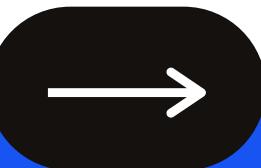
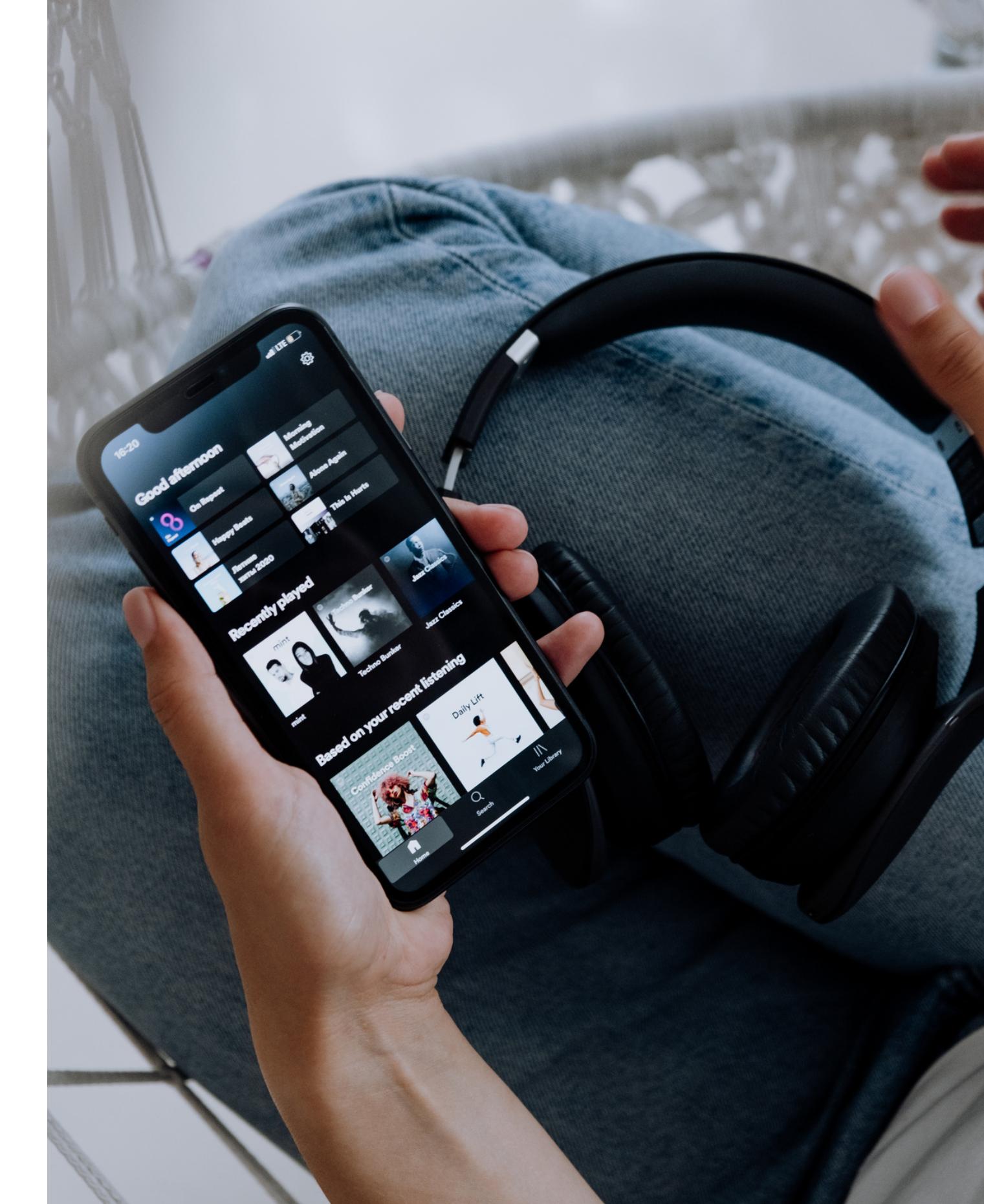


Análisis de Datos

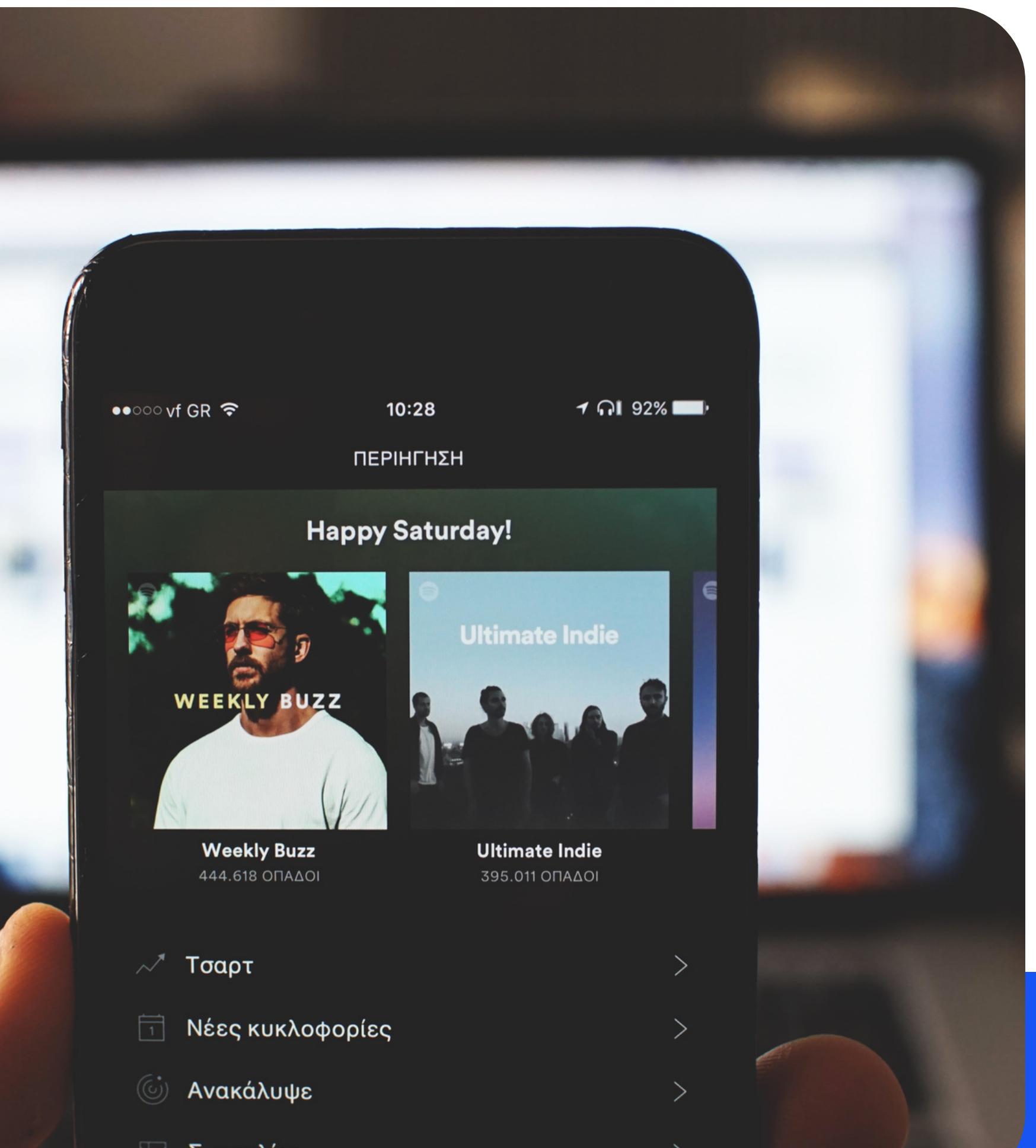
# Predicción de popularidad usando data de Spotify



# Objetivos

Descubrir si existe una correlación entre la popularidad y alguna característica de las canciones

Desarrollar un modelo analítico que nos indique que tan popular puede ser una canción según las características que esta posea





# Dataset

**Origen del Dataset** Spotify for Developers API - Kaggle

**Variables** Popularidad, duration\_ms, danceability

# Metodología KDD

## (Knowledge Discovery in Databases)

# Steps

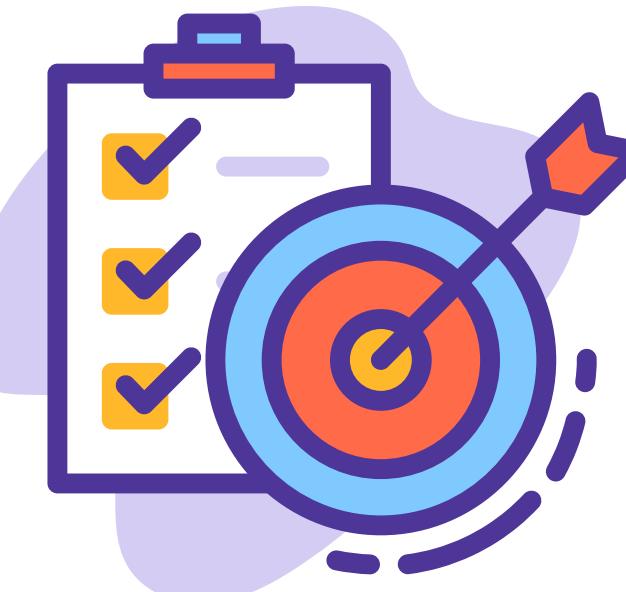
1

Entendimiento  
del negocio



2

Creación de  
dataset objetivo



# Steps

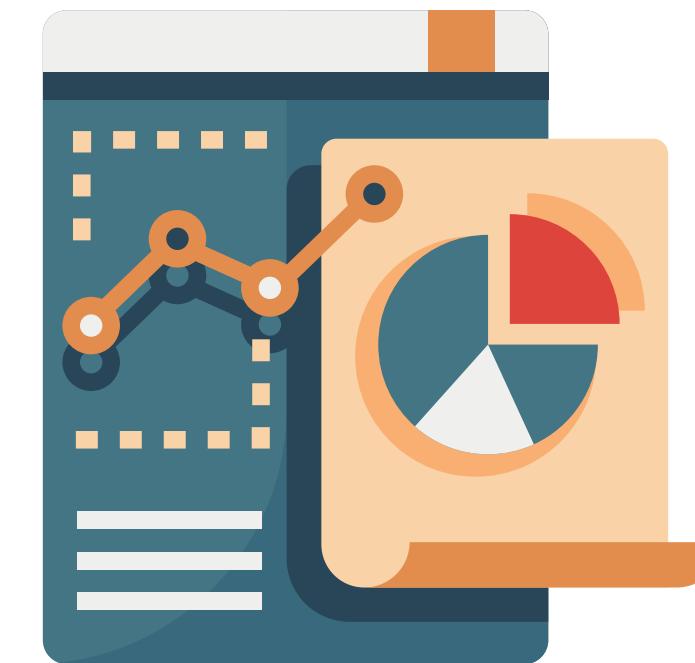
3

Limpieza de data y  
preprocesamiento



4

Reducción y  
transformación



# Steps

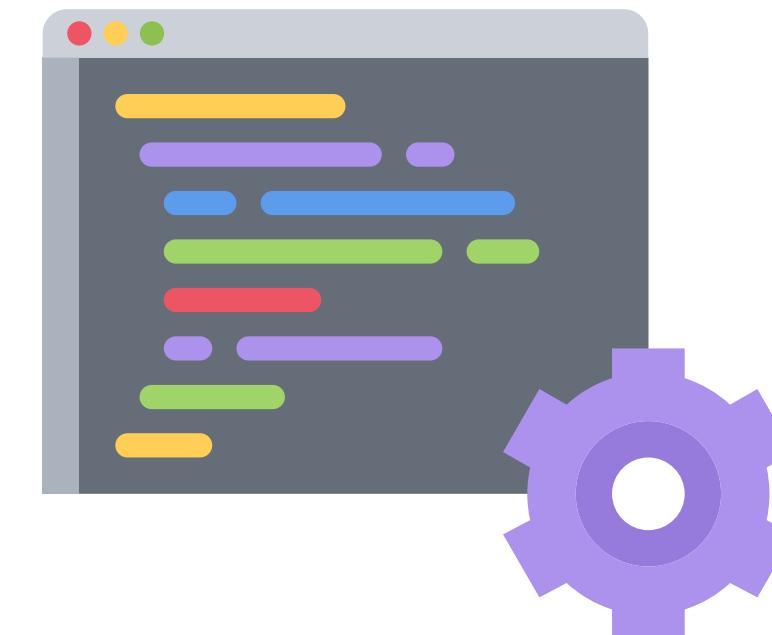
5

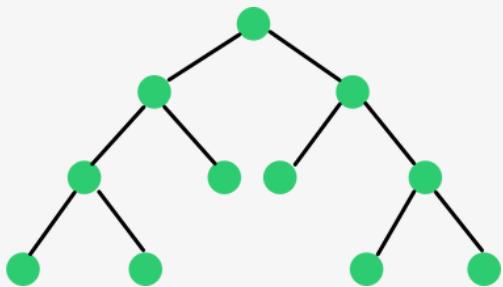
Elección del  
modelo analítico



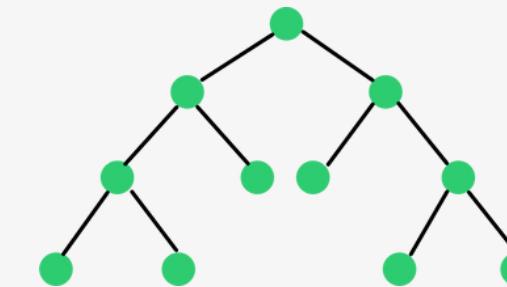
6

Elección de  
algoritmos y pruebas





# Modelos Analíticos



Nº	Modelo	Balanceo de datos	Tipo de balanceo	Grid Search	Accuracy	Accuracy - Cross Validation	Standar Deviation - Cross Validation	Sensibilidad	Especificidad	Parámetros
1	Árbol de clasificación CART	No	-	No	0.696	0.731	0.047	0.456	0.784	semilla aleatoria, índice de Gini
2	Árbol de clasificación CART	Si	Smote	No	0.671	0.696	0.002	0.535	0.721	semilla aleatoria, índice de Gini
3	Árbol de clasificación CART	Si	Smote	Si	0.675	0.701	0.002	0.539	0.725	semilla aleatoria, índice de entropía, splitter best
4	Regresión Logística	No	-	Si	0.758	0.756	0.001	0.263	0.940	max_iter:200, semilla aleatoria
5	Regresión Logística	Si	Smote	Si	0.679	0.756	0.001	0.694	0.674	max_iter:200, semilla aleatoria
6	Árbol de clasificación CART	Si	RandomUnderSampler	No	0.634	0.696	0.002	0.635	0.634	semilla aleatoria, índice de Gini
7	Árbol de clasificación CART	Si	RandomUnderSampler	Si	0.636	0.701	0.002	0.637	0.636	'criterion': 'entropy', 'random_state': 0, 'splitter': 'best'
8	Regresión Logística	Si	RandomUnderSampler	Si	0.729	0.756	0.001	0.080	0.968	'max_iter': 200, 'random_state': 0
9	Random Forest Classifier	Si	RandomUnderSampler	Si	0.732	-	-	0.013	0.996	{'criterion': 'entropy', 'random_state': 0}
10	Random Forest Classifier	Si	Smote	Si	0.747	-	-	0.587	0.806	{'criterion': 'entropy', 'random_state': 0}

A blue-tinted photograph of a person sitting at a desk. They are wearing a light-colored shirt and are focused on a laptop screen. On the desk in front of them is a keyboard and a white mug. The background is slightly blurred, showing what appears to be a window or a wall with some decorations.

**¡Muchas gracias!**