

MODELOS

Modelos Dicotómicos

Logit

- Basado en la función logística
- Útil para distribuciones no normales
- Aplicación

Probabilidad de que un cliente compre un producto

Probit

- Basado en la función acumulativa normal
- Ideal para precisar umbrales
- Aplicación

Aprobación de un crédito

Modelos supervisados

LDA (Análisis Discriminante Lineal)

- Clasificador lineal basado en la maximización de separación entre clases
- Asume normalidad y varianza homogénea
- Aplicación

Diagnóstico médico, análisis financiero

QDA (Análisis Discriminante Cuadrático)

- Extensión de LDA para casos no lineales
- Permite diferentes matrices de covarianza por clase
- Aplicación

Clasificación con clases más complejas

Modelos de Conglomerados

Clustering Jerárquico

- Agrupar datos en niveles anidados
- Ventajas
- Visualización

- Útil para estructuras jerárquicas
- Mejor para conjuntos pequeños
- Dendrogramas

Clustering Particional

- K-Means
- K-Medoids

- Divide datos en grupos homogéneos basados en centroides
- Aplicación
- Similar a K-Means pero robusto ante outliers
- Aplicación

Segmentación de clientes

Datos con métricas no euclídeas

Distancia

- Distancia euclídeana
- Distancia matricial
- Manhattan

- $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}$
- Elevar al cuadrado

HDBSCAN

- Basado en densidad jerárquica.
- Maneja outliers automáticamente
- No requiere especificar el número de clusters
- Aplicación

Análisis geoespacial y detección de anomalías

Árboles de Decisión

CART (Classification and Regression Trees)

- Divide datos en ramas binarias
- Ideal para clasificación y regresión
- Ejemplo

Decisión sobre otorgar un crédito

Árboles Aleatorios (Random Forests)

- Conjunto de múltiples árboles con votación
- Reduce el sobreajuste
- Ejemplo

Predicción de fraudes

Árboles de Gradiente (Gradient Boosting)

- Optimiza secuencialmente el error de los árboles anteriores

Redes Neuronales

Perceptrón

- Red más simple
- Clasifica datos linealmente separables

Redes Multicapa

- Capturan patrones no lineales complejos
- Aplicaciones

- Predicción de series temporales
- Análisis de imágenes y texto

Procesamiento de Lenguaje Natural (NLP)

Técnicas clave

- Tokenización
- Embeddings
- Corpus

- Dividir texto en palabras o frases
- Representar palabras en vectores
- Modelos de embedding

- Codificación
- Neutralizar

- UTF-8
- ASCII
- Latín
- Números
- Artículos
- Conectores
- Minúsculas
- Stop words
- Positivo
- Neutro
- Negativo

Aplicaciones

- Análisis de sentimientos
- Clasificación automática de textos
- Transformers
- Bases de datos vectoriales

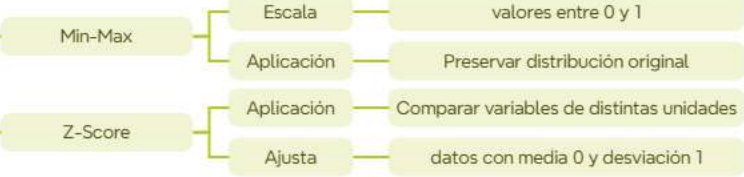
Diccionario

ROADMAP

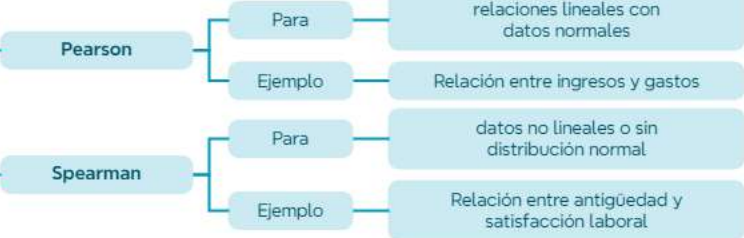
1. Realizar análisis exploratorio

2. Dividir datos en entrenamiento y prueba

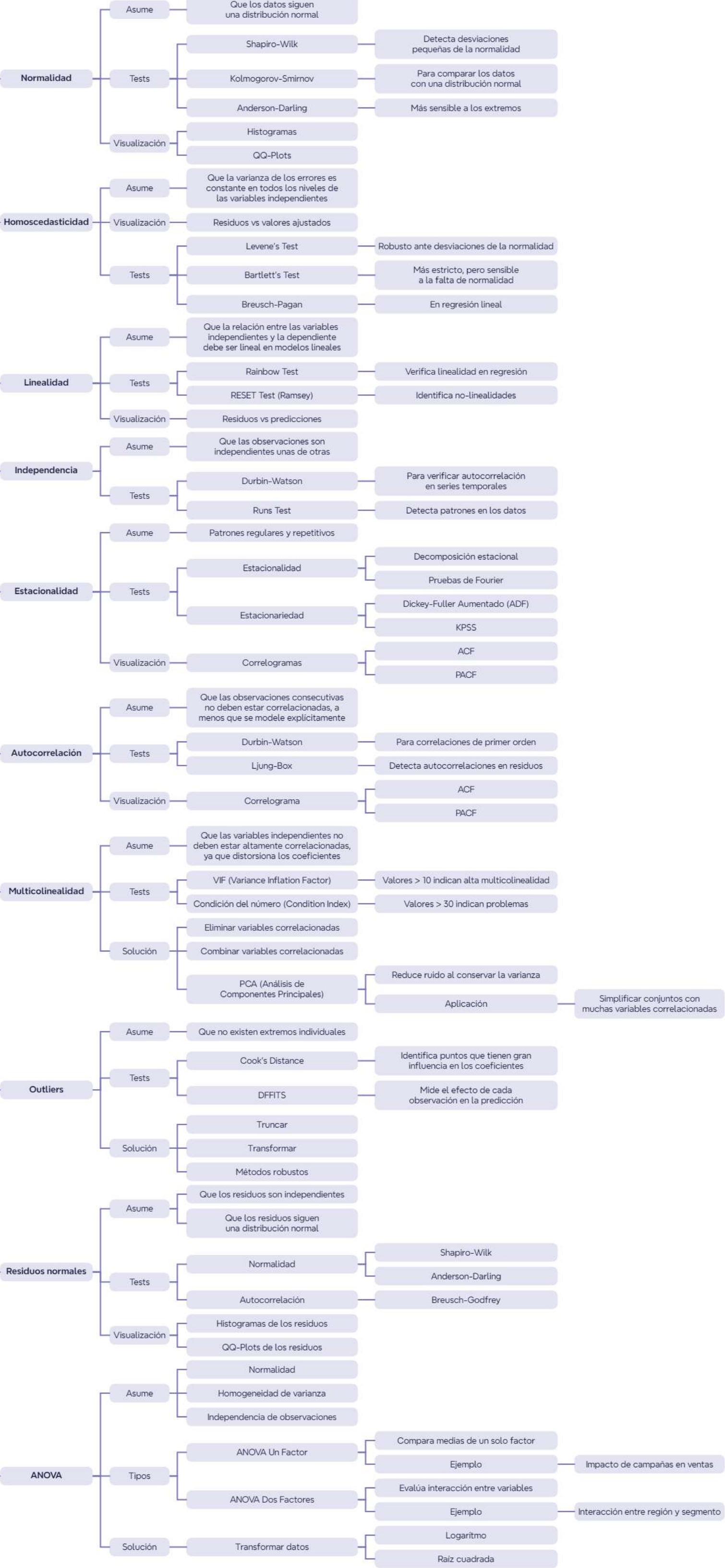
3. Normalizar y escalar variables



4. Identificar correlaciones relevantes



5. Probar supuestos estadísticos



6. Entrenar el modelo seleccionado

7. Evaluar residuos y métricas de desempeño

8. Ajustar hiperparámetros según resultados