

Datos e inteligencia para encontrar tu match ideal



Inteligencia
artificial

Afinix

Integrantes:

Johan Santiago Rojas Naranjo - 2225005

Nancy Liliana Saenz Moreno - 2224510

Leyson David Celis Acelas - 2225002



XXX

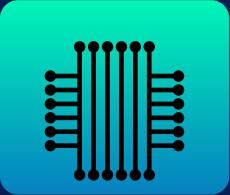
Tabla de Contenido

01
02
03
04
05
06

- DESCRIPCIÓN Y MOTIVACIÓN
- POTENCIALES USUARIOS O BENEFICIARIOS
- SOLUCIONES EXISTENTES
- OBTENCIÓN DE LOS DATOS
- DESCRIPCIÓN DE LOS DATOS
- RELACIÓN CON TEMÁTICA DE LA MATERIA



DESCRIPCIÓN Y MOTIVACIÓN



En la actualidad, **millones de personas utilizan aplicaciones de citas en línea** para buscar pareja. Sin embargo, **la mayoría de estas plataformas se apoyan en filtros muy básicos**, como la edad, la ubicación o algunos intereses generales, lo que provoca que **muchas de las recomendaciones no reflejen una verdadera compatibilidad** entre los usuarios.

Compatibles: Personas con suficientes aspectos en común o complementarios para que haya buena conexión.

DESCRIPCIÓN Y MOTIVACIÓN

Ante esta limitación surge la **necesidad de un sistema más avanzado** que vaya más allá de los filtros tradicionales. El reto es **predecir la compatibilidad entre dos personas considerando múltiples factores (hábitos, creencias, intereses y personalidad)** con el fin de mejorar la calidad de las recomendaciones y favorecer conexiones más afines y duraderas...



POTENCIALES USUARIOS



Usuarios de
aplicaciones de citas



Empresas de
plataformas de citas



Investigadores en
psicología social y
ciencias de datos



Investigadores y estudiantes que estudian el
comportamiento y la compatibilidad en relaciones
humanas usando inteligencia artificial.

SOLUCIONES ACTUALES



Tinder, Bumble o Hinge

Filtros básicos: edad, género, distancia.
Algoritmos de “swipe” con modelos de popularidad
(mostrar primero a usuarios con más “likes”).
Cuestionarios de personalidad para calcular un
porcentaje de compatibilidad simple.





Objetivos

01

Desarrollar un modelo de predicción de compatibilidad de pareja que, a partir de la información contenida en los perfiles de los usuarios de **OKCupid**.

02

Analizar qué variables (edad, educación, intereses, hábitos, texto en biografías) impactan más en la compatibilidad.

03

Explorar técnicas de NLP para comparar similitudes en biografías.

DESCRIPCIÓN DE LOS DATOS

Aspecto Clave	Descripción
Nombre del dataset	OkCupid Profiles
Fuente	Perfiles públicos de usuarios de OkCupid
Registros totales	~60,000 perfiles de usuario
Variables	31 columnas: numéricas, categóricas y texto libre
Variables principales	Edad, género, orientación sexual, ubicación, hábitos, educación, trabajo, ensayos de texto
Aplicaciones en IA	EDA (Análisis Exploratorio de Datos), NLP (Procesamiento de Lenguaje Natural), clustering, sistemas de recomendación, análisis de comportamiento
Consideraciones éticas	Datos anonimizados, requiere manejo responsable y ético

LARXEL - UPDATED 5 YEARS AGO

OkCupid Profiles

Explore dating app profiles in search for love

Data Card Code (30) Discussion (3) Suggestions (0)

About Dataset

Abstract

Find out what OkCupid users are like. Not really this dataset is voluminous, but you can explore data points through this dataset.

Usability 10.00

License Other (specified in description)

The image shows a digital interface for a dataset named "OkCupid Profiles". At the top, there's a header with the dataset name, a small profile picture, and some navigation buttons. Below the header, there's a section titled "About Dataset" with a sub-section "Abstract" containing a brief description of the data. To the right, there are two numerical ratings: "Usability" at 10.00 and "License" as "Other (specified in description)". The background of the slide features a collage of three images: a glowing blue cloud with internal circuitry, a wireframe model of a human brain, and two stylized human profiles facing each other, all set against a dark background with red heart icons.

RELACIÓN CON TEMATICA DE CLASE



- El primer paso **será limpiar y explorar el dataset de OKCupid**, identificando valores **faltantes**, **variables relevantes** y **relaciones entre los atributos** (**edad, educación, hábitos, biografías**).
- Se aplicará análisis exploratorio **de datos (EDA)** para descubrir patrones y correlaciones que influyen en la compatibilidad.
- Aprendizaje de máquina supervisado
- Con el dataset procesado, se pueden crear modelos **supervisados** de clasificación para predecir si dos usuarios son compatibles (**compatible vs. no compatible**).
- Aprendizaje de máquina no supervisado
- Se pueden aplicar técnicas de **clustering (K-means, DBSCAN, clustering jerárquico)** para agrupar usuarios con características e intereses similares.



Muchas Gracias

RECUERDA: LA COMPATIBILIDAD ES LA CHISPA QUE CONVIERTÉ EL
ENCUENTRO EN DESTINO.