



Proyecto IA: VibeMatch

Integrantes:

Carlos Alejandro Cepeda - 2202011
Santiago Galvis Saavedra - 2212938
Carlos Andrés Galeano - 2215629

MOTIVACIÓN



Identificar canciones con potencial de éxito antes de que se vuelvan virales puede brindar una ventaja estratégica a artistas, productores y plataformas.

Este proyecto busca aprovechar el análisis de datos y el aprendizaje automático para descubrir patrones ocultos en la música que predigan popularidad.

OBJETIVOS

Identificar canciones con alto potencial de convertirse en éxitos basadas en sus características musicales.



Clasificar canciones según su popularidad usando técnicas de aprendizaje supervisado y no supervisado.

DATASET

Para nuestro proyecto escogimos un dataset que agrupa las canciones más famosas del 2023 en Spotify, reuniendo composiciones de todos los géneros y artistas.



CARACTERÍSTICAS

Nuestro dataset cuenta con 954 filas y 24 columnas,
algunas de las más importantes son:

- **track_name:** nombre de la canción
- **artist(s)_name:** nombre del artista.
- **streams:** cantidad de reproducciones en Spotify.
- **bpm:** pulsos/golpes por minuto.
- **danceability_%:** que tan adecuada es la canción para bailar.

CARACTERÍSTICAS

- **valence_%**: que tan positiva/alegre es la canción
- **energy_%**: intensidad de la canción.
- **acousticness_%**: sonidos acusticos (guitarras, pianos, etc..) en la canción.
- **instrumentalness_%**: que tantas voces o instrumentos tiene la canción.
- **liveness_%**: presentación en vivo.
- **speechiness_%**: cantidad de palabras en la canción.

ALGORITMO GENÉTICO



Ejecutando algoritmo genético para encontrar combinación óptima...

Generación 0: Mejor fitness = 0.2504
Generación 10: Mejor fitness = 0.4219
Generación 20: Mejor fitness = 0.4941
Generación 30: Mejor fitness = 0.4941
Generación 40: Mejor fitness = 0.4941

Mejor combinación encontrada de características musicales:

danceability_%: 65.3%
energy_%: 75.8%
valence_%: 0.0%
acousticness_%: 70.6%
liveness_%: 64.5%

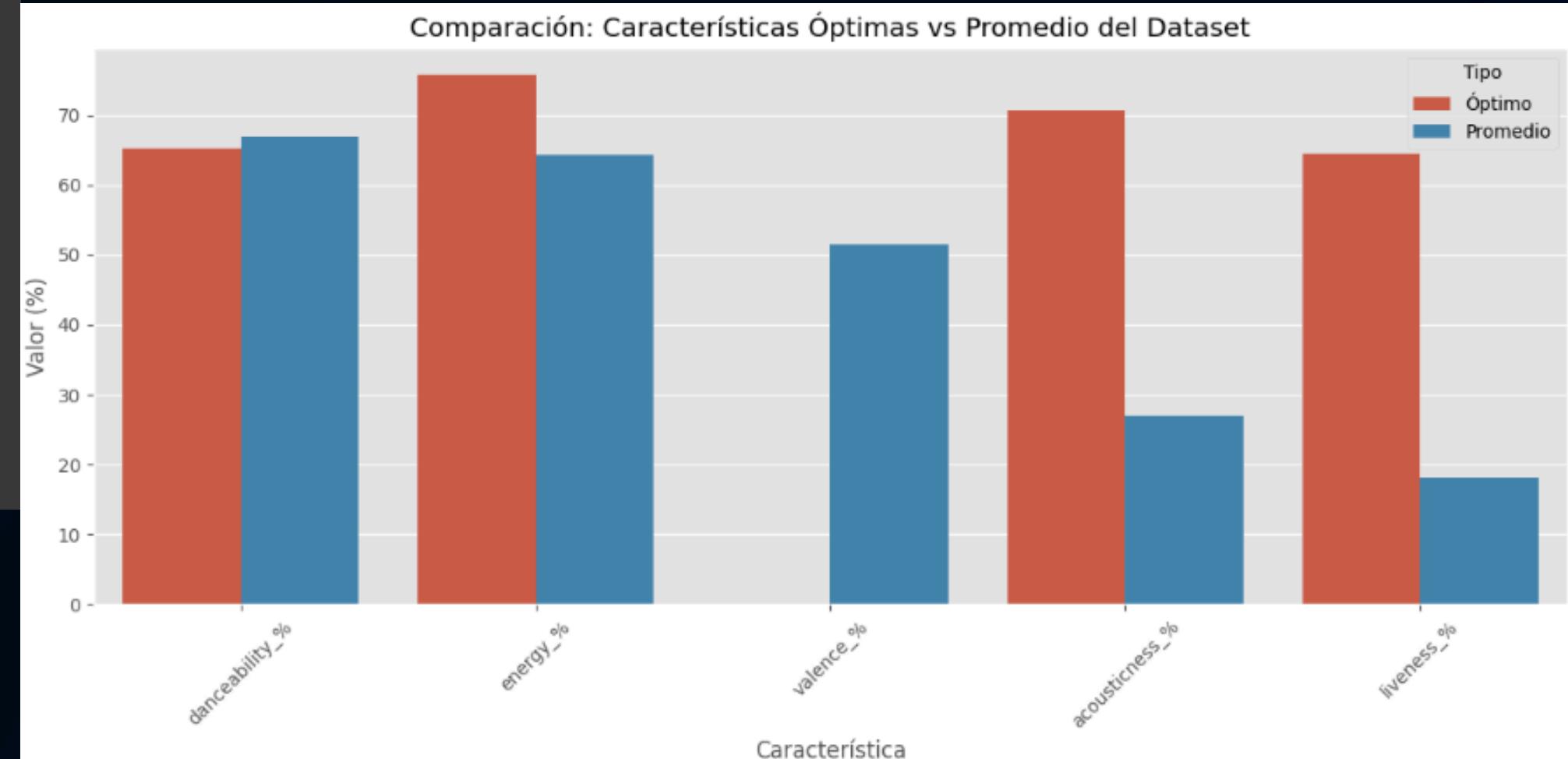
Fitness (éxito estimado): 0.4941

Canciones reales más similares a la combinación óptima:

	track_name	artist(s)_name	streams \
332	Tormenta (feat. Bad Bunny)	Gorillaz, Bad Bunny	149,778,242
72	golden hour	JVKE	751,134,527
565	deja vu	Olivia Rodrigo	1,256,880,657
249	Oi Balde - Ao Vivo	ZiçXic Neto & Crist	145,458,418
790	Callaita	Bad Bunny, Tainy	1,304,313,953

	danceability_%	energy_%	valence_%	acousticness_%	liveness_%
332	64	77	38	46	38
72	51	59	14	65	25
565	44	68	22	61	42
249	67	67	55	60	80
790	61	62	24	60	24

RESULTADOS



REGRESIÓN LINEAL



RESULTADOS

❖ Modelo de Regresión Lineal Optimizado (con Ridge + log-streams)

MSE: 1.9224

RMSE: 1.3865

R² Score: -0.0334

Mejor alpha (regularización): 79.06043210907701

Importancia de características:

	Feature	Importance
3	acousticness_%	0
0	danceability_%	0
5	liveness_%	0
6	speechiness_%	0
1	energy_%	0
4	instrumentalness_%	0
7	bpm	0
2	valence_%	0

🎧 Recomendaciones de canciones con potencial no explotado:

Que Vuelvas por Carin Leon, Grupo Frontera – Streams actuales: 2,762, Predichos: 395,218,112

Back To December (Taylor's Version) por Taylor Swift – Streams actuales: 39,228,929, Predichos: 405,135,356

Enchanted (Taylor's Version) por Taylor Swift – Streams actuales: 39,578,178, Predichos: 392,135,256

Man Made A Bar (feat. Eric Church) por Morgan Wallen, Eric Church – Streams actuales: 58,890,931, Predichos: 380,832,559

Crown por Kendrick Lamar – Streams actuales: 42,485,571, Predichos: 364,350,577

RANDOM FOREST



RESULTADOS

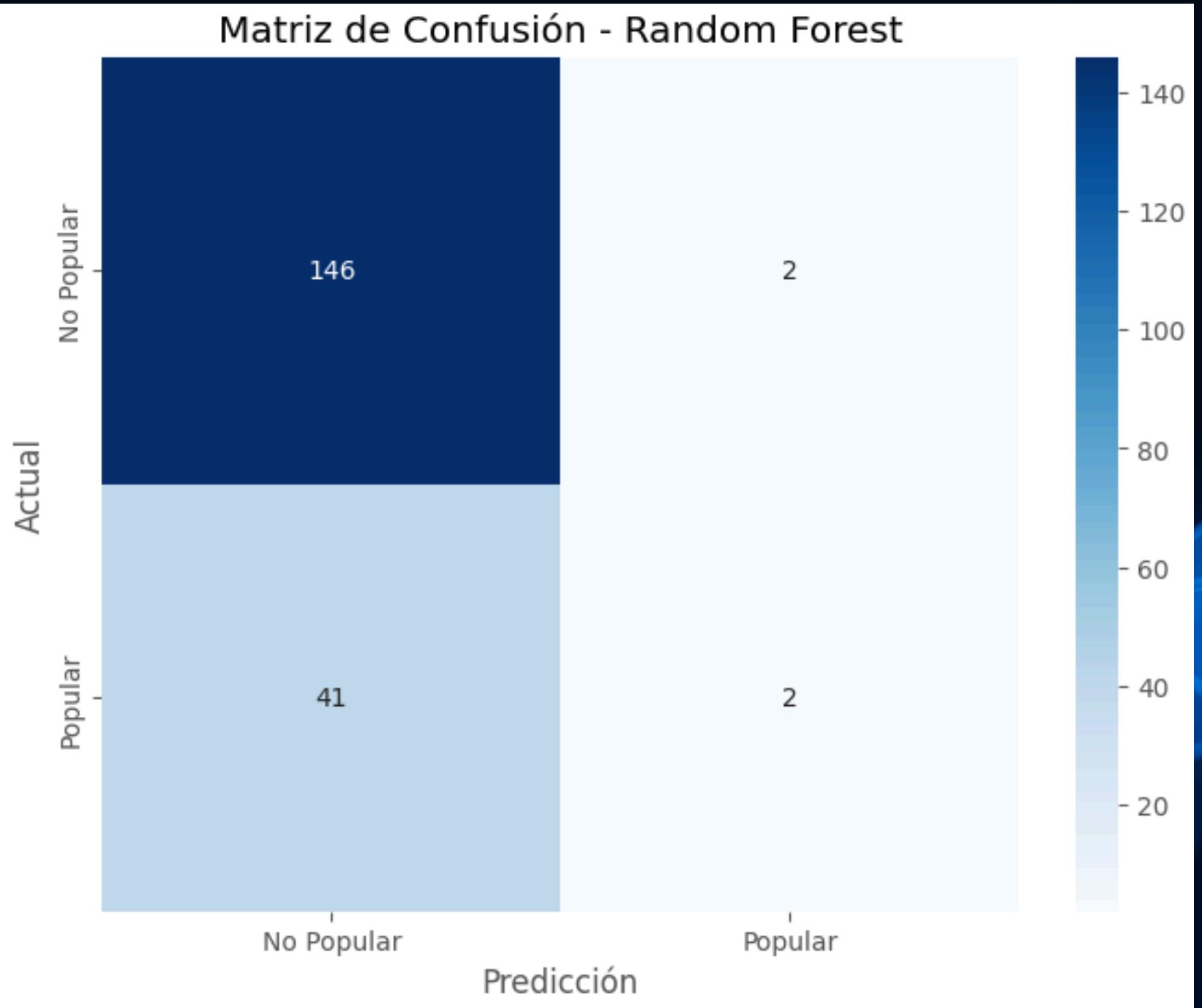
Métricas del Modelo Random Forest:

Precisión: 0.7749

Reporte de clasificación:

	precision	recall	f1-score	support
0	0.78	0.99	0.87	148
1	0.50	0.05	0.09	43
accuracy			0.77	191
macro avg	0.64	0.52	0.48	191
weighted avg	0.72	0.77	0.69	191

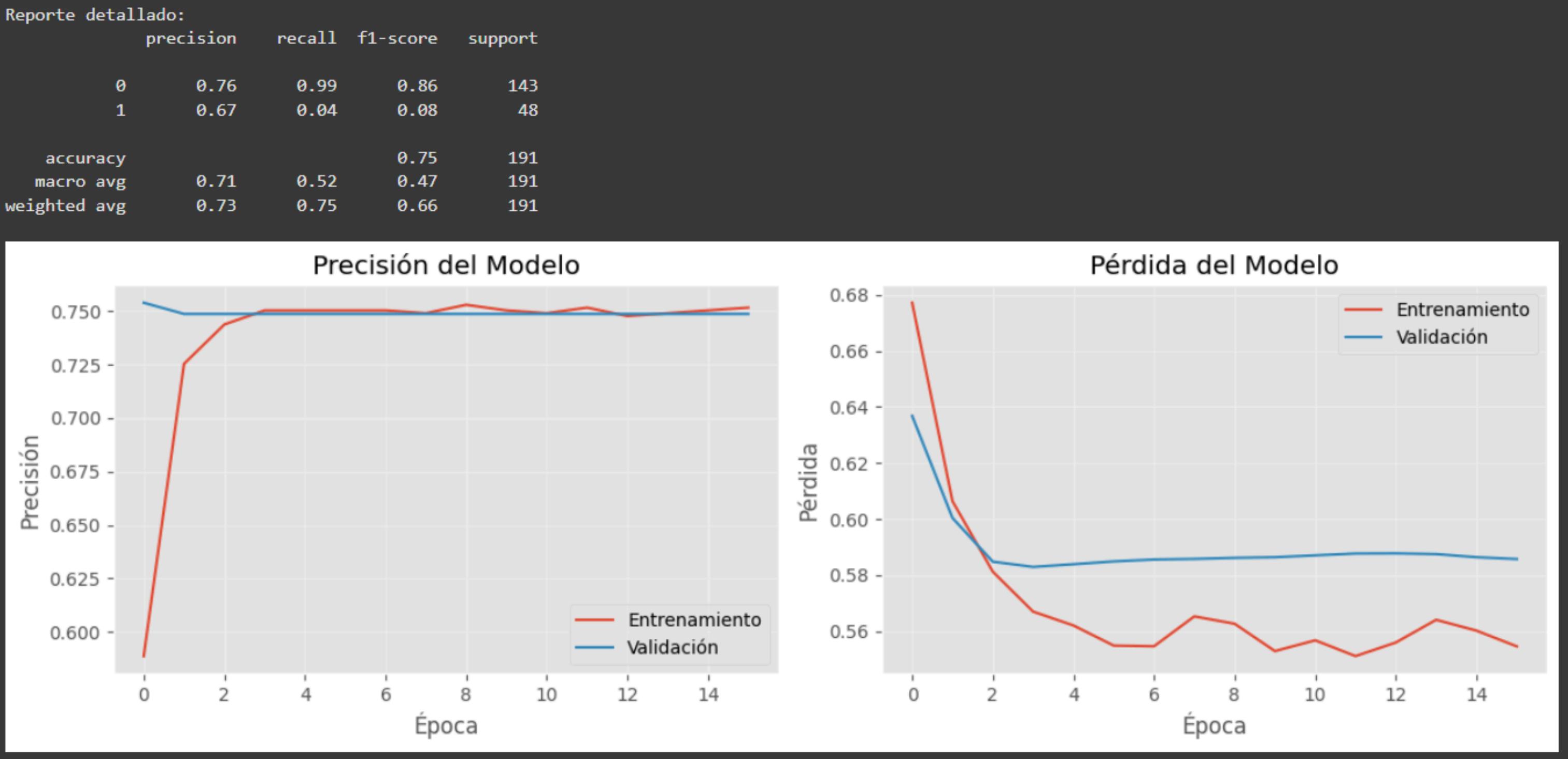
Matriz de Confusión - Random Forest



RED NEURONAL



RESULTADOS



MODELO BASADO EN PCA





RESULTADOS

🏆 TOP 5 CANCIONES ACTUALES (Confirmadas):
1. 'Blinding Lights' - The Weeknd (3,703,895,074 streams)
2. 'Shape of You' - Ed Sheeran (3,562,543,890 streams)
3. 'Someone You Loved' - Lewis Capaldi (2,887,241,814 streams)
4. 'Dance Monkey' - Tones and I (2,864,791,672 streams)
5. 'Sunflower - Spider-Man: Into the Spider-Verse' - Post Malone, Swae Lee (2,808,096,550 streams)

📊 ANÁLISIS COMPLETO FINALIZADO
Varianza explicada total del modelo: 96.13%

MÉTRICAS DE PRECISIÓN:

Exactitud (Accuracy): 58.19%

Precisión (Precision): 58.25%

Sensibilidad (Recall): 58.19%

F1-Score: 58.22%

◆ TOP 5 CANDIDATOS CON POTENCIAL (Según PCA):

1. 'Slut Me Out' - NLE Choppa
Streams: 190,490,915
Confianza PCA: 100.0%
Estado actual: No Popular
Potencial: Alto
2. 'On BS' - Drake, 21 Savage
Streams: 170,413,877
Confianza PCA: 99.9%
Estado actual: No Popular
Potencial: Alto
3. 'Cartiñ%íñ%o B' - MC Caverinha, KayBlack
Streams: 71,573,339
Confianza PCA: 99.8%
Estado actual: No Popular
Potencial: Alto
4. 'Doja' - Central Cee
Streams: 482,257,456
Confianza PCA: 99.7%
Estado actual: No Popular
Potencial: Alto
5. 'Me Arrepentiñ' - Ak4:20, Cris Mj, Pailita
Streams: 118,381,354
Confianza PCA: 99.6%
Estado actual: No Popular
Potencial: Alto

MODELO DBSCAN



RESULTADOS

Top 5 canciones más escuchadas:

	track_name	artist(s)_name
55	Blinding Lights	The Weeknd
179	Shape of You	Ed Sheeran
86	Someone You Loved	Lewis Capaldi
620	Dance Monkey	Tones and I
41	Sunflower - Spider-Man: Into the Spider-Verse	Post Malone, Swae Lee

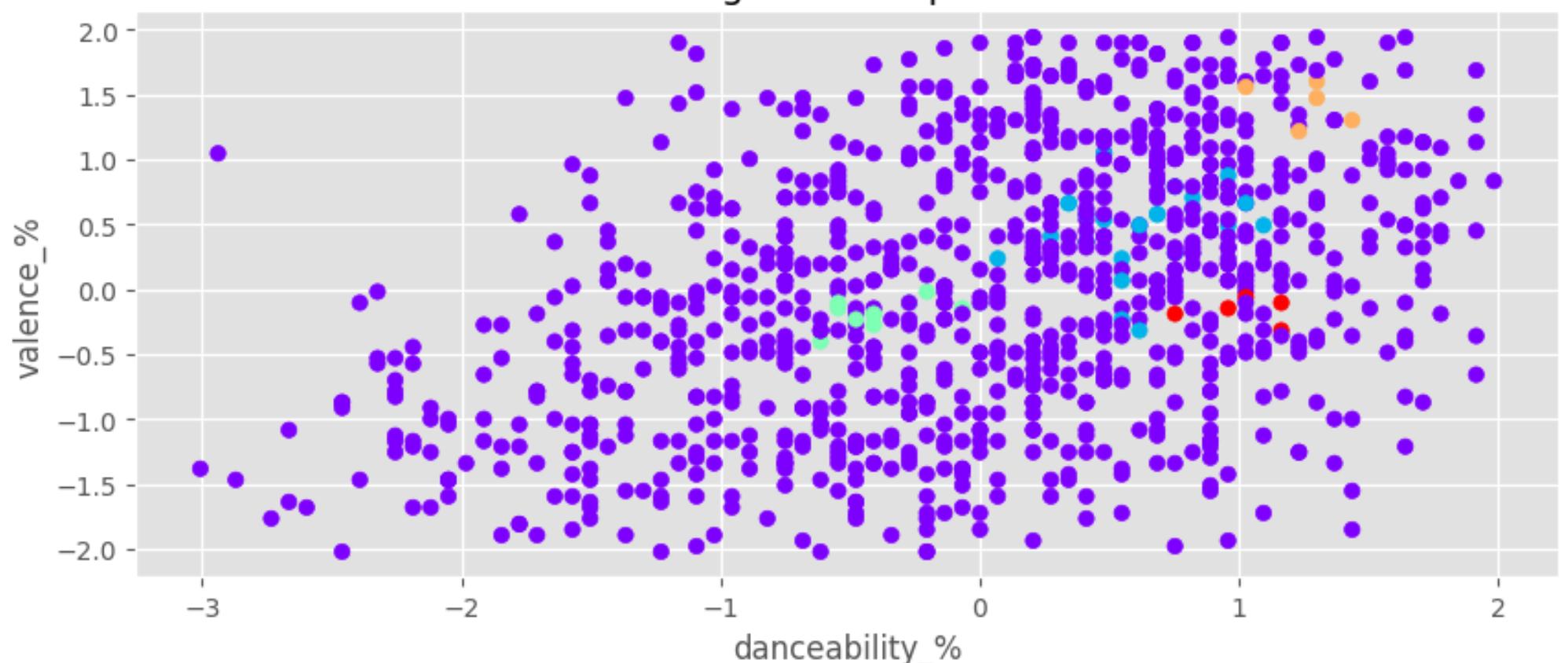
streams

55	3,703,895,074
179	3,562,543,890
86	2,887,241,814
620	2,864,791,672
41	2,808,096,550

5 canciones que podrían entrar al top (por similitud):

	track_name	artist(s)_name	streams
0	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	141,381,703
2	vampire	Olivia Rodrigo	140,003,974
3	Cruel Summer	Taylor Swift	800,840,817
4	WHERE SHE GOES	Bad Bunny	303,236,322
5	Sprinter	Dave, Central Cee	183,706,234

Clústeres generados por DBSCAN



Modelos	Accuracy	Precisión	Recall	F1_score	MSE	RMSE
Regresión lineal	-	-	-	-	1.9224	1.3865
Random Forest	0.77	0.64	0.52	0.48	-	-
Red Neuronal	0.75	0.71	0.52	0.47	-	-
PCA	0.58	0.58	0.58	0.58	-	-

CONCLUSIÓN

Los modelos de Machine Learning aplicados (Regresión, Random Forest, Red Neuronal) y las técnicas como PCA y Clustering fueron efectivos para entender y predecir el éxito musical en Spotify 2023. Se logró identificar la influencia de diversas características en la popularidad y predecir el potencial de las canciones. Los resultados de precisión y los análisis de importancia de características confirman que estos enfoques basados en datos son útiles para descifrar qué atributos impulsan los streams y la popularidad, permitiendo incluso descubrir canciones con potencial aún no reconocido.

GRACIAS!

POR SU ATENCIÓN