

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

MÉTODO DE NAIVE BAYES

Variables	
Edad	Age
Género	Gender
Orientación sexual	Sexual orientation
Estado civil	Marital status
Tipo de cuerpo	Body type
Dieta	Diet
Consumo de alcohol	Alcohol consumption
Consumo de drogas	Drug use
Educación	Education
Altura	Height
Ingresos	Income
Ocupación	Occupation
Hijos	Offspring
Mascotas	Pets
Religión	Religion
Signo zodiacal	Zodiac sign
Hábito de fumar	Smoking habit
Idiomas	Languages
Ensayos de texto libre (personalidad, intereses, etc.)	Free-text essays (personality, interests, etc.)

Naive Bayes es un modelo de clasificación que usa el teorema de Bayes para calcular probabilidades a partir de los datos. Asume que las características son independientes, lo que lo hace rápido y eficiente. Aprende patrones en la información y estima la probabilidad de compatibilidad entre dos personas de forma sencilla y precisa.



METODO DE NAIVE BAYES

× × ×

Resultados principales del modelo

TUS TOP 5 MEJORES MATCHES

#1 → MATCH: 99.2%

33 años | F | straight

Trabajo: student

Educación: graduated from college/university

168 cm | Fuma: no | Bebe: socially

#2 → MATCH: 99.1%

31 años | M | gay

Trabajo: student

Educación: graduated from college/university

183 cm | Fuma: no | Bebe: not at all

#3 → MATCH: 99.1%

25 años | F | straight

Trabajo: student

Educación: graduated from college/university

160 cm | Fuma: no | Bebe: socially

#4 → MATCH: 99.0%

25 años | F | straight

Trabajo: student

Educación: graduated from college/university

160 cm | Fuma: no | Bebe: socially

#5 → MATCH: 98.9%

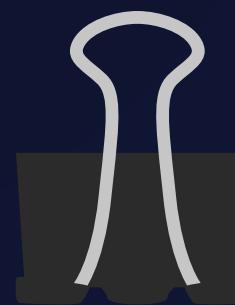
27 años | F | straight

Trabajo: student

Educación: graduated from college/university

155 cm | Fuma: no | Bebe: socially





METODO DE NAIVE BAYES

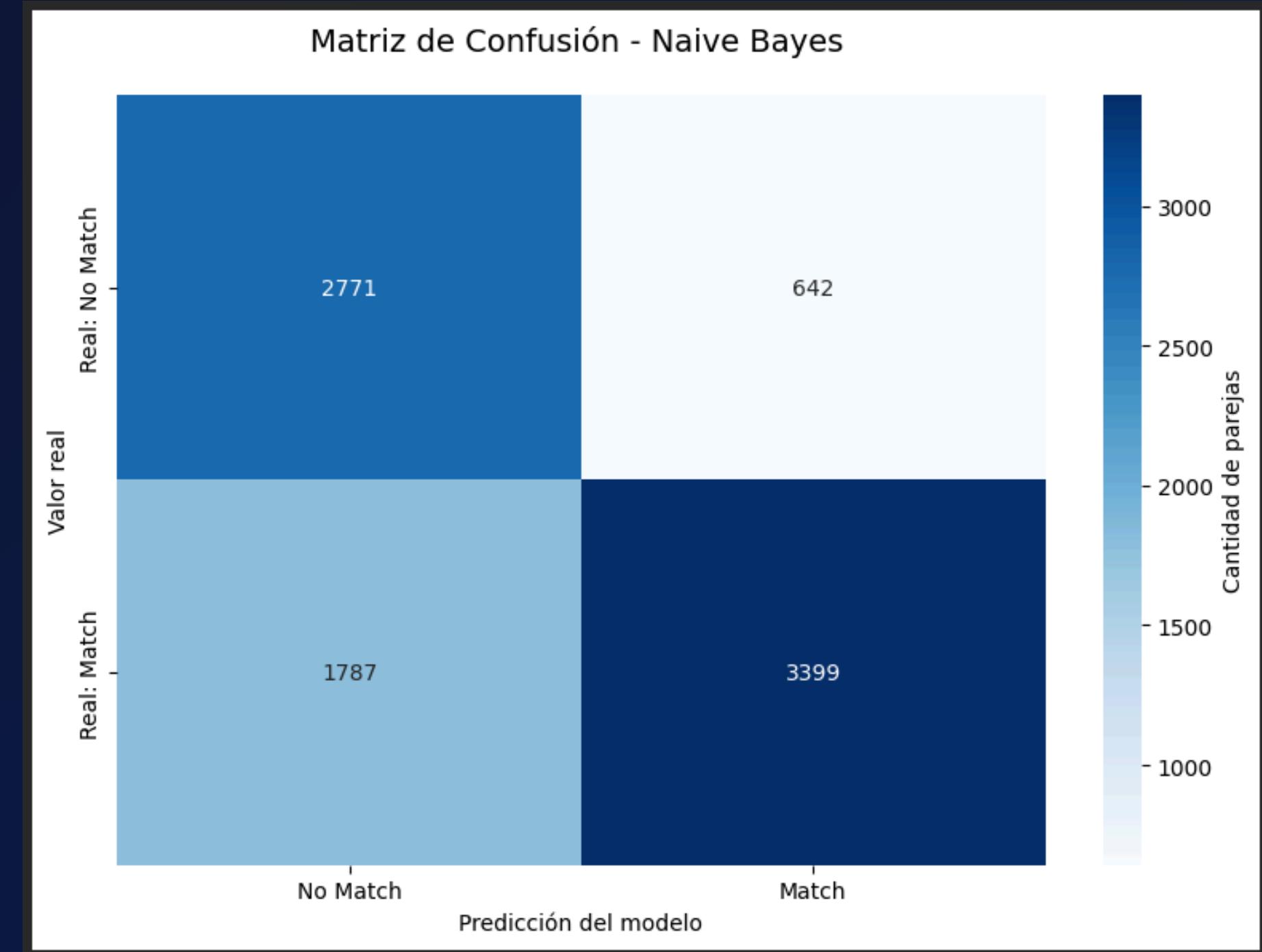
REPORTE FINAL

	precision	recall	f1-score	support
0	0.61	0.81	0.70	3413.00
1	0.84	0.66	0.74	5186.00
accuracy	0.72	0.72	0.72	0.72
macro avg	0.72	0.73	0.72	8599.00
weighted avg	0.75	0.72	0.72	8599.00

Accuracy: 0.7175 (71.75%)

DE 3.413 PAREJAS
QUE NO ERAN MATCH,
EL MODELO ACERTÓ
EN 2.771

1.787 FALSOS
NEGATIVOS



HUBO 642
FALSOS
POSITIVOS

DE 5.186 QUE SÍ
ERAN MATCH,
ACERTÓ EN 3.399

Resultados principales del modelo



METODO DE DECISIÓN TREE

=====

TOP 5 MEJORES MATCHES (ÁRBOL CON LÍMITES)

=====

#1 MATCH: 81.5%
40 años | F | bisexual
Trabajo: medicine / health
Educación: graduated from two-year college
163 cm | Fuma: sometimes | Bebe: socially

#2 MATCH: 81.5%
49 años | F | straight
Trabajo: executive / management
Educación: graduated from college/university
170 cm | Fuma: sometimes | Bebe: socially

#3 MATCH: 81.5%
40 años | F | straight
Trabajo: artistic / musical / writer
Educación: working on masters program
168 cm | Fuma: sometimes | Bebe: rarely

#4 MATCH: 81.5%
41 años | F | straight
Trabajo: medicine / health
Educación: graduated from college/university
152 cm | Fuma: sometimes | Bebe: socially

#5 MATCH: 81.5%
42 años | F | straight
Trabajo: missing
Educación: graduated from college/university
165 cm | Fuma: sometimes | Bebe: socially



Resultados principales del modelo

RESULTADOS



REPORTE DE CLASIFICACIÓN DETALLADO

	precision	recall	f1-score	support
0	0.69	0.66	0.68	1152
1	0.71	0.74	0.73	1307
accuracy			0.70	2459
macro avg	0.70	0.70	0.70	2459
weighted avg	0.70	0.70	0.70	2459

Matriz de confusión (valores):

```
[[766 386]
 [344 963]]
```

PRECISIÓN (ACCURACY) DEL MODELO

Accuracy: 0.7031 (70.31%)

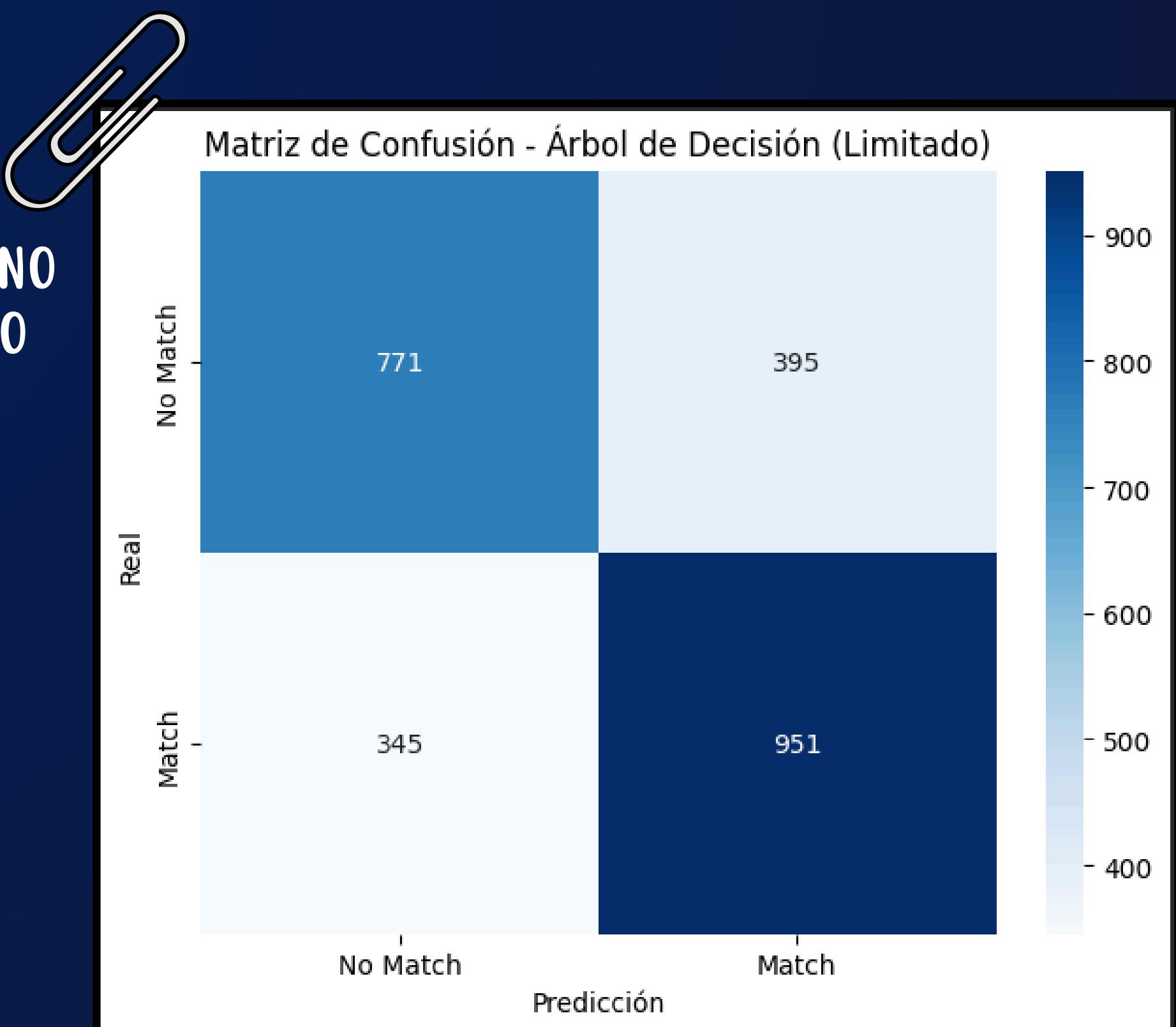


DE 1.166 PAREJAS QUE NO
ERAN MATCH, EL MODELO
ACERTÓ EN 771

HUBO 345 FALSOS
POSITIVOS

395 FALSOS
NEGATIVOS

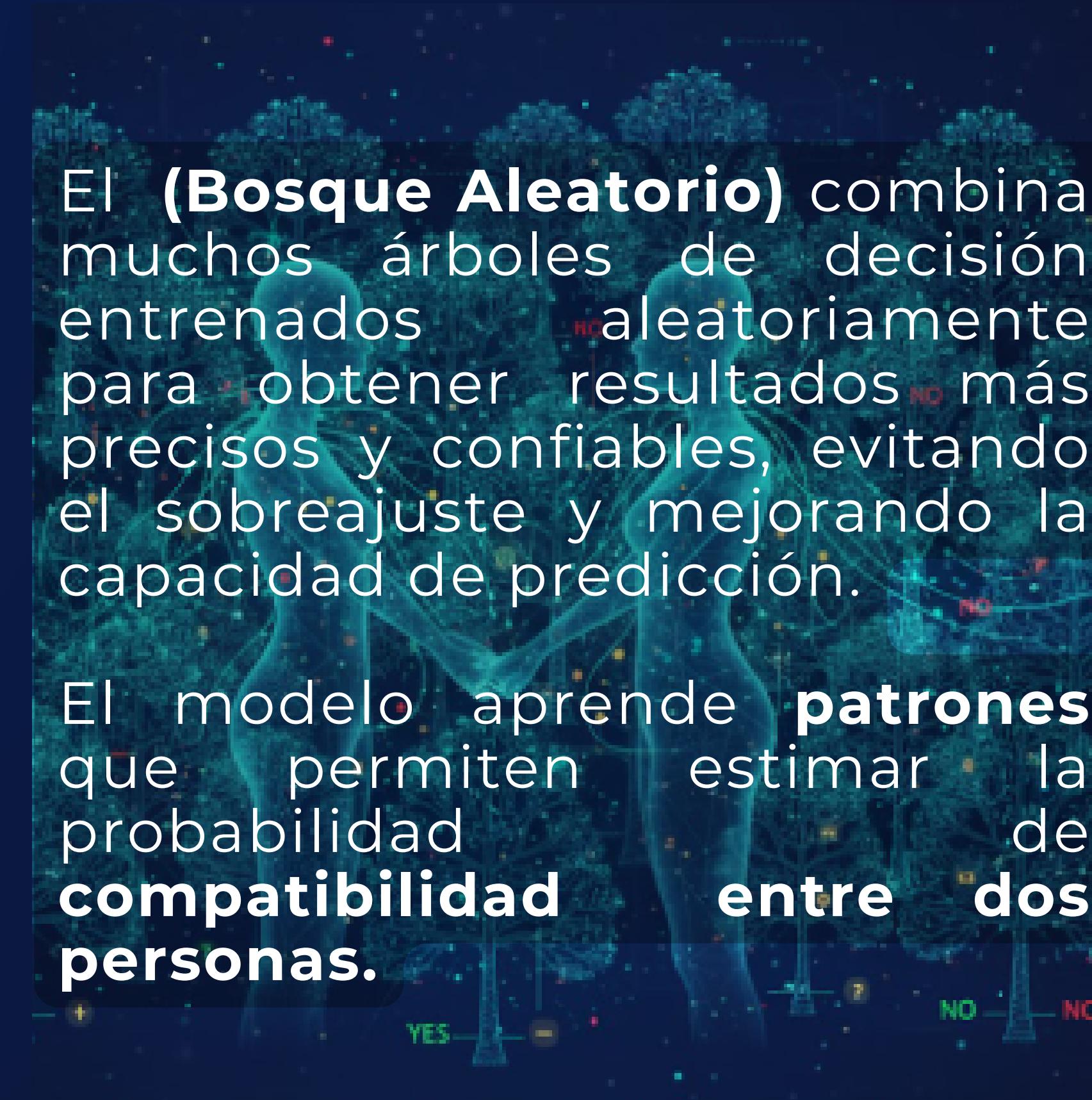
DE 1.296 QUE SÍ ERAN
MATCH, ACERTÓ EN 951



MÉTODO DE RANDOM FOREST

Canva

Variables	
Edad	Age
Género	Gender
Orientación sexual	Sexual orientation
Estado civil	Marital status
Tipo de cuerpo	Body type
Dieta	Diet
Consumo de alcohol	Alcohol consumption
Consumo de drogas	Drug use
Educación	Education
Altura	Height
Ingresos	Income
Ocupación	Occupation
Hijos	Offspring
Mascotas	Pets
Religión	Religion
Signo zodiacal	Zodiac sign
Hábito de fumar	Smoking habit
Idiomas	Languages
Ensayos de texto libre (personalidad, intereses, etc.)	Free-text essays (personality, interests, etc.)



El **(Bosque Aleatorio)** combina muchos árboles de decisión entrenados aleatoriamente para obtener resultados más precisos y confiables, evitando el sobreajuste y mejorando la capacidad de predicción.

El modelo aprende **patrones** que permiten estimar la probabilidad **compatibilidad** entre **dos personas**.

METODO DE RANDOM FOREST

Etapas de desarrollo del modelo

- **Carga y preparación de datos:** Se usa un dataset de perfiles (**sample_df**).
- Formamos **pares de usuarios:** Cada combinación de usuarios se evalúa por sus diferencias y similitudes.
- **Extracción de características:** calculan variables como **age_diff, height_diff, same_smokes, etc.** modelo pueda ver esas similitudes
- Entrenamiento del **modelo Random Forest:**

Hiperparámetros principales:

- **n_estimators=100** (100 árboles de decisión)
- **max_depth=10** (profundidad máxima para evitar sobreajuste)
- **class_weight='balanced'** (para compensar clases desiguales)

Evaluación y visualización:

- Se divide el conjunto en **entrenamiento** (70%) y **prueba** (30%).
- Se calcula la precisión (Accuracy) y se genera la matriz de confusión.



METODO DE RANDOM FOREST

- buscó los usuarios más compatibles.
- Se obtuvieron 5 coincidencias principales con porcentajes entre 89 % y 92 %.
- Las variables más influyentes fueron:
- edad, género, orientación, educación, profesión y hábitos (fumar / beber).
- El modelo identificó patrones de similitud en estilo de vida y formación.

× × ×

Resultados principales del modelo



TOP 5 MATCHES con Random Forest

#1 → 92.5%

27 años | M | straight **Edad, género y orientación**.
Other | graduated from college/univers **Profesión y educación**
190cm | Fuma: no | Bebe: socially **Hábitos (fuma / bebe)**

#2 → 91.8%

27 años | M | straight
Entertainment / Media | graduated from college/univers
188cm | Fuma: no | Bebe: socially

#3 → 91.2%

28 años | M | straight
Science / Tech / Engineering | graduated from college/univers
188cm | Fuma: no | Bebe: socially

#4 → 90.2%

25 años | M | bisexual
Education / Academia | working on space camp
188cm | Fuma: no | Bebe: socially

#5 → 89.4%

27 años | F | straight
Science / Tech / Engineering | working on ph.d program
170cm | Fuma: no | Bebe: socially



METODO DE RANDOM FOREST

Resultados principales del modelo



TOP 5 PEORES (pero compatibles)

#1 → 8.3% → 42 años | banking / financial / real estate

#2 → 10.5% → 39 años | other

#3 → 10.9% → 38 años | education / academia

#4 → 11.4% → 43 años | political / government

#5 → 11.5% → 51 años | education / academia

- El modelo identificó los 5 perfiles con menor nivel de coincidencia.
- Los porcentajes de compatibilidad están entre 8 % y 11 %.
- En general, presentan mayor diferencia de edad y profesiones distintas
- El modelo distingue bien entre similitudes y diferencias, reconociendo qué factores influyen en la compatibilidad.

MÉTODO DE RANDOM FOREST



EVALUACIÓN DEL MODELO RANDOM FOREST					
Reporte de Clasificación:					
	precision	recall	f1-score	support	
No Compatible = 0	0.57	0.59	0.58	49	
Compatible = 1	0.73	0.71	0.72	77	
accuracy			0.67	126	
macro avg	0.65	0.65	0.65	126	
weighted avg	0.67	0.67	0.67	126	

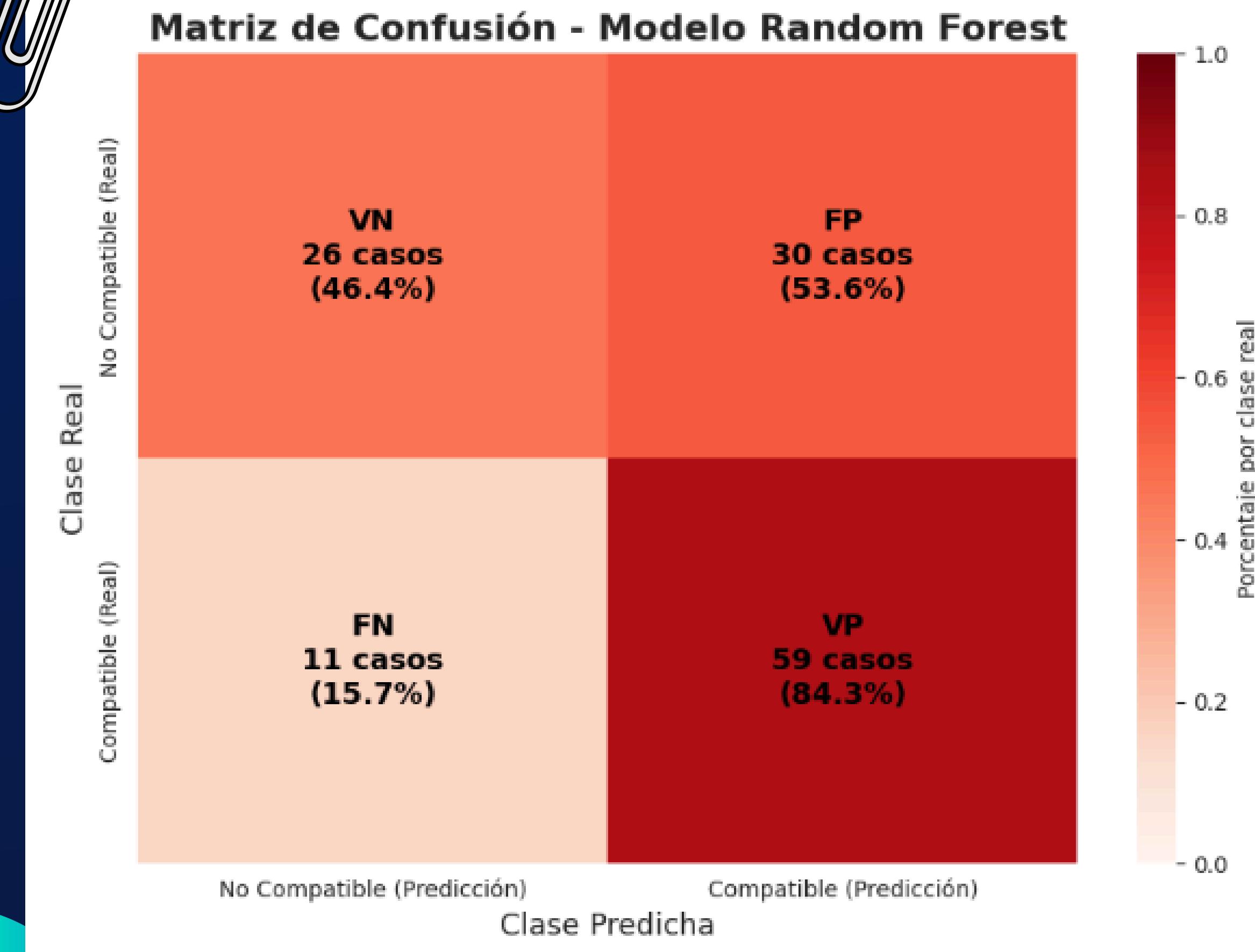
Resultados principales del modelo

- Precisión general (accuracy) del 66.67 %, clasificó correctamente dos de cada tres casos.
- En la clase Compatible (1), el modelo tuvo mejor desempeño con una precisión de 0.73 (73 %) y un recall de 0.71 (71 %), mostrando que identifica correctamente la mayoría de los perfiles compatibles.
- En cambio, la clase No compatible (0) obtuvo valores más bajos (precisión 0.57, recall 0.59), el modelo tiende a confundir algunos casos no compatibles como compatibles.
- El total de la evaluación se realizó sobre 126 pares válidos, divididos en 49 no compatibles y 77 compatibles.
- Los promedios (0.65 y 0.67) indican un rendimiento estable, con buena precisión en ambas clases.

METODO DE RANDOM FOREST

Matplotlib y Seaborn

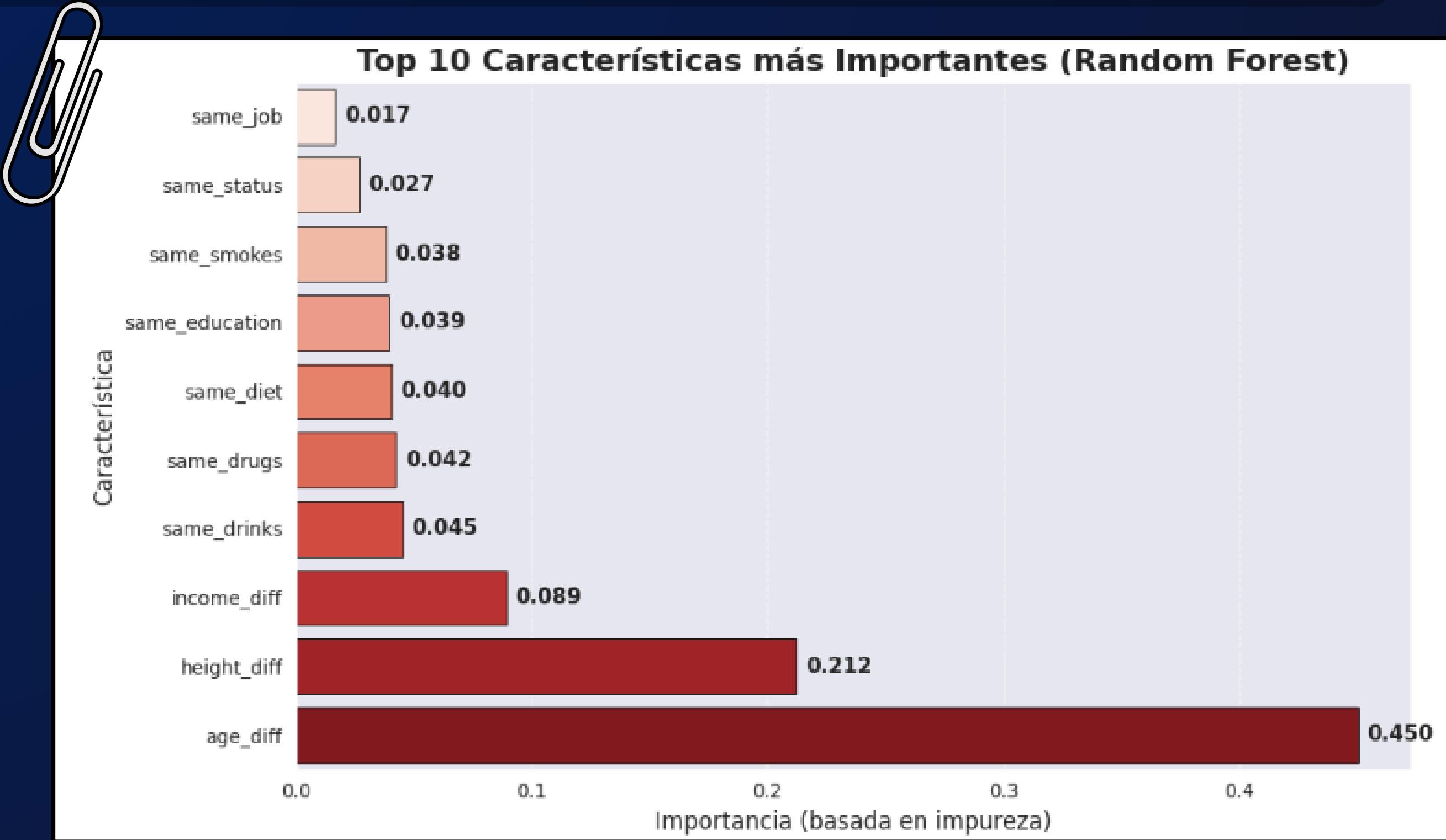
Variables más influyentes en la compatibilidad



El modelo acierta principalmente en los perfiles compatibles (VP), aunque comete algunos errores al predecir compatibilidad en casos que no lo son (FP).

METODO DE RANDOM FOREST

características más importantes que el modelo Random Forest usó para decidir si dos personas eran compatibles o no.



METODO DE SVM

Canva

Variables	
Edad	Age
Género	Gender
Orientación sexual	Sexual orientation
Estado civil	Marital status
Tipo de cuerpo	Body type
Dieta	Diet
Consumo de alcohol	Alcohol consumption
Consumo de drogas	Drug use
Educación	Education
Altura	Height
Ingresos	Income
Ocupación	Occupation
Hijos	Offspring
Mascotas	Pets
Religión	Religion
Signo zodiacal	Zodiac sign
Hábito de fumar	Smoking habit
Idiomas	Languages
Ensayos de texto libre (personalidad, intereses, etc.)	Free-text essays (personality, interests, etc.)

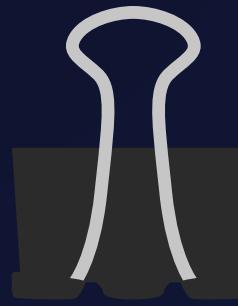
SVM (Máquinas de Vectores de Soporte) es un modelo de clasificación que busca la mejor frontera que separe los datos en diferentes categorías. Funciona incluso cuando los datos no son lineales, usando funciones que permiten separar los casos más complejos. Aprende patrones y estima la probabilidad de compatibilidad entre dos personas encontrando la separación más precisa entre perfiles compatibles y no compatibles.

METODO DE SVM

TOP 5 MATCHES

#1 → 98.9%		27 años		Sin trabajo		180cm
#2 → 98.7%		21 años		Artistic / Musical / Writer		170cm
#3 → 98.1%		29 años		Hospitality / Travel		183cm
#4 → 97.6%		25 años		Executive / Management		178cm
#5 → 97.3%		27 años		Sin trabajo		168cm





RESULTADOS DE SVM

REPORTE SVM

	precision	recall	f1-score	support
0	0.63	0.74	0.68	2279.00
1	0.83	0.74	0.78	3862.00
accuracy	0.74	0.74	0.74	0.74
macro avg	0.73	0.74	0.73	6141.00
weighted avg	0.75	0.74	0.74	6141.00

Accuracy: 0.7396 (73.96%)

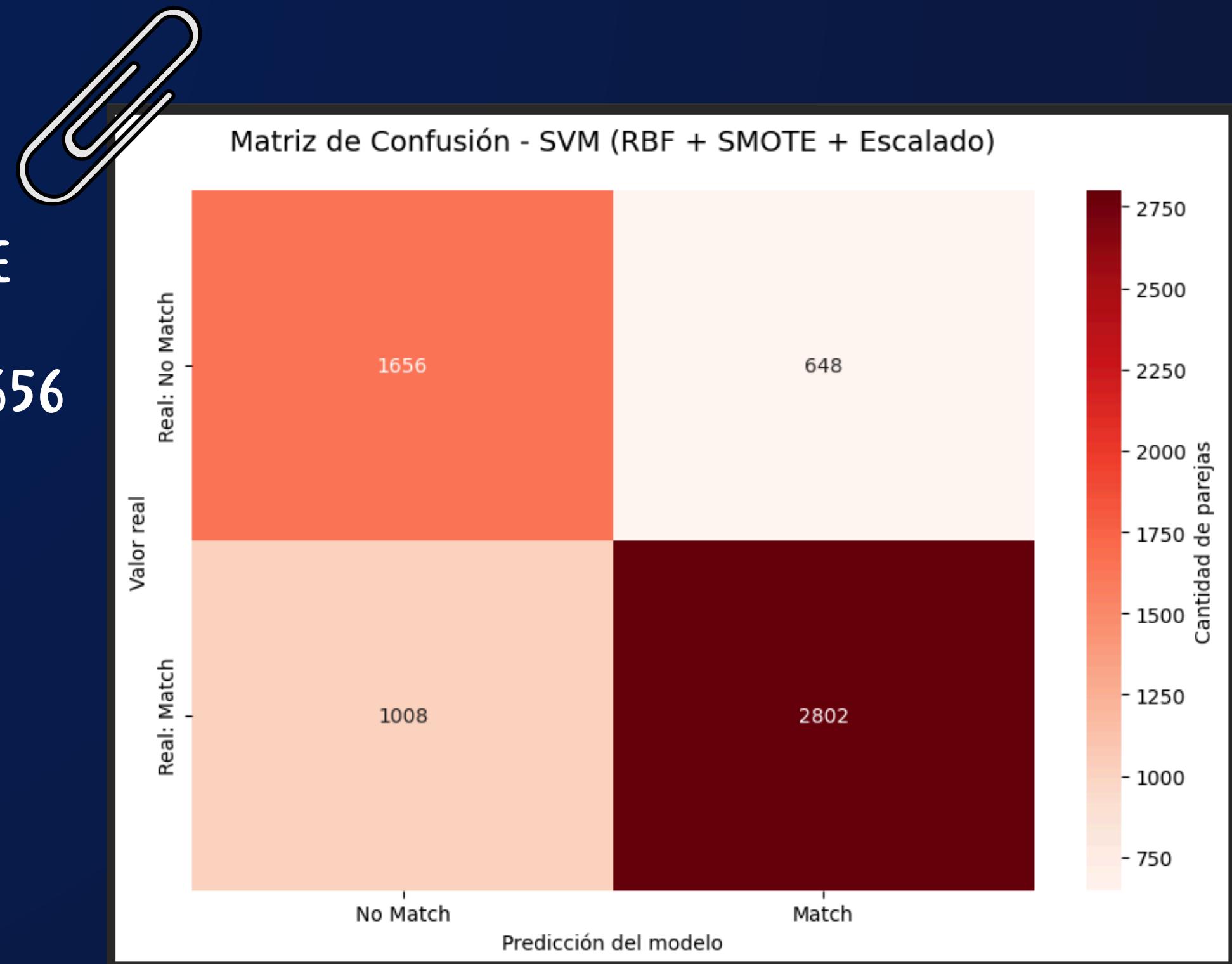


XXX



DE 2.304 PAREJAS QUE
NO ERAN MATCH, EL
MODELO ACERTÓ EN 1.656

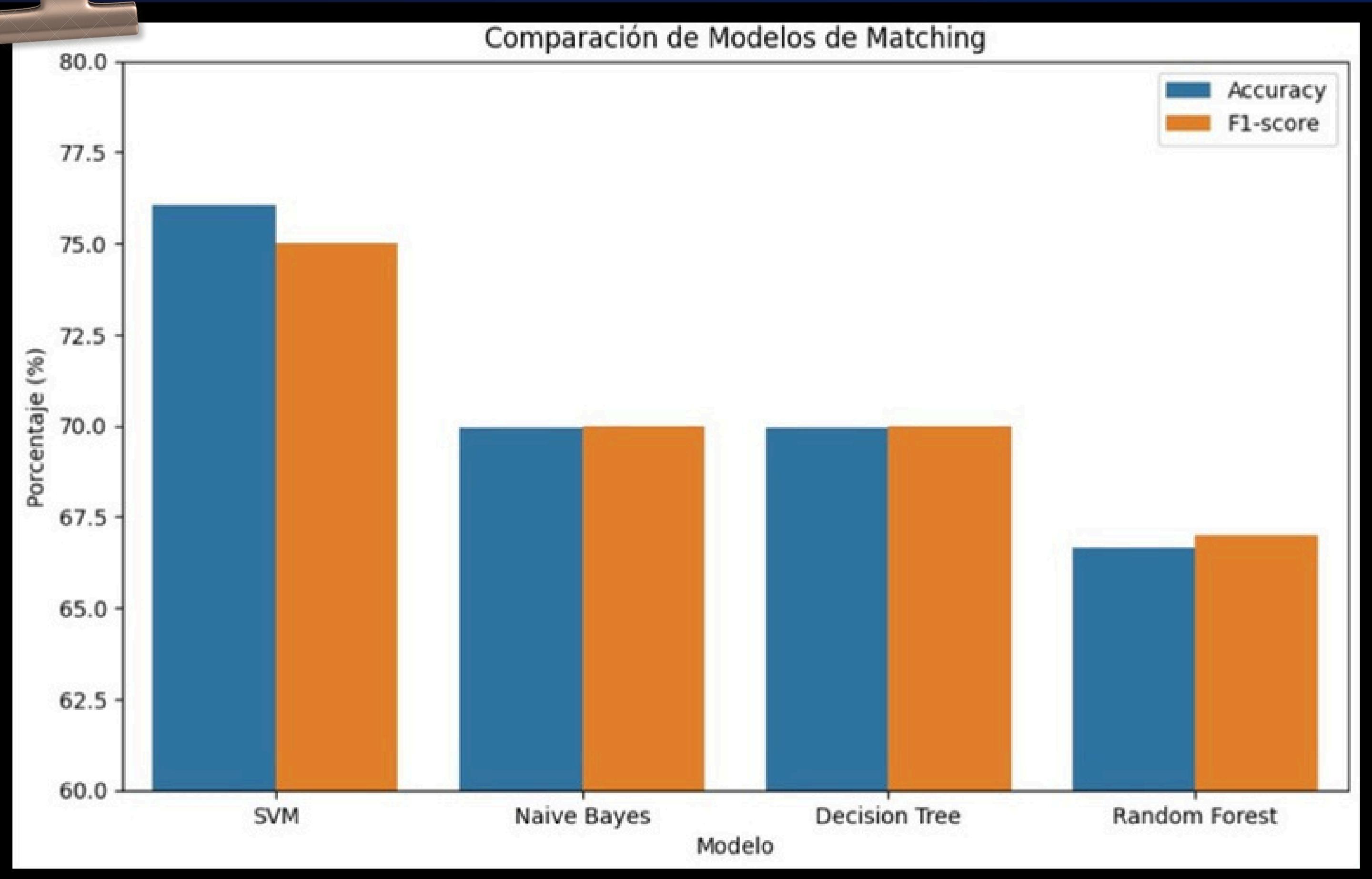
1.008 FALSOS
NEGATIVOS



SOLO HUBO 648
FALSOS POSITIVOS

DE 3.810 QUE SÍ ERAN
MATCH, ACERTÓ EN 2.802

COMPARACIÓN





Resultados principales del modelo

ELECCIÓN



REPORTE SVM				
	precision	recall	f1-score	support
0	0.63	0.74	0.68	2279.00
1	0.83	0.74	0.78	3862.00
accuracy	0.74	0.74	0.74	0.74
macro avg	0.73	0.74	0.73	6141.00
weighted avg	0.75	0.74	0.74	6141.00

Accuracy: 0.7396 (73.96%)



Muchas gracias

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