

Universidad Autónoma Gabriel Rene Moreno

Facultad de Ingenieria en Ciencias de la Computación y Telecomunicaciones

**Informe de Proyecto:**

**Análisis de Credito.arff**

**Materia:**

**INF-432, Soporte para la toma de decisiones**

**Integrantes:**

**Escalante Ustariz Eddy ------------ 212057774**

**Grimaldo Bravo Paul ---------------214040801**

**Taboada Frias Nanet----------------214056228**

**Docente:** **Ing. Peinado Jesús Miguel**

*Minería de datos*

***Materia****:Soporte para la toma de decisiones*

***01-2017***

***-***

Análisis de credit-g.arff

En este ejemplo vamos a utilizar el archivo “credit-g.arff”.

Este conjunto de datos contiene 1.000 ejemplos que representan clientes de una entidad bancaria que demandaron un crédito.

Existen siete atributos numéricos y 13 nominales.

Los atributos de cada objeto indican información sobre el cliente en cuestión: estado civil, edad, parado, etc., así como información del crédito: propósito del crédito, cantidad solicitada, etc.

La clase es binaria e indica si el cliente puede ser considerado como fiable como para concederle el crédito o no

# ¿Cuál es el objetivo a analizar?

Conceder un crédito o no a un determinado cliente a partir de las características propias del cliente.

# ¿Cuál es la técnica que se podría utilizar?

Técnica: **Clasificación**

|  |  |  |
| --- | --- | --- |
|  | Real | |
| Predicho | 0.0 | 1.0 |
| 5.0 | 0.0 |

Matriz de costes:

Esta tabla indica que es 5 veces más costoso si se otorga un crédito a una persona que no lo devuelve, que la situación contraria.

**=== Run information ===**

Scheme: weka.classifiers.rules.ZeroR

Relation: german\_credit

Instances: 1000

Attributes: 21

checking\_status

duration

credit\_history

purpose

credit\_amount

savings\_status

employment

installment\_commitment

personal\_status

other\_parties

residence\_since

property\_magnitude

age

other\_payment\_plans

housing

existing\_credits

job

num\_dependents

own\_telephone

foreign\_worker

class

Test mode: 10-fold cross-validation

Evaluation cost matrix:

0 1

5 0

# ¿Qué resultados se obtuvieron?

Se obtiene una precisión del 70 %, y un coste de 1500 unidades. Era de esperar pues nos equivocamos 300 veces, pero en el peor de los casos.

**=== Classifier model (full training set) ===**

ZeroR predicts class value: good

Time taken to build model: 0 seconds

**=== Stratified cross-validation ===**

=== Summary ===

Correctly Classified Instances 700 70 %

Incorrectly Classified Instances 300 30 %

Kappa statistic 0

Total Cost 1500

Average Cost 1.5

Mean absolute error 0.4202

Root mean squared error 0.4583

Relative absolute error 100 %

Root relative squared error 100 %

Total Number of Instances 1000

**=== Detailed Accuracy By Class ===**

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

1,000 1,000 0,700 1,000 0,824 0,000 0,500 0,700 good

0,000 0,000 0,000 0,000 0,000 0,000 0,500 0,300 bad

Weighted Avg. 0,700 0,700 0,490 0,700 0,576 0,000 0,500 0,580

**=== Confusion Matrix ===**

a b <-- classified as

700 0 | a = good

300 0 | b = bad