

Tipo de artículo: Artículo original  
Temática: Inteligencia Artificial  
Recibido: 21/02/19 | Aceptado: 10/07/19 | Publicado: 25/09/19

## Capacidad predictiva de las Máquinas de Soporte Vectorial. Una aplicación en la planificación financiera.

### *Predictive power of the Support Vector Machine. An application to the financial planning*

Ing. Victor Manuel Cuevas Soto <sup>1\*</sup>, Lic. Susana Alvarez Iriarte <sup>2</sup>, Lic. Mileydis Azcona Romero <sup>3</sup>, Msc. A. Israel Rodríguez Rogert<sup>4</sup>

<sup>1</sup> Empresa de Mantenimiento a Centrales Eléctricas. 100 y 69, Marianao, La Habana, Cuba. [victor@oc.emce.une.cu](mailto:victor@oc.emce.une.cu)

<sup>2</sup> Empresa de Mantenimiento a Centrales Eléctricas. 100 y 69, Marianao, La Habana, Cuba. [iriarte@oc.emce.une.cu](mailto:iriarte@oc.emce.une.cu)

<sup>3</sup> Empresa de Mantenimiento a Centrales Eléctricas. 100 y 69, Marianao, La Habana, Cuba. [mileydis@oc.emce.une.cu](mailto:mileydis@oc.emce.une.cu)

<sup>4</sup> Oficina Central de la Unión Eléctrica. Ave Salvador Allende # 666, La Habana, Cuba. [aisrael@oc.une.cu](mailto:aisrael@oc.une.cu)

\* Autor para correspondencia: [victor@oc.emce.une.cu](mailto:victor@oc.emce.une.cu)

---

#### Resumen

En los últimos años el uso de modelos no lineales para predicción y clasificación a través de algoritmos y técnicas computacionales de avanzada (Machine Learning), han surgido como alternativa eficiente para modelar procesos económicos. El presente trabajo fue desarrollado con el objetivo de construir un modelo para pronosticar la proyección de los ingresos en la Empresa de Mantenimiento a Centrales Eléctricas (EMCE). La modelación fue realizada a través del algoritmo Support Vector Machine (SVM) obteniéndose 95.8 % de aciertos con función kernel RBF. Se propone una herramienta gráfica (graph of prognostic projections) para la interpretación de pronóstico sobre proyecciones de variables de naturaleza económica.

**Palabras clave:** pronóstico, inteligencia artificial, series de tiempo, máquinas de soporte vectorial, aprendizaje automático.

#### Abstract

*In recent years, the use of non-linear models for prediction and classification through algorithms and advanced computational techniques (Machine Learning) have emerged as an efficient alternative for modeling economic processes. The present work was developed with the objective of building a model to forecast the projection of revenues in the Maintenance Company to Power Plants (EMCE). The modeling was carried out through the Support Vector Machine (SVM) algorithm, obtaining 95.8% of hits with the RBF kernel function. A graphical tool (graph of prognostic projections) is proposed for the interpretation of forecasts about projections of variables of an economic nature.*

**Keywords:** *Forecasting, Artificial Intelligence, Time Series, Support Vector Machine, Machine Learning*

---

## Introducción

Las herramientas de pronóstico son una necesidad ante los contextos cada vez más cambiantes y competitivos a los que se enfrentan las organizaciones; en ello existe un gran potencial para gestionar los riesgos inducidos por la incertidumbre en los escenarios en los que se convive. Como consecuencia, una aspiración real de muchas empresas es mejorar el proceso de toma de decisión basado en información de calidad, exacta, oportuna y relevante.

La Inteligencia Artificial (IA) trata de conseguir que los ordenadores simulen la inteligencia humana; es una rama de las ciencias computacionales encargada de estudiar modelos de cómputo capaces de realizar actividades propias de los seres humanos, en base a dos de sus características primordiales: el razonamiento y la conducta. La inteligencia artificial es el “*esfuerzo de desarrollar sistemas basados en computadora que se pueden comportar como los humanos con la capacidad de aprender lenguajes naturales, efectuar tareas físicas coordinadas, utilizar un aparato perceptor y de emular la experiencia y la toma de decisiones*” (Laudon K, 2004). En el momento actual la IA se aplica a numerosas actividades humanas, y como líneas de investigación más explotadas destacan el razonamiento lógico, la traducción automática y comprensión del lenguaje natural, la robótica, la visión artificial y, especialmente, las técnicas de aprendizaje y de ingeniería del conocimiento. Estas dos últimas ramas son las más directamente aplicables al campo de las finanzas pues, desde el punto de vista de gestión organizacional, lo que interesa es construir sistemas que incorporen conocimiento y, de esta manera, sirvan de ayuda a los procesos de toma de decisiones.

En los últimos años, en el ámbito empresarial gracias a la IA, se ha pasado de las herramientas que facilitan o simulan el procesamiento manual de la información a robustos algoritmos que sirven como soporte para la toma de decisiones adecuadas, eficaces y oportunas (Aboueldahab, 2011; Asadi, 2012; Dai, 2012); siendo los más notables los *sistemas expertos* y las *redes neuronales artificiales*. Desde hace tiempo los economistas han estudiado e intentado comprender los movimientos de los índices financieros derivados de series económicas. Tanto investigadores como profesionales reconocen que el pronóstico en contexto económico-financiero es una tarea desafiante ya que las series de datos siguen dinámicas complejas, no lineales, caóticas, ruidosas y cambiantes en el tiempo, haciéndolas volátiles, no estacionarias, multifractales, de memoria larga, inestables y de alto riesgo (H. Jiang, 2010; C.Cai, 2012). En las empresas, los estados financieros son objetos de interés desde la óptica predictiva. En especial la modelación de la

disponibilidad real de efectivo es una actividad vital para garantizar la permanencia y crecimiento económico de una organización. Para predecir los flujos de efectivo de una forma más acertada, estos se deben desagregar y se debe tener en cuenta el ingreso como variable básica para proyectar la disponibilidad.

## **Materiales y métodos o Metodología computacional**

Una característica distintiva en la mayoría de los procesos de las organizaciones, es la generación de eventos sometidos a influencias de factores no predecibles, en los que sus atributos toman forma de series de tiempo. En la predicción de series de tiempo financieras, los modelos lineales tienen una considerable ventaja sobre el resto por la fácil interpretación de sus elementos. A principio de la década de 1980, los modelos empleados para modelar la media de un activo eran los ARIMA o ARIMAX, pero entrada la década de los 90 con los modelos ARCH y Garch, se le da más importancia a la volatilidad que a la media de los datos (Alonso, 2009).

El modelo autorregresivo integrado de media móvil ARIMA introducido por Box y Jenkin ha sido uno de los enfoques más populares utilizado en la predicción (Pai, 2005). En este enfoque se parte del supuesto de que el valor futuro de la variable es una combinación lineal de valores y errores pasados; sin embargo, cualquier grupo de características no lineales lo limita. En 2005 se sugiere el uso de un nuevo modelo denominado Arfima (Bhardwaj, 2006). Las estimaciones usan una variedad de procedimientos estándares que proporcionan mejores predicciones que los modelos de la familia ARIMA-ARIMAX y Garch.

En estudios realizados sobre la volatilidad diaria del índice S&P-100 de Estados Unidos (Liu y Hung, 2010), se compararon los modelos Garch-N, Garch-t, Garch-HT y Garch-SGT con modelos del tipo asimétrico como GJR-Garch y Egarch, arribando a la conclusión de que estos últimos mejoran la volatilidad del pronóstico y demuestran que el componente asimétrico es más importante que la especificación de la distribución cuando hay presencia de asimetría, leptocurtosis y efectos Leverage.

## **Métodos no lineales de predicción. Machine Learning en el ámbito contable-financiero**

El análisis de datos económicos en la era actual está caracterizado por el creciente uso de modelos no lineales de predicción. La teoría econométrica destaca diferentes fuentes potenciales para la presencia de no linealidades y ciclos en los activos financieros. Desde el enfoque del aprendizaje automático, en cada momento  $t$  de una data histórica, es posible aplicar modelos de predicción; y en ello la IA y sus aportes desde el Machine Learning (ML) están liderando desde hace unas pocas décadas.

## Redes Neuronales Artificiales

En los últimos años se reporta un incremento del uso de las Redes Neuronales Artificiales (RNA) en la predicción de series financieras, ya que parece ser muy eficaz y presenta multitudes de variantes. Una RNA es un modelo de procesamiento de información que es inspirado en el funcionamiento del sistema nervioso biológico. Es un procesador distribuido en paralelo de forma masiva que tiene una tendencia natural para almacenar conocimiento de forma experimental y lo hace disponible para su uso. La ventaja de usar redes neuronales está en el hecho que se pueden separar regiones no lineales de decisión tan complicadas como se deseen dependiendo del número de neuronas y capas.

(Zemke, 1999) utilizó cuatro técnicas del ML para analizar la Bolsa de valores de Varsovia. La tarea consistió en predecir el valor del índice a través de decisiones binarias; es decir pronosticar la proyección (aumento o disminución respecto al valor actual) del índice WIG en la semana siguiente. Las técnicas utilizadas fueron: redes neuronales artificiales, clasificador bayesiano, K-nearest neighbor y K-nearest neighbor prediction scrutinized. Se concluyó que con las RNA se obtuvieron las mejores predicciones debido a que la bolsa de valores está dominada por la no linealidad en los datos. En el mismo año, Lee y Jo desarrollaron un intérprete gráfico para predecir el mercado, al que llamaron Candlestick Chart Analysis Expert System (Lee y Jo, 1999). Los resultados experimentales obtenidos revelaron que el modelo tenía un porcentaje de aciertos promedio de 72%, lo cual ayuda a los inversores a obtener mayores beneficios de su inversión en acciones. Guresen, Kayakutlu y Daim observaron que en la mayoría de los casos los modelos de redes neuronales permiten obtener mejores resultados que otros métodos (Guresen, 2011). Los modelos que se compararon en su trabajo son multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) y un modelo híbrido de redes neuronales y Garch, con los valores diarios del índice Nasdaq desde el 7 de octubre de 2008 hasta el 26 de junio de 2009. El rendimiento de los modelos se evaluó a través del error cuadrado medio y la desviación absoluta media, obteniéndose que el clásico modelo MLP supera a los demás al obtener mejores resultados. Existe una gran cantidad de estudios que incluyen modelos de redes neuronales artificiales, sin embargo, las RNA tienen limitaciones en el aprendizaje de patrones debido a que los datos de las series financieras por lo general tienen dimensionalidad compleja, gran ruido y en algunos casos volatilidad considerable (Kim y Han, 2000).

## Máquinas de Soporte Vectorial

Las Máquinas de Vectores Soporte (*SVM*) es una de las técnicas más poderosas del *aprendizaje automático*, que a pesar de su sencillez ha demostrado ser un algoritmo robusto y que generaliza bien en problemas de la vida real (Gala

García, 2013). Se considera una alternativa eficiente ante las limitaciones de las *RNA* frente a la dimensionalidad y ruido en los datos. Hay estudios que reportan que las *SVM* superan en precisión a los modelos autorregresivos de medias móviles, las redes neuronales artificiales y los sistemas adaptativos de inferencia neuro-difusos (Jaramillo, 2015). Fueron desarrolladas por Vapnik y sus colaboradores en el marco de la Teoría de Aprendizaje Estadístico (*SLT*) (Vapnik et al, 1999). Usa modelos lineales para implementar clases con separaciones no lineales al transformar el espacio de entrada en un nuevo espacio. Han sido estudiadas intensamente en los últimos años y aplicadas exitosamente en gran variedad de temas como: estimación de densidades probabilísticas, y la predicción de series de tiempo (Cuevas, 2010).

En principio las *SVM* están indicadas para ser usadas como clasificadores binarios. La tarea de clasificación está dividida en dos etapas: la *fase de aprendizaje automático* y la *fase de reconocimiento*. En la primera se selecciona el *conjunto de entrenamiento*, se extraen los atributos y características del espacio de entrada y se entrena el clasificador. El entrenamiento da como resultado un conjunto de parámetros  $w$  que define al clasificador y un hiperplano de separación óptimo (*HSO*) de margen máximo que representa las regiones de clasificación. En la *fase de reconocimiento* el modelo del clasificador entrenado, asigna a los datos que se le presenta una de las clases según la región de clasificación en la que hayan sido mapeado los nuevos datos. El aprendizaje según (Farías, 2011), se logra mediante la búsqueda de alguna dependencia funcional entre un conjunto de vectores con los datos de entrada y salida, permitiendo así encontrar un espacio lo más amplio posible con lo cual se puedan separar los datos pertenecientes a una clase u otra. La esencia de las *SVM* puede ser entendida sin el uso de fórmulas, para lo cual es necesario el conocimiento de cuatro conceptos básicos: el hiperplano de separación, el hiperplano óptimo, el margen suave y la función *kernel* o núcleo.

#### *SVM lineales. Hiperplano de separación óptimo y margen máximo*

Las *SVM* poseen una gran capacidad para aprender a partir de un conjunto de  $N$  muestras experimentales denominado *conjunto de entrenamiento*:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

Donde cada muestra  $(x_i, y_i)$  para  $i=1 \dots N$  está formado por un vector de  $n$  características  $x_i \in R^n$  y a una etiqueta  $y_i \in R$  que indica la clase  $\{\pm 1\}$  a la que pertenece cada muestra. El objetivo es encontrar la función  $f: R^n \rightarrow \{\pm 1\}$  que separe los datos en dos clases y que clasifique correctamente nuevas muestras perteneciente al *conjunto de pruebas*.

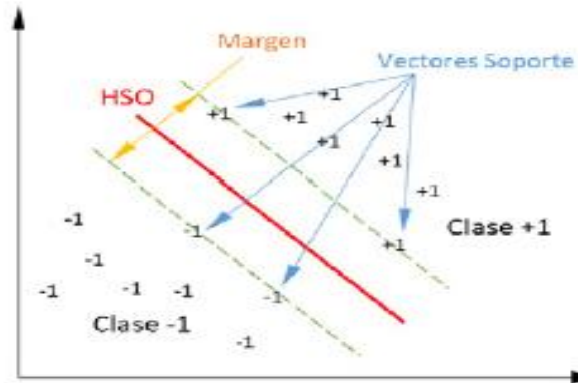


Figura 1. Hiperplano de separación óptimo desarrollado por las SVM. Imagen tomada de (Gonzales, 2017)

En un problema linealmente separable existen muchos hiperplanos que pueden clasificar los datos, pero las SVM no hayan uno cualquiera sino el único que maximiza la distancia entre el y el dato más cercano de cada clase (Figura 1). El *hiperplano de separación óptimo (HSO)* está definido por el margen máximo de separación entre las dos clases. Tomando como referencia la notación de la figura 1, existen dos hiperplanos que definen las muestras a ambos lados de cada clase:  $w \cdot x + b = +1$  y  $w \cdot x + b = -1$  que son paralelos al  $HSO$   $w \cdot x + b = 0$ . El margen máximo está dado por la distancia entre los hiperplanos paralelos y el  $HSO$ , cuyo resultado geométrico equivale a  $2/\|w\|$ . El vector de pesos  $w$  contiene la ponderación de cada atributo, indicando qué tanto aportan en el proceso de clasificación.

Hallar el mejor hiperplano de separación es un clásico problema de maximización con restricciones lineales, quedando como

$$\min \frac{\|w\|^2}{2} \text{ sujeto a : } y_i [(w \cdot x_i) + b] \geq 1$$

pudiendo ser resuelto aplicando los multiplicadores de Lagrange:

$$L(w, b, \alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^N \alpha_i [y_i [(w \cdot x_i) + b] - 1]$$

La búsqueda de la solución es preferible en el espacio dual buscando que se dependa únicamente del producto escalar de los patrones de entrada, lo cual simplifica los cálculos:

$$\begin{aligned} \min : L_d(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \cdot \mathbf{x}_j) \\ \text{sujeto a : } \sum_{i=1}^N \alpha_i y_i &= 0 \text{ y } \alpha_i \geq 0, i \in \{1, \dots, N\} \end{aligned}$$

Esta formulación del problema satisface las condiciones de Karush-Kuhn-Tucker (*KKT*) y por tanto se tienen las condiciones necesarias y suficientes para que un valor extremo exista (H. Mora, 2001). De la expresión en forma matricial

$$\min: L_d(\alpha) = \frac{1}{2} \alpha^T H \alpha - \mathbf{f}^T \alpha \text{ s.a.: } \mathbf{y}^T \alpha = 0, \alpha \geq 0$$

donde se utiliza el vector unitario  $\mathbf{f} = [1 \dots 1]^T$  y mediante cualquier método de optimización se halla el vector de multiplicadores  $\alpha_0 = (\alpha^0_1, \alpha^0_2, \alpha^0_3 \dots \alpha^0_N)$  para finalmente determinar el vector normal  $w_0$  y el bias  $b_0$  del *HSO*

$$\mathbf{w}_0 = \sum_{i=1}^N y_i \alpha^0_i \mathbf{x}_i \text{ y } b_0 = -\frac{1}{2} \mathbf{w}_0 \cdot [\mathbf{x}_r + \mathbf{x}_s]$$

esto indica que  $w_0$  se puede expresar como combinación lineal de  $N$  vectores de entrada. De las condiciones *KKT* se desprende que gran parte de los multiplicadores de Lagrange  $\alpha_0$  son cero, por lo que una menor cantidad de vectores del conjunto de entrada  $N_{sv}$  participan de la combinación lineal que origina a  $w_0$ : los vectores de soporte. En la expresión anterior  $\mathbf{X}_r$  y  $\mathbf{X}_s$  son un par de vectores soporte, uno de cada clase. La expresión final del clasificador buscado finalmente sería

$$f(x) = \text{sign}(\mathbf{w}_0 \cdot \mathbf{x} + b_0) = \text{sign} \left( \sum_{i \in SV} y_i \alpha^0_i (\mathbf{x}_i \cdot \mathbf{x}) + b_0 \right)$$

donde el signo resultante indicará la clase a la cual pertenece un dato determinado. Nótese que la sumatoria no se realiza sobre todos los puntos de entrenamiento  $N$ , sino sobre los que son vectores soporte (*SV*) siendo el número de estos por lo general mucho menor.

### *SVM no lineales con Kernel*

La mayoría de los eventos reales no son separables linealmente por lo que se dificulta la definición del *HSO*. La figura 2 muestra un conjunto de datos donde no pueden ser separados linealmente por un hiperplano en  $R^n$ , pero si en un espacio de mayor dimensión  $R^h$  (Statnikov et al., 2011).

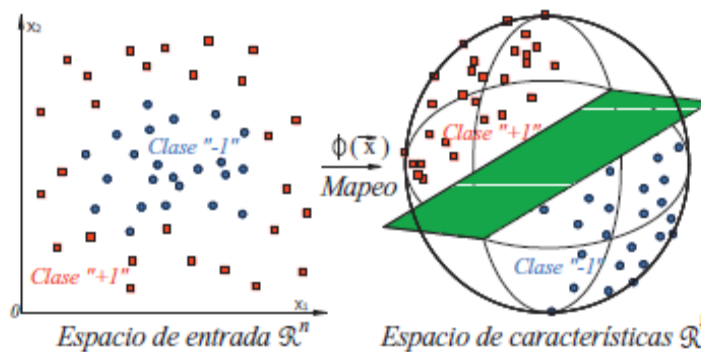


Figura 2. Mapeo de datos a una mayor dimensión y separación lineal de las clases en el nuevo espacio

Cuando no existe una apropiada superficie lineal de decisión en el espacio original de los datos, se considera un mapeo del vector de entrada en un espacio de mayor dimensión  $R^h$  llamado espacio de características, que está dotado de producto escalar. Eligiendo la transformación adecuada  $R^n \rightarrow R^h$  se realiza el mapeo y se busca el HSO siguiendo la metodología del apartado anterior que será lineal en  $R^h$ , pero representa un espacio no lineal en  $R^n$ . Este tipo de proyección hacia un espacio de característica es a través de las funciones denominadas *kernels*:  $K(x_i, x_k) = \phi(x_i) * \phi(x_k)$ . Las funciones *kernel* permiten realizar las operaciones algebraicas en  $R^h$  de manera eficiente y sin conocer a la transformación  $\phi$ . Así, en principio cualquier técnica de análisis multivariado para datos  $x \in R^n$  que se pueda reformular en un algoritmo computacional en términos de productos escalares, se puede generalizar a los datos transformados utilizando las funciones núcleos.

Existen varias funciones *kernel* o núcleos, destacando las siguientes cuatro, consideradas básicas:

**Kernel lineal:**

$$K(x_i, x_j) = x_i^T * x_j$$

**Kernel polinomial:**

$$K(x_i, x_j) = (p + \gamma x_i^T * x_j)^d, \gamma > 0$$

**Kernel gaussiano RBF:**

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2), \gamma > 0$$



donde  $Y$ ,  $d$  y  $p$  son parámetros de los núcleos y  $*$  producto escalar. Finalmente para el caso no lineal la expresión del clasificador buscado usando un *kernel* conveniente  $K$  sería

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i \in SV} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \right)$$

En general, el núcleo *RBF* es una razonable primera elección. Este núcleo no lineal mapea las muestras a un espacio de dimensión más grande, y a diferencia del núcleo lineal puede ayudar cuando la relación entre las etiquetas de las clases y las características es no lineal (Lio, 2016).

### Método de trabajo desarrollado

La Empresa en la que se realiza la investigación es perteneciente a la Unión Eléctrica. La Empresa de Mantenimiento a Centrales Eléctricas (*EMCE*), como su nombre lo indica, su responsabilidad en el sistema electro-energético cubano (*SEN*) fundamentalmente es ofrecer mantenimiento y servicios técnicos a las *Centrales Térmicas* desplegadas en el territorio nacional. El desplazamiento continuo y poco predecible de las acciones de mantenimiento en el *SEN* producto al comportamiento fortuito de las averías, exige de la dirección de la entidad una planificación acertada de los financiamientos hacia las acciones de mantenimiento con una frecuencia semanal. La hipótesis que conduce la investigación es la siguiente: *Un pronóstico adecuado sobre la proyección que experimentarán los ingresos, proporcionará un punto de partida adecuado para el diseño de estrategias efectivas para financiar las acciones de mantenimiento en el SEN, sin poner en riesgo el pago de las obligaciones con proveedores, el apoyo de las inversiones en proceso, y la disponibilidad corriente de efectivo para abordar eventualidades.* La metodología de trabajo está inspirada en un proceso de *KDD* o *minería de datos*: *“proceso no trivial de identificación en los datos de patrones válidos, nuevos, potencialmente útiles y finalmente comprensibles”*. La minería de datos busca la extracción de conocimiento explícito, potencialmente útil que permanece oculto en grandes volúmenes de datos, con el fin de encontrar patrones o reglas y resumirlos de forma que sean fácilmente entendibles por parte de un usuario (Hand, 2001).

### Selección del conjunto de datos y pre- procesamiento

Los datos necesarios fueron obtenidos del estado de cuentas bancarias de la organización en el periodo 2016-2017. La ocurrencia de las operaciones de débitos es aleatoria, para lo cual fue necesario normalizar la serie de tiempo hacia una frecuencia semanal a través del cómputo de los ingresos acumulados al final de cada semana, arribando al arreglo

ordenado y espaciado con frecuencia semanal  $x_i$ . Para completar la etapa de pre-procesamiento se trabajó en la *especificación* de  $y_i$ : etiquetas para cada vector de  $n$  característica  $x_i \in R^n$ . Las etiquetas  $y_i \in R$  toma los valores  $\{\pm 1\}$  identificando los incrementos de los ingresos respecto a la semana anterior ( $+1$ ), en lo adelante *proyección alcista*; y las disminuciones ( $-1$ ), en lo adelante *proyección bajista (PB)*, calculándose con la siguiente expresión:

$$y_i = \text{sign}(x_i - x_{i-1})$$

Finalmente, el conjunto de entrenamiento se conformó con las etiquetas y los ingresos con retardos de hasta cuatro semanas ( $x_{t-1}, \dots, x_{t-4}, y_t$ ), tomando como supuesto fundamental que la proyección de los ingresos para un periodo  $t+1$  está en función de los valores pasados de  $t$ .

### *Diseño e implementación de la máquina de soporte vectorial*

El diseño de la SVM pasa por la elección del tipo de *kernel*, el algoritmo para resolver el problema de programación cuadrática que nos lleva a la construcción adecuada del HSO, y la heurística para determinar el ajuste óptimo de los parámetros del núcleo. Para el caso que nos ocupa se utilizará el método de entrenamiento denominado SMO (Optimización Mínima Secuencial o Sequential Minimal Optimization). Su rendimiento es muy bueno para el caso de SVM no lineales y pocos datos en el conjunto de entrenamiento (Castro, 2010). El *kernel* usado será el gaussiano o RBF siendo  $\gamma$  el parámetro a sintonizar a partir de métodos de validación cruzada (*crossvalidation*), por ser el más utilizado y con mejor resultados. Se hicieron comparaciones del desempeño con otros tipos de *kernel* como el lineal y polinomial de varios órdenes.

## **Resultados y discusión**

Los resultados obtenidos están en función de las  $n$  característica  $x_i$  del conjunto de entrenamiento y el conjunto de prueba. En la fase de aprendizaje se entrenaron cuatro SVM con *kernel* diferentes, algoritmo SMO y método de validación cruzada para sintonizar el parámetro del núcleo. En el conjunto de entrenamiento para cada uno de los modelos entrenados se incrementó la dimensionalidad desde dos hasta cuatro rezagos de la variable *input*.

### **Elección del mejor clasificador para pronóstico**

El modelo óptimo resultó con un *kernel* de tipo gaussiano para  $\gamma = 5.7$  y dimensionalidad  $R^2$ ,  $R^3$  y  $R^4$  indistintamente; es decir, el uso de dos hasta cuatro variables *delay* como input al modelo no hace la diferencia en su precisión. La elección definitiva hacía del modelo construido sobre  $R^2$  obedece a los criterios siguientes (Flores, 2015):

1. Una representación con menos atributos realza el poder predictivo del modelo de clasificación disminuyendo su

complejidad y reduciendo el riesgo de *Overfitting* (*sobreajuste*) causado por la *maldición de la dimensionalidad*.

2. Permite una mejor interpretación del clasificador, lo que es particularmente importante debido a que muchos profesionales consideran que las técnicas de *machine learning* son cajas negras y se rehúsan a emplear estos métodos debido a su complejidad.

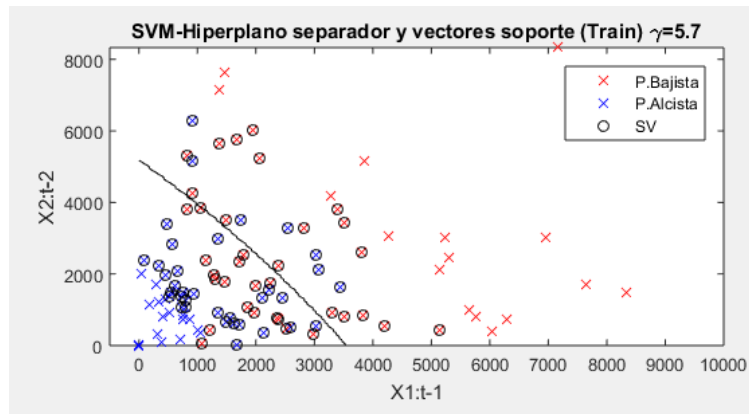


Figura 3. Etapa de entrenamiento de la SVM con kernel gaussiano y espacio de entrada  $R^2$

La figura 3 muestra el hiperplano sobre los datos de entrenamiento en un espacio determinado por los ingresos de las dos últimas semanas, resultando el error de generalización de 23.31 %. La SVM construye dos regiones de clasificación a partir del HSO, permitiendo un determinado número de muestras mal clasificadas a ambos lados del hiperplano sin comprometer el margen máximo de separación, como principal garantía de una buena generalización. Los SV que determinan la construcción del HSO se muestran circulados en la gráfica.

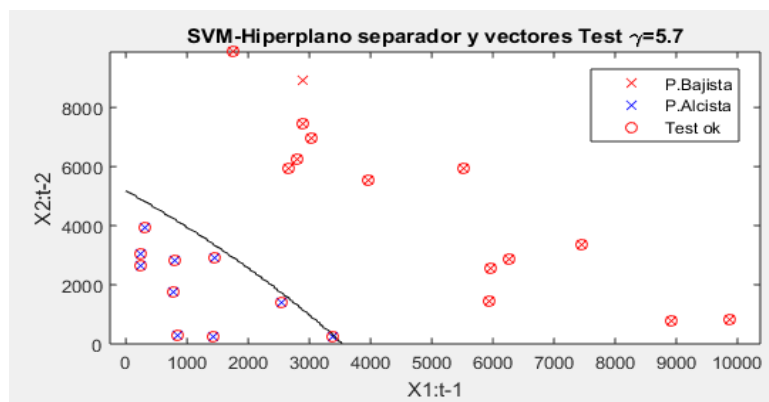


Figura 4. Etapa de prueba de la SVM con kernel gaussiano. Solo muestra no fue clasificada correctamente

En la etapa de prueba los resultados del clasificador son mejores que en la etapa de aprendizaje. Se le presentó 24 muestras nuevas al clasificador seleccionado para evaluar su desempeño y como se puede apreciar en la Figura 4, solo confunde una proyección *bajista* con *alcista*, para un error de generalización de 4.2 %. De manera que el modelo encontrado es capaz de predecir si los ingresos en el próximo periodo se incrementarán (*alcista*) o disminuirán (*bajista*) respecto al periodo anterior, con una tasa de error inferior al 5%. La tabla 1 muestra el mejor desempeño sobre el conjunto de entrenamiento de diferentes *SVM* con *kernel* del tipo *lineal* y *polinomio* en comparación con el *RBF* o *gaussiano*.

Tabla 1. Máquinas de Soporte Vectorial con diferentes tipos de kernel

| <i>Kernel</i>        | <i>Parámetros del kernel</i><br>Y | <i>variables input</i> | <i>Área curva ROC</i> | <i>Desempeño conjunto test</i><br>% |
|----------------------|-----------------------------------|------------------------|-----------------------|-------------------------------------|
| Lineal               | 0.723                             | 4                      | 0.86                  | 87.5                                |
| Polinómico (orden 2) | 0.547                             | 3                      | 0.86                  | 83.3                                |
| Polinómico (orden 3) | 0.368                             | 3                      | 0.83                  | 87.5                                |
| Gaussiano o RBF      | 5.70                              | 2,3 y 4                | 0.87                  | <b>95.8</b>                         |

Se evidencia el buen desempeño de todos siendo más notorio el construido sobre el *kernel* RBF con 95.8%. La *SVM* con núcleo *lineal* alcanzó un menor error de generalización a mayor dimensionalidad, sin embargo, con el uso de *kernel* polinómico de orden 3 se obtiene similar desempeño y la dimensionalidad es menor. Para el modelo con *kernel* gaussiano, el aumento de la dimensionalidad en la etapa de entrenamiento no aportó mejoras en el rendimiento final del modelo. La figura 5 muestra como la complejidad del *HSO* en ocasiones no es indicativa de la capacidad discriminativa de las *SVM*. El hiperplano del modelo con *kernel* polinómico de tercer orden presenta un grado de complejidad mayor dando la sensación de tener mejores propiedades de generalización; no obstante, su error de generalización (12.5 %) es superior al de la *SVM* con *kernel* RBF, la que exhibe un *HSO* visualmente menos complejo y un error de generalización menor.

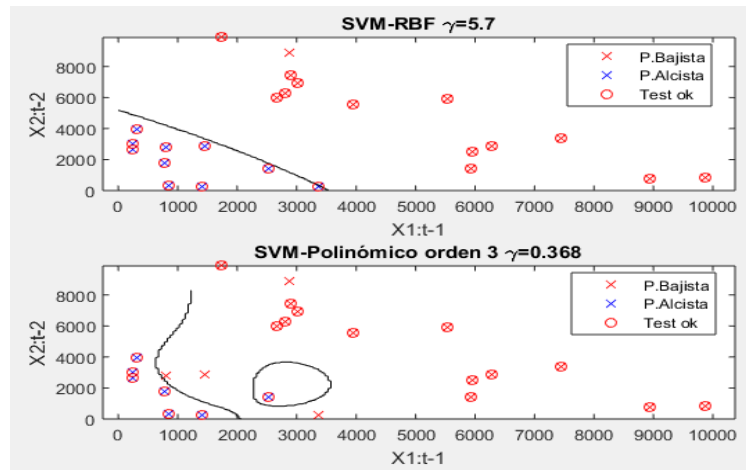


Figura 5. SVMs con los kernels no lineales de mejor desempeño

**Herramienta gráfica para visualizar resultados de pronósticos sobre la proyección de variables económicas:  
 graph of prognostic projections**

La interpretación de una proyección *alcista* o *bajista* de un indicador económico a través de gráficos como los mostrados en las figuras 3 y 4 pueden no ser de fácil comprensión para los decisores en una organización que no están familiarizados con los conceptos del *Machine Learning* y de algoritmos como las SVMs. Es por ello que se hace necesario el uso de otra herramienta gráfica que muestre de forma compacta los resultados de un modelo de pronóstico como los obtenidos en esta investigación. Las proyecciones o movimientos que experimentan los ingresos en nuestro caso de estudio, están en función de combinaciones no lineales de los valores pasados;

$$y_i = \text{sign}(f(x_{i-1}, x_{i-2}, \dots, x_{i-n})),$$

$y_i \in R$  indica la clase  $\{\pm 1\}$ , donde  $+1$  significa que los ingresos en el periodo  $x_i > x_{i-1}$  y  $-1$  verifica que  $x_i < x_{i-1}$ ; y  $M = f(x_{i-1}, \dots, x_{i-n})$  es el modelo de predicción diseñado. De manera que podemos definir un pronóstico en función del periodo anterior y el modelo de predicción de la siguiente forma:

$$P_i = y_i * x_{i-1}$$

donde  $x_{i-1}$  determinará el límite inferior para las proyecciones alcistas, o límite superior para las proyecciones bajistas. Este límite aportará información en el proceso de interpretación.

En la figura 6, se visualiza la interpretación de los conceptos anteriores en la gráfica que denominamos: *graph of prognostic projections (GPP)*. Se proyecta sobre el eje de las ordenadas el valor  $P$ , donde el signo obtenido refiere la dirección de la proyección (alcista o bajista) del pronóstico, y el  $|P|$  el límite que servirá de referencia para la

interpretación final del resultado.

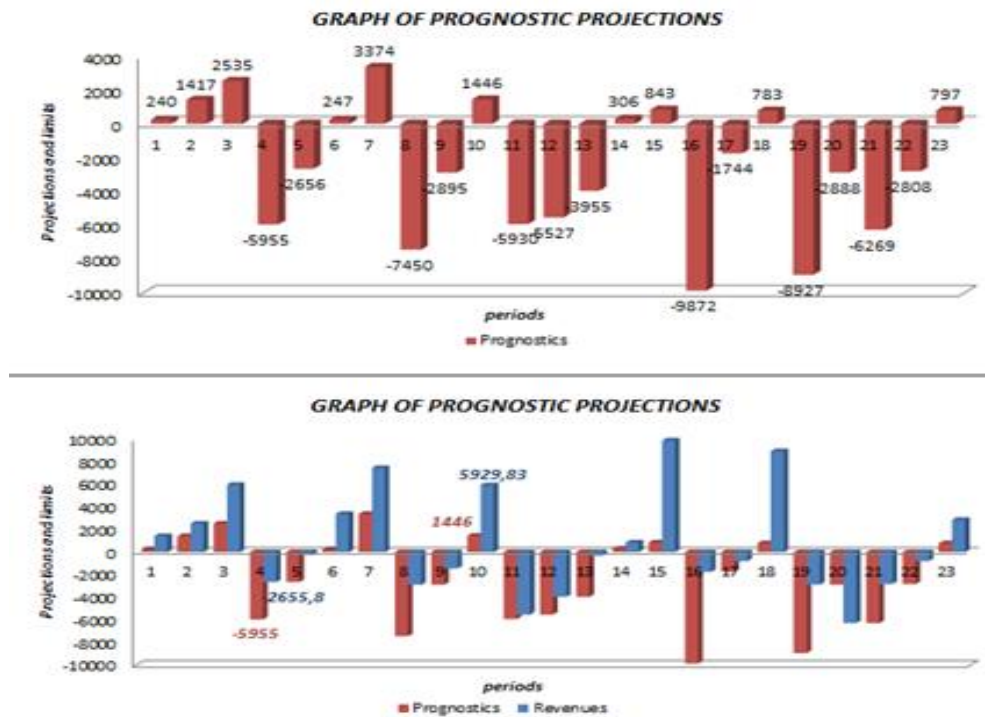


Figura 6. Gráfico para el pronóstico de proyecciones sobre variables económicas , GPP. a) Proyecciones pronósticos sobre el conjunto de prueba. b) Validación de las proyecciones pronósticos sobre el conjunto de prueba.

En a) el pronóstico resultante para la séptima semana es 3,374 millones de pesos, lo que significa que se pronostica para ese momento un incremento del ingreso respecto al periodo anterior siempre mayor a este valor. Para el periodo siguiente  $P = -7450$ , significando así una disminución de los ingresos respecto al periodo anterior y fijándose un límite máximo de ingreso de 7,450 millones de pesos. En b) se comparan los resultados de cada una de las predicciones con lo valor reales obtenidos  $R_i$  (barras azules) en cada semana que compone el conjunto de prueba. Se verifica que para cada  $P_i > 0$  (proyecciones alcistas),  $R_i > P_i$ , obteniéndose un acierto del 100%; no así en la proyección bajista ( $P_i < 0$ ) donde se cumple para todos los casos la condición  $R_i < P_i$  excepto el vigésimo periodo, para un acierto de 95.8%.

## Conclusiones

La *Inteligencia Artificial* como rama revolucionaria de la Ciencia de la Computación, constituye una herramienta

robusta para enfrentar tareas de pronóstico y clasificación en actividades tan sensibles como la planificación financiera. En ello, la variable ingreso constituye un indicador económico indispensable para modelar la disponibilidad de efectivo en una organización y proyectar sus movimientos futuros, en aras de lograr un proceso de toma de decisiones efectivo con un mínimo de riesgo.

Utilizando elementos del análisis univariado de series de tiempo, se implementa el algoritmo Máquinas de Soporte Vectorial (*SVM*) para pronosticar el movimiento que percibirán los ingresos en la Empresa de Mantenimiento a Centrales Eléctricas para el próximo periodo, y se concluye:

- Las *SVM* son adecuadas para enfrentar problemas de clasificación y predicción sobre series de tiempo de carácter financiero.
- El modelo seleccionado, construido sobre el *kernel* del tipo *gaussiano* supera a los del tipo lineal y *polinómico*.
- Para el caso de estudio abordado, se pudo constatar que la complejidad del *HSO* obtenido por algún tipo de *kernel* en el entrenamiento, no garantiza ser el de mejor desempeño frente a nuevos datos.
- Como novedad, se diseña una herramienta visual (graph of prognostic projections) para interpretar los resultados obtenidos por el algoritmo *SVM* para pronosticar la proyección de los ingresos. Su aplicación es independiente del modelo de predicción que se use sobre variables económicas.

## Referencias

- Aboueldahab, T. y Fakhreldin, M. *Prediction of Stock Market Indices using Hybrid Genetic Algorithm/ Particle Swarm Optimization with Perturbation Term*. International Conference on swarm intelligence. Cergy. France: 2011.
- Alonso, J. y Garcia, J. ¿Qué tan buenos son los patrones del IGBC para predecir su comportamiento? Universidad Icesi, 2009: pp 13-36.
- Asadi, S.; Hadavandi, E.; Mehmanpazir, F. y Nakhostin, M. (2012). Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge-Based Systems*, 2012: p. 245-58.
- Bhardwaj, G. y Swanson, N. An empirical investigation of the usefulness of Arfima models for predicting macroeconomic and financial time series. *Journal of Econometrics*, 2006, 539-78.

- C.Cai, Q.Ma, and S.Lv. Research on Support Vector regression in the stock market forecasting. *Adv. Intell. Soft Comput.*, 2012, vol.148: pp. 607-612.
- Castro Fuentes, Félix. LS-SVM Basado en optimización por enjambres de partículas para clasificación de accidentes de tránsito. Tesis de Grado. Pontificia Universidad Católica de Valparaíso, Chile, 2010.
- Cuevas Alfaro, E. A. Máquinas de Soporte Vectorial con algoritmos basado en poblaciones para el pronóstico del precio de acciones LAN en Chile. Tesis de Grado. Universidad Católica de Valparaíso, Facultad de Ingeniería, Escuela de Ingeniería Informática, Chile, 2010.
- Dai, W.; Wu, J.-Y. y Lu, C.-J. (2012). Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. *Expert Systems with Applications*, 2012, 4444-52.
- Farías Concha, M.N. Máquinas Vectoriales Híbridas para clasificar accidentes de tránsito en la región metropolitana. Tesis de Grado. Pontificia Universidad Católica de Valparaíso. Facultad de Ingeniería, Escuela de Ingeniería Informática, Chile, 2011.
- Flores, A. Selección de atributos y Support Vector Machine adaptado al problema de fuga de clientes. *Revista Ingeniería de Sistemas*, 2015, Vol. XXIX.
- Gala Garcia, Y. Algoritmos SVM para problemas sobre Big Data. Tesis de Maestría, Universidad Autónoma de Madrid, Madrid, 2013.
- Gonzales, R. Aplicación de SVM al diagnóstico clínico de la enfermedad del parkinson. *Revista Iberoamericana de Automática e informática Industrial*, 2017, 394-405.
- Guresen, E.; Kayakutlu, G. y Daim, T. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 2011, 10389-97.
- Hand, D. J., P. Smyth y H. Mannila. *Principles of data mining*. MIT Press Cambridge, 2001.
- H. Jiang and Z. Wang. GRMVm-SVR model for financial time series forecasting. *Expert Systems. Appl.*, 2010, vol.37, no. 12: pp. 7813-7818.
- H. Mora. *Optimización No Lineal y Dinámica*. Bogotá, Editorial Unibiblos, 2001.
- Jaramillo, Johana A. Tendencias Recientes en el Pronóstico de Series de Tempo Financieras usando Máquinas de Vectores de Soporte. Tesis de Maestría, Universidad Nacional de Colombia, Facultad de Minas, Colombia, 2015.
- Kim, K.-j. y Han, I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, 2000, 125-32.



- Nievas Lio, E. Aplicando Maquinas de Soporte Vectorial al análisis de pérdidas no técnicas de energía eléctrica. Trabajo Especial. Universidad Nacional de Córdoba, Argentina, 2016.
- Laudon K, & Laudon, J. Sistemas de Información Gerencial. Administración de la Empresa Digital, 2004, Octava Edición: p. 327.
- Lee, K. y Jo, G. Expert system for predicting stock market timing using a candlestick chart. Expert Systems with Applications, 1999, 357-64.
- Liu, H.-C. y Hung, J.-C. Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in Garch models. Expert Systems with Applications, 2010, 4928-34.
- Statnikov, A., Aliferis, C., Hardin, D., Guyon, I. A Gentle Introduction to Support Vector Machines in Biomedicine. Volume 1: Theory and Methods. USA, 2011: chapter 3, pp. 59-63.
- Pai, P.-F. y Lin, C.-S. A hybrid Arima and support vector machines model in stock price forecasting. Omega, Elsevier, 2005: pp. 497-505.
- Vapnik, V.N. An overview of statistical learning theory. Neural Netw, 1999, IEEE Trans., Vol. 10, n° 5: pp. 988–999.
- Zemke, S. Nonlinear index prediction. Physica A: Statistical Mechanics and its Applications, 1999, 177-83.