

SISTEMA EXPERTO EN APOYO A TOMA DE DECISIONES PARA APROBACIÓN DE LÍNEAS DE CRÉDITO

Jimmy Josue Peña Koo

Tecnológico Nacional de México/Instituto Tecnológico Superior del Sur del Estado de Yucatán
jimjpk@hotmail.com

Orlando Adrian Chan May

Tecnológico Nacional de México/Instituto Tecnológico Superior del Sur del Estado de Yucatán
oacmay@hotmail.com

Cinthia del Carmen Balam Almeida

Tecnológico Nacional de México/Instituto Tecnológico Superior del Sur del Estado de Yucatán
cinthia_220880@hotmail.com

Resumen

El presente estudio aborda el preproceso y análisis de la información obtenida de una base de datos de un banco de Japón, en el cual profesionales expertos han tomado decisiones para aprobación o rechazo de líneas de crédito para 690 usuarios; para el análisis de la información se emplearon herramientas de minería de datos hasta obtener reglas de producción con base al algoritmo de clasificación J48; a partir de las reglas obtenidas se empleó una técnica de inteligencia artificial, sistemas expertos, para modelar el comportamiento en la aprobación o rechazo de líneas de crédito, obteniendo una herramienta confiable al 91 por ciento para emular la resolución de expertos humanos en la toma de decisiones, dirigida a profesionales no expertos en el área, cuyos resultados pueden ser empleados en otras áreas que requieran de esta metodología para su implementación.

Palabra(s) Clave: Inteligencia artificial, Minería de datos, Sistema experto.

Abstract

The present investigation deals with the pre-processing and analysis of information obtained from a database of a bank in Japan, in which expert

professionals have made decisions for approval or rejection of credit lines for 690 users; data mining tool are used for the analysis of the information to obtain production rules based on the classification algorithm J48; from the obtained rules was used an artificial intelligence technique, expert systems, to model the behavior in credit approval or rejection, obtaining a reliable tool at 91 percent to emulate the resolution of human experts in decision making, aimed at professionals who are not experts in the area, whose results can be used in other areas that require this methodology for its implementation.

Keywords: *Artificial intelligence, Data mining, Expert system.*

1. Introducción

Han sido muchos los avances realizados para obtener máquinas con las capacidades cognitivas del ser humano. La inteligencia artificial, es el área encargada de dirigir los desarrollos realizados; pretende construir sistemas y máquinas inteligentes que presenten un comportamiento como si fuera llevado a cabo por una persona.

Los sistemas expertos, siendo una técnica de la inteligencia artificial, enlazan reglas de razonamiento, obtenidas de un preproceso y análisis de datos, con el objetivo de llegar a una conclusión. Estos surgen de la necesidad de automatizar sistemas de diagnóstico y toma de decisiones, basados en la información aportada por profesionales expertos, todo ello enmarcado en diferentes elementos tecnológicos integrados de tal manera que permitan evaluar, relacionar y determinar resultados con un alto grado de calidad.

En este contexto de aplicación de los sistemas expertos, empleados para la toma de decisiones basada en información de profesionales expertos, se obtuvo como fuente de información una base de datos libre de un banco de Japón relacionada con la aprobación o rechazo de las líneas de crédito; cuyo preproceso, análisis y modelado del comportamiento para el apoyo en la toma de decisiones con un 91% de precisión por medio de un sistema experto, lo cual fue el motivo del presente estudio.

2. Metodología

De acuerdo con la clasificación publicada por Hernández en 2010, la investigación por su propósito se denomina aplicada, por su enfoque cuantitativa, se basó en un diseño de tipo preexperimental, en el cual se manipula intencionalmente la variable independiente (el proceso de decisión), para analizar las consecuencias sobre la variable dependiente (aprobación del crédito).

Para el diseño y programación del sistema experto se empleó la metodología Buchanan, basada en el ciclo de vida en cascada, el cual indica que el proceso de construcción de un sistema se plantea como un proceso de revisión constante e implica la redefinición de los conceptos, de las representaciones o el refinamiento del sistema implementado [Delgado *et al.*, 2015]. La metodología Buchanan define que la adquisición del conocimiento de un sistema inteligente, y por extensión la construcción de todo el sistema se divide en cinco fases: identificación, conceptualización, formalización, implementación y prueba, tal como se puede observar en la figura 1.

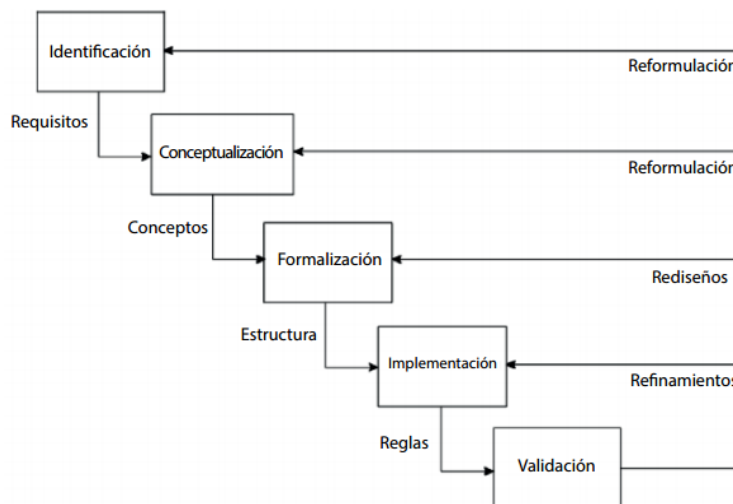


Figura 1 Ciclo de vida de la Metodología Buchanan.

Esta metodología fue creada por Bruce G. Buchanan. La característica más importante es la constante relación que debe existir entre el ingeniero de conocimiento y el experto humano, pues la metodología se enfoca en la adquisición del conocimiento [Miranda, 2011]:

- Identificación. Se obtuvo la base de datos de 690 personas que fueron sujetos a evaluación para su aprobación de tarjetas de crédito, de los cuales a 307 personas (44.5%) se les aprobó el crédito, mientras que a 383 personas (55.5%) se les rechazó el crédito. La información obtenida es relacionada con el género, estado civil, sueldo, nivel de crédito, entre otros, sin embargo, la descripción de cada uno de los quince atributos obtenidos fue confidencial por política de la fuente de información.
- Conceptualización. En esta fase se procedió a la adquisición del conocimiento del sistema experto; se recopiló la información cualitativa y cuantitativa del modelo probabilístico a partir de la fuente de información obtenida de un archivo en formato CSV para su estudio, análisis y procesamiento.

En la etapa de conceptualización se empleó la herramienta de minería de datos Weka, la cual permitió realizar el preproceso de las 690 instancias, observable en la figura 2.

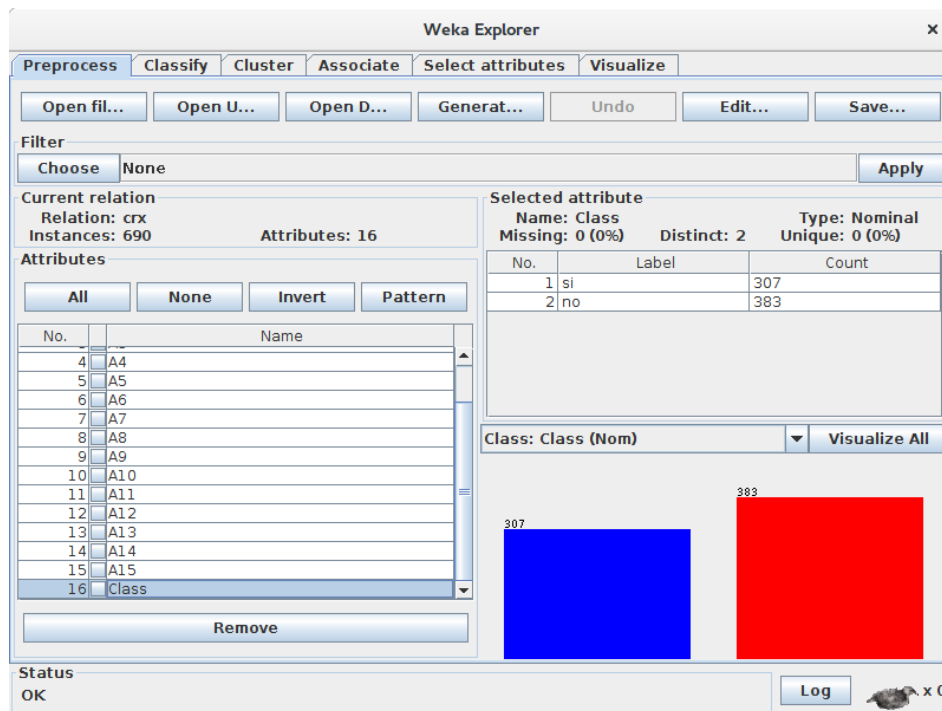


Figura 2 Preproceso de información.

- Formalización. Se construyó el modelo gráfico probabilístico basado en la clasificación por medio del algoritmo J48 de árboles. Del resumen del resultado de esta salida, se observa de acuerdo con la figura 3, que 626 instancias fueron clasificadas de forma correcta, representando un 90.7 % de la población, lo cual se refleja con una alta precisión para la toma de decisiones de aprobación (89%) o rechazo (92%) de la línea de crédito.

```
=== Summary ===
Correctly Classified Instances      626           90.7246 %
Incorrectly Classified Instances    64            9.2754 %
Kappa statistic                    0.8125
Mean absolute error                 0.1564
Root mean squared error             0.2781
Relative absolute error             31.6573 %
Root relative squared error         55.9554 %
Total Number of Instances          690

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.902   0.089   0.891     0.902   0.896     0.937   si
                0.911   0.098   0.921     0.911   0.916     0.937   no
Weighted Avg.   0.907   0.094   0.907     0.907   0.907     0.937

=== Confusion Matrix ===
  a  b  <-- classified as
277 30 |  a = si
 34 349 |  b = no
```

Figura 3 Salida del algoritmo de clasificación J48.

La ejecución del algoritmo permitió generar el árbol de la figura 4. Los nodos representan las variables (características del sujeto de estudio) y los arcos sus dependencias (factores de cumplimiento para la aprobación o rechazo del crédito).

- Implementación. Tomando como entrada para esta etapa el resultado de la formalización, se construyó el núcleo del sistema en el lenguaje de programación Prolog, a partir de la definición de las reglas para las inferencias que apoyen a las decisiones. Este código se puede observar en la figura 5.

El desarrollo de la interfaz de usuario se realizó en el lenguaje de programación C, para pedir datos al usuario, enviarlos a Prolog y posterior

visualizar la respuesta recibida desde el núcleo del sistema. La rutina principal de este procesamiento se puede observar en la figura 6.

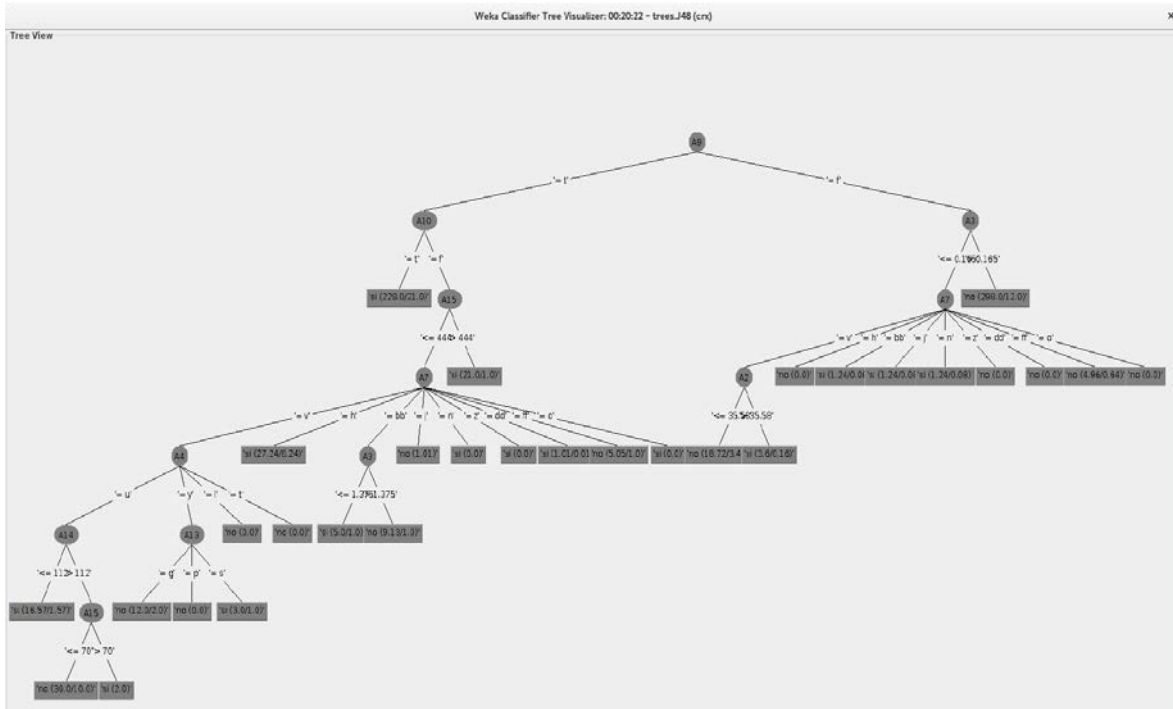


Figura 4 Árbol de clasificación.

```
es_cierto(X) :- =(X,t).
es_igual(X,Y) :- =(X,Y).
es_mayor(X,Y) :- >(X,Y).
```

```
ruta(A9,A10) :- es_cierto(A9),es_cierto(A10).
ruta(A9,A10,A15,A7,A4,A14) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,v),es_igual(A4,u),es_mayor(A14,112).
ruta(A9,A10,A15,A7,A4,A14,A15) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,v),es_igual(A4,u),\+es_mayor(A14,112),es_mayor(A15,70).
ruta(A9,A10,A15,A7,A4,A13) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,v),es_igual(A4,t),es_igual(A13,g).
ruta(A9,A10,A15,A7) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,h).
ruta(A9,A10,A15,A7,A3) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,bb),\+es_mayor(A3,1.375).
ruta(A9,A10,A15,A7) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,n).
ruta(A9,A10,A15,A7) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,z).
ruta(A9,A10,A15,A7) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,dd).
ruta(A9,A10,A15,A7) :- es_cierto(A9),\+es_cierto(A10),\+es_mayor(A15,444),es_igual(A7,o).
ruta(A9,A10,A15) :- es_cierto(A9),\+es_cierto(A10),es_mayor(A15,444).
ruta(A9,A3,A7,A2) :- \+es_cierto(A9),\+es_mayor(A3,0.165),es_igual(A7,v),es_mayor(A2,35.58).
ruta(A9,A3,A7) :- \+es_cierto(A9),\+es_mayor(A3,0.165),es_igual(A7,bb).
ruta(A9,A3,A7) :- \+es_cierto(A9),\+es_mayor(A3,0.165),es_igual(A7,j).
ruta(A9,A3,A7) :- \+es_cierto(A9),\+es_mayor(A3,0.165),es_igual(A7,n).
```

```
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10,A15,A7).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10,A15,A7,A4,A14).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10,A15,A7,A4,A14,A15).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10,A15,A7,A4,A13).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10,A15,A7,A3).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A10,A15).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A3,A7).
aprobado(A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13,A14,A15) :- ruta(A9,A3,A7,A2).
```

Figura 5 Código en Prolog del sistema experto.

```

Start_Prolog(argc, argv);
func = Find_Atom("aprobado");
for (;;)
{
    printf("\nA1 {valores: a,b} (o 'fin' para terminar): ");    scanf("%s", A1);
    if (strcmp(A1, "fin") == 0)
        break;

    printf("\nA2 {flotante: referencia 35.58}: ");    scanf("%lf", &A2);
    printf("\nA3 {flotante: referencia 0.165}: ");    scanf("%lf", &A3);
    printf("\nA4 {valores: u, y, l, t}: ");    scanf("%s", A4);
    printf("\nA7 {valores: v, h, bb, j, n, z, dd, ff, o}: ");    scanf("%s", A7);
    printf("\nA9 {valores: t, f}: ");    scanf("%s", A9);
    printf("\nA10 {valores: t, f}: ");    scanf("%s", A10);
    printf("\nA13 {valores: g, p, s}: ");    scanf("%s", A13);
    printf("\nA14 {entero: referencia 112}: ");    scanf("%d", &A14);
    printf("\nA15 {entero: referencias 70, 444}: ");    scanf("%d", &A15);

    PL_Query_Begin(TRUE);
    arg[0] = Mk_String(A1);    arg[1] = Mk_Float(A2);    arg[2] = Mk_Float(A3);
    arg[3] = Mk_String(A4);    arg[4] = Mk_String(A5);    arg[5] = Mk_String(A6);
    arg[6] = Mk_String(A7);    arg[7] = Mk_String(A8);    arg[8] = Mk_String(A9);
    arg[9] = Mk_String(A10);    arg[10] = Mk_String(A11);    arg[11] = Mk_String(A12);
    arg[12] = Mk_String(A13);    arg[13] = Mk_Integer(A14);    arg[14] = Mk_Integer(A15);
    result = PL_Query_Call(func, 15, arg);
    PL_Query_End(PL_KEEP_FOR_PROLOG);

    if(result) printf("\n*** Crédito Aprobado ***\n");
    else printf("\n*** Crédito No Aprobado ***\n");
}
Stop_Prolog();

```

Figura 6 Código en C del Sistema Experto.

- Prueba. Para realizar las pruebas de funcionalidad del sistema experto se emplearon algunas instancias de la base de datos original, verificando el cumplimiento de la aprobación o rechazo del crédito según corresponda con la información original. Uno de las instancias evaluadas con resolución afirmativa fue (b,30.83,0,u,g,w,v,1.25,t,t,01,f,g,00202,0, sí), cuyo resultado se puede observar en la figura 7.

```

jimmy@headquarter:~/Documentos/doctorado/Seminario18-SistemaExperto$ ./crx
A1 {valores: a,b} (o 'fin' para terminar): b
A2 {flotante: referencia 35.58}: 30.83
A3 {flotante: referencia 0.165}: 0
A4 {valores: u, y, l, t}: u
A7 {valores: v, h, bb, j, n, z, dd, ff, o}: v
A9 {valores: t, f}: t
A10 {valores: t, f}: t
A13 {valores: g, p, s}: g
A14 {entero: referencia 112}: 202
A15 {entero: referencias 70, 444}: 0
*** Crédito Aprobado ***
A1 {valores: a,b} (o 'fin' para terminar): fin
jimmy@headquarter:~/Documentos/doctorado/Seminario18-SistemaExperto$ _

```

Figura 7 Prueba de ejecución del Sistema Experto.

3. Resultados

A partir de la base de datos obtenida para el estudio de la aprobación o rechazo de créditos se aplicó el preproceso de información. Esta fase fue interesante por la variedad de tipos de atributos.

Las quince variables almacenadas fueron continuas y nominales con pocos y múltiples valores, las cuales permitieron la identificación de los 307 casos favorables y los 383 casos para los cuales se negó el servicio.

Los quince atributos se estudiaron por medio del modelo de clasificación generado a partir de la aplicación del algoritmo J48; de ese proceso resultaron 626 instancias clasificadas de manera correcta que representa un 91%.

Por otra parte resultaron 64 instancias clasificadas de forma incorrecta lo cual representa un 9%, de esos 64 casos clasificados incorrectamente: 30 resultaron ser falsos positivos, mientras que 34 resultaron ser falsos negativos.

Se observó una mayor correlación entre las variables A9 de tipo nominal dicotómica, la variable A3 de tipo continua y la variable A7 de tipo nominal con múltiples valores, las cuales permitieron la identificación de las principales reglas definidas en programación lógica para el proceso de resolución.

Con la aceptación del 91% como nivel de confiabilidad del modelo de clasificación obtenido a partir del entrenamiento del conjunto completo de instancias, se empleó como fundamento para la programación de la base de conocimiento del sistema experto en el lenguaje de programación Prolog, con interfaz en C para la interacción con el tomador de decisiones, con la finalidad de apoyar en el proceso de aceptación o rechazo de créditos.

El uso de la aplicación desarrollada favorece proporcionalmente en el nivel de confianza del modelo de clasificación en apoyo a la toma de decisiones, sin embargo, es de interés comentar que del 9% de error el 4% corresponde a falsos positivos, esto quiere decir, que se asignan créditos a personas que no cumplen el perfil para dicha aceptación, lo cual puede reducirse con el posterior entrenamiento del sistema experto con la información obtenida de las nuevas instancias almacenadas.

4. Discusión

A partir de la implementación de la metodología Buchanan se puede trabajar de forma organizada para la construcción de sistemas expertos que apoyen en la toma de decisiones en diferentes áreas empresariales, tal como comenta Delgado al hacer mención de un proceso de refinación constante, de manera similar como se ha de comportar el sistema experto a medida que continúa aprendiendo.

La herramienta de minería de datos Weka, por medio de los algoritmos de clasificación tal como emplea Azoumana para realizar un análisis de deserción de alumnos, aporta entradas válidas para la construcción de reglas que apoyen a la construcción de sistemas expertos, para los cuales se desconoce el contenido de la información inicial obtenida para su estudio y análisis. Por medio del algoritmo J48, el cual es una implementación de árboles de decisión, se generaron las reglas del sistema experto; de la matriz de confusión obtenida se observó una mejora con un 91% de precisión la toma de decisiones en la aprobación o rechazo de líneas de crédito, por lo que puede ser empleado para apoyo en las decisiones por personas no expertas en el área.

Se propone como extensión a este proyecto integrador el desarrollo de una interfaz gráfica de usuario por medio de la integración de la librería Boost para el paso de parámetros y el lenguaje de programación Python para el diseño de la interfaz, permitiendo un desarrollo libre y portable.

El software desarrollado es un ejemplo de la forma de aplicar los sistemas expertos para la toma de decisiones, pudiendo ser aprovechado en cualquier disciplina que requiera de un experto humano.

5. Bibliografía y Referencias

- [1] Azoumana, K. (2013, febrero 28). Analysis of Student desertion at Universidad Simón Bolívar, Faculty of Systems Engineering, with data mining techniques. *Pensamiento Americano*, Vol. 6, pp. 41-51.
- [2] Badaro, S., Ibañez, L. & Agüero, M. (2013, octubre). Sistemas Expertos: Fundamentos, Metodologías y Aplicaciones. *Ciencia y Tecnología*, Vol. 13, pp. 349-364.

- [3] Delgado, L, Cortez, A. & Ibáñez, E. (2015, enero 31). Aplicación de metodología Buchanan para la construcción de un sistema experto con redes bayesianas para apoyo al diagnóstico de la Tetralogía de Fallot en el Perú. *Industrial Data*, Vol. 18, pp. 135-148.
- [4] Gutiérrez, M., Barroso, Y. & Bedoya, J. (2011, Septiembre 29). Expert System for medical diagnosis of genetic diseases with dysmorphic. *Serie Científica de la Universidad de las Ciencias Informáticas*, No. 9, Vol. 4.
- [5] Hernández, R., Fernández, C. & Baptista, P. (2010). *Metodología de la Investigación*. México: Mc Graw Hill.
- [6] Kushmerick, N. (2002). *Machine Learning Repository*. Credit Approval Data Set. Octubre 20, 2016, de University of California, Irvine, School of Information and Computer Sciences: <https://archive.ics.uci.edu/ml/datasets.html>.
- [7] Miranda, N. (2011). *Diseño e implementación de un prototipo de sistema experto de información turística*. Tesis de Ingeniería. Universidad Austral de Chile.