



ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE FORMACIÓN DE TECNÓLOGOS



FUNDAMENTOS INTELIGENCIA ARTIFICIAL

ASIGNATURA:

Fundamentos de Inteligencia Artificial

PROFESOR:

Ing. Vanessa Guevara

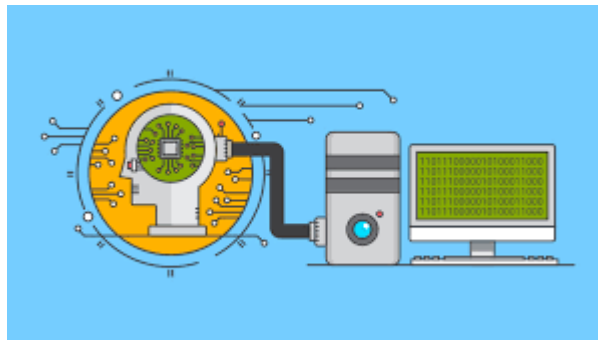
PERÍODO ACADÉMICO:

2024-A

TAREA

TÍTULO:

Aprendizaje Supervisado. Regresión Lineal



Vela David

1. PROPÓSITO DE LA PRÁCTICA

El propósito de esta tarea es construir un modelo de regresión que pueda predecir **el costo de un seguro de salud** para un individuo en función de sus características personales.

Para ello, se proporciona un conjunto de datos que contiene información sobre diferentes individuos.

El conjunto de datos está en formato CSV y contiene las siguientes columnas:

- age: edad del beneficiario principal
- sex: genero del beneficiario principal (0: masculino, 1: femenino)
- bmi: índice de masa corporal (kg/m)
- children: número de niños cubiertos por el seguro de salud
- smoker: fumador (0: No, 1: Sí)
- charges: costos facturados por el seguro de salud

2. OBJETIVOS

- Entender el proceso para crear un modelo de regresión y cómo evaluar su rendimiento utilizando métricas adecuadas.
- Desarrollar habilidades en la realización de predicciones utilizando el modelo entrenado y evaluar su precisión comparándolas con los valores reales.

3. INSTRUCCIONES

- Importar el conjunto de datos y realizar una exploración inicial para familiarizarte con la estructura y los tipos de datos.
- Dividir el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba. **(70% entrenamiento, 30% prueba)**
- Construir un modelo de regresión utilizando el algoritmo de regresión lineal.
- Entrenar el modelo utilizando el conjunto de entrenamiento y evaluar su rendimiento utilizando el conjunto de prueba.
- Evaluar las predicciones del modelo con las métricas de evaluación correspondiente
- Realizar predicciones del costo del seguro para un conjunto de datos de prueba adicional.

4. DESARROLLO DE LA PRÁCTICA

Aquí detallar y explicar todas las etapas realizadas para construir el modelo. Incluir captura de pantallas.

1. Importar datos

```
import pandas as pd
datos=pd.read_csv("seguro_salud.csv")
print(datos.head())
print(datos.describe())
```

2. Dividir el conjunto de datos (70% entrenamiento, 30% prueba)

```
#dividir los datos de conjuntos de entrenamientos y prueba
x_train, x_test, y_train,y_test = train_test_split(x,y,test_size = 0.3,random_state=70)
from sklearn.linear_model import LinearRegression
```

3. Construir un modelo

```
# crear modelo de regresion lineal
modelo = LinearRegression()
```

4. Entrenar el modelo

```
#entrenar el modelo
modelo.fit(x_train,y_train)

#Evaluar el modelo con los datos de prueba
y_obt=modelo.predict(x_test)
from sklearn.metrics import mean_squared_error,r2_score
```

5. Métricas

```

#metricas de evaluacion
error=mean_squared_error(y_test,y_obt)
print("Error cuadrático medio es:",error)
r2 =r2_score(y_test,y_obt)
print("el coeficiente de determinacion es:",r2)

#realizar predicción
y_obt=modelo.predict([[55,1,35,2,1]])
print("La Predicción de los cargos del seguro es: ",y_obt)

#coeficientes e intercepcion
print("coeficientes :",modelo.coef_)
print("intercepcion :",modelo.intercept_)

```

5. RESULTADOS

- Con el modelo desarrollado, predecir el costo de seguro de salud para los siguientes casos:

- Mujer, 30 años, sin hijos, no fumador, con IBM de 30 kg/m

La Predicción de los cargos del seguro es: [5013.68430854]

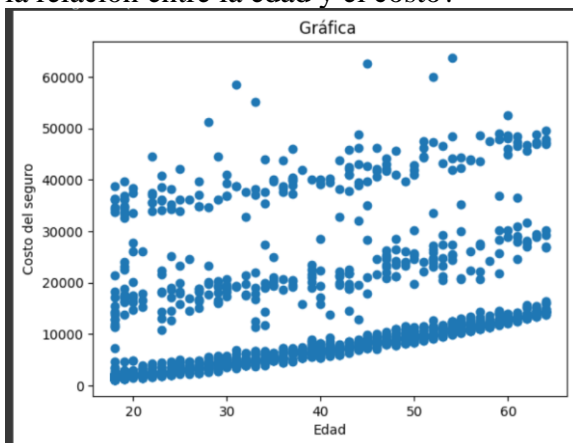
- Hombre, 42 años, 3 hijos, fumador, con IBM de 45 kg/m

La Predicción de los cargos del seguro es: [38906.38842388]

- Mujer, 55 años, 2 hijos, fumador, con IBM de 35 kg/m

La Predicción de los cargos del seguro es: [38079.09198285]

- Elaborar un diagrama de dispersión con la edad (age) como variable independiente y el costo del seguro como variable dependiente. ¿Qué indica el diagrama de dispersión sobre la relación entre la edad y el costo?



- Explique si el modelo creado proporcionará una buena predicción del costo del seguro.

El coeficiente de determinacion es: 0.7262067194293922 por lo que es relativamente bueno, debido a que mientras más próximo a cero es, es mejor la predicción

6. CONCLUSIONES Y RECOMENDACIONES

Conclusiones:

- **Eficacia del Modelo de Regresión:** El modelo de regresión lineal desarrollado demostró una capacidad relativamente buena para predecir el costo de un seguro de salud basado en las características personales proporcionadas. Esto se evidencia en el coeficiente de determinación (R^2) de 0.7262067194293922, el cual indica que aproximadamente el 72.62% de la variabilidad en los costos del seguro puede ser explicada por el modelo.
- **Importancia de las Características Personales:** Las características personales como la edad, el género, el índice de masa corporal (IMC), el número de hijos y si se es fumador o no, tienen un impacto significativo en los costos del seguro de salud. Estos factores deben ser considerados cuidadosamente al calcular las primas de seguro.
- **Potencial de Mejora:** Aunque el modelo ofrece una buena predicción en general, aún existen áreas para mejorar, como la inclusión de más variables predictoras relevantes o la exploración de otros algoritmos de regresión que puedan capturar mejor la complejidad de los datos.

Recomendaciones:

- **Refinamiento del Modelo:** Se recomienda explorar técnicas avanzadas de modelado, como la regularización o el uso de modelos no lineales, para mejorar la precisión y la generalización del modelo. Además, la inclusión de interacciones entre las variables predictoras puede ayudar a capturar relaciones más complejas.
- **Actualización y Monitoreo Continuo:** Dado que el costo de los seguros de salud puede verse afectado por cambios en la legislación, en los costos médicos o en los comportamientos de los asegurados, se recomienda monitorear y actualizar regularmente el modelo para garantizar su relevancia y precisión a lo largo del tiempo.

7. BIBLIOGRAFÍA

PRESENTACIÓN

Una vez culminada la tarea, subir con el nombre del archivo:
Tarea_IA_2024A_NApellido.