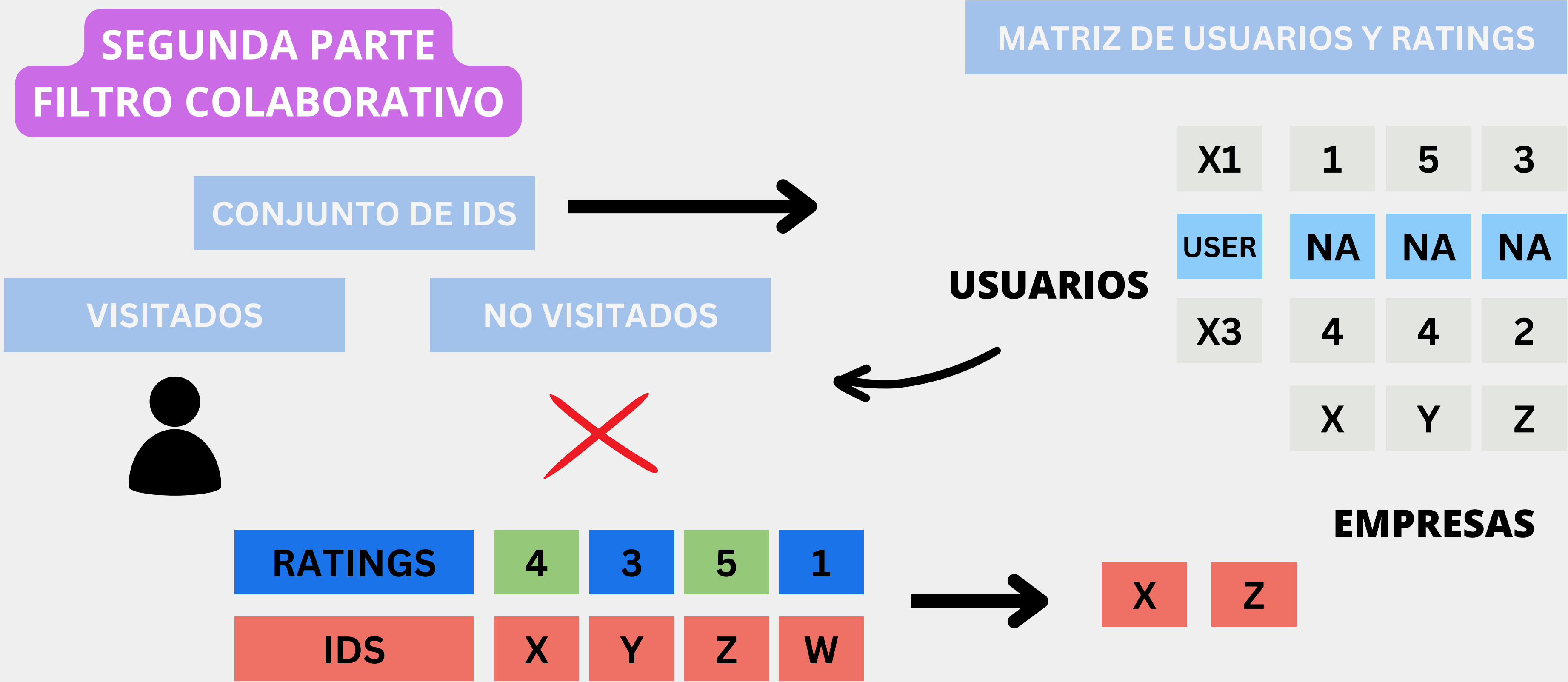


**INSTITUCIONES NO CONSIDERADAS**



# SISTEMA DE RECOMENDACION

## SEGUNDA PARTE FILTRO COLABORATIVO



# SISTEMA DE RECOMENDACION

TERCERA PARTE  
FILTRO POR CONTENIDO

MATRIZ DE CONTENIDO

X	0.6	0.1	0	0.3
Z	0.4	0.3	0	0.1
X1	0	0.3	0.5	0.2
X2	0.1	0	0	0.9
X3	0	0.3	0.5	0.2



AGRUPAMOS

X	0.6	0.1	0	0.3
Z	0.4	0.3	0	0.1
XN	0.1	0.2	0.3	0.4



MATRIZ DE SIMILITUD

	XN	Z	X
XN	1	0.6	0
Z	0.6	1	0.2
X	0	0.2	1

# RESUMEN INFORMATIVO



## DESDE 0

Necesario cuando no existe el modelo.

Libertad y control del dataset

Requiere mucho trabajo y tiempo.

Numerosos enfoques.

Alto coste energetico, en infraestructura, y ambiental.

**GENSIM**  
**SPACY**



## OPENAI

Interfaz sencilla.

Altamente flexible.

No requiere infraestructura.

0GB ocupa de memoria

Es pago y lento.

Nuevo enfoque.

Requiere conexión a internet.

**API OPENAI**  
**LANGCHAIN**



## HUGGING FACE

API sencilla

Enorme cantidad de opciones.

Se le puede realizar fine-tuned.

Comunidad establecida.

Modelos OpenSource.

Ocupa mucho espacio.

Requiere infraestructura

**TRANSFORMERS**  
**DATASET**

# RESUMEN INFORMATIVO

---

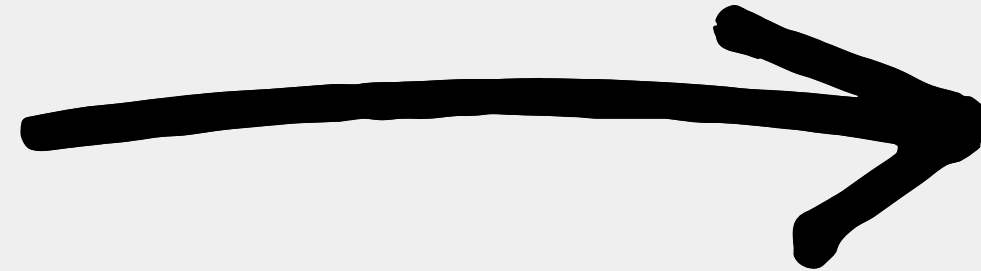
## PROBLEMAS

OVERFLOW DE  
MEMORIA

FALTA DE GPU

POCAS EPOCHS

SMALL DATASET EN  
EL FINE-TUNNED



## SOLUCIONES

ZERO-COPY  
MEMORY MAPPING

GOOGLE COLAB Y  
KAGGLE

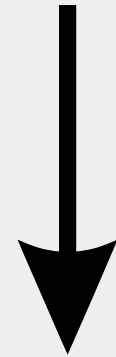


# RESUMEN INFORMATIVO

---



ID DE LA EMPRESA

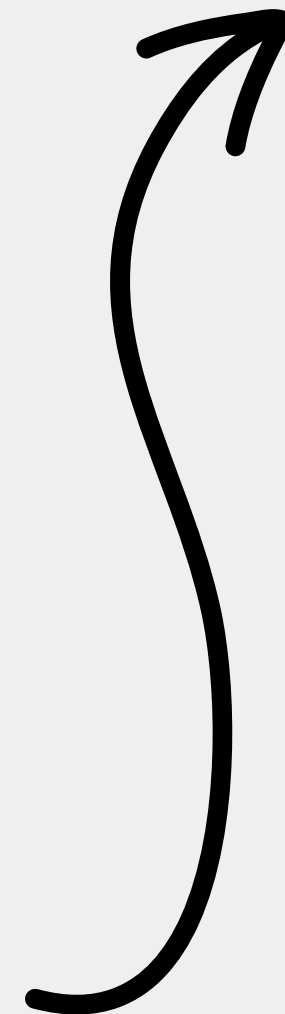


RESEÑAS

.....

.....

.....



MODELOS PRE-ENTRENADOS

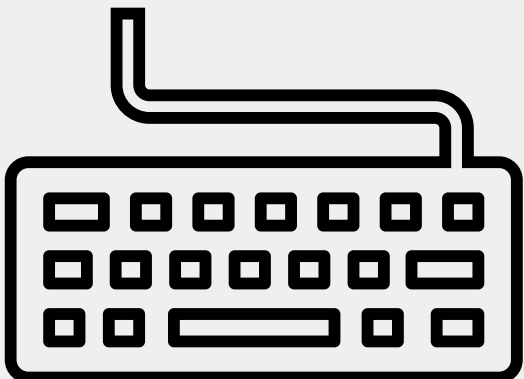
knkarthick/  
MEETING\_SUMMARY

FINE-TUNED MODELOS

bert-base-uncased

roberta-base

facebook/bart-large





# RESUMEN INFORMATIVO

---

**NOMBRE DEL MODELO:** knkarthick/MEETING\_SUMMARY

**MODELO BASE:** facebook/bart-large-xsum

**FINALIDAD:** Realizar un resumen sobre las reseñas.

**METRICA:** ROUGE-1 on samsum 53.188

**DATASET:** AMI Meeting Corpus, SAMSUM Dataset, DIALOGSUM Dataset, XSUM Dataset

---

**MODELO BASE:** bert-base-uncased

**FINALIDAD:** Predice el rating de una reseña entre 1 a 5 puntos.

**METRICA:** ACCURACY 0.88

**DATASET:** GOOGLE REVIEWS

---

**MODELO BASE:** roberta-base

**FINALIDAD:** Clasifica las reseñas entre cool | funny | useful.

**METRICA:** ACCURACY 0.43

**DATASET:** YELP REVIEWS

---

**MODELO BASE:** facebook/bart-large

**FINALIDAD:** Responder a las reviews realizadas teniendo en cuenta el contexto de la misma.

**METRICA:** CrossEntropyLoss 2.01

**DATASET:** GOOGLE REVIEWS

# ESTRUCTURA FINAL

---

SISTEMA DE  
RECOMENDACION

RESUMEN INFORMATIVO

API

STREAMLIT

