INSTITUTO POLITÉCNCO NACIONAL
ESCUELA SUPERIOR DE CÓMPUTO
MATERIAL EDUCATIVO PARA LA UNIDAD DE APRENDIZAJE DE MINERÍA DE DATOS. 2023–1

Práctica para los alumnos del grupo: 3CV11

Dra. Fabiola Ocampo Botello

No. Equipo: _____

Nombre de los integrantes del equipo:

1) _____

2) _____

3) _____

1. Descripción del conjunto de datos.

Ejercicio adaptado con fines educativos de:
Portal IBM. SPSS Statistics 23.0.0. Casos de estudio. Disponible en:
https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/tutorials/trees_scoring_intro1.html

Contiene dos archivos: uno de entrenamiento (carros.csv) y el de prueba (carros_prueba.csv).

2. Objetivo de la práctica.

Considerando un conjunto de datos que contiene información demográfica y el precio de compra del vehículo. Construir tres modelos que se puede usar para:

2a) Predecir cuánto es probable que las personas con características demográficas similares gasten en un automóvil nuevo. El modelo creado podrá ser aplicado a otros archivos de datos donde la información demográfica está disponible, pero no la información sobre compras anteriores de vehículos.

2b) Crear un modelo de clasificación (árbol) considerando sólo datos nominales.

2c) Crear un modelo de clasificación (árbol) considerando datos nominales y numéricos.

Crear una portada general del proyecto que incluya: datos del curso, participantes, propósito, fecha de entrega (27 de noviembre de 2022) y lo que considere conveniente.

Para cada uno de estos tres modelos (2a, 2b y 2c) debe establecer lo siguiente:

A) Crear una portada que separe cada uno de los tres modelos de generalización de la clasificación.
B) Objetivo
C) Tipo de árbol a aplicar.
D) Aplicar el método de validación cruzada.

E) Porción del diccionario de datos que incorpora, identificando el atributo objetivo
F) Las reglas generadas, según el modelo.
G) Resultados. Descripción de medidas obtenidas, según el archivo anexo
Incluir las siguientes medidas y la explicación del significado de cada una de ellas:
- Matriz de confusión
- Sensibilidad
- Precisión
- Tasa de error
- Especificidad
- Explicación de los positivos verdaderos y positivos falsos
- Exactitud

H) Conclusiones
I) Agregar pantallas de configuración del flujo de trabajo

Diccionario de datos:

Nombre	Descripción	Tipo	Dominio
coche	Precio del vehículo principal	Numérico	
edad	Edad en años	Numérico	
sexo	sexo	Cadena	f = femenino m = masculino
cating	Categoría de ingresos en miles	Ordinal	1.00 = "Menos de $25" 2.00 = "$25 - $49" 3.00 = "$50 - $74" 4.00 = "$75+"
educ	de estudios	Ordinal	1 = "No completó el bachillerato" 2 = "Bachillerato" 3 = "Estudios universitarios" 4 = "Licenciado" 5 = "Estudios de post-grado"
e_civil	Estado civil	Nominal	0 = "Sin casar" 1 = "Casado"

Archivos a entregar:
- Entregará un archivo empaquetado (.zip o .rar) que contenga los siguientes aspectos:
  - El nombre del archivo que colocará en la asignación debe ser de la forma: eq*Numero*_pClasifIngreso.zip
    Por ejemplo: eq1_ pClasifIngreso.zip
  - Todos los archivos de datos y los flujos de trabajo creados en Knime
  - Un reporte en formato pdf, con el formato: eq*Numero*_pCIngreso.pdf