**PRACTICA CALIFICADA 2**

**Mendoza Medrano Adrián Esteban 20131281H Morales Oyola Diego Antonio 20124100A**

**Romero Velazque Manuel Rossell 20111389H Saavedra Alvarez Sebastian 20134560E**

04 de octubre de 2017

Universidad Nacional de Ingeniería

Facultad de Ingeniería Industrial y de Sistemas

Inteligencia Artificial

Ciclo 2017-II

**Resumen**

El presente trabajo del curso de Inteligencia Artificial aborda el tema del “credit scoring”, el cual es un método que se está empleando cada vez más entre las instituciones financieras para medir el riesgo de otorgarle financiamiento a una persona, ya sea natural o jurídica. Como paso inicial se representará el conocimiento declarativo, explicar todas las variables que se requieren para representar cualquier problema algorítmico buscando que esta debe de ser genérica para que pueda ser implementada en cualquier lenguaje de programación. Luego de haber logrado representar el conocimiento declarativo se pasa a representar el conocimiento procedural (mediante bloques con entradas y salidas).

**Palabras claves:** *credit scoring, modelo, probabilidad, ranking, discretización, entropía, modelo de Naive Bayes.*

**INTRODUCCIÓN**

Durante muchos años los otorgantes de crédito han estado utilizando los sistemas de puntaje o puntuación de crédito (credit scoring) para determinar si una persona o empresa es un candidato que posee un nivel de riesgo de crédito aceptable para otorgarle tarjetas de crédito, préstamos para la compra de automóviles e hipotecas.

En la actualidad, hay otros tipos de negocios —particularmente las compañías de seguros de carros y propietarios de vivienda y las compañías telefónicas— que están utilizando puntajes de crédito para decidir si le emitirán una póliza o le otorgarán un servicio, y en qué términos. Si el solicitante de crédito posee un puntaje de crédito alto se considerará que representa un menor riesgo para la compañía, lo que a su vez significa que es más factible que le otorguen el crédito o el seguro —o que pague menos.

**PROBLEMA**

*ESPECIFICACIÓN DEL PROBLEMA ORIGINAL*

El problema planteado para esta segunda práctica consiste en la construcción de un modelo de credit scoring para obtener la calificación crediticia de un conjunto de clientes. Con esta información se podrá saber si un cliente tiene la posibilidad de obtener un crédito financiero o no. Para esto se usará un dataset con información sobre las características de los clientes como su edad o sexo y se creará el modelo de credit scoring que se ajuste más a los resultados otorgados en ese conjunto de datos.

**MARCO CONCEPTUAL**

El 'credit scoring' consiste en asignarle un puntaje al potencial deudor que, a su vez, representa una estimación del desempeño del crédito para el banco.

Los modelos scoring que una institución financiera desarrolla pueden distinguirse en cuanto al tiempo de cobertura o ventana de información. Lo más común es tomar un diseño muestral en forma de sección cruzada, considerando una selección de crédito al consumo en el tiempo t, para posteriormente seguir su comportamiento de pago sobre k periodos en el futuro. Igualmente, este tipo de modelos se desarrollan para predecir el comportamiento del cliente sobre el intervalo [t;t+k], como una función de las características observadas en el tiempo t.

Respecto a los cálculos necesarios, el modelo suele desarrollarse a partir del denominado análisis discriminante, que se realiza en dos etapas:

Una primera etapa denominada Análisis factorial discriminante, cuyo objetivo es identificar las variables que se utilizarán en el sistema de puntuación.

Para ello, se necesita un conjunto de prestatarios que integran la muestra a utilizar y se ha de conseguir que la muestra sea lo más representativa posible. También se ha de definir los comportamientos de los que se quieren formular predicciones (pago puntual de los créditos o impago de los créditos). A través del análisis estadístico se identifican las diferencias entre los diversos grupos de prestatarios. Por un lado, están los que han devuelto los préstamos sin incidencias (No Default, o cumplimiento de las obligaciones de pago). Por otro lado, están los que han tenido retrasos, los que han impagado los préstamos… (Default, o falta de pago). Estas diferencias que muestran las características que discriminan más suelen ser de tipo personal (edad, número de hijos…), socioeconómicas (situación laboral, tipo de contrato laboral, años de antigüedad en la empresa, barrio de residencia…) y financieras (ingresos, vivienda, patrimonio, titularidad de otros préstamos, historial de pagos con la entidad, incidencias de morosidad…). Hay informaciones de tipo personal que en determinados países no se pueden utilizar para confeccionar los scorings, como la información sobre sexo, estado civil o raza, por ejemplo, ya que se considera que pueden ser discriminatorias.

La segunda etapa, denominada Diseño del sistema de puntuación se utilizan técnicas estadísticas tales como la regresión o las redes neuronales. La puntuación se confecciona de manera que clasifique a los individuos analizados en grupos diferentes en relación con la probabilidad de pago de los préstamos. La puntuación que se otorga a cada variable estudiada se obtiene tras realizar un análisis minucioso de operaciones análogas, a fin de valorar las características del riesgo, según las probabilidades de morosidad que se hayan asignado a cada una de las características consideradas en la morosidad.

El sistema de puntuación ha de permitir que el máximo número posible de clientes analizados queden correctamente clasificados en función de su comportamiento de pago. Por ejemplo, si la muestra se divide entre los clientes que han pagado puntualmente los créditos y los que no, la puntuación asignada a cada cliente debería clasificarlo en el grupo acorde con su comportamiento de pago. Si el sistema de puntuación es útil debería minimizar los errores (clientes buenos clasificados como malos y viceversa).

**SOLUCIÓN DEL PROBLEMA**

1. **Especificación del conocimiento declarativo.**
2. **Especificación del conocimiento procedural**
   1. **Lectura de datos del data set**

* El dataset para el ejercicio es “h2.dataset.xls” el cual contiene datos de los clientes de un banco
* El dataset posee 7100 filas
* El dataset posee 41 columnas las cuales son:
  + Cliente
  + Edad
  + Renta
  + Sexo
  + EstadoCivil
  + NivelSocioEconomico
  + Departamento
  + AntiguedadCuenta\_Meses
  + RegistraAdicionales
  + NumeroAdicionales
  + AntiguedadAdicionales\_Meses
  + LineaCreditoNormal
  + SituacionCuenta
  + PromediodePagos
  + SaldoExtra
  + SaldoTotal
  + HabitodePago1
  + HabitodePago2
  + HabitodePago3
  + HabitodePago4
  + HabitodePago5
  + HabitodePago6
  + HabitodePago7
  + HabitodePago8
  + PeorHabitoPago
  + FlagRefinanciadoHistorico
  + FlagRefinanciado
  + DiasMora200812
  + DiasMora200901
  + DiasMora200902
  + DiasMora200903
  + DiasMora200904
  + DiasMora200905
  + CódigoInstitucion
  + PagaDespues
  + DispEfectivo
  + CalificacionSBS
  + CompraAnual\_200806\_200905
  + NroTransaccionesAnual\_200806\_200905
  + CompraMensualMaxima\_200806\_200905
  + Flag\_Cliente

Los roles son los siguientes:

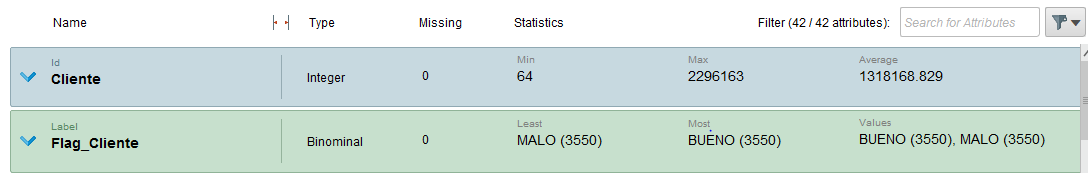
* Cliente : Id
* Flag\_Cliente : Label

También se hace el cambio de los siguientes atributos para que estén de acuerdo a lo que significan:

* Sexo: Binominal
* RegistraAdicionales: Binominal
* NumeroAdicionales: Polynominal
* SituacionCuenta: Binominal
* FlagRefinanciado: Binominal
* CalificacionSBS: Polynominal
* PagaDespues: Polynominal
* Flag\_cliente: Binominal

El resto de atributos lo reconoció correctamente Rapidminer, obteniéndose el siguiente resultado:

* 1. **Frecuencia inicial**



* 1. **Muestreo Inicial**
  2. **Balanceo de los datos**
  3. **Ranking de atributos**
  4. **Descripción de los datos**
  5. **Discretización de los datos**
  6. **Construcción de modelo regresión lineal**
  7. **Extracción de coeficientes del modelo**
  8. **Construcción del polinomio**
  9. **Ecuación del cálculo de la probabilidad**
  10. **Probabilidad marginal de cada atributo**

1. **Resultados obtenidos**
2. **Conclusiones**
3. **Referencias usadas**

Rayo S., Lara J., Camino D. (2010). A Credit Scoring Model for Institutions of Microfinance under the Basel Ii Normative. *Journal of Economics, Finance and Administrative Science*, pp 91-124

Gutiérrez, M. (2007). Modelos de Credit Scoring- Qué, Cómo, Cuándo y Para Qué. Buenos Aires: Banco Central de la República de Argentina.

1. **Código Fuente**