



ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE FORMACIÓN DE TECNÓLOGOS



FUNDAMENTOS INTELIGENCIA ARTIFICIAL

ASIGNATURA:

Fundamentos de Inteligencia Artificial

PROFESOR:

Ing. Vanessa Guevara

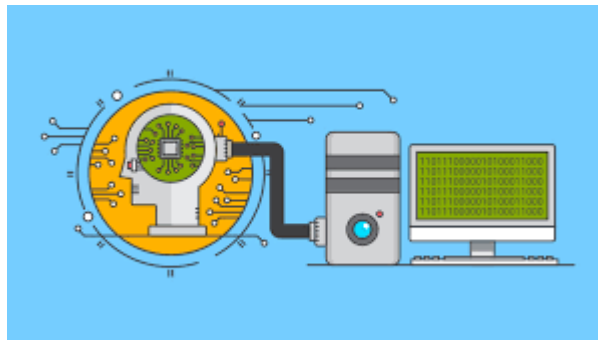
PERÍODO ACADÉMICO:

2024-A

TRABAJO EN CLASE

TÍTULO:

Aprendizaje Automático



Vela David

1. OBJETIVOS

- Proporcionar una introducción práctica a los fundamentos del aprendizaje automático utilizando Python.

2. INDICACIONES

- Incluir captura de pantalla solo de lo que respalde la respuesta a las preguntas y de las nuevas predicciones.

3. EJERCICIOS

➤ Ejercicio 1: Regresión Lineal

- **Objetivo Específico:** Predecir los precios de las viviendas utilizando características como la superficie, el número de habitaciones y si tiene garaje o no.
- **Dataset:** viviendas.csv
- **Preguntas:**

¿Qué nos indican el Error cuadrático medio (MSE) y el coeficiente de determinación sobre el rendimiento del modelo?

```
4 60000.00 40 1 0
Error cuadrático medio es: 2028620708.9849699
el coeficiente de determinación es: 0.29202641088678527
La Predicción del precio: [111997.88797805]
```

Sabiendo que el error cuadrático nos dice la diferencia existente entre la predicción y el valor real. Y viendo en nuestro ejemplo que este valor es muy elevado, se concluye que el modelo no es preciso.

El coeficiente de determinación señala qué tan bien se ajusta el modelo a los datos, varía entre 0 y 1, donde si se acerca a 1 es un mejor modelo y viceversa. En el ejemplo está muy cerca a cero, concluimos que es un mal modelo.

- **Predicciones:**

Predecir el valor de una vivienda con 2 habitaciones, superficie de 52 m² con garaje.

```
el coeficiente de determinación es: 0.29202641088678527
La Predicción del precio: [111997.88797805]
/usr/local/lib/python3.10/dist-packages/sklearn
```

Predecir el valor de una vivienda con 1 habitaciones, superficie de 60 m² sin garaje.

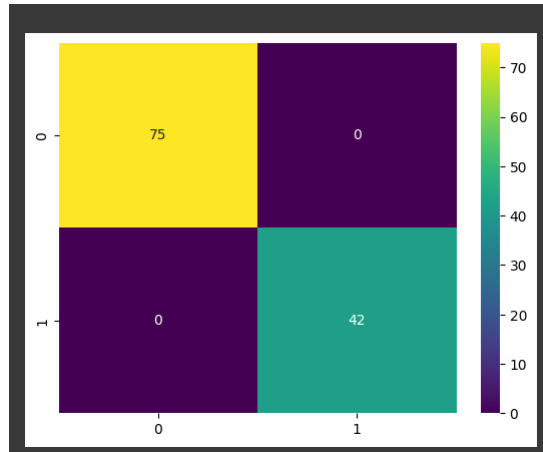
```
el coeficiente de determinación es: 0.29202641088678527
La Predicción del precio: [112195.63193565]
/usr/local/lib/python3.10/dist-packages/sklearn
```

Predecir el valor de una vivienda con 3 habitaciones, superficie de 80 m² sin garaje.

```
el coeficiente de determinación es: 0.29202641088678527
La Predicción del precio: [118548.42674169]
/usr/local/lib/python3.10/dist-packages/sklearn
```

➤ Ejercicio 2: Clasificación

- **Objetivo Específico:** Predecir si un pasajero sobrevivió o no en el Titanic utilizando características como la edad, el sexo y la clase del boleto.
- **Dataset:** titanic.csv
- **Preguntas:**
¿Qué nos dice la matriz de confusión sobre el rendimiento del modelo?



El rendimiento del modelo es bueno, viendo los supuestos, verdaderos y falsos.

El gráfico nos dice que ha acertado en todas las predicciones de sobrevivientes y no sobrevivientes.

¿Qué otras métricas puedes utilizar para evaluar un modelo de clasificación y por qué son importantes?

La exactitud que varía entre 0 y 1, si se acerca a 1 es un buen modelo, y viceversa.

- **Predicciones:**

Predecir la supervivencia de un pasajero de 30 años, femenino, en clase 1.

```
Supervivencia del pasajero es: [1]
*****
```

Predecir la supervivencia de un pasajero de 10 años, femenino, en clase 2.

```
*****
Supervivencia del pasajero es: [1]
*****
```

Predecir la supervivencia de un pasajero de 65 años, masculino, en clase 3.

```
Supervivencia del pasajero es: [0]
*****
```

➤ Ejercicio 3: Clustering

- **Objetivo Específico:** Agrupar los clientes de un centro comercial en segmentos utilizando características como el género, la edad, el ingreso anual y el puntaje de gasto.
- **Dataset:** mall_customers.csv
- **Preguntas:**

¿Por qué es importante estandarizar las características antes de aplicar K-Means?

Esto se debe a que el algoritmo de K-Means, para agrupar, usa la distancia Eucladiana para determinar la distancia hacia el cluster, normalizando las características nos aseguramos de tener las características a la misma escala.

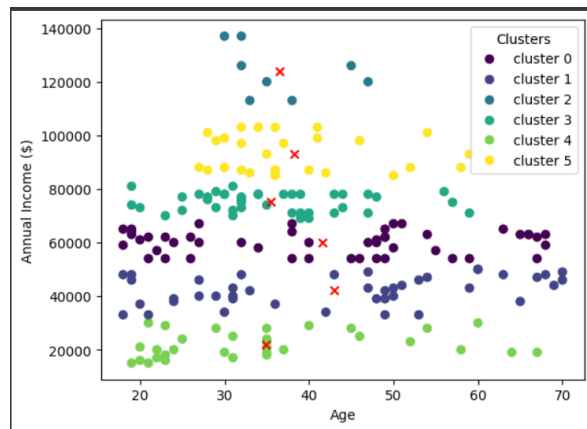
¿Qué criterio utilizaste para determinar el número óptimo de clusters y por qué?

Usamos dos métodos, el método del código que en el eje x tiene los posibles números de cluster y en el eje y la inercia, la inercia calcula distancia entre nodos, a menor inercia, mejor modelo. Debemos observar el punto de inflexión que es donde se minimiza la varianza en ese punto para los clusters seleccionados.

Además usamos el método de la silueta, que oscila entre -1 y 1, donde una puntuación cercana a 1 indica que los puntos de datos se ajustan bien a su propio clúster y no a los clústeres vecinos.

Con ambos métodos llegamos a la conclusión que el número de clusters óptimos es de 5.

¿Qué patrones puedes observar en los clusters obtenidos?



Podemos observar que los clusters se ubican teniendo en cuenta los ingresos anuales de diferentes edades.

- **Predicciones:**

Asignar un nuevo cliente femenino de 40 años, con un ingreso anual de 70000 y un puntaje de gasto de 50.

Un cliente con género Femenino, edad 40, ingreso anual 70000 y puntaje de gasto 50 pertenece al cluster 2.

Asignar un nuevo cliente masculino de 80 años, con un ingreso anual de 80000 y un puntaje de gasto de 70.

```
Un cliente con género Masculino, edad 80, ingreso anual 80000 y puntaje de gasto 70 pertenece al cluster 2.
```

Asignar un nuevo cliente femenino de 20 años, con un ingreso anual de 60000 y un puntaje de gasto de 10.

```
Un cliente con género Femenino, edad 20, ingreso anual 60000 y puntaje de gasto 10 pertenece al cluster 2.
```

Link del Google Colab: <https://colab.research.google.com/drive/1ZnU4eoEZiD1loG-71Ygr05D-qazwGHkt?usp=sharing>