**Uso de MVS para el pronóstico de incumplimiento del pago bancario de clientes**

RESUMEN

Las Máquinas de Vectores Soporte (SVM por sus siglas en inglés), frecuentemente han sido utilizadas para resolver problemas de reconocimiento de patrones, sin embargo, en los últimos años se las ha empleado para resolver problemas de clasificación de tráfico de red, reconocimiento de voz, categorización de texto, detección facial, detección de tarjeta defectuosa, por las ventajas que han manifestado poseer frente a estos problemas, aunque sean complejas de desarrollar.

El objetivo es resolver el problema de pronóstico de pago de los agentes de un banco mediante el uso de Máquinas de Vectores Soporte, para ello, se han hecho revisiones bibliográficas acerca de problemas de incumplimiento de pago por parte de los clientes de un banco, en las revisiones se comparó otros algoritmos que generalmente han sido utilizados para este propósito con el algoritmo de SVM, resaltando este último frente a estos algoritmos.

**Palabras Clave:** Máquinas de Vectores Soporte, Pronóstico de pago, etc

CAPÍTULO 1

# Introducción

La teoría de las Máquinas de Vectores Soporte (SVM en inglés) fue desarrollada por V. Vapnik a principios de los 80, se conoce que las SVM son un tipo de redes neuronales que se utilizan en problemas de reconocimiento de patrones, problemas no lineales de clasificación, problemas de regresión, predicción de series temporales, entre otros debido a su capacidad de generalización, su objetivo es dar solución al problema general que surge en distintos campos, donde se estudia la relación entre sesgo y varianza, el sobreajuste de datos, el control de la capacidad, etc. Este problema consiste en buscar una adecuada función que permita llevar a cabo una buena generalización que resulte de una adecuada relación entre la precisión alcanzada con un particular conjunto de entrenamiento y la capacidad del modelo.

Las SVM presentan ciertas ventajas con respecto a otras técnicas no lineales:

* Su especificación se basa en el principio de minimización del riesgo estructural que equivale a minimizar el límite superior del error de generalización del modelo.
* La estimación de los parámetros de una SVM es equivalente a la solución de un modelo de programación cuadrática con restricciones lineales.

Se han hecho revisiones de varios trabajos donde se emplean SVMs y se compara su efectividad para resolver problemas con otros algoritmos que, hasta ahora, se han usado comúnmente para resolver estos problemas con muy buenos resultados, de estos, hablaremos brevemente de los más resaltantes:

Pronóstico del pago de los agentes mediante SVM:

Existe un problema de clasificación relacionado con la capacidad de predicción del incumplimiento de las obligaciones financieras de los clientes de un banco, para resolver este problema, se emplea la información de dichos clientes, tal como su edad, estrato económico, nivel de ingresos, educación, etc. El banco debe considerar el comportamiento que ha observado de sus antiguos clientes para realizar la clasificación en base a la información de cada cliente.

Mapeo de los recursos naturales:

El mapeo de recursos naturales reviste gran importancia para la generalización acerca de las cantidades existentes, calidad o estado de los mismos y su distribución en el espacio. Uno de los métodos más utilizados para transformar los datos contenidos en estas imágenes en información temática es la clasificación digital y uno de los principales usos de esta tecnología es el mapeo de los distintos tipos de cubiertas.

## Descripción del problema

Se desea realizar el pronóstico de incumplimiento de pagos de los agentes de un banco, para ello se compararán diferentes algoritmos estadísticos, matemáticos y métodos de inteligencia artificial.

## Justificación del problema

Se ha escogido como problemática a resolver el pronóstico de incumplimiento de pagos de los clientes de un banco debido a que muchos no están al tanto de sus graves consecuencias, por lo que, al involucrarnos en cómo administrar mejor una deuda, nos será más sencillo entender:

* El significado de un pago incumplido.
* Lo que se debe hacer si se cae en un incumplimiento de pago
* Como es que omitir un pago del préstamo puede resultar un problema.
* Las consecuencias de incurrir en un incumplimiento de pago

## Objetivos

### Objetivo General

* + Resolver el problema de incumplimiento de pagos de los agentes de un banco mediante el uso de SVM.

### Objetivos Específicos

* 1. Describir todos los pasos necesarios en el proceso descrito como objetivo general.
  2. Describir que tan útil fue la implementación de la MVS para el problema planteado.

## Marco teórico

El problema de clasificación se puede plantear como xk = (x1k, x2k, x3k, …, xnk), donde este conjunto describe la información del cliente, de cada individuo obtenemos {xk, yk} donde k = 1, 2, …, N.

Para poder entender el proceso de aprendizaje podemos plantear el problema de identificación de voz, que manifiesta su relación con el perceptrón. Se debe buscar una forma de almacenar, detectar y reconocer una voz en una máquina para así poder identificarla, llevándonos al objetivo de hacer que la máquina aprenda y reciba voces, esto nos lleva al uso de minería de datos. La máquina de aprendizaje, que debe ser capaz de aprender de cada evento, se aplicará sobre los datos antes citados y así mejorará su rendimiento a partir de eventos ya realizados.

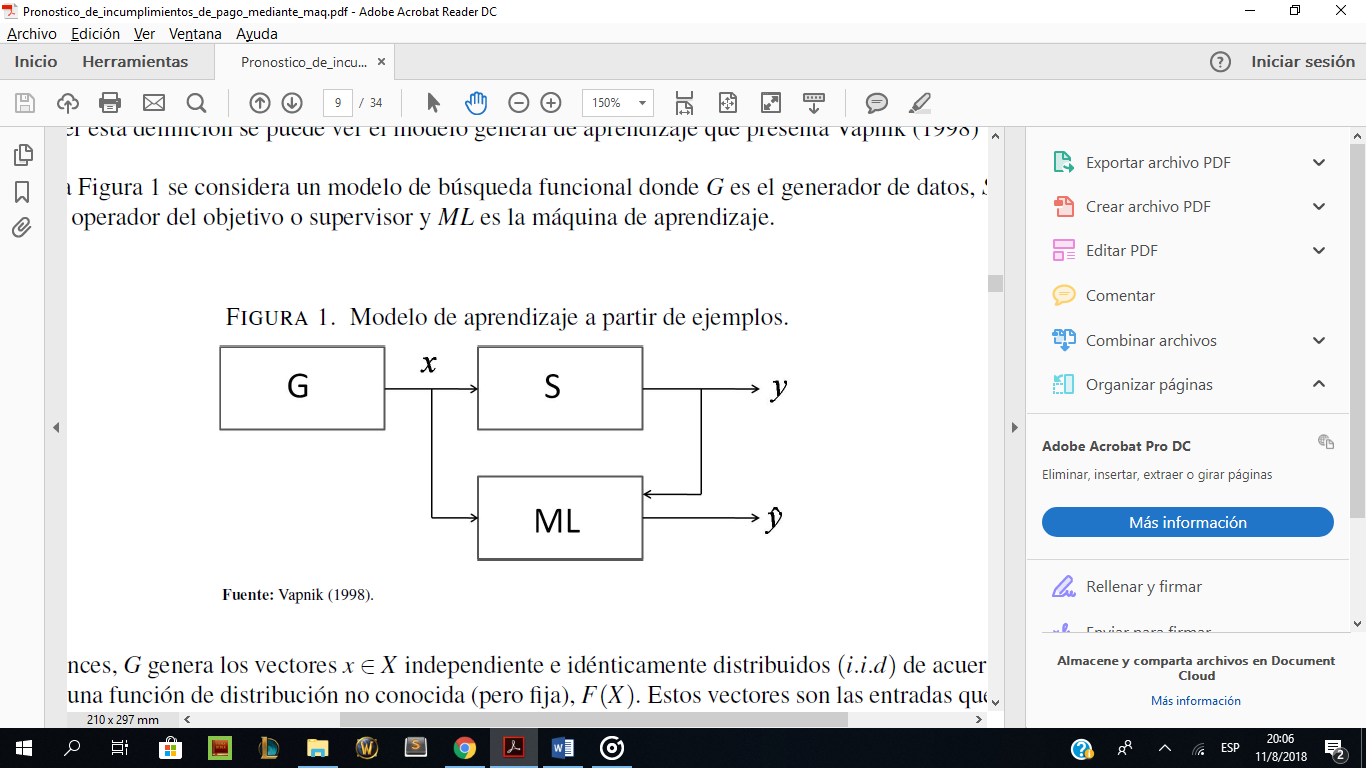


Figura 1.1 Modelo de aprendizaje a partir de ejemplos

G genera los vectores x ∈ X independiente e idénticamente distribuidos

ML observa N parejas (x1, y1), …, (xn, yn) o conjunto de entrenamiento.

S produce la salida y para el vector x de acuerdo a la distribución condicional

Por lo tanto, el proceso de aprendizaje selecciona la función más apropiada de un conjunto de funciones dadas.

Al implementar la MVS podemos toparnos con 2 casos: Caso linealmente separable y caso linealmente no separable.

Caso linealmente separable: Es el caso más sencillo, permite entender el conjunto de problemas que se desea abordar desde esta metodología. Las características de entrada pueden ser la edad del cliente y su nivel de ingresos y la salida el pago o no pago de sus obligaciones. El conjunto de entrenamiento CE = {xk, yk} | xk ∈ RN, yk ∈ {-1, 1}}k = 1, …, N contiene las características y el comportamiento de los clientes respecto a sus obligaciones.

Los puntos que se encuentren el hiperplano de separación satisfacen la siguiente relación:

xw + b = 0 (1.1)

Donde w es un vector normal y perpendicular al hiperplano y xw corresponde al producto punto entre 2 vectores.

El margen se puede expresar como:

yk(xkw + b) - 1 ≥ 0 (1.2)

Caso linealmente no separable: Con el fin de extender la metodología de SVM a casos no linealmente separables, se deben reflejar ciertas restricciones mediante la introducción de una variable positiva:

Xkw + b ≥ + 1 – ek; para yk = +1 (1.3)

Xkw + b ≥ + 1 + ek; para yk = -1 (1.4)

Se tiene la fórmula solución de un caso no lineal:

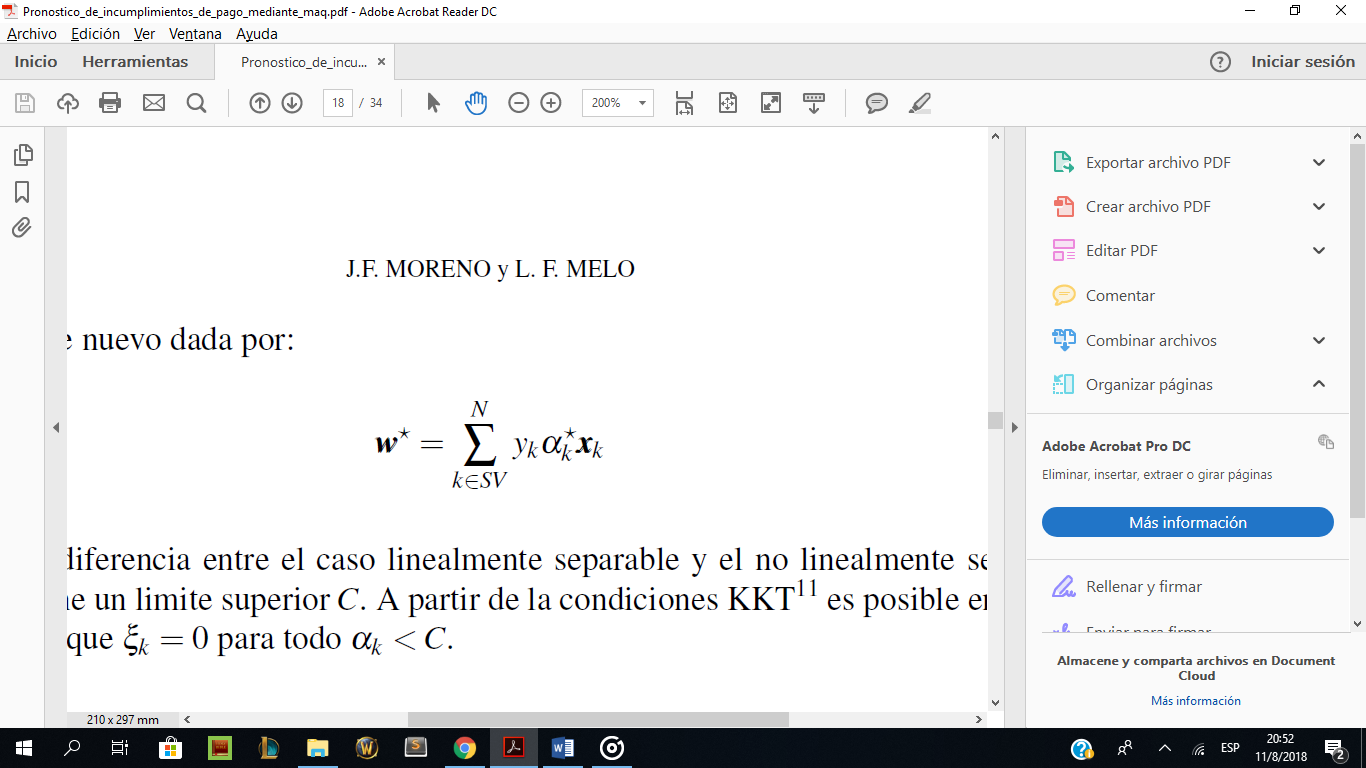


Figura 1.2 Fórmula solución MVS no lineal

**CAPÍTULO 2**

# Metodología

El problema de pronóstico de incumplimiento de pago puede realizarse mediante el uso de métodos estadísticos y matemáticos. Aquí destacan técnicas de regresión (LM), análisis discriminante (LDA), y árboles de decisión (DT), como métodos tradicionales, y como métodos de inteligencia artificial, tenemos redes neuronales (NN), algoritmos evolutivos (EA), máquinas de vectores soporte (SVM), entre otras.

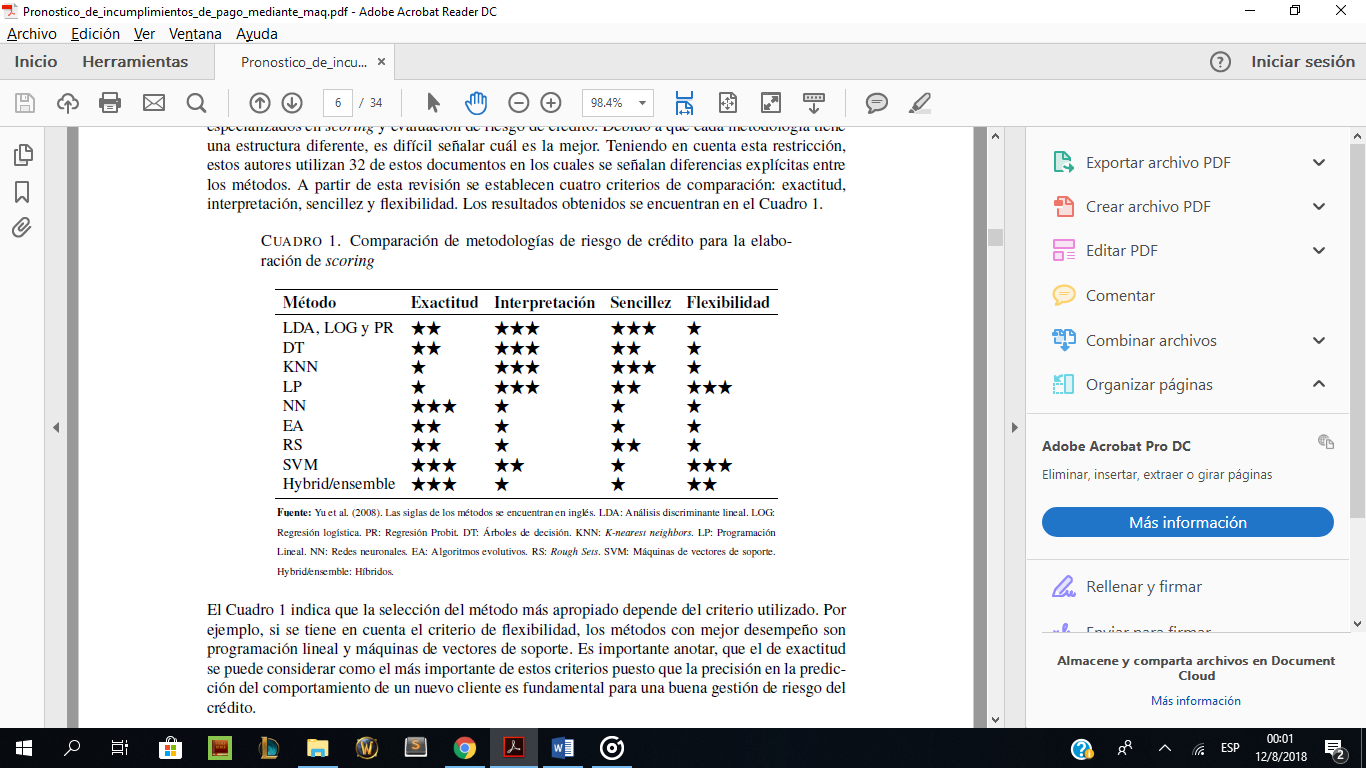


Figura 1.3 Comparación de metodologías de riesgo de crédito

El método más destacado frente a las demás es el de Máquinas de Vectores Soporte, escogido por sobre los demás, aparte, este método presenta ciertas ventajas frente a otros, las cuales describimos a continuación:

* Comparada con otras metodologías requiere menos supuestos sobre los datos de entrada.
* Su desempeño no depende del tamaño o dimensión de la muestra.
* Se resuelve mediante programación cuadrática, lo que hace de su solución única, global y óptima.
* El algoritmo se puede ajustar a problemas no lineales, lo que hace que sea un método generalizable.

Para realizar la estimación mediante SVM se llevan a cabo los siguientes pasos:

* Dividimos las bases de datos en muestras de pronósticos y muestras de entrenamiento, para ello se toma aleatoriamente el 80% de los datos como muestra de entrenamiento y el restante como muestra de pronóstico.
* Los métodos empleados para simular la MVS son:

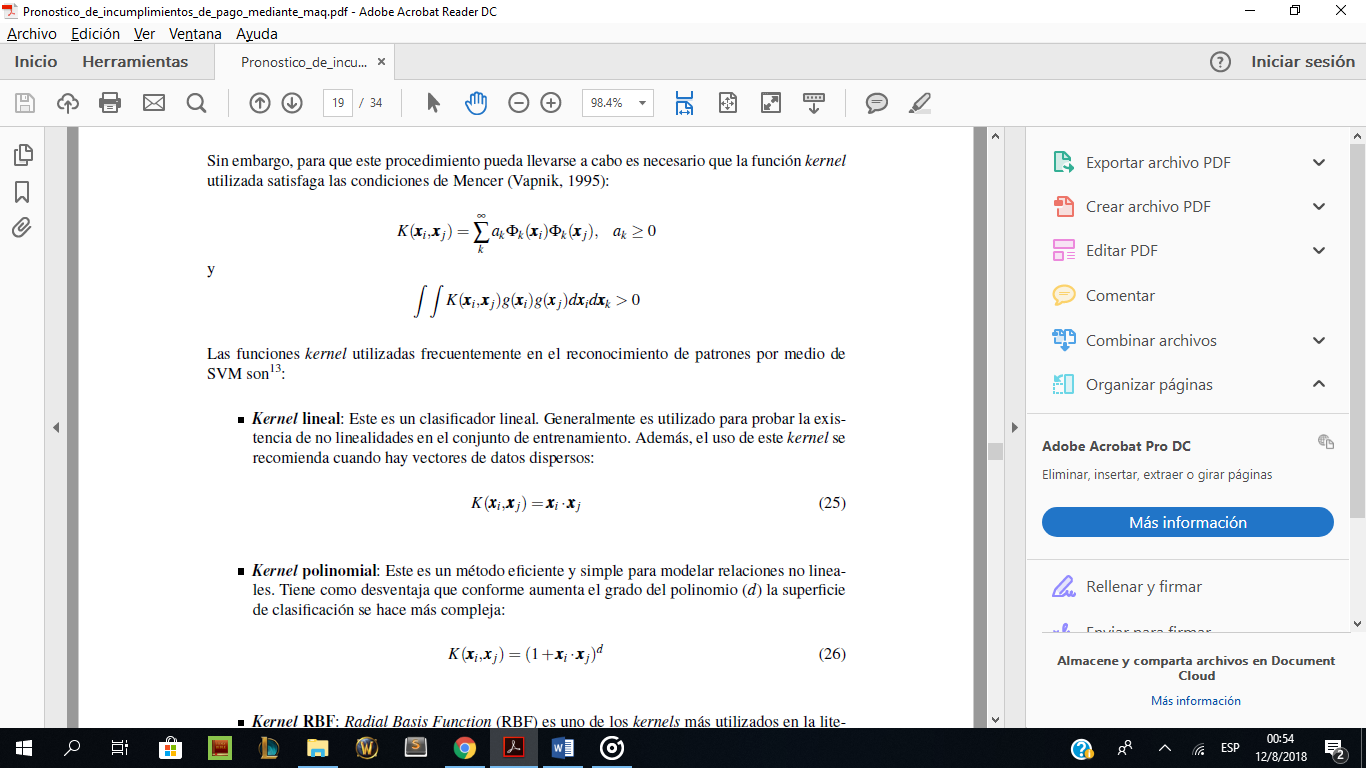
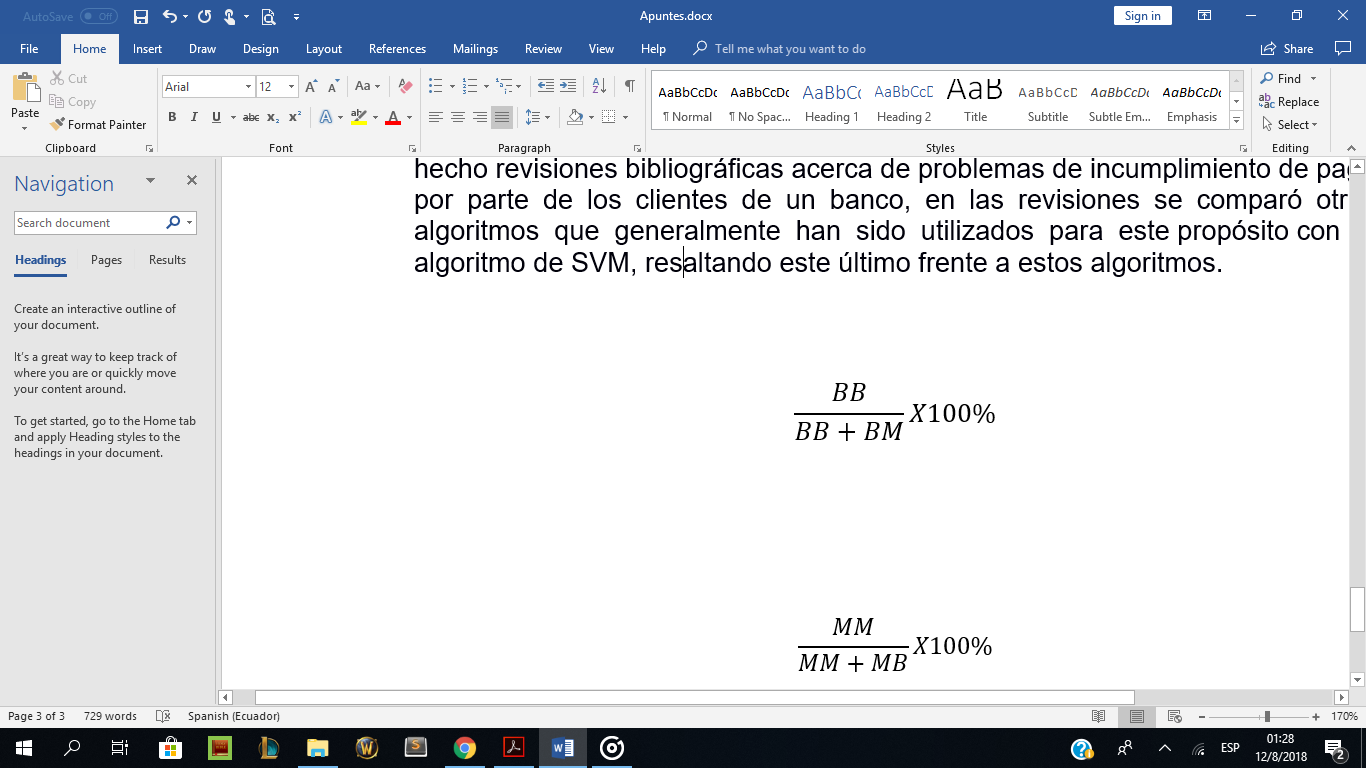


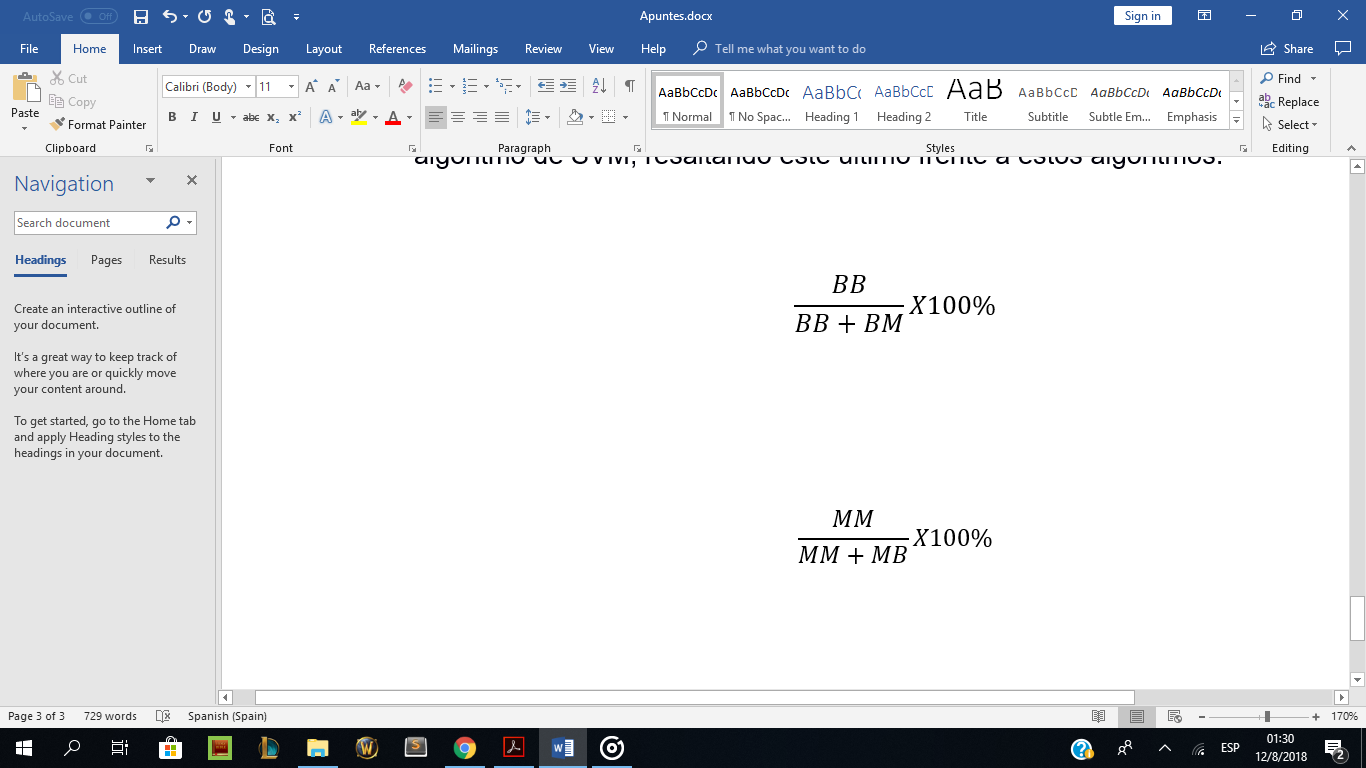
Figura 1.4 Funciones para estimar MVS

* La calibración de los parámetros de los kernels se lleva a cabo mediante búsqueda directa, seleccionando aquellos parámetros que generan el menor error cross-validation.
* Se utiliza el método numérico de minimización de funciones de Nelder y Mead para calibrar los parámetros utilizados.
* Se establece el umbral de clasificación según la probabilidad que asigna la MVS de estar en una u otra clase.
* Se estima la MVS con los parámetros seleccionados, se clasifica la muestra y se compara los valores observados. Se realiza la validación del modelo que establece el poder de pronóstico de la MVS para nuevos clientes.
* Para validar el modelo se utilizan medidas que cuantifican el desempeño del pronóstico en las 2 muestras, como:
* Para validar el modelo se utilizan medidas que cuantifican el desempeño del pronóstico en las 2 muestras:

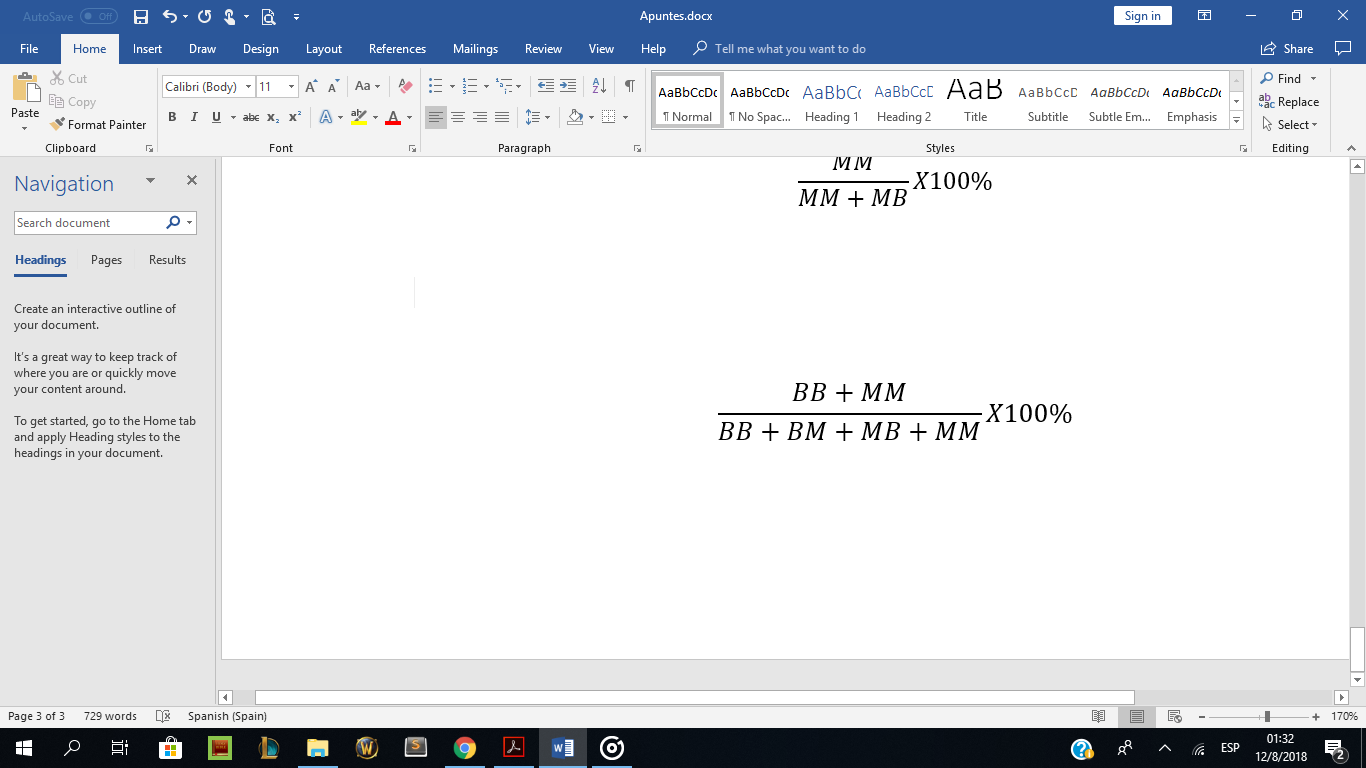
Precisión en la clasificación de buenos clientes:

 (2.1)

Precisión en la clasificación de malos clientes:

 (2.2)

Precisión en la clasificación total:

 (2.3)

Siendo BB la cantidad de buenos clientes clasificados como buenos, BM son buenos clientes clasificados como malos, MB son malos clientes clasificados como buenos y MM son malos clientes clasificados como malos.

CAPÍTULO 3

# Resultados Y ANÁLISIS

En esta sección se describe exhaustivamente los resultados de la solución obtenida, pero sin repetir los detalles ya descritos en el Capítulo 2 (Metodología). Presente los datos más representativos, si son datos estadísticos, presente los que sean significativos. Los resultados deben redactarse en pasado.

En este capítulo también se incluirá una sección de Análisis de Costos, en donde se realizará una descripción de los costos relacionados con el desarrollo de la solución o alternativa seleccionada (desde el diseño/rediseño hasta su construcción si fuera el caso del trabajo).

Es importante considerar la viabilidad económica en el sentido, que la solución planteada debe ser económicamente viable y tecnológicamente factible.

Se sabe que, producto de investigaciones pasadas acera de pronósticos de cumplimiento de pagos mediante MVS, los resultados variarán de acuerdo a la base de datos que se use y que aquellos modelos que presenten un mejor rendimiento dentro de la muestra de entrenamiento serán los de mayor precisión en la muestra de pronóstico.

Por criterio de comparación, se conoce que clasificar mal a los buenos clientes no es equivalente a clasificar mal a los malos clientes, ya que este último error se considera más grave que el primero, debido a las pérdidas que estos representan para el banco.

Si tuviéramos que escoger entre varios modelos de MVS, deberíamos escoger en base al criterio que utilicemos.

CAPÍTULO 4

# Conclusiones Y RECOMENDACIONES

## Conclusiones

Se concluye que la metodología MVS es superior a otros modelos generalmente utilizado en este tipo de análisis (como por ejemplo regresión logística y análisis discriminante lineal) en cuestión de desempeño.

## Recomendaciones

Se recomienda a futuro utilizar algoritmos híbridos que ayuden a mejorar la estimación de los parámetros de estos modelos.

Se puede recomendar los algoritmos de MVS para resolver problemas de pronósticos de series económicas de interés.

BIBLIOGRAFÍA

[1] BELLOTTI, T., Y J. CROOK (2009): “Support vector machines for credit scoring and discovery of significant features.,” Expert Systems with Applications: An international Journal, 36(2), 3302–3308.

[2] V. Fernandez. “Wavelet- and SVM-based forecasts: An analysis of the U.S. metal and materials

manufacturing industry”. Resources Policy. Vol 32 No 1-2, pp. 80-89. March-June 2007.

[3] BURGES, C. J. (1998): “A Tutorial on Support Vector Machines for Pattern Recognition,” Data Mining and Knowledge Discovery, 2, 121–167.

[4] VELÁSQUEZ, J. D., Y. OLAYA, Y C. J. FRANCO (2010): “Predicción de series temporales usando máquinas de vectores de soporte,” Ingeniare. Revista chilena de ingeniería, 18, 64 – 75.

[5] GA Betancourt - Scientia et technica, 2005 - revistas.utp.edu.co

[6] LG Abril - Asociación Científica Europea de Economía Aplicada, 2003 - researchgate.net

[7] [JD Velásquez](https://scholar.google.com/citations?user=kS3VWpEAAAAJ&hl=es&oi=sra), Y Olaya, [CJ Franco](https://scholar.google.com/citations?user=-crU08EAAAAJ&hl=es&oi=sra) - Ingeniare. Revista chilena de …, 2010 - scielo.conicyt.cl

[8] JF Moreno-Gutiérrez… - Borradores de Economía; …, 2011 - researchgate.net

[9] YU, L., W. YUE, S. WANG, Y K. K. LAI (2010): “Support vector machine based multiagent ensemble learning for credit risk evaluation,” Expert Systems with Applications: An international Journal, 37(2), 1351–1360.

[10] JI MARÍN, PA Muñoz, FJ IBARGÜEN - Scientia et technica, 2006 - revistas.utp.edu.co

[11] [JC Riquelme Santos](https://scholar.google.com/citations?user=YWzbLeEAAAAJ&hl=es&oi=sra), R Ruiz, [K Gilbert](https://scholar.google.com/citations?user=w_04Fl8ILQoC&hl=es&oi=sra) - Inteligencia artificial: Revista …, 2006 - idus.us.es