Juan Pablo Padilla Martin
Examen Mineria de datos

① Categóricas

Nominal: simbolos o nombre de las cosas

Ordinal: Orden o posición en el conjunto de datos, evalua objetivamente cualidades del objeto

Binario: representa solo dos categorias, simetricas y asimetricas (falso-verdadero, macho-hembra etc

Numéricos

Intervalo: medida en escala de igual tamaño, tienen orden, pueden ser positivos o negativos, no poseen cero verdadero

Razón: Posee cero verdadero, es un valor multiplo de otros, los valores estan ordenados y se pueden calcular diferencias entre ellos.

②

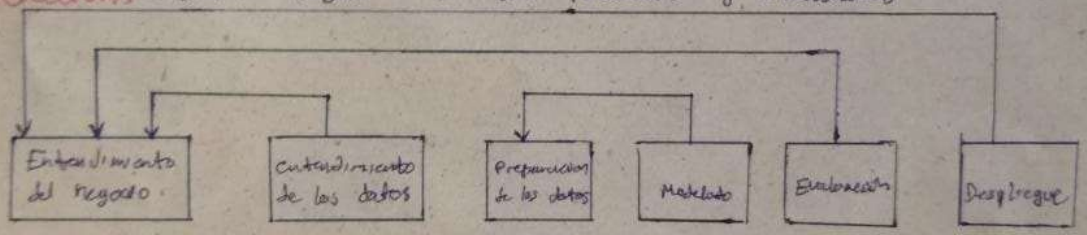
Clasificación: clasifica y etiqueta datos en clases predefinidas

Regresión: Partiendo de los datos genera una función predictiva

Agrupamiento: creación de grupos y conjuntos

Asociación: encuentra conjuntos de datos frecuentes y relacionados

③

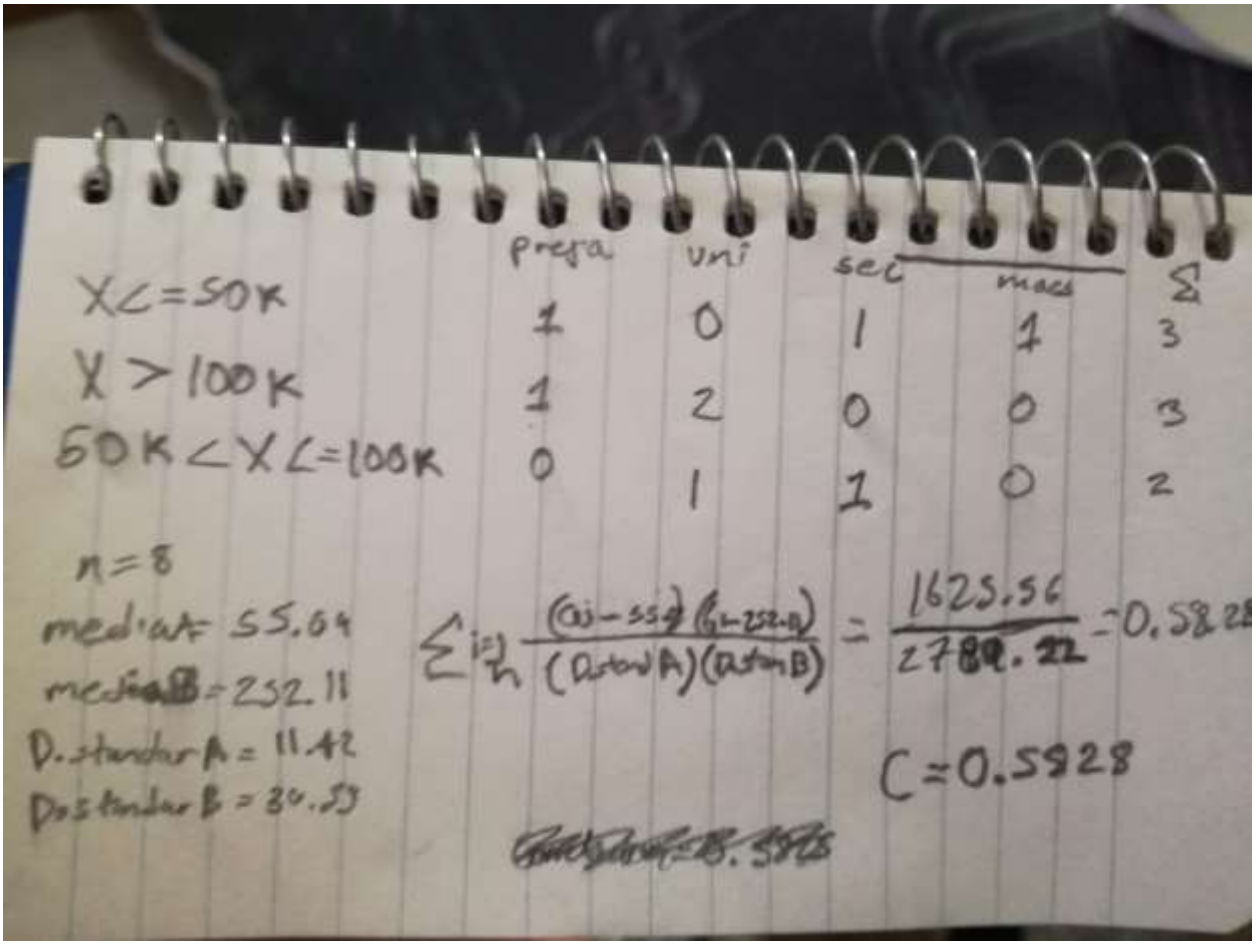


CRISP-MD

⑤

1- Si hay correlación

2- Escolaridad e ingreso tienen poca correlación
Es pertinente eliminar los de 37 años ya que es el mas lejano al conjunto de datos



	preja	uni	sec	macs	Σ
$X \leq 50K$	1	0	1	1	3
$X > 100K$	1	2	0	0	3
$50K < X \leq 100K$	0	1	1	0	2

$n = 8$

media A = 55.64

media B = 252.11

D. standar A = 11.42

D. standar B = 30.53

$$\sum_{i=1}^{n} \frac{(x_i - \bar{x}_A)(x_i - \bar{x}_B)}{(\text{D.stand A})(\text{D.stand B})} = \frac{1625.56}{2789.22} = 0.5828$$

$C = 0.5928$