

Índice

[Investigación teórica: 3](#_Toc190948720)

[ Explica qué es el análisis predictivo y su relación con la minería de datos. 3](#_Toc190948721)

[ Describe al menos tres técnicas utilizadas en minería de datos para el análisis predictivo (ejemplo: árboles de decisión, regresión logística, redes neuronales). 3](#_Toc190948722)

[ Investiga casos de uso reales donde se haya aplicado análisis predictivo en Big Data. 4](#_Toc190948723)

[Aplicación práctica: 4](#_Toc190948724)

[ Elige un conjunto de datos abierto (ejemplo: Kaggle, UCI Machine Learning Repository). 4](#_Toc190948725)

[ Preprocesa los datos (limpieza, transformación y reducción si es necesario). 5](#_Toc190948726)

[ Implementa un modelo de análisis predictivo usando Python y una herramienta de minería de datos como Scikit-learn, TensorFlow o Weka. 5](#_Toc190948727)

[ Evalúa el rendimiento del modelo con métricas adecuadas (precisión, recall, F1-score). 6](#_Toc190948728)

# Investigación teórica:

## Explica qué es el análisis predictivo y su relación con la minería de datos.

El **análisis predictivo** es un enfoque analítico que utiliza datos históricos y actuales para predecir eventos futuros o tendencias.  
La **minería de datos** juega un papel fundamental en el análisis predictivo, ya que es el proceso de descubrir patrones, relaciones y tendencias significativas en grandes conjuntos de datos.

## Describe al menos tres técnicas utilizadas en minería de datos para el análisis predictivo (ejemplo: árboles de decisión, regresión logística, redes neuronales).

1. **Árboles de decisión**:
   * Los árboles de decisión son estructuras de flujo lógico que dividen iterativamente un conjunto de datos en subconjuntos más pequeños basados en atributos específicos.
   * Cada nodo del árbol representa una decisión o criterio, y las hojas representan los resultados finales. Es una técnica popular debido a su facilidad de interpretación y visualización.
2. **Regresión logística:**
   * Este método se utiliza para modelar relaciones entre una variable dependiente binaria (como "sí" o "no") y una o más variables independientes.
   * Se basa en el uso de una función logística para predecir probabilidades.
3. **Redes neuronales:** 
   * Las redes neuronales son modelos inspirados en el cerebro humano que consisten en capas de neuronas artificiales interconectadas.
   * Son capaces de modelar relaciones complejas y no lineales en los datos, lo que las hace extremadamente poderosas en el análisis predictivo.

## Investiga casos de uso reales donde se haya aplicado análisis predictivo en Big Data.

**Algunos casos de uso reales donde se haya aplicado el análisis predictivo:**

1. **Sector de la salud:**

Instituciones médicas utilizan análisis predictivo para predecir la probabilidad de enfermedades en pacientes basándose en datos médicos históricos.

* Por ejemplo, los algoritmos predictivos ayudan a identificar pacientes con alto riesgo de enfermedades cardiovasculares.

1. **Sector financiero**

Bancos y entidades financieras aplican análisis predictivo para detectar fraudes en tiempo real.

* Por ejemplo, sistemas como los utilizados por **Visa** analizan patrones de transacciones para identificar actividades sospechosas.

1. **Plataformas digitales**

Plataformas como Amazon o Netflix aplican análisis predictivo para ofrecer recomendaciones personalizadas a los usuarios basadas en su historial de compras o preferencias.

* Esto mejora la experiencia del usuario y aumenta la conversión.

# Aplicación práctica:

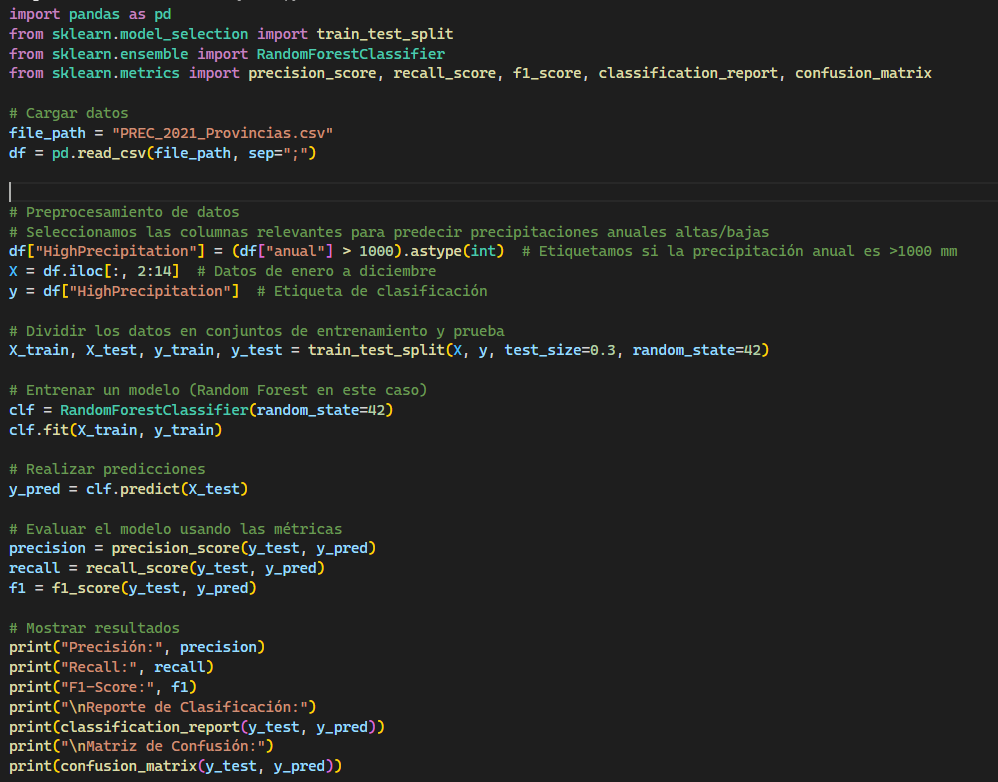
## Elige un conjunto de datos abierto (ejemplo: Kaggle, UCI Machine Learning Repository).

Elegimos un dataset sobre las precipitaciones de las regiones de España de 2021, de enero a diciembre y una columna Anual.

Puedes verlo [pinchando aquí](https://www.kaggle.com/datasets/josepblasco/hidro-data?select=PREC_2021_Provincias.csv), o en el archivo adjunto a la práctica.

## Preprocesa los datos (limpieza, transformación y reducción si es necesario).

## Implementa un modelo de análisis predictivo usando Python y una herramienta de minería de datos como Scikit-learn, TensorFlow o Weka.

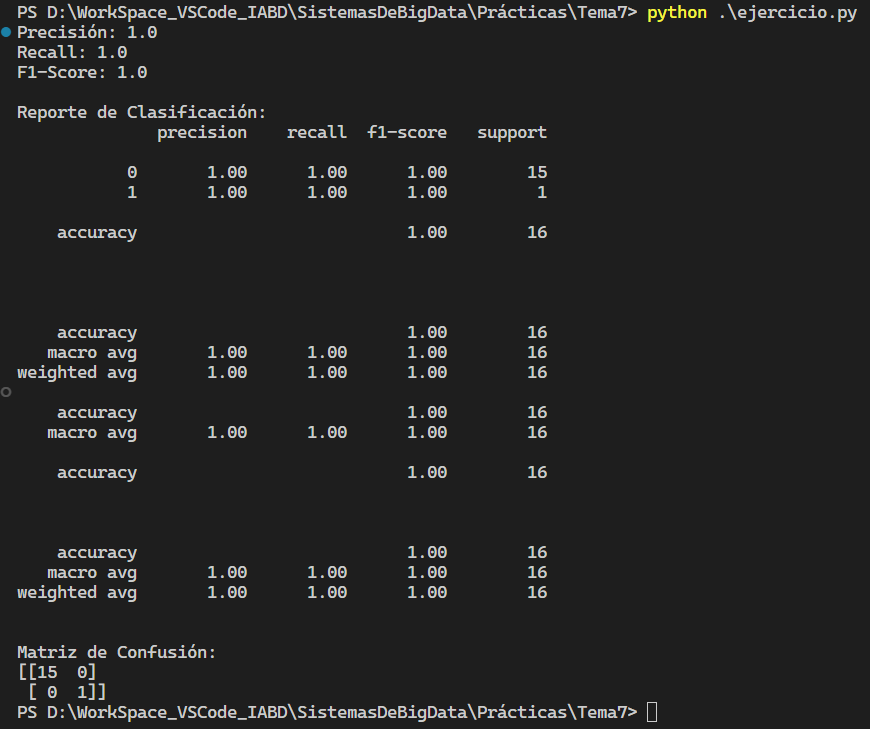


Cargamos los datos, preprocesamos y seleccionamos lo relevante.

Posteriormente, dividimos los datos y entrenamos un modelo.

Luego, hacemos predicciones y lo evaluamos mostrando finalmente los resultados.

## Evalúa el rendimiento del modelo con métricas adecuadas (precisión, recall, F1-score).



Al ejecutar, nos salen los datos.