

# Aprendizaje Automático y Modelos de Clasificación. Aplicación en la Calificación Crediticia de los Gobiernos Autónomos Descentralizados Municipales Como Clientes del Banco del Estado

Cervantes J.\*; Andrade J.\*

\*Escuela Politécnica Nacional, Facultad de Ciencias, Quito, Ecuador  
jairomcervantes@hotmail.com; jaime.andrade@epn.edu.ec

**Resumen:** El documento presenta las características más importantes de los modelos de clasificación basados en análisis discriminante y máquinas de vectores soporte, siendo el análisis discriminante un método que permite identificar las características que diferencian a dos o más grupos. Por otro lado las máquinas de vectores de soporte constituyen nuevas estructuras de aprendizaje automático que han demostrado un excelente desempeño en aplicaciones de clasificación. La información utilizada para el entrenamiento y validación de las metodologías propuestas, está constituida por un conjunto de indicadores de gestión administrativa y financiera correspondientes a 221 Gobiernos Autónomos Descentralizados Municipales.

**Palabras clave:** Análisis discriminante. Máquinas de vectores de soporte.

**Abstract:** The paper presents the main features of classification models based on discriminant analysis and support vector machines, with the discriminant analysis method to identify the characteristics that differentiate two or more groups, and create a function that can distinguish the accurately as possible to members of either group. On the other hand, support vector machine are machine learning new structures that have demonstrated excellent performance in classification applications. The information used for training and validation of the methodologies proposed, consists of a set of administrative and financial management for Decentralized Autonomous Municipal Government 221.

**Keywords:** Discriminant analysis. Support vector machines.

## 1. INTRODUCCION

En estudios de riesgos de crédito se han usado distintos métodos de clasificación para estimar la calificación crediticia, tales como el Análisis Discriminante, modelos Probit, Logísticos y Redes Neuronales[11]. Sin embargo muy poco se ha hecho respecto al estudio de riesgo crediticio en las entidades públicas en el Ecuador y la literatura formal al respecto es prácticamente inexistente. En razón a lo mencionado surge la inquietud de realizar una investigación tendente a utilizar Máquinas de Vectores de Soporte como técnica de clasificación de la calificación de riesgo de crédito y compararla con el método de Análisis Discriminante.

## 2. MARCO TEÓRICO

Las bases matemáticas que se va a revisar son el análisis discriminante [6] y las máquinas de vectores de soporte[15].

### 2.1 Análisis Discriminante

Sean  $\Omega_1, \Omega_2$  dos conjuntos, y  $X_1, \dots, X_p$  variables observables, donde  $\mathbf{x} = (x_1, \dots, x_p)$  las observaciones de las variables sobre un individuo  $\omega$ .

#### 2.1.1 Análisis discriminante para dos grupos

Discriminador lineal: Sean  $\mu_1, \mu_2$  los vectores de medias de las variables en  $\Omega_1, \Omega_2$  respectivamente, y se supone que la matriz de covarianzas  $\Sigma$  es común. Las distancias de Mahalanobis[10] de las observaciones  $\mathbf{x} = (x_1, \dots, x_p)$  de un individuo  $\omega$  a las poblaciones son:

$$M^2(\mathbf{x}, \mu_i) = (\mathbf{x} - \mu_i)' \Sigma^{-1} (\mathbf{x} - \mu_i), \quad i = 1, 2.$$

La primera regla de clasificación consiste en asignar  $\omega$  a la población más próxima:

$$\text{Si } M^2(\mathbf{x}, \mu_1) \begin{cases} < M^2(\mathbf{x}, \mu_2) & \rightarrow \omega \text{ a } \Omega_1, \\ > M^2(\mathbf{x}, \mu_2) & \rightarrow \omega \text{ a } \Omega_2 \end{cases} \quad (1)$$

La regla (1) expresada como una función discriminante, está dada por:

$$M^2(\mathbf{x}, \mu_2) - M^2(\mathbf{x}, \mu_1) = (\mu_2 - \mu_1)\Sigma^{-1}(\mu_2 + \mu_1) + 2\mathbf{x}'\Sigma^{-1}(\mu_1 - \mu_2)$$

Definimos la función discriminante

$$L(\mathbf{x}) = \left[ \mathbf{x} - \frac{1}{2}(\mu_1 + \mu_2) \right]' \Sigma^{-1}(\mu_1 - \mu_2) \quad (2)$$

y la regla de clasificación es:

$$\text{Si } L(\mathbf{x}) \begin{cases} > 0 & \rightarrow \omega \text{ a } \Omega_1, \\ < 0 & \rightarrow \omega \text{ a } \Omega_2 \end{cases}$$

### 2.1.2 Generalización del análisis discriminante en el caso de $k$ grupos

Si se supone que el individuo  $\omega$  ahora puede provenir de  $k$  grupos  $\Omega_1, \Omega_2, \dots, \Omega_k$  donde  $k \geq 3$ . Es preciso construir una regla que permita asignar  $\omega$  a una de los  $k$  grupos sobre la base de las observaciones  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  de  $p$  variables.

Si la media de las variables en  $\Omega_i$  es  $\mu_i$ , y la matriz de covarianzas  $\Sigma$  es común.

Considerando las distancias de Mahalanobis de  $\omega$  a los grupos

$$M^2(\mathbf{x}, \mu_i) = (\mathbf{x} - \mu_i)' \Sigma^{-1}(\mathbf{x} - \mu_i), \quad i = 1, \dots, k$$

un criterio de clasificación consiste en asignar  $\omega$  al grupo más próximo:

Si  $M^2(\mathbf{x}, \mu_i) = \min\{M^2(\mathbf{x}, \mu_1), \dots, M^2(\mathbf{x}, \mu_k)\}$  asignamos  $\omega$  a  $\Omega_i$ . (3)

Introduciendo las funciones discriminantes lineales

$$L_{ij}(\mathbf{x}) = (\mu_i - \mu_j)' \Sigma^{-1} - \frac{1}{2}(\mu_i - \mu_j)' \Sigma^{-1}(\mu_i + \mu_j)$$

la función de decisión equivale a: si  $L_{ij}(\mathbf{x}) > 0$  para todo  $j \neq i$ , asignamos  $\omega$  a  $\Omega_i$ .

### 2.2 Máquinas de vectores de soporte

Las máquinas de vectores de soporte son métodos relativamente nuevos para clasificación binaria. La idea básica radica en encontrar un hiperplano que separe los datos perfectamente en dos clases. Y luego la clasificación consiste en observar de qué lado del hiperplano queda el vector que representa un nuevo caso no conocido previamente[3, 4, 8].

#### 2.2.1 Caso linealmente separable

Dado un conjunto de entrenamiento de tamaño  $N$  compuesto de pares atributo-etiqueta  $(x_i, y_i), 1 \leq i \leq N$ , siendo  $x_i \in \mathbb{R}^n$  y  $y_i \in \{-1, 1\}$ , se desea obtener una ecuación para un hiperplano que divida dicho conjunto de entrenamiento, de manera que aquellos puntos con igual etiqueta queden del mismo lado del hiperplano. Si existe un hiperplano que aísle los puntos positivos de los negativos, se dice que los datos son linealmente separables.

El hiperplano que separe los datos es:

$$\pi: w \cdot x + b = 0 \quad (4)$$

Donde  $w \in \mathbb{R}^d$  es un vector normal (vector perpendicular) al hiperplano  $\pi$ ,  $|b|/\|w\|$  es la distancia perpendicular del hiperplano  $\pi$  al origen, y  $\|w\|$  es la norma euclídea de  $w$ .

A partir del hiperplano separador definido por  $w$  y  $b$  se definen dos hiperplanos paralelos a éste de tal forma que los puntos más cercanos al hiperplano cumplan que  $\pi_1: x \cdot w + b = 1$  y  $\pi_2: x \cdot w + b = -1$ . Así se tiene que los hiperplanos  $\pi_1$  y  $\pi_2$  son paralelos, y el margen de separación entre ellos es  $2/\|w\|$  y cumplen que no existe ningún vector del conjunto de entrenamiento que se encuentre entre ellos dos (ver Fig. 1).

Entonces para un conjunto de entrenamiento linealmente separable, se puede encontrar un clasificador lineal con máximo margen resolviendo el siguiente problema:

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} \|w\|^2 \quad (5)$$

$$\text{s. a } y_i(x_i \cdot w + b) - 1 \geq 0, \forall i = 1, \dots, n,$$

En la Fig. 1 podemos observar que algunos de los vectores de ensayo se encuentran sobre el hiperplano  $\pi_1$  o sobre el hiperplano  $\pi_2$ . Estos vectores son llamados vectores de soporte, y son los únicos que intervienen en la definición de los hiperplanos separadores ya que si sobre la Fig. 1 se mueven los restantes vectores, sin desplazar ninguno entre los hiperplano  $\pi_1$  y  $\pi_2$ , los hiperplano separadores no cambian, es decir, la solución del problema permanece invariante.

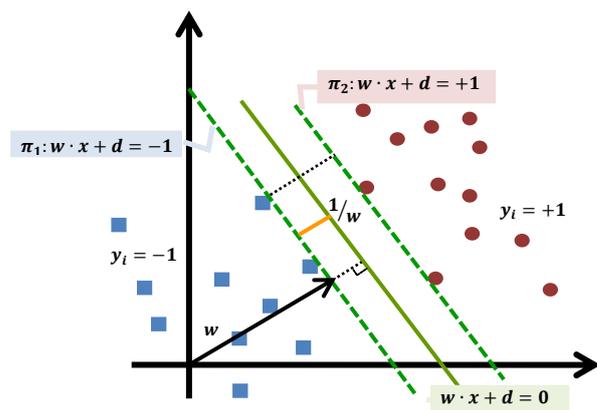


Figura 1. Concepto básico para entender el problema de clasificación a partir de las máquinas de vectores de soporte.

Para la resolución del problema de optimización en (5) se utiliza la técnica de los multiplicadores de Lagrange. Así la función objetivo es:

$$L_p(w, b, \alpha_i) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i(x_i \cdot w + b) - 1) \quad (6)$$

Una vez conocidos los multiplicadores de Lagrange se determina una solución de la forma:

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

donde  $w$  son los pesos del vector,  $\alpha_i$  son los multiplicadores de Lagrange.

La búsqueda de los valores de  $w$  y  $\alpha_i$  se convierte en un problema de programación cuadrática, el cual puede ser

solucionado por medio del algoritmo de secuencia mínima de optimización. La constante  $b$  se obtiene de las condiciones de Karush – Kunh – Tucker.

Una vez obtenido el vector  $w$  y la constante  $b$  la solución al problema de optimización en (5) consiste en determinar en que lado del hiperplano frontera (el hiperplano que se encuentran justamente en la mitad de los hiperplanos  $\pi_1$  y  $\pi_2$ ) se encuentra la nueva entrada de  $x$  y asignar la correspondiente clase (1 ó -1), es decir, asignar la clase mediante:

$$y = \text{signo}(w \cdot x + b)$$

### 2.2.2 Caso linealmente no separable

Para datos que son linealmente no separable, carece completamente de sentido buscar un hiperplano de separación. Sin embargo, no es difícil emplear las ideas generales del caso separable al caso no separable. Se castigará con un coste adicional la función objetivo cuando no se cumpla con las restricciones. Este castigo se puede indicar de forma explícita introduciendo una variable de holgura  $\xi_i$ ,  $i = 1, \dots, n$  en las restricciones, planteando los siguientes conjuntos de restricciones:

$$\begin{aligned} x_i \cdot w + b &\geq +1 - \xi_i \text{ para } y_i = +1 \\ x_i \cdot w + b &\leq -1 + \xi_i \text{ para } y_i = -1 \\ \xi_i &\geq 0 \forall i = 1, \dots, n \end{aligned} \quad (7)$$

Ahora tenemos que para producir un error es necesario que el valor correspondiente  $\xi_i$  sea superior a la unidad (ver Fig. 2). Luego para que se produzca un error se tiene que cumplir que  $\xi_i \geq 1$ , entonces  $\sum_i \xi_i$  es una cota superior del número de errores que se comete dentro del conjunto de ensayo, lo que significa que el castigo está acotado por  $\sum_i \xi_i$  y de aquí que el castigo medio por cada vector de ensayo es menor que  $\frac{1}{n} \sum_{i=1}^n \xi_i$ .

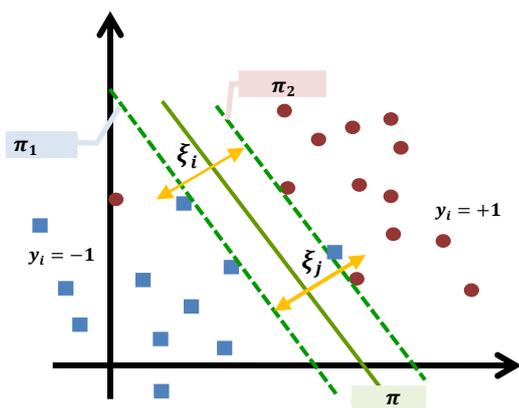


Figura 2. Presentación del parámetro de error  $\xi_i$ .

Por lo expuesto se puede plantear el siguiente problema de optimización:

$$\begin{aligned} \min_{w \in R^d} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s. a } & \begin{cases} y_i \cdot (x_i \cdot w + b) - 1 + \xi_i \geq 0, & \forall i \\ \xi_i \geq 0, & \forall i \end{cases} \end{aligned} \quad (8)$$

### 2.2.3 Máquinas no lineales de vectores de soporte

El principio de las máquinas de vectores de soporte no lineales consiste en mapear el espacio de entrada a un espacio de representación de dimensión mayor a través de una función núcleo elegida a priori. Para desarrollar el problema se sigue el siguiente desarrollo.

Sea una aplicación, a la que se denota por  $\Phi$ , del conjunto de entrada  $\mathcal{X} \subset R^d$  en un espacio vectorial  $H$  dotado de un producto escalar:

$$\Phi: \mathcal{X} \subset R^d \rightarrow H$$

Ahora, se consideran los vectores transformados  $\{\Phi(x_1), \dots, \Phi(x_n)\}$  en lugar de considerar el conjunto de vectores  $\{x_1, \dots, x_n\}$  y se tiene que, si se plantea el problema de optimización original a estos vectores, es decir, se cambia de vectores de entrada y de espacio de entrada, entonces se tiene que los nuevos vectores entran a formar parte de la solución del problema, solo a través del producto escalar definido en  $H$  como funciones de la forma  $\langle \Phi(x_1), \Phi(x_n) \rangle$ . Por tanto si se considera una función  $k: H \times H \rightarrow R$  a la que se le denomina función núcleo, tal que

$$k(x_i, x) = \Phi(x_i) \cdot \Phi(x) = \langle \Phi(x_i), \Phi(x) \rangle \quad (9)$$

Solo es necesario conocer la función núcleo para resolver el algoritmo y no se necesita tener la forma explícita de la aplicación  $\Phi$ .

Por tanto si se reemplaza  $\langle x_i, x \rangle$  por  $k(x_i, x)$  en la solución de los problemas de optimización, se habrá conseguido una máquina de vectores de soporte planteada en un espacio de dimensión superior (ver Fig. 3).

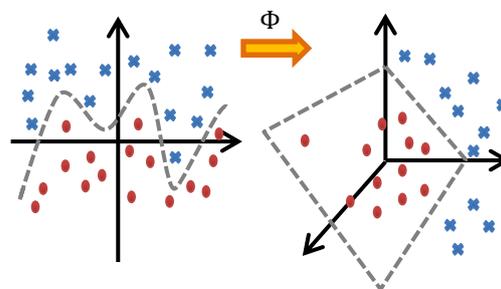


Figura 3. Idea del uso de una función para la transformación del espacio de los datos.

Resolviendo el problema de la máquina de vectores de soporte en un espacio de dimensión superior al del espacio de entrada, el cual trabaja con un conjunto de funciones lineales, el resultado es lineal en este espacio, pero no es necesariamente lineal en el espacio de entrada  $\mathcal{X}$ , con lo cual se está generalizando el problema a conjuntos de funciones no lineales.

Si transformamos los datos, el vector solución  $w$  queda  $w = \sum_{i=1}^N \alpha_i y_i \Phi(x_i)$  y en términos de los vectores de soporte  $w = \sum_{i=1}^N \alpha_i y_i \Phi(s_i)$ . Donde por  $\Phi(s_i)$  denota los vectores de soporte del conjunto  $\{\Phi(x_1), \dots, \Phi(x_n)\}$ .

Se puede ver que la aplicación  $\Phi$  aparece explícitamente utilizada en la solución del vector  $w \in H$  dada en términos de

los vectores de soporte, no es necesario tener identificada la transformación  $\Phi$  ya que la solución queda:

$$f(x) = \sum_{i=1}^{N_{VS}} \alpha_i y_i \langle \Phi(s_i), \Phi(x) \rangle + b$$

### 2.2.4 Máquinas de vectores de soporte para la multclasificación

Se genera los problemas de clasificación a etiquetas múltiples. Así sea  $\{x_1, \dots, x_n\} \subset \mathcal{X} \subset \mathbb{R}^n$  un conjunto de vectores de entrada,  $\mathcal{Y} = \{\theta_1, \theta_2, \dots, \theta_\ell\}$  el conjunto de todas las posibles etiquetas, con  $\ell > 2$ .

Para resolver problemas de multclasificación, la forma más habitual de utilización de las máquinas de vectores de soporte admite dos tipos de arquitecturas:

La primera arquitectura para los problemas de multclasificación son las máquinas de vectores de soporte biclasificadoras generalizadas, en la cual se construyen una función clasificadora global a partir de un conjunto de funciones dicotómicas.

El segundo tipo de arquitectura son las máquinas de vectores de soporte multclasificadoras, generando una función global directamente considerando todas las clases a la vez.

La solución al problema de multclasificación en las máquinas de vectores de soporte biclasificadoras generalizadas es transformar las  $\ell$  particiones del conjunto de entrenamiento en el conjunto  $L$  biparticiones, en las cuales construye la correspondiente función discriminadora obteniendo  $f_1, \dots, f_L$  clasificadores dicotómicos o biclasificadores. A continuación mediante un esquema de reconstrucción, se realiza la función de los biclasificadores  $f_i, i = 1, \dots, L$  con el objeto de proporcionar como salida final, a una de las  $\ell$  clases posibles,  $\{\theta_1, \dots, \theta_\ell\}$ .

En el esquema de descomposición, las máquinas más utilizadas son las:

Máquinas de vectores de soporte  $1 - c - r$  (iniciales de uno-contr-resto), donde cada función clasificadora parcial  $f_i$  enfrenta la clase  $\theta_i$  contra el resto de las clases.

Máquinas de vectores de soporte  $1 - c - 1$  (iniciales de uno-contr-uno), donde cada función clasificadora parcial  $f_{ij}$  enfrenta la clase  $\theta_i$  contra la clase  $\theta_j$ , sin considerar las restantes clases.

La máquina de vectores de soporte  $1 - c - r$  fue dada por Vapnik. Estas máquinas multclasificadoras construyen  $\ell$  biclasificadores donde la función discriminante  $f_k, k = 1, \dots, \ell$  discrimina los vectores de entrenamiento de la clase  $Z_k$ , del resto de vectores de las otras clases  $Z_k$ , esto es, si el biclasificador  $f_k$  lleva a cabo la discriminación de las clases sin error, entonces  $sign(f_k(x_i)) = 1$ , si el vector  $x_i \in Z_k$  y  $sign(f_k(x_i)) = -1$ , si el vector  $x_i \in Z_k$ .

Entonces dado una nueva entrada  $x$ , la salida numérica de la máquina  $f_k(x)$  se interpreta de la siguiente forma:

$$\theta(f_k(x)) = \begin{cases} \theta_k & sign(f_k(x_i)) = 1 \\ \theta_0 & sign(f_k(x_i)) = -1 \end{cases} \quad (11)$$

Para el problema de multclasificación las máquinas de vectores de soporte  $1 - c - 1$  se construyen  $L = \frac{\ell(\ell-1)}{2}$  biclasificadores donde la función discriminante  $f_{kh}, 1 \leq k \leq h \leq \ell$  discriminan los vectores de entrenamiento de la clase  $Z_k$ , de los vectores de entrenamiento de la clase  $Z_h$ , esto es, si el biclasificador  $f_{kh}$  lleva a cabo la discriminación de las clases sin error, entonces  $sign(f_{kh}(x_i)) = 1$ , si el vector  $x_i \in Z_k$  y  $sign(f_{kh}(x_i)) = -1$ , si el vector  $x_i \in Z_h$ . Los restantes vectores de entrenamiento  $Z/\{Z_k \cup Z_h\}$  no se consideran en la construcción del problema de optimización.

Entonces dada una nueva entrada  $x$ , la respuesta de la máquina  $f_{kh}(x_i)$  es de la siguiente forma:

$$\theta(f_{kh}(x)) = \begin{cases} \theta_k & sign(f_{kh}(x)) = 1 \\ \theta_h & sign(f_{kh}(x)) = -1 \end{cases} \quad (12)$$

## 3. METODOLOGÍA

A continuación se desarrolla el análisis empírico, que comprende una serie de etapas que van desde la búsqueda de la información, la recolección de los datos, el análisis de las variables, hasta la interpretación de los principales resultados obtenidos.

### 3.1 Análisis de datos y definición de variable

Con la información recabada se construyeron indicadores de gestión administrativa y financiera, que se presentan a continuación [1, 2]:

- Ingreso propio / Gasto corriente ( $Ip/Gc$ );
- Transferencias / Ingresos totales ( $T/It$ );
- Gasto corriente / Ingreso total ( $Gc/It$ );
- Inversiones / Gasto total ( $I/Gt$ );
- Inversiones / Transferencias de capital ( $I/Tk$ );
- Ingresos propios / Ingresos totales ( $Ip/It$ );
- Servicio de la deuda / Ingreso propio ( $Sd/Ip$ );
- Cartera vigente / Ingresos totales;  $Cv/It$
- Cartera vigente / Ingresos propios ( $\frac{Cv}{Ip}$ );
- Servicio de la deuda total / (Ingreso total - Crédito público) ( $Sd/(It - Cp)$ );
- Deuda pública / (Ingreso total - Crédito público) ( $Dp/(It - Cp)$ ).

### 3.2 Selección de las variables más representativas

Se aplica la metodología de análisis de componentes principales a los 221 Gobiernos Autónomos Descentralizados Municipales (GADM), con sus respectivas variables de gestión administrativa y financiera.

En la tabla 1 se presenta las correlaciones entre las variables originales y cada uno de las componentes principales, donde

se puede apreciar que la primera componente está constituida por las variables ingresos propios sobre ingresos totales e ingresos propios sobre gasto corriente. Estas variables reflejan, en la parte positiva de la primera componente la autosuficiencia que tienen los GADM. La parte negativa de la primera componente está constituida por las variables de transferencias sobre ingresos totales, servicio de la deuda sobre ingresos propios y cartera vigente sobre ingresos propios, reflejando la dependencia y el nivel de endeudamiento que tienen los GADM.

En la segunda componente se observa que las variables con más peso son: cartera vigente sobre ingresos totales, servicio de la deuda y deuda pública sobre ingresos totales sin crédito público. Por lo tanto, esta componente refleja el peso de la deuda que tienen los GADM.

La tercera componente principal está formada por las variables que miden las inversiones sobre gasto total y sobre las transferencias de capital que tienen los GADM, representando el grado de inversión que tienen las entidades. En la parte negativa se encuentra la variable gasto corriente sobre ingreso total, representando la gestión del gasto de funcionamiento de las entidades.

Tabla 1. Matriz de componentes.

Variable	Componente		
	1	2	3
$I_p/I_t$	0,89	0,2	-0,13
$I_p/G_c$	0,8	0,24	0,29
$T/I_t$	-0,75	-0,18	-0,03
$S_d/I_p$	-0,66	0,52	-0,02
$C_v/I_p$	-0,63	0,56	0,04
$C_v/I_t$	0,04	0,87	-0,06
$S_d/(I_t - C_p)$	-0,04	0,84	-0,14
$D_p/(I_t - C_p)$	0,12	0,56	-0,14
$I/G_t$	-0,13	-0,04	0,94
$G_c/I_t$	0,31	-0,05	-0,81
$I/T_k$	0,44	0,3	0,6

La tabla 2 muestra los puntajes obtenidos por varios GADM en las tres primeras componentes. Al estudiar los puntajes se puede notar las diferencias que existen entre los GADM de 24 de Mayo y Aguarico, el GADM de 24 de Mayo presenta un valor negativo en la primera componente reflejando una mayor dependencia de las transferencias por parte del Estado en comparación con el GADM de Aguarico que presenta un valor positivo en la primera componente reflejando una mayor autosuficiencia. Los puntajes en la segunda componente determinan que el GADM de 24 de Mayo tiene mayores niveles de endeudamiento que el GADM de Aguarico. En la tercera componente el puntaje del GADM de Aguarico evidencia que tiene una mayor inversión y menor gasto corriente que el GADM de 24 de Mayo.

Al examinar las primeras puntuaciones que obtiene el GADM de Olmedo (Loja) en las dos primeras componentes, se observa que es una entidad con alta dependencia de las transferencias del Gobierno y presenta un elevado nivel de endeudamiento, el puntaje de la tercera componente

representa que monto de las inversiones que realiza el GADM es mayor al gasto corriente.

Analizando los valores obtenidos por los GADM en las tres componentes principales, se aprecia que existen GADM con puntajes altos, tanto en la parte positiva como negativa. En la tabla 2, se puede observar la autosuficiencia de ingresos propios que tiene el GADM de Samborondón, por otra parte el sobre endeudamiento que tiene el GADM de Cuenca, el abuso del gasto corriente del GADM de Chimbo, etc.

Tabla 2. Matriz de puntajes en las tres componentes principales.

GADM	Componente		
	1	2	3
24 De Mayo	-0,78	0,24	-0,82
Aguarico	2,08	-0,84	0,49
Baquerizo Moreno	-2,25	0,96	-1,29
Olmedo (Loja)	-3,33	3,09	0,9
Samborondón	4,33	1,00	1,68
Cuenca	1,97	4,44	-0,38
Chimbo	-0,29	-0,52	-3,17
Rumiñahui	2,81	0,1	0,43
Salcedo	0,72	0,19	0,37

### 3.3 Determinación del grupo de pertenencia de los GADM por la similitud de la información

Partiendo de los resultados del análisis de componentes principales, se considera las 3 componentes y los 192 GADM que no presentan datos atípicos en las componentes ya que los GADM con valores atípicos en cualquiera de las componentes, puede originar una clasificación pobre y grupos con pocos elementos.

Los GADM con valores atípicos que no son considerados en esta etapa son los que presentaron los puntajes más altos, tanto en la parte positiva como negativa de las componentes principales.

Se formaron 5 conglomerados, ya que representa el número de categorías utilizadas por la Superintendencia de Bancos y Seguros para calificar y provisionar un crédito.

En la tabla 3, se presenta los centros de los conglomerados finales. Esta tabla es de gran utilidad para interpretar la constitución de los conglomerados, pues resume los valores centrales de cada conglomerado en las 3 componentes principales.

Tabla 3. Centros de los conglomerados finales.

		Componente		
		1	2	3
Conglomerado	1	-0,135	1,476	0,488
	2	1,761	1,332	-1,3
	3	1,69	-0,458	0,653
	4	-1,243	-0,976	1,149
	5	-1,755	1,423	-1,511

En la tabla 4, se expone el número de casos asignado a cada conglomerado. En el proyecto, el conglomerado 4 es el de mayor tamaño con el 33% del total de los casos considerados

y el conglomerado 5 es el más pequeño de los conglomerados con el 8%.

Tabla 4. Número de GADM en cada conglomerado.

Conglomerado	Nº GADM
1	37
2	29
3	48
4	63
5	15
Total	192

Una vez culminada la agrupación de los GADM, se aplica el siguiente criterio para determinar la calificación de cada grupo según el comportamiento de las componentes. Se utilizan las medias de las tres componentes dentro de cada grupo otorgando 5 puntos al grupo con una mejor media, 1 punto al grupo con la peor media en cada una de las componentes. En la tabla 5, se presenta los puntajes obtenidos en cada conglomerado para cada una de las componentes.

Analizando la tabla 3 y 5 se observa que el conglomerado 2, obtiene un puntaje de 5 en la componente (1) que manifiesta el comportamiento de autosuficiencia de los GADM, debido a que su centro es el valor más positivo de todos los conglomerados. En la componente (2) que representa el nivel de endeudamiento de los GADM, el conglomerado 4 obtiene un puntaje de 5, ya que presenta el valor más negativo. Y en la componente (3) que representa el grado de inversión que tiene los GADM, el conglomerado 4 obtiene un puntaje de 5, ya que presenta el valor más positivo.

Tabla 5. Puntajes obtenidos en cada conglomerado para cada componente.

		Componente		
		1	2	3
Conglomerado	1	3	1	3
	2	5	3	2
	3	4	4	4
	4	2	5	5
	5	1	2	1

### 3.4 Análisis Discriminante

El objetivo final del análisis discriminante es encontrar la combinación lineal de las variables independientes que permiten diferenciar a los grupos. Una vez encontrada esa combinación, podrá ser utilizada para la asignación de nuevos casos[9].

#### 3.4.1 Selección de las variables independientes y dependiente

En el proyecto se debe tener en cuenta que las variables para el cálculo de las funciones discriminantes, son los 11 indicadores de gestión administrativa y financiera utilizados en el análisis de componentes principales. La pertenencia a los grupos, determinada con el análisis de conglomerados se utiliza como variable dependiente.

#### 3.4.2 Validación de las funciones discriminantes

En la tabla 6, se presenta los valores propios, en la tabla se puede observar qué porcentaje de la varianza explica cada eje discriminante. En el estudio la primera función explica el 44,69% de la varianza total de los datos, la segunda función aporta el 37,36%, la tercera función aporta con el 11,47% y la cuarta aporta con el 6,48% de variabilidad entre grupos. El coeficiente de correlación canónica de la tabla 6, mide el poder discriminante de la  $i$ -ésima función discriminante, ya que es el porcentaje de la variación total en dicha función, que es explicada por las diferencias entre grupos. Valores cercanos a 1 indican mayor potencia discriminante de la  $i$ -ésima función, en nuestro estudio las funciones tienen valores aceptables.

Tabla 6. Valores propios.

Función	Valores propios	% de varianza	Correlación canónica
1	2,47	44,69	0,84
2	2,06	37,36	0,82
3	0,63	11,47	0,62
4	0,36	6,48	0,51

En la tabla 7, se determina si las funciones generadas son estadísticamente significativas, es decir, si el conjunto de las funciones permiten que las medias de los grupos estén separadas. El valor de la lambda de Wilks para el conjunto formado por las cuatro funciones es cercano a cero (0,04), por lo que se puede concluir que las cuatro funciones tienen un alto poder discriminatorio. Esto se corrobora con los  $p$ -valores asociados al estadístico  $J_i$  - cuadrado, son menores a 0,05

Tabla 7. Lambda de Wilks.

Contraste funciones	Lambda Wilks	Ji-cuadrado	gl	Sig.
1 a la 4	0,04	578,1	44	0

#### 3.4.3 Valoración de la capacidad predictiva

En la tabla de centroides (8), se muestra que los GADM del grupo C obtiene, en términos generales, mayores puntuaciones que los demás GADM de los otros grupos en la primera función discriminante. En la segunda función, los GADM del grupo A tienen su centroide en la parte negativa y los GADM del grupo E tiene su centroide ubicado en la parte positiva. En la tercera función, el centroide de los GADM del grupo D se sitúa en la parte positiva, por lo que esta función ayuda a discriminar mejor esta calificación, y en la cuarta función, el centroide de los GADM del grupo C se sitúa en la parte negativa, y el centroide de los GADM del grupo A está ubicado en la parte positiva, es razonable que la cuarta función ayude a discriminar mejor a dichos grupos.

Tabla 8. Funciones en los centroides de los grupos.

Calificación de Riesgo	Función			
	1	2	3	4
A	0,06	-1,58	0,43	0,71
B	-1,53	-0,11	-0,73	-0,26
C	2,9	-0,82	-0,18	-0,78
D	-0,55	1,58	1,26	-0,31
E	1,98	3,2	-1,04	1,07

En la Fig 4, se presenta el diagrama de dispersión de todos los casos utilizados en el análisis sobre el plano definido por las dos funciones discriminantes. Los casos están identificados por la calificación de los GADM.

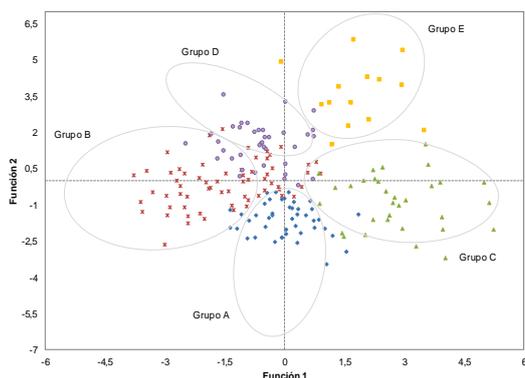


Figura 4. Idea del uso de una función para la transformación del espacio de los datos.

### 3.5 Máquinas de vectores de soporte

Para la aplicación de las máquinas de vectores de soporte en el presente proyecto se recurre al software Weka [5].

#### 3.5.1 Selección de núcleo

El núcleo que se utiliza en el presente proyecto, es la función núcleo de base radial, se emplea este núcleo ya que diversos autores lo sugieren como una elección razonable para clasificación con MVS. Además este núcleo hace corresponder de manera no lineal ejemplos en un espacio de mayor dimensión y tiene menos parámetros asociados a él que otros núcleos, lo que hace menos complejo el modelo[7]. La función del núcleo de base radial es la siguiente:  $k(x) = \exp(-\gamma \|x\|^2)$

Un parámetro adicional en la aplicación de la MVS, es el de penalidad para el error (C). El parámetro C, permite tener cierto error de clasificación a cambio de una mejor generalización de los datos. Sin embargo, el valor del parámetro varía de acuerdo a los datos que se desean clasificar. Por lo tanto, se realiza una búsqueda exhaustiva de un valor adecuado que permita clasificar correctamente la mayor cantidad de casos como sea posible. Es importante tener en cuenta que con un valor muy elevado de C, tiene una alta penalización para puntos no separables y tiende a tener muchos vectores soporte, lo que puede llevar al sobreajuste. Por otra parte, un valor muy pequeño de C, hace que el modelo sea muy rígido, lo que puede conducir a un subajuste.

La determinación de los parámetros es una etapa trascendental en la aplicación de las MVS, por cuanto la obtención de resultados acertados dependerá, en gran parte, de apropiados valores de los parámetros. Por lo tanto, se propone varias configuraciones con la intención de determinar que configuración obtiene mejores resultados en cuanto a errores de clasificación. Tomando esto en cuenta se decidió probar las MVS con configuraciones donde se variaba el valor de C entre 1 hasta 100 y el valor de  $\gamma$  entre 0,01 y 10.

#### 3.5.2 Entrenamiento

Con los datos seleccionados, se realizan los distintos entrenamientos de las MVS, por medio con un clasificador que utiliza algoritmos de optimización mínima secuencial (SMO) por sus siglas en ingles. Este clasificador resuelve los problemas de multclasificación utilizando clasificaciones por pares uno contra uno, determinando el modelo a partir de todas las posibles combinaciones de los diferentes grupos.

Para el clasificador SMO se selecciona el núcleo de base radial, configurado de acuerdo a los distintos parámetros asociados. Con el entrenamiento del clasificador seleccionado se experimentan las diferentes configuraciones propuestas, luego se calcula el porcentaje de casos clasificados correctamente para cada una de las configuraciones, de cada grupo y para todos los grupos. Este porcentaje permite determinar las configuraciones que alcanzan mejores resultados. Los resultados señalan que con la configuración 31 obtiene el 91% de GADM clasificados correctamente. A continuación se presentan los resultados de varias configuración es generadas.

Tabla 9. Resultado de varias configuraciones.

Configuración	C	$\gamma$	% clasificados
1	1	0,01	86%
7	10	0,01	90%
8	10	0,05	88%
14	30	0,05	85%
20	50	0,05	85%
25	70	0,01	90%
31	100	0,01	91%

En la tabla 10 se presenta la información de los GADM escogidos como vectores de soporte para obtener la MVS como función de clasificación entre los grupos C y E. El modelo genera 9 vectores de soporte, 5 son GADM del grupo C y 4 son GADM del grupo E.

Tabla 10. Vectores de soporte para los grupos C-E.

Vectores de soporte	
C	E
Quinsaloma	24 De Mayo
Quinindé	Flavio Alfaro
Las Naves	Bolívar (Manabí)
Suscal	Santa Rosa
Naranjal	

La Fig. 5 muestra la dispersión entre los grupos C y E, donde se puede observar que no existen vectores de soporte que

presentan características similares en sus indicadores de gestión administrativa y financiera. También se observa que los vectores de soporte generan una aceptable frontera de separación entre los dos grupos.

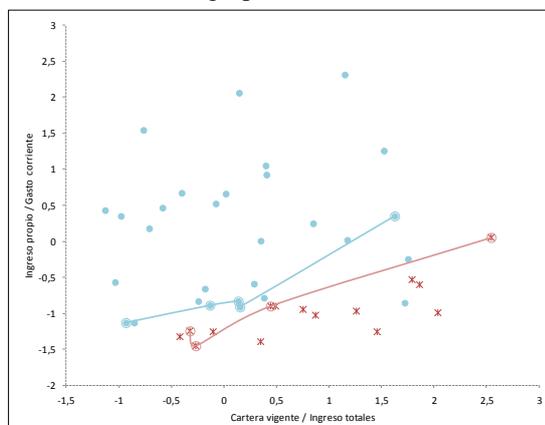


Figura 5. Solución gráfica a la clasificación de los grupos C y E. Los puntos (●) en celeste representan al grupo C y los puntos (\*) en rosado al grupo E. Los puntos en círculos representan los vectores de soporte.

#### 4. RESULTADOS Y DICUSIONES

Una vez aplicada la teoría que debe seguirse para desarrollar un modelo de análisis discriminante y de máquinas de vectores de soporte, se procede a discutir los resultados obtenidos. Para determinar la precisión de asignación de la calificación de las metodologías propuestas en el presente proyecto, se analiza la curva ROC y la matriz de confusión como medidas de capacidad de predicción [12].

##### 4.1 Pruebas y análisis de resultados del análisis discriminante

La tabla 11 resume la capacidad predictiva de las funciones discriminantes. Los GADM del grupo A son correctamente clasificados en el 83,3% de los casos; los GADM del grupo B en el 92,1%; los GADM del grupo C en el 96,6%; los GADM del grupo D en el 94,6% y los GADM del grupo E en el 86,7%. En general, las funciones discriminantes clasifican correctamente en promedio el 90,6% de los GADM, por lo tanto se puede esperar que las funciones discriminantes, clasifiquen correctamente el 90,6% de los futuros casos nuevos que se intenten clasificar.

Tabla 11. Resultado de la clasificación del análisis discriminante.

Calificación	Grupo Pronosticado					Total	
	A	B	C	D	E		
Real	A	40	4	3	1	0	48
	B	3	58	0	2	0	63
	C	0	1	28	0	0	29
	D	1	0	0	35	1	37
	E	0	1	1	0	13	15
%	A	83,3	8,3	6,3	2,1	0	100
	B	4,8	92,1	0	3,2	0	100
	C	0	3,4	96,6	0	0	100
	D	2,7	0	0	94,6	2,7	100
	E	0	6,7	6,7	0	86,7	100

##### 4.2 Pruebas y análisis de resultados de las máquinas de vectores de soporte

La matriz de confusión (tabla 12) obtenida para las máquinas de vectores de soporte presenta los siguientes resultados. Los GADM del grupo A son correctamente clasificados en el 97,9% de los casos; los GADM del grupo B en el 88,9%; los GADM del grupo C en el 93,1%; los GADM del grupo D en el 91,9% y los GADM del grupo E en el 66,7%. En general, las máquinas de vectores de soporte clasifican correctamente en promedio el 90,6% de los GADM, se puede esperar que los hiperplanos óptimos, clasifiquen correctamente el 90,6% de los futuros casos nuevos que se intenten clasificar.

Tabla 12. Resultado de la clasificación de las máquinas de vectores de soporte.

Calificación	Grupo Pronosticado					Total	
	A	B	C	D	E		
Real	A	47	0	0	1	0	48
	B	3	56	0	4	0	63
	C	0	1	27	0	1	29
	D	1	1	0	34	1	37
	E	0	0	5	0	10	15
%	A	97,9	0	0	2,1	0	100
	B	4,8	88,9	0	6,3	0	100
	C	0	3,4	93,1	0	3,4	100
	D	2,7	2,7	0	91,9	2,7	100
	E	0	0	33,3	0	66,7	100

##### 4.3 Análisis comparativo de los resultados del análisis discriminante y de las máquinas de vectores de soporte.

Comparando el funcionamiento del análisis discriminante con las máquinas de vectores de soportes, se determina que ambas predicen adecuadamente la calificación de riesgo de los GADM. La Fig. 6 representa la curva ROC del análisis discriminante y de las máquinas de vectores de soporte, se puede observar que las curva del análisis discriminante (AD) está más cerca de la esquina superior izquierda del diagrama, obteniendo la mayor área bajo la curva y por lo tanto mayor la discriminación en relación con la curva de las máquinas de vectores de soporte.

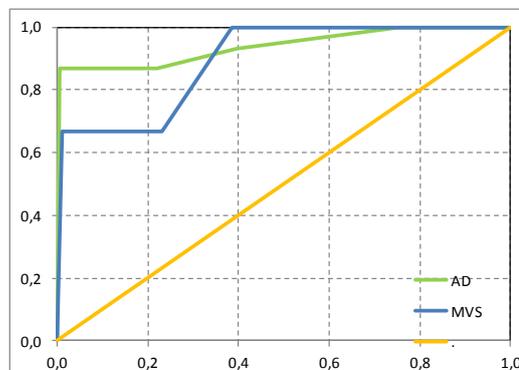


Figura 6. Curva ROC para el análisis discriminante y las máquinas de vectores de soporte.

Los promedios de efectividad de las pruebas anteriores son buenos, sin embargo si se compara por ejemplo los resultados

de las tablas de confusión, se observa que el análisis discriminante tiene menos errores de clasificación en los GADM del grupo C y las máquinas de vectores de soporte tiene menos errores de clasificación en los GADM del grupo A.

La predicción de la calificación de riesgo B, C y D dada por las máquinas de vectores y el análisis discriminante tienen resultados muy similares. En cambio se comportan de manera distinta las dos metodologías aplicadas en la predicción de las calificaciones A y D.

Las diferentes predicciones dadas por los dos modelos, sirven para realizar un enfoque de su utilidad en el cálculo del monto de provisión que realiza el Banco del Estado, tomando como base la normativa de la Superintendencia de Bancos y Seguros.

La determinación del monto de provisiones son de especial importancia en las instituciones financieras, de ahí, la calidad de estimar las provisiones de una manera adecuada de modo que no sean insuficientes tal que sobrestimen el valor de los activos, y que no sean excesivas tal que afecten a la rentabilidad de la institución.

La Superintendencia de Bancos y Seguros dispone a las instituciones financieras, realizar la medición adecuada de las provisiones a través de la cuantificación de los siguientes 3 factores para la calificación de la cartera crediticia y contingentes: capacidad de pago, experiencia de pago y riesgo de entorno económico [13, 14].

Dado estos antecedentes y considerando aquellos GADM mal clasificados en cada una de los modelos propuestos, se procede a calificar la cartera crediticia y el monto de provisiones que debe hacer el Banco del Estado para dichas entidades con fecha de corte al 31 de diciembre del 2011.

Para determinar la capacidad de pago y situación financiera del deudor, se utiliza la calificación de riesgo determinada por el análisis discriminante y las máquinas de vectores de soporte. Las ponderaciones de experiencia de pago y riesgo de entorno económico de las entidades se obtuvieron del Reporte R04, generado por el Banco del Estado para el Organismo de Control al 31 de diciembre del 2011.

La tabla 13 presenta la información de un GADM que fueron calificados incorrectamente utilizando las funciones discriminantes y los hiperplanos óptimos de separación. En la tabla se considera la calificación real con sus respectivos valores en cada factor de riesgo y la cartera vigente al 31 de diciembre del 2011. La tabla considera la calificación obtenida con las funciones discriminantes y los hiperplanos óptimos de separación para dichas entidades. La calificación y el porcentaje de provisión se determinan atendiendo el procedimiento establecido por el Organismo de Control.

Comparando las calificaciones obtenidas por las metodologías propuestas, se observa que el GADM de Bolívar presenta una calificación real de E, con un porcentaje de provisión del 20% generando un monto de provisión de 163 mil dólares, utilizando la calificación asignada por las funciones discriminantes, el GADM de Bolívar es calificado como B, con un porcentaje de provisión del 1% generando un

monto de provisión de 8 mil dólares y aplicando la calificación asignada por los hiperplanos óptimos de separación, el GADM de Bolívar escalificado como C, con un porcentaje de provisión del 10% generando un monto de provisión de 81 mil dólares.

Tabla 13: Calificación de cartera Ejemplo simple de tabla.

		Bolívar
<b>Real</b>	Calificación	E
	Calificación de cartera	C1
	% de Provisión	20%
	Cartera (miles \$)	819
	Provisión (miles \$)	163
<b>Análisis Discriminante</b>	Calificación	B
	Calificación de cartera	A3
	% de Provisión	1%
	Cartera (miles \$)	819
	Provisión (miles \$)	8
<b>Máquinas de vectores de soporte</b>	Calificación	C
	Calificación de cartera	B2
	% de Provisión	10%
	Cartera (miles \$)	819
	Provisión (miles \$)	81

## 5. CONCLUSIONES Y RECOMENDACIONES

Cabe señalar que el estudio tiene como objetivo principal contar con una herramienta de pronóstico de calificación de riesgo, evaluando su desempeño para la calificación de riesgo de los Gobiernos Autónomos Descentralizados Municipales en el año 2010.

Al comparar la capacidad predictiva del análisis discriminante y las máquinas de vectores de soporte, se determina que las tasas de clasificación errónea son muy similares y bajas en la mayoría de los escenarios estudiados, esto demuestra que las dos metodologías son apropiadas para la valoración de la calificación de riesgo de crédito.

La conclusión general que se extrae es que, tanto los modelos de análisis discriminantes como las máquinas de vectores de soporte son adecuados para el estudio y predicción de la calificación de riesgo de los Gobiernos Autónomos Descentralizados Municipales, consiguiendo mediante las dos metodologías una elevada eficacia predictiva.

La predicción de las calificaciones realizadas por los modelos propuestos, presentan especial importancia en la medición de la cartera crediticia y el monto de provisiones para el Banco del Estado, puesto que éste realiza el análisis para contar con el capital suficiente para afrontar posibles pérdidas, ya que la estimación de las provisiones sirve para realizar el presupuesto de la entidad.

Trabajos futuros deben enfocarse en comparar el desempeño de las metodologías estudiadas considerando otro tipo de escenarios en los cuales se pueden estudiar aspectos como: mayor número de grupos a clasificar, distintos Gobiernos Autónomos Descentralizados (Provinciales y Parroquiales) de estudio con sus respectivas variables explicativas y medidas de desempeño diferentes a la tasa de clasificación errónea.

En este proyecto las máquinas de vectores de soporte obtuvieron resultados satisfactorios en términos de predicción de clasificación, para nuevas aplicaciones se debe responder la pregunta de cómo seleccionar adecuadamente los parámetros del núcleo con miras a mejorar los resultados obtenidos, ya que la selección de los parámetros tiene un efecto significativo sobre el clasificador propuesto.

Finalmente y considerando que el organismo de control, en este caso la Superintendencia de Bancos y Seguros, no posee una metodología específica para este tipo de clientes, se recomienda el análisis y discusión de la metodología propuesta como modelo de medición de la calificación de riesgo de crédito para los Gobiernos Autónomos Descentralizados Municipales.

## REFERENCIAS

- [1] E. Arriaga, *Finanzas públicas en México*. México, Instituto Politécnico Nacional, 2001.
- [2] A. Barreiros, *Introducción General: Gerencia financiera*. Ecuador, Escuela Politécnica Nacional, 2005.
- [3] P. Barrera, Máquinas de vectores de soporte para procesado de secuencias: el kernel de Fisher en secuencias bioinformáticas. España, Universidad Carlos III de Madrid, 2002.
- [4] G. Betancourt, *Las máquinas de soporte vectorial*. Colombia, Pereira, 2005
- [5] M. Césari, *Aprendizaje automático con WEKA*. España, 2007.
- [6] C Cuadras, *Nuevos métodos de análisis multivariante*. España, Barcelona, 2001.
- [7] B. Estrada, C. Mera, *Support Vector Machine*. Colombia: Universidad Nacional de Colombia, 2006.
- [8] S. Gunn, *Support vector machines for classification and regression*. EEUU, 1977.
- [9] E. López, *Modelo mixto de Credit Scoring construido con Análisis discriminante y el Algoritmo de Kohonen*. Valoración de las componentes de riesgo según Basilea II. España, Universidad de Sevilla, 2005.
- [10] A. Mahalanobis, *Prasanta Chandra Mahalanobis*. India, 1930.
- [11] A. Mendoza, P. Gómez, *Herramientas para el pronóstico de calificación crediticia de las finanzas Públicas en México: Redes Neuronales Artificiales, Modelo Probit Ordenado y Análisis Discriminante*. México, Tecnológico de Monterrey, 2009.
- [12] D. Peña, *Análisis de datos multivariante*. España, Madrid, 2002.
- [13] Superintendencia de Bancos y Seguros. *De la gestión y administración de riesgos*. Resolución N° JB-2003-601.27
- [14] Superintendencia de Bancos y Seguros. *De la administración del riesgo de crédito*. Resolución N° JB-2003-602.28
- [15] Shawe, Taylor, *Na introduction to support vector machines*. Inglaterra, 2000.