



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

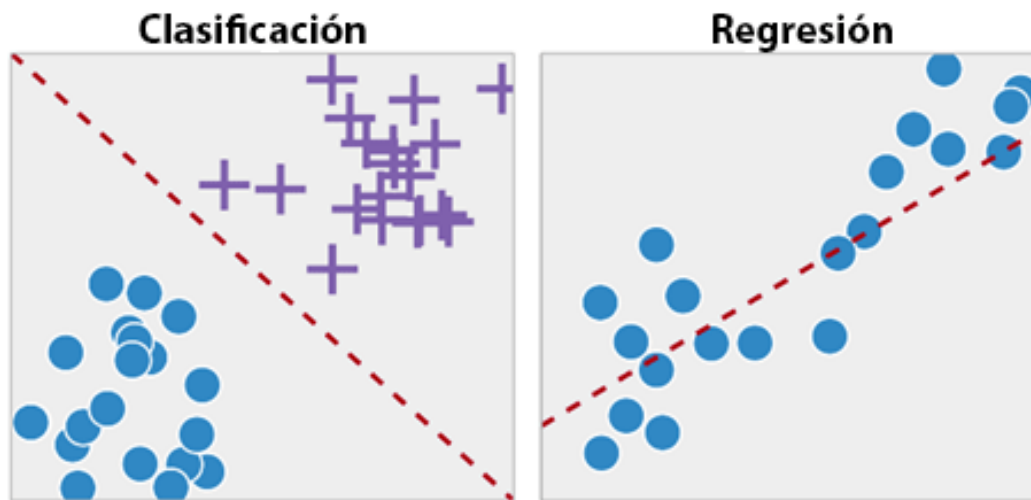
INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Facultad de Ingeniería Asignatura Machine Learning

Unidad de aprendizaje 2: Supervisado – Regresión

Introducción

Hola, estamos en la segunda unidad de la asignatura denominada: **“Aprendizaje Supervisado - Regresión”**. En esta unidad aprenderemos qué son los algoritmos de regresión. También realizaremos algunas pruebas con los algoritmos: Regresión Lineal, Máquinas de Soporte Vectorial (SVM) y



Bosque Aleatorio.

Figura: <https://masterdatascience.online/aprendizaje-automatico-supervisado/>



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

2.1 ¿Qué son los algoritmos de regresión?

Los algoritmos de regresión son un conjunto de técnicas en el campo del aprendizaje automático y la estadística utilizadas para modelar y analizar la relación entre una variable dependiente (también llamada variable objetivo o de respuesta) y una o más variables independientes (también llamadas características o predictores). El objetivo principal de la regresión es predecir o estimar el valor de la variable dependiente en función de los valores de las variables independientes. Esta tarea es comúnmente utilizada en problemas de predicción y análisis de datos en los que se busca entender cómo una o más variables afectan a otra.

Los algoritmos de regresión se utilizan principalmente en dos tipos de problemas:

1. Regresión Lineal: Este es uno de los métodos más simples y ampliamente utilizados en regresión. Se asume que la relación entre las variables dependientes e independientes es lineal. Un ejemplo clásico es la regresión lineal simple, que involucra una variable independiente y una variable dependiente. En regresión lineal múltiple, se consideran múltiples variables independientes.

2. Regresión No Lineal: Cuando la relación entre las variables no es lineal, se utilizan algoritmos de regresión no lineal. Estos modelos pueden tener formas más complejas y flexibles para capturar la relación entre las variables. Algunos ejemplos incluyen la regresión polinómica, la regresión logística y los modelos de regresión basados en árboles, como los árboles de decisión y los bosques aleatorios.

Los algoritmos de regresión se utilizan en una amplia variedad de aplicaciones, incluyendo:

- **Predicción de precios de bienes raíces en función de características como ubicación, tamaño y antigüedad.**
- **Pronóstico de ventas en función de factores como la temporada, la publicidad y el precio.**
- **Modelado de la relación entre la edad y la presión arterial en estudios médicos.**
- **Predicción del rendimiento de un estudiante en función de variables como el tiempo de estudio, el sueño y la asistencia a clases.**

Los algoritmos de regresión permiten analizar y modelar relaciones entre variables y hacer predicciones útiles en una amplia gama de campos, desde la ciencia de datos y la investigación médica hasta la economía y la ingeniería. Cada algoritmo de regresión tiene sus propias suposiciones y características, por lo que la elección del método adecuado depende de la naturaleza de los datos y el problema en cuestión.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

2.2 Regresión Lineal

La regresión lineal es un método estadístico y matemático utilizado en el campo del análisis de datos y el aprendizaje automático para modelar la relación entre una variable dependiente (también conocida como variable objetivo) y una o más variables independientes (también conocidas como características o predictores). El objetivo principal de la regresión lineal es encontrar una relación lineal que describa cómo los cambios en las variables independientes están relacionados con los cambios en la variable dependiente.

La forma más simple de regresión lineal se llama "regresión lineal simple", que involucra una única variable independiente y una única variable dependiente. La ecuación de regresión lineal simple tiene la forma:

$$Y = aX + b$$

Donde:

- **Y es la variable dependiente.**
- **X es la variable independiente.**
- **a es la pendiente de la línea (coeficiente de regresión).**
- **b es la intersección en el eje Y cuando X = 0 (también llamada término de intersección o sesgo).**

El objetivo en la regresión lineal es encontrar los valores óptimos de *a* y *b* de manera que la línea se ajuste de la mejor manera posible a los datos observados. Esto se hace minimizando la suma de los errores cuadrados, que mide la discrepancia entre los valores predichos por la línea de regresión y los valores reales en los datos.

La regresión lineal es útil para varios propósitos, incluyendo:

1. Predicción: Puedes utilizar un modelo de regresión lineal para hacer predicciones sobre la variable dependiente en función de los valores de las variables independientes. Por ejemplo, predecir el precio de una casa en función de su tamaño y ubicación.

2. Análisis de relaciones: La regresión lineal puede ayudar a analizar y cuantificar la relación entre variables. Puedes determinar si existe una relación positiva o negativa entre las variables y qué tan fuerte es.

3. Control de variables: En investigaciones científicas y experimentos, la regresión lineal puede utilizarse para controlar el efecto de las variables independientes en la variable dependiente, teniendo en cuenta otras variables que puedan influir.

La regresión lineal es una técnica poderosa y ampliamente utilizada, pero tiene limitaciones. Funciona mejor cuando la relación entre las variables es aproximadamente lineal, y puede no ser adecuada para modelar relaciones más complejas. En tales casos, se pueden explorar técnicas de regresión no lineal o modelos más avanzados de aprendizaje automático.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Ejemplo en Python:

✓ Ejercicio Práctico: Regresión lineal

En este ejercicio vamos a predecir la concentración del Glifosato en el agua.

✓ Paso 1: Importar las librerías

```
[1] # Librería para operaciones matemáticas o estadísticas
import numpy as np
# Librería para el manejo de datos
import pandas as pd
# Librerías para gráficas
import matplotlib.pyplot as plt
# Librería para separar los datos de entrenamiento y pruebas
from sklearn.model_selection import train_test_split
# Librería para el algoritmo de regresión lineal
from sklearn.linear_model import LinearRegression
# Librería para las métricas de regresión
from sklearn.metrics import mean_squared_error, r2_score
```

✓ Paso 2: Cargar los datos del archivo plano a un DataFrame

```
[2] # Se cargan los datos de entrada
data = pd.read_csv("2.3-glifosato.csv", sep=";")
# Se observa la cantidad de registros y columnas del archivo
data.shape
```

(626, 9)

```
[3] # Se observan los primeros registros
data.head()
```

	date(m/d/y)	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
0	25/01/05	12.30	1	0.0	17.02	13.17	747	8.68	0.1
1	25/01/05	12.31	1	0.0	17.12	13.20	747	8.67	0.1
2	25/01/05	12.32	1	0.0	17.12	13.25	747	8.68	0.1
3	25/01/05	12.33	1	0.0	17.11	13.19	747	8.68	0.1
4	25/01/05	12.34	1	0.0	17.03	13.20	747	8.68	0.1

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

```
[4] # Se observan las estadísticas de los datos, para valores numéricos  
data.describe()
```

	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
count	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000
mean	13.869968	1.504792	1.161981	17.323882	13.348083	739.704473	8.652077	0.068371
std	1.857436	0.500377	1.164567	0.511636	0.133306	7.442400	0.091153	0.101617
min	10.050000	1.000000	0.000000	16.550000	12.980000	717.000000	8.240000	-0.100000
25%	12.562500	1.000000	0.000000	16.880000	13.280000	737.000000	8.650000	0.000000
50%	14.120000	2.000000	0.400000	17.105000	13.400000	740.000000	8.670000	0.000000
75%	15.287500	2.000000	1.500000	17.880000	13.450000	746.000000	8.700000	0.100000
max	17.300000	2.000000	3.000000	18.350000	13.530000	753.000000	8.730000	0.500000

✓ Paso 3: Procesamiento de los datos

```
[5] # Eliminar filas con valores nulos  
data = data.dropna()  
  
# Mostrar las primeras filas  
data.head()
```

	date(m/d/y)	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
0	25/01/05	12.30	1	0.0	17.02	13.17	747	8.68	0.1
1	25/01/05	12.31	1	0.0	17.12	13.20	747	8.67	0.1
2	25/01/05	12.32	1	0.0	17.12	13.25	747	8.68	0.1
3	25/01/05	12.33	1	0.0	17.11	13.19	747	8.68	0.1
4	25/01/05	12.34	1	0.0	17.03	13.20	747	8.68	0.1

✓ Paso 3: Selección de variables y preparación para el modelo

```
[6] # Seleccionar variables dependientes e independientes  
X = data[['chloride(mg/L)', 'DO(mg/L)', 'ORP(mV)', 'pH', 'turbidity(NTU+)']  
y = data['concentration']  
  
# Dividir los datos en conjuntos de entrenamiento (80%) y prueba (20%)  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

✓ Paso 4: Crear y entrenar el modelo de regresión lineal

```
[7] # Crear el modelo de regresión lineal
    model = LinearRegression()

    # Entrenar el modelo
    model.fit(X_train, y_train)

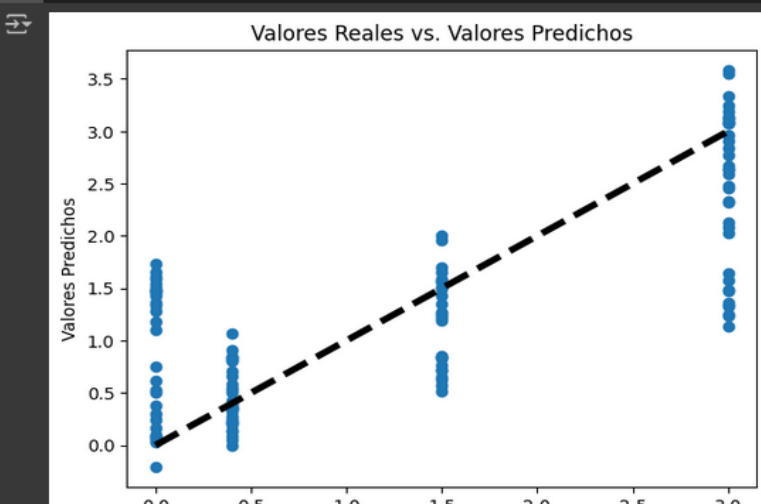
    # Realizar predicciones
    y_pred = model.predict(X_test)

    # Evaluar el modelo
    mse = mean_squared_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)
    print(f"Error cuadrático medio (MSE): {mse}")
    print(f"Coeficiente de determinación (R^2): {r2}")
```

➡ Error cuadrático medio (MSE): 0.6208290472239829
Coeficiente de determinación (R²): 0.5714080489262698

✓ Paso 5: Generar la gráfica de los resultados

```
# Generar la gráfica
plt.scatter(y_test, y_pred)
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=4)
plt.xlabel('Valores Reales')
plt.ylabel('Valores Predichos')
plt.title('Valores Reales vs. Valores Predichos')
plt.show()
```



📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



Evaluación del Modelo

- Error cuadrático medio (MSE): 0.62
- Coeficiente de determinación (R^2): 0.57

El MSE mide el promedio de los cuadrados de los errores, es decir, la diferencia promedio entre los valores reales y los valores predichos. Un valor más bajo de MSE indica un mejor ajuste del modelo.

El R^2 , también conocido como coeficiente de determinación, indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Un R^2 de 0.57 significa que aproximadamente el 57% de la variabilidad en la concentración de glifosato puede explicarse por las variables seleccionadas.

Conclusiones del Modelo

1. Precisión del Modelo:

El modelo tiene un error cuadrático medio (MSE) de 0.62 y un coeficiente de determinación (R^2) de 0.57. Esto sugiere que el modelo puede explicar aproximadamente el 57% de la variabilidad en la concentración de glifosato a partir de las variables independientes seleccionadas.

2. Interpretación de la Gráfica:

La gráfica de dispersión muestra los valores reales de la concentración de glifosato frente a los valores predichos por el modelo. La línea roja representa la línea de referencia donde los valores reales son iguales a los valores predichos.

La mayoría de los puntos están cerca de esta línea, lo que indica que el modelo tiene una capacidad razonable para predecir la concentración de glifosato. Sin embargo, hay algunos puntos alejados de la línea, lo que indica errores de predicción.

3. Mejora del Modelo:

Para mejorar el modelo, se podrían considerar las siguientes acciones:

- Incluir más variables independientes que puedan influir en la concentración de glifosato.
- Probar otros algoritmos de machine learning que puedan capturar mejor las relaciones no lineales entre las variables.
- Realizar una exploración de datos más profunda para identificar y tratar valores atípicos y posibles correlaciones entre variables.

4. Aplicación Práctica:

Este modelo puede ser útil para predecir la concentración de glifosato en función de variables como el cloruro, el oxígeno disuelto, el ORP, el pH y la turbidez del agua. Esto podría ser valioso para monitorear la calidad del agua y tomar decisiones informadas en la gestión ambiental y agrícola.

Puede ver el código fuente de este ejemplo en:

https://github.com/jose-llanos/ML_hidroinformatica/blob/main/2.2-RegresionLineal.ipynb

video: Ejemplo Regresión Lineal en Python

<https://www.youtube.com/watch?v=SZyH6YkQqIk>

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

2.3 Máquinas de Soporte Vectorial (SVM)

Las Máquinas de Soporte Vectorial (Support Vector Machines, SVM por sus siglas en inglés) son un conjunto de algoritmos de aprendizaje automático utilizados tanto en problemas de clasificación como en problemas de regresión. Fueron desarrolladas originalmente por Vladimir Vapnik y Alexey Chervonenkis en la década de 1960 y han demostrado ser altamente efectivas en una variedad de aplicaciones, incluyendo reconocimiento de imágenes, procesamiento de texto, diagnóstico médico y más.

El principio fundamental detrás de las SVM es encontrar un hiperplano (una superficie en dimensiones más altas) que mejor separe dos clases diferentes de datos en un espacio multidimensional. En otras palabras, las SVM buscan la "máxima separación" entre las clases. Este hiperplano se elige de tal manera que maximiza la distancia entre los puntos de datos más cercanos de ambas clases, conocidos como vectores de soporte, de ahí el nombre "Máquinas de Soporte Vectorial".

Las SVM pueden utilizarse tanto en problemas de clasificación binaria. Además, se pueden aplicar a datos linealmente separables (cuando las clases pueden separarse mediante un hiperplano) y a datos no linealmente separables utilizando técnicas de transformación de características y el uso de "kernels" (funciones que mapean los datos a un espacio dimensional superior).

Las características clave de las SVM incluyen:

1. Margen Máximo: Las SVM buscan encontrar el hiperplano con el margen máximo entre las clases, lo que proporciona una mayor robustez en la clasificación de nuevos datos.

2. Funciones de Kernel: Las SVM utilizan funciones de kernel (como el kernel lineal, el kernel polinómico y el kernel radial) para manejar datos no linealmente separables al mapearlos a un espacio dimensional superior donde puedan ser linealmente separables.

3. Vectores de Soporte: Los vectores de soporte son los puntos de datos más cercanos al hiperplano de separación y desempeñan un papel crucial en el algoritmo SVM.

4. Regularización: Las SVM incorporan la regularización para evitar el sobreajuste y mejorar la generalización del modelo.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

5. Clasificación Robusta: Las SVM son efectivas incluso en conjuntos de datos ruidosos o con solapamiento de clases.

Ejemplo en Python:

Ejercicio Práctico: SVM

Paso 1: Importar las librerías

```
1 # Librería para operaciones matemáticas o estadísticas
2 import numpy as np
3 # Librería para el manejo de datos
4 import pandas as pd
5 # Librerías para gráficas
6 import seaborn as sb
7 import matplotlib.pyplot as plt
8 # Librería de árbol de regresión para realizar pruebas con Eli5
9 from sklearn.tree import DecisionTreeRegressor
10 # Librería para las mejores características Eli5
11 from eli5 import show_weights
12 # Librería para dividir los datos de entrenamiento y pruebas
13 from sklearn.model_selection import train_test_split
14 # Librería para SVM con regresión
15 from sklearn.svm import SVR
16 # Se importa la librería para la métrica R cuadrado y RMSE
17 from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
```

Paso 2: Cargar los datos a un DataFrame

```
1 # Se cargan los datos de entrada
2 data = pd.read_csv("data/2.3-glifosato.csv", sep=";")
3 # Se observa la cantidad de registros y columnas del archivo
4 data.shape
```

(626, 9)

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Paso 3: Preprocesamiento de los datos

```
1 # Se buscan registros NaN para eliminarlos
2 print('Columna          Cantidad NaN')
3 print(data.isnull().sum(axis = 0))
4 print(data.shape)
```

Columna	Cantidad NaN
date(m/d/y)	0
time	0
day	0
concentration	0
chloride(mg/L)	0
DO(mg/L)	0
ORP(mV)	0
pH	0
turbidity(NTU+)	0

dtype: int64
(626, 9)

```
1 # Se eliminan los registros NaN del DataFrame
2 #data = data.dropna()
3 #data.shape
```

Nota: Como no existen registros NaN no es necesario utilizar 'dropna()'

```
1 # Se observan los primeros registros
2 data.head()
```

	date(m/d/y)	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
0	25/01/05	12.30	1	0.0	17.02	13.17	747	8.68	0.1
1	25/01/05	12.31	1	0.0	17.12	13.20	747	8.67	0.1
2	25/01/05	12.32	1	0.0	17.12	13.25	747	8.68	0.1
3	25/01/05	12.33	1	0.0	17.11	13.19	747	8.68	0.1
4	25/01/05	12.34	1	0.0	17.03	13.20	747	8.68	0.1

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

```
1 # Se observan las estadísticas de los datos, para valores numéricos
2 data.describe()
```

	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
count	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000
mean	13.869968	1.504792	1.161981	17.323882	13.348083	739.704473	8.652077	0.068371
std	1.857436	0.500377	1.164567	0.511636	0.133306	7.442400	0.091153	0.101617
min	10.050000	1.000000	0.000000	16.550000	12.980000	717.000000	8.240000	-0.100000
25%	12.562500	1.000000	0.000000	16.880000	13.280000	737.000000	8.650000	0.000000
50%	14.120000	2.000000	0.400000	17.105000	13.400000	740.000000	8.670000	0.000000
75%	15.287500	2.000000	1.500000	17.880000	13.450000	746.000000	8.700000	0.100000
max	17.300000	2.000000	3.000000	18.350000	13.530000	753.000000	8.730000	0.500000

```
1 # Visualizo las características de la entrada en un histograma, sin las columnas: time, day, concentration
2 data.drop(['time','day','concentration'], 1).hist()
3 plt.rcParams['figure.figsize'] = (10,9)
4 plt.show()
```

Paso 4: Selección de las mejores características

Antes de generar la predicción se debe:

- Seleccionar las mejores características para el modelo
- Separar los datos de entrenamiento y de pruebas
- Generar la predicción
- Obtener los valores para las métricas

```
1 # Se seleccionan todas las características:
2 features = ['chloride(mg/L)', 'DO(mg/L)', 'ORP(mV)', 'pH', 'turbidity(NTU+)']
3 X = data[features]
4 # Variable Objetivo
5 y = data['concentration']
6
7 # Se dividen los datos para el entrenamiento (80% entrenamiento y 20% pruebas)
8 X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.8, random_state = 1)
9
10 # Se crea el modelo
11 dtr = DecisionTreeRegressor()
12 # Se entrena el modelo
13 dtr.fit(X_train, y_train)
14 # Se genera la predicción
15 prediccion = dtr.predict(X = X_test)
16
17 # Técnica de mejores características con Eli5 (con todas las características)
18 show_weights(dtr, feature_names = features)
```

Weight	Feature
0.3523	DO(mg/L)
0.2687	ORP(mV)
0.1724	chloride(mg/L)
0.1721	pH
0.0345	turbidity(NTU+)

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Paso 5: Predicción con mejores características y SVM

```
1 # Se seleccionan las mejores características por su peso: DO(mg/L), ORP(mV), pH, chloride(mg/L)
2 features = ['DO(mg/L)', 'ORP(mV)', 'pH', 'chloride(mg/L)']
3 X = data[features]
4 # Variable Objetivo
5 y = data['concentration']
6
7 # Se dividen los datos para el entrenamiento (80% entrenamiento y 20% pruebas)
8 X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.8, random_state = 1)
9
10 # Se crea el modelo con SVM
11 svr = SVR()
12 # Se entrena el modelo
13 svr.fit(X_train, y_train)
14 # Se genera la predicción
15 prediccion = svr.predict(X = X_test)
16
17 # Métricas
18 # Se utiliza la métrica: R2
19 print('R2-árbol de regresión:', r2_score(y_test, prediccion))
20 # Se utiliza la métrica: raíz del error cuadrático medio (rmse)
21 rmse = mean_squared_error(y_true = y_test, y_pred = prediccion, squared = False)
22 print('RMSE-árbol de regresión:', rmse)
23 # Se utiliza la métrica: MAE
24 mae = mean_absolute_error(y_test, prediccion)
25 print('MAE-árbol de regresión:', mae)
```

R2-árbol de regresión: -0.3404448365976993
RMSE-árbol de regresión: 1.4202423455224715
MAE-árbol de regresión: 1.080533257702623

Nota: A partir de las mejores características se genera la predicción. De momento el modelo no alcanzó muy buenos resultados para las métricas seleccionadas, pero en el siguiente ejemplo (Random Forest) aplicaremos Hiperparámetros para mejorar los resultados del modelo.

Puede ver el código fuente de este ejemplo en:

https://github.com/jose-llanos/ML_hidroinformatica/blob/main/2.3-SVM.ipynb

Video: Ejemplo SVM en Python

https://www.youtube.com/watch?v=X38yCdQ_cWw

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107.584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

2.4 Bosque Aleatorio

Un Bosque Aleatorio (Random Forest en inglés) es un algoritmo de aprendizaje automático que se utiliza tanto para tareas de clasificación como para tareas de regresión. Es una técnica de conjunto que combina múltiples árboles de decisión individuales para obtener un modelo más robusto y preciso.

El concepto fundamental detrás de un Bosque Aleatorio es la idea de que la sabiduría de la multitud puede superar la de un solo experto. En lugar de depender de un solo árbol de decisión para tomar decisiones, un Bosque Aleatorio construye un conjunto (o "bosque") de árboles de decisión y combina sus resultados para obtener predicciones más confiables.

A continuación, se explican algunos aspectos clave de los Bosques Aleatorios:

1. Muestreo Bootstrap: Para construir cada árbol en el bosque, se utiliza un muestreo aleatorio con reemplazo (bootstrap) del conjunto de datos de entrenamiento. Esto significa que se generan múltiples conjuntos de datos de entrenamiento, cada uno de los cuales es una muestra diferente del conjunto de datos original.

2. Muestreo de Características: Además del muestreo de datos, en cada nodo de decisión de cada árbol, se selecciona un subconjunto aleatorio de características para dividir el nodo. Esto introduce más aleatoriedad y diversidad en el proceso de construcción del árbol.

3. Votación o Promedio: Para realizar predicciones, los Bosques Aleatorios combinan las predicciones individuales de todos los árboles en el bosque. En el caso de la clasificación, se utiliza una votación para determinar la clase final, y en la regresión, se promedian las predicciones para obtener el valor final.

Las ventajas de los Bosques Aleatorios incluyen:

- Reducción del sobreajuste: Al construir múltiples árboles y combinar sus predicciones, los Bosques Aleatorios tienden a ser más resistentes al sobreajuste en comparación con un solo árbol de decisión.
- Alta precisión: Debido a la combinación de múltiples modelos, los Bosques Aleatorios suelen tener una alta precisión en tareas de clasificación y regresión.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

- Lidar con características importantes: Los Bosques Aleatorios pueden medir la importancia de las características en la clasificación o regresión, lo que ayuda a identificar las características más relevantes en los datos.
- Robustez a valores atípicos: Los Bosques Aleatorios son robustos ante valores atípicos y datos ruidosos.

Los Bosques Aleatorios se utilizan en una amplia gama de aplicaciones, como la detección de spam, la clasificación de imágenes, la predicción de precios de acciones y muchas otras tareas de aprendizaje automático. Su capacidad para producir resultados sólidos y su flexibilidad los convierten en una elección popular en la comunidad de aprendizaje automático.

Ejemplo en Python:

Ejercicio Práctico: Random Forest

Paso 1: Importar las librerías

```
1 # Librería para operaciones matemáticas o estadísticas
2 import numpy as np
3 # Librería para el manejo de datos
4 import pandas as pd
5 # Librerías para gráficas
6 import seaborn as sb
7 import matplotlib.pyplot as plt
8 # Librería para dividir los datos de entrenamiento y pruebas
9 from sklearn.model_selection import train_test_split
10 # Librería para Random Forest
11 from sklearn.ensemble import RandomForestRegressor
12 # Librería para las métricas
13 from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
14 # Librerías para Hiperparámetros
15 from sklearn.preprocessing import StandardScaler
16 from sklearn.model_selection import RepeatedKFold
17 #from sklearn.model_selection import GridSearchCV
18 from sklearn.model_selection import RandomizedSearchCV
```

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Paso 2: Cargar los datos a un DataFrame

```
1 # Se cargan los datos de entrada
2 data = pd.read_csv("data/2.3-glifosato.csv", sep=";")
3 # Se observa la cantidad de registros y columnas del archivo
4 data.shape
```

(626, 9)

Paso 3: Preprocesamiento de los datos

```
1 # Se buscan registros NaN para eliminarlos
2 print('Columna          Cantidad NaN')
3 print(data.isnull().sum(axis = 0))
4 print(data.shape)
```

Columna	Cantidad NaN
date(m/d/y)	0
time	0
day	0
concentration	0
chloride(mg/L)	0
DO(mg/L)	0
ORP(mV)	0
pH	0
turbidity(NTU+)	0
dtype: int64	
(626, 9)	

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

```
1 # Se eliminan los registros NaN del DataFrame
2 #data = data.dropna()
3 #data.shape
```

Nota: Como no existen registros NaN no es necesario utilizar 'dropna()'

```
1 # Se observan los primeros registros
2 data.head()
```

	date(m/d/y)	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
0	25/01/05	12.30	1	0.0	17.02	13.17	747	8.68	0.1
1	25/01/05	12.31	1	0.0	17.12	13.20	747	8.67	0.1
2	25/01/05	12.32	1	0.0	17.12	13.25	747	8.68	0.1
3	25/01/05	12.33	1	0.0	17.11	13.19	747	8.68	0.1
4	25/01/05	12.34	1	0.0	17.03	13.20	747	8.68	0.1

```
1 # Se observan las estadísticas de los datos, para valores numéricos
2 data.describe()
```

	time	day	concentration	chloride(mg/L)	DO(mg/L)	ORP(mV)	pH	turbidity(NTU+)
count	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000	626.000000
mean	13.869968	1.504792	1.161981	17.323882	13.348083	739.704473	8.652077	0.068371
std	1.857436	0.500377	1.164567	0.511636	0.133306	7.442400	0.091153	0.101617
min	10.050000	1.000000	0.000000	16.550000	12.980000	717.000000	8.240000	-0.100000
25%	12.562500	1.000000	0.000000	16.880000	13.280000	737.000000	8.650000	0.000000
50%	14.120000	2.000000	0.400000	17.105000	13.400000	740.000000	8.670000	0.000000
75%	15.287500	2.000000	1.500000	17.880000	13.450000	746.000000	8.700000	0.100000
max	17.300000	2.000000	3.000000	18.350000	13.530000	753.000000	8.730000	0.500000

```
1 # Visualizo las características de la entrada en un histograma, sin las columnas: time, day, concentration
2 data.drop(['time','day','concentration'], 1).hist()
3 plt.rcParams['figure.figsize'] = (10,9)
4 plt.show()
```

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Paso 4: Predicción con mejores características y Random Forest

```
1 # Se seleccionan las mejores características por su peso: DO(mg/L), ORP(mV), pH, chloride(mg/L)
2 features = ['DO(mg/L)', 'ORP(mV)', 'pH', 'chloride(mg/L)']
3 X = data[features]
4 # Variable Objetivo
5 y = data['concentration']
6
7 # Se dividen los datos para el entrenamiento (70% entrenamiento y 30% pruebas)
8 X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.7, random_state = 1)
9
10 # Se crea el modelo con Random Forest
11 rfr = RandomForestRegressor()
12 # Se entrena el modelo
13 rfr.fit(X_train, y_train)
14 # Se genera la predicción
15 prediccion = rfr.predict(X = X_test)
16
17 # Métricas
18 # Se utiliza la métrica: R2
19 print('R2-árbol de regresión:', r2_score(y_test, prediccion))
20 # Se utiliza la métrica: raíz del error cuadrático medio (rmse)
21 rmse = mean_squared_error(y_true = y_test, y_pred = prediccion, squared = False)
22 print('RMSE-árbol de regresión:', rmse)
23 # Se utiliza la métrica: MAE
24 mae = mean_absolute_error(y_test, prediccion)
25 print('MAE-árbol de regresión:', mae)
```

R2-árbol de regresión: 0.7783651082956048

RMSE-árbol de regresión: 0.563871703989697

MAE-árbol de regresión: 0.70705007508784706

Paso 5: Hiperparámetros para Random Forest

```
1 # RandomForestRegressor: Ajuste Hiperparámetros (Random Search)
2 scaler = StandardScaler()
3 X_train = scaler.fit_transform(X_train)
4 X_test = scaler.transform(X_test)
5
6 rfr = RandomForestRegressor()
7
8 # Número de árboles en un bosque aleatorio
9 n_estimators = [int(x) for x in np.linspace(start=20, stop=200, num = 10)]
10 # Número de características a considerar en cada división
11 max_features = [1.0]
12 # Número máximo de niveles en el árbol
13 max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
14 max_depth.append(None)
15 # Número mínimo de muestras requeridas para dividir un nodo
16 min_samples_split = [2, 5, 10, 15, 20]
17 # Número mínimo de muestras requeridas en cada nodo hoja
18 min_samples_leaf = [1, 2, 4, 6]
19 # Método de selección de muestras para entrenar cada árbol
20 bootstrap = [True, False]
21 grid = dict(n_estimators=n_estimators,
22            max_features=max_features,
23            max_depth=max_depth,
24            min_samples_split=min_samples_split,
25            min_samples_leaf=min_samples_leaf,
26            bootstrap=bootstrap)
27 cvFold = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
28 # Hiperparámetros con Random Search
29 randomSearch = RandomizedSearchCV(estimator= rfr,
30                                   grid=grid,
31                                   cv=cvFold,
32                                   n_jobs=-1,
```

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

```
Mejores Parametros (RandomSearch): RandomForestRegressor(max_depth=80, max_features=1.0, min_samples_split=5,  
n_estimators=60)
```

```
R2 = 0.79  
RMSE = 0.55  
MAE = 0.30
```

Nota: Al aplicar Hiperparámetros el modelo mejoró en 2% en la métrica R2 al pasar del 0.77 a 0.79.

Puede ver el código fuente de este ejemplo en:

https://github.com/jose-llanos/ML_hidroinformatica/blob/main/2.4-RandomForest.ipynb

Video: Ejemplo Bosque Aleatorios en Python

<https://www.youtube.com/watch?v=AFpJVWAJiTA>

Bibliografía

- Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O'Reilly Media.
- Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3, 1157-1182.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. Springer.
- Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models. CRC Press.
- Müller, A. C., & Frank, E. (2016). Getting started with WEKA. University of Waikato.
- Müller, A. C., & Guido, S. (2017). Introduction to machine learning with Python: A guide for data scientists. O'Reilly Media.
- Raschka, S., & Mirjalili, V. (2019). Python machine learning (3rd ed.). Packt Publishing.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). Data mining: Practical machine learning tools and techniques (3rd ed.). Morgan Kaufmann.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"