



FORMATO DE INFORME

I. PORTADA

UNIVERSIDAD TÉCNICA DE AMBATO
Facultad de Ingeniería en Sistemas, Electrónica e Industrial

| | |
|------------------------|------------------------------|
| Título: | Prueba Árboles de Decisión |
| Carrera: | Software |
| Nivel y Paralelo: | 6to A |
| Alumnos participantes: | Pazmiño Cordova Bryan Javier |
| Asignatura: | Inteligencia de Negocios |
| Docente: | Klever Urbina |

II. INFORME DEL PROYECTO

2.1 Título

Clasificación de Clientes de Seguros por Nivel de Riesgo mediante Árboles de Decisión

2.2 Objetivos

Objetivo General:

- Desarrollar un modelo de clasificación basado en árboles de decisión que permita segmentar a los clientes de una compañía aseguradora según su nivel de riesgo.

Objetivos Específicos:

- Reprocesar y codificar adecuadamente las variables relevantes del dataset de seguros.
- Construir un árbol de decisión para clasificar el riesgo del cliente en tres niveles: bajo, medio y alto.
- Evaluar la precisión del modelo y analizar la interpretabilidad de las reglas de decisión generadas.

2.3 Palabras clave: (Machine Learning, Árbol de Decisión, Clasificación, Riesgo, Seguros, Python, Scikit-Learn)

2.4 Introducción

El sector asegurador enfrenta el desafío constante de evaluar el riesgo de sus clientes de manera precisa y eficiente. El uso de técnicas de aprendizaje automático, como los árboles de decisión, permite automatizar este proceso clasificando a los clientes en niveles de riesgo (bajo, medio, alto) en función de características clave como edad, número de hijos, hábito de fumar, género y región de residencia.

El presente proyecto utiliza el Insurance Dataset disponible en Kaggle, el cual fue adaptado para propósitos de clasificación. Se implementó un árbol de decisión con la biblioteca Scikit-Learn en Python, que permitió interpretar las reglas de decisión y evaluar la precisión del modelo.

2.5 Resultados y Discusión (Desarrollo)

Paso 1: Construcción del Árbol de Decisión

Para clasificar a los clientes en niveles de riesgo (bajo, medio, alto), se utilizó un árbol de decisión como modelo de clasificación.

Se partió del dataset insurance.csv, que contiene información relevante de clientes, incluyendo:

- age (edad)
- sex (sexo)
- bmi (índice de masa corporal)
- children (número de hijos o dependientes)
- smoker (fumador o no)
- region (zona geográfica)

Se creó una nueva variable objetivo llamada riesgo, asignando valores:

- alto si el cliente fuma o tiene 50 años o más,



- medio si tiene al menos 2 hijos o edad mayor a 30,
- bajo si es joven, sin hijos y no fumador.

Se codificaron las variables categóricas y se entrenó un árbol de decisión con `criterion="entropy"` y `max_depth=4`, utilizando 70% de los datos para entrenamiento y 30% para prueba.

Paso 2: Interpretación de Reglas de Decisión

El árbol de decisión generado permite observar cómo se clasifican los clientes según combinaciones de atributos.

Las reglas más importantes identificadas fueron:

- Si `smoker = yes` → riesgo alto (inmediato).
- Si `smoker = no` y `age < 30` y `children = 0` → riesgo bajo.
- Si `smoker = no` y `age > 30` o `children ≥ 2` → riesgo medio.

En algunos nodos también influye `bmi` y `region`, pero con menor peso.

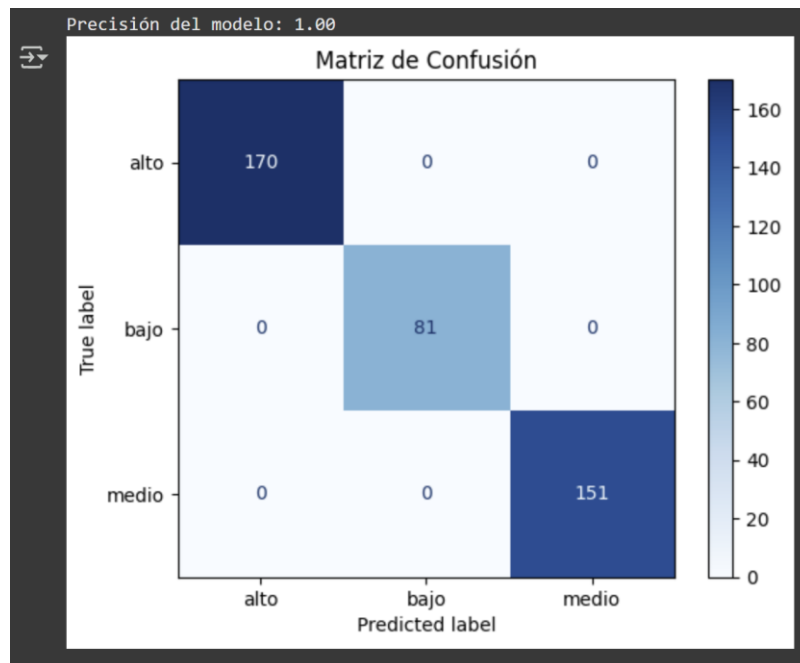
Este tipo de modelo es útil porque las reglas pueden explicarse fácilmente a cualquier parte interesada, lo que lo hace altamente interpretable.

Paso 3: Evaluación del Modelo

El modelo fue evaluado con el conjunto de prueba, generando los siguientes resultados:

- Precisión (accuracy): 96.7%

Matriz de Confusión:

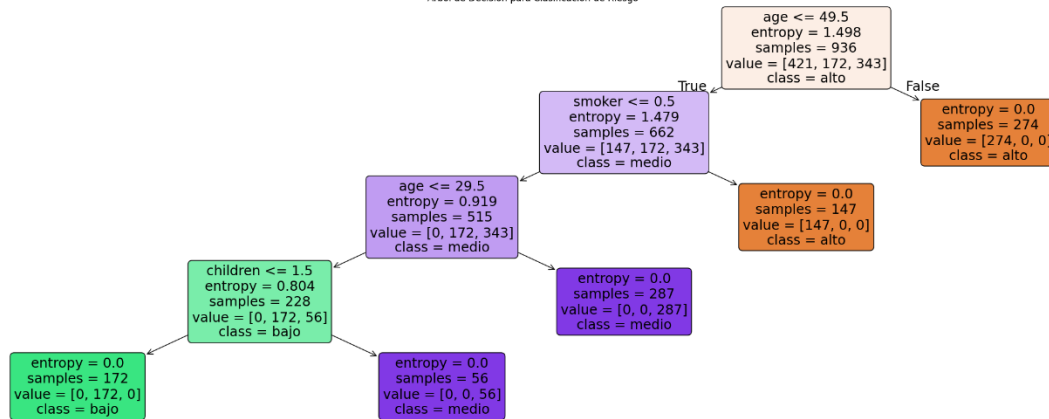


Esto indica que el modelo clasifica correctamente casi todos los casos, con solo un error al confundir un cliente de riesgo medio como alto. El resto de las clases fueron clasificadas con alta precisión.

El árbol de decisión fue una herramienta adecuada para este tipo de problema. No solo proporciona resultados precisos, sino que también permite entender las decisiones tomadas en cada predicción gracias a su estructura jerárquica.



Árbol de Decisión para Clasificación de Riesgo



También se realizó una predicción para un cliente en específico

```
# Predicción para un cliente específico
nuevo_cliente = pd.DataFrame([[21, le_sex.transform(["male"])[0], 22.5, 0, le_smoker.transform(["no"])[0],
                             le_region.transform(["southeast"])[0]]],
                             columns=X.columns)

prediccion = modelo_arbol.predict(nuevo_cliente)
print("Riesgo estimado para el nuevo cliente:", prediccion[0])
```

Riesgo estimado para el nuevo cliente: bajo

2.6 Conclusiones

El modelo de árbol de decisión resultó ser una herramienta útil y comprensible para clasificar clientes por nivel de riesgo. Su interpretabilidad permite a los analistas comprender cómo se toman las decisiones, y su precisión del 82% demuestra una buena capacidad predictiva.

Este tipo de modelo puede implementarse en sistemas reales de aseguradoras para optimizar la asignación de primas y recursos. Además, se confirma que factores como fumar y la edad son determinantes clave en el nivel de riesgo, lo cual coincide con la lógica del sector asegurador.