**EVALUACION DEL RIESGO CREDITICIO CON MINERIA DE DATOS Y SISTEMAS EXPERTOS**

*Credit risk assessment and Data Mining Expert Systems*

Víctor Guillermo Andrade Soto[[1]](#footnote-1)\*

**RESUMEN**

Se evaluó el uso de algoritmos de Arboles de Decisión sobre una minería de Datos para la creación de un Sistema Experto que permitiera evaluar el riesgo crediticio en base la información de la minería.

Se empleó una data de 23 mil créditos evaluados. Por las reglas encontradas el algoritmo ha utilizado 4 atributos y no se ha tomado en cuenta los atributos con cantidad. La aplicación diseñada, y que valida si el crédito se aprueba o no, es muy sencilla. Luego se debe extender a una que permita generar un histórico de las evaluaciones de crédito y que permita emitir reportes. El sistema deberá en un futuro cercano permitir usar la minería u almacenar en la misma.

**Palabras clave:** *Arboles de Decisión, Minería de Datos, Sistemas Expertos.*

**ABSTRACT**

Using Decision Tree algorithms on a data mining for the creation of an expert system that would assess credit risk based information mining was evaluated.

Employment data is one of 23 thousand credits evaluated. The rules found by the algorithm used four attributes are not taking into account the attributes quantity. The designed application, which validates the claim is approved or not, is very simple. Then should extend to providing scope for a historic credit evaluations and reports enabling the broadcast. The system shall allow in the near future mining use or store in it.

**Keywords:** *Decision Trees, Data Mining, Expert Systems.*

**INTRODUCCIÓN**

Una característica común de los procesos de evaluación de riesgo crediticio actuales es el constante y rápido crecimiento de su capacidad para almacenar datos normalmente históricos, de los agentes de crédito. Sin embargo, al aumentar la cantidad de datos almacenados, la capacidad para asimilarlos disminuye, por lo que se hace necesario el uso de herramientas que permitan extraer conocimiento útil de esos grandes conjuntos de datos. Es ahí donde tienen cabida la Minería de Datos y los Sistemas Expertos, técnicas que dan respuesta al análisis automático de "masas de datos", con el fin de servir como ayuda a la toma de decisiones y extracción de conocimiento.

La minería de datos (DM, Data Mining) consiste en la extracción no trivial de información que reside de manera implícita en los datos. Dicha información era previamente desconocida y podrá resultar útil para algún proceso. En otras palabras, la minería de datos prepara, sondea y explora los datos para sacar la información oculta en ellos.Las bases de la minería de datos se encuentran en la inteligencia artificial y en el análisis estadístico. Mediante los modelos extraídos utilizando técnicas de minería de datos se aborda la solución a problemas de predicción, clasificación y segmentación.

**MATERIALES Y MÉTODOS**

**Material**

1. Minería de datos de 23000 registros en una hoja de cálculo Excel
2. Software Rapid Miner para minería de Datosy creación del algoritmo de árbol de decisión
3. Software Visual Prolog para la construcción del Sistema Experto.

**Metodología**

1. **Recolección de datos:**

Nuestra minería de datos tendrá un repositorio de datos hecho en Excel (pero puede ser cualquier otro repositorio, como SQL server, Oracle, un archivo de texto, etc.). Se muestra en la figura Nro 1, la parte superior de una data de 23 mil créditos evaluados.



Figura 1

1. **Determinación de atributos:**

Los atributos de los datos a evaluar son:

|  |  |
| --- | --- |
| ESTADOCIVIL | Estado Civil de solicitante del crédito (Dato binominal) |
| OCUPACION | Ocupación (Dato binominal) |
| MONTO | Monto solicitado para aprobación (Dato entero) |
| OCUP\_GAR | Ocupación del garante (Dato polinominal) |
| SUELDO\_GAR | Sueldo del garante (Dato entero) |
| CREDITO\_GAR | Historia de crédito del garante en esta institución (Dato binominal) |
| OTORGA\_CREDITO | Fin de evaluación, se otorga o no en función de los atributos anteriores (Dato binominal) – Clase Objetivo |

Existen diversas técnicas de minería de datos por ejemplo, las Redes neuronales, la Regresión lineal y los Árboles de decisión entre otros. Nosotros tomaremos este último, que es un modelo de predicción utilizado en el ámbito de la inteligencia artificial (ámbito también de los Sistemas expertos); dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. Usaremos esta técnica. La herramienta para gestionar los datos será el RapidMiner. La Figura 2 muestra la forma en que se definen los atributos de los datos.



Figura 2

1. **Desarrollo del Modelo**

La siguiente figura muestra el modelo creado a partir de los datos:



Figura 3

La parte de la izquierda nos permite seleccionar los algoritmos de gestión de los datos. Al centro se aprecia el “modulo” que captura los datos (Read Excel) y el módulo que generará el árbol de decisiones (decisionTree). Una vez ejecutado el modelo obtendremos el siguiente árbol.

1. **Creación del Árbol de Decisiones:**



Este árbol muestra la forma en que se opera cada vez que se otorgó uno de los 23000 créditos solicitados. Construye las siguientes reglas:

OCUPACION = ESTUDIANTE: NO
OCUPACION = TRABAJADOR
| CREDITO\_GAR = NO: NO

| CREDITO\_GAR = SI
| | ESTADOCIVIL = CASADO
| | | OCUP\_GAR = OTRO: NO
| | | OCUP\_GAR = TRABAJADOR PRIVADO: SI

| | | OCUP\_GAR = TRABAJADOR PUBLICO: SI
| | ESTADOCIVIL = DIVORCIADO: SI

| | ESTADOCIVIL = SOLTERO: NO
| | ESTADOCIVIL = VIUDO: SI

Por las reglas encontradas el algoritmo ha discrecionalizado 4 atributos no tomando en cuenta los atributos con cantidad ya sea porque la información no es suficiente o porque realmente funciona de esa manera. Nosotros vamos a presumir lo segundo, aun cuanto una investigación más exhaustiva, digamos con el doble de datos podría tener más fiabilidad.

1. **Creación del Sistema Experto:**

Con estas reglas, la construcción de un sistema experto se hace necesaria pero sencilla. Este permitirá que a partir del ingreso de cada solicitud se indique si el crédito se otorga o no sin necesidad de ocupar a ningún experto crediticio. Un sistema experto es aquel que reemplaza a un experto humano en un dominio en particular. La herramienta para la construcción del sistema experto puede ser cualquiera pero se sugiere el uso de Visual Prolog que permite tomar las reglas generadas por el RapidMiner y convertirlas en predicados del tipo “SI y solo Si se cumplen las condiciones señaladas entonces se procede con el crédito”.

La aplicación básica, es decir aquella que valida si el crédito se aprueba o no, es muy sencilla. Luego se debe extender a una que permita generar un histórico de las evaluaciones de crédito y que permita emitir reportes.



**RESULTADOS**

Se puede apreciar en el análisis de los 23000 registros que se evaluó como Minería, que no existen pesos diferenciados entre atributos, es decir todos debieran influir de la misma manera en la solución, sin embargo el RapidMiner puede discrecionalizar algunos y omitir otros en la solución. Estos datos nos han permitido generar conocimiento –que es el objetivo del uso de las Minerías-, es decir aportar una forma de tomar la decisión de otorgar o no los créditos a partir de los datos de manera tal que se pudo construir un sistema experto con la solución hallada.

**CONCLUSIONES**

* Por las reglas encontradas el algoritmo tomó 4 atributos desechando aquellos con cantidades. Esto es el resultado de:
1. La información no es suficiente o
2. Realmente funciona de esa manera.
* Nosotros vamos a presumir lo segundo, aun cuanto una investigación más exhaustiva, digamos con el doble de datos podría tener más fiabilidad.
* La aplicación desarrollada es bastante básica, por lo que luego deberá extenderse a una que permita generar un histórico de las evaluaciones de crédito y que permita emitir reportes.

**REFERENCIAS**

* Mc Graw Hill Alex Berson, Stephen J. Smith. DATA WAREHOUSE, DATA MINING AND OLAP. USA, 1997.
* María José Ramírez Quintana José Hernández Orallo. EXTRACCIÓN AUTOMÁTICA DE CONOCIMIENTO EN BASES DE DATOS E INGENIERÍA DEL SOFTWARE. España, 2003.
* Colin J. White. IBM ENTERPRISE ANALYTICS FOR THE INTELLIGENT E-BUSINESS. IBM Press, USA, 2001.
* Edward Feigenbaum Chair, Peter E. Friedland, Bruce B. Johnson, H. Penny Nii, Herbert Schorr, Howard Shrobe, Robert S. Engelmore (Ed.). KNOWLEDGE-BASED SYSTEMS IN JAPAN. Mayo 1993. Del Japanese Technology Evaluation Center (JTEC).
* Durkin, J. “EXPERT SYSTEMS: DESIGN AND DEVELOPMENT”. New York. Maxwell Macmilan. 1994
* Sánchez, J. “SISTEMAS EXPERTOS: UNA METOLOGIA DE PROGRAMACION”.Prentice Hall. México. 1991
1. \* Psicólogo con grado de Magister en Ingeniería de Sistemas. Docente de la UAP.

Correo: v\_andrade@doc.uap.edu. [↑](#footnote-ref-1)