1 de abril de 2025

Actividad de Aprendizaje 6 – Ciencia de Datos con Python

ANÁLISIS Y DESARROLLO DE SOFTWARE

Actividad 2

eSTEFANÍA BETANCUR LEÓN

ANJHELL REYES CUFIÑO

SERGIO MORALES

ANDY RIOS

Análisis Predictivo del Largo del Pétalo usando Regresión Lineal en el Conjunto de Datos Iris

Resumen

Este informe presenta un análisis predictivo utilizando el conjunto de datos Iris, ampliamente utilizado en el ámbito del aprendizaje automático. Se aplicó un modelo de regresión lineal con el fin de predecir la longitud del pétalo (petal\_length) en función de otras características como sepal\_length, sepal\_width y petal\_width. Se utilizaron métricas de evaluación como el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R²), obteniendo un rendimiento predictivo alto. Este modelo demuestra ser útil para entender la influencia de ciertas características en la longitud del pétalo, lo cual puede ser aplicado como base para modelos más complejos.

1. Introducción

El aprendizaje automático ha revolucionado la manera en la que se analizan grandes volúmenes de datos, permitiendo realizar predicciones precisas en diversos campos. Uno de los conjuntos de datos más utilizados para la experimentación es el conjunto Iris, introducido por Ronald Fisher (1936), que contiene mediciones de flores de tres especies diferentes. En este estudio se aplica un modelo de regresión lineal con el objetivo de predecir la longitud del pétalo (petal\_length) basándose en otras variables del conjunto.

2. Metodología

2.1 Conjunto de Datos

El conjunto de datos Iris contiene 150 muestras de flores, con cuatro características: sepal\_length, sepal\_width, petal\_length y petal\_width, así como la especie correspondiente. En este estudio se utilizó la versión del conjunto de datos incluida en la librería `Seaborn` de Python.

2.2 Preprocesamiento de Datos

Se seleccionaron como variables predictoras (features) las columnas sepal\_length, sepal\_width y petal\_width. La variable objetivo (target) fue petal\_length. Los datos se dividieron en dos subconjuntos: 80% para entrenamiento y 20% para prueba.

2.3 Modelo Utilizado

Se aplicó un modelo de regresión lineal utilizando la clase `LinearRegression()` de la biblioteca `scikit-learn`. El modelo fue entrenado sobre los datos de entrenamiento y luego evaluado sobre el conjunto de prueba.

2.4 Métricas de Evaluación

Para evaluar el modelo, se utilizaron dos métricas comunes:

- Error Cuadrático Medio (MSE): mide la diferencia promedio al cuadrado entre los valores predichos y los reales.

- Coeficiente de Determinación (R²): indica qué proporción de la varianza en la variable objetivo es explicada por el modelo.

3. Resultados

Los resultados obtenidos fueron los siguientes:

- Error Cuadrático Medio (MSE): 0.13

- Coeficiente de Determinación (R²): 0.96

Estos resultados indican que el modelo tiene un buen rendimiento, ya que explica el 96% de la variabilidad en la longitud del pétalo, con un error medio bajo.

4. Conclusiones

La regresión lineal demostró ser un modelo eficiente para predecir la longitud del pétalo en función de otras características morfológicas de la flor. La alta puntuación en R² y el bajo error cuadrático medio sugieren que el modelo es confiable y preciso. Este análisis puede ser un punto de partida para proyectos más complejos, como la clasificación de especies de flores o la incorporación de redes neuronales para mayor precisión.

