

Desarrollo de una metodología para el análisis y el pronóstico de acciones de la Bolsa Mexicana de Valores basada en optimización multiobjetivo

Development of a methodology for the analysis and forecasting for stocks of the Mexican Stock Exchange

Juan Andrés Martínez-Escobar*

Silvia Beatriz González Brambila**

Román Anselmo Mora-Gutiérrez***

Rubén Caudillo Félix****

(Fecha de recepción: 20 de marzo de 2020. Fecha de aceptación: 27 de mayo de 2020)

RESUMEN

En este trabajo, se presenta una nueva metodología para analizar y predecir el comportamiento de acciones de la Bolsa Mexicana de Valores basada en la concatenación sinérgica de estrategias estadísticas no paramétricas y modelos multiobjetivos de optimización. Esta metodología involucra dos fases, la primera, de filtrado, constituye un proceso automatizado para el análisis, evaluación y selección de la

* Maestría en Ciencias de la Computación,
Universidad Autónoma Metropolitana-Unidad Azcapotzalco,
al2173803820@azc.uam.mx. ORCID: 0000-0003-4116-4506

** Departamento de Sistemas,
Universidad Autónoma Metropolitana-Unidad Azcapotzalco,
sgb@azc.uam.mx. ORCID: 0000-0001-7298-4094

*** Departamento de Sistemas,
Universidad Autónoma Metropolitana-Unidad Azcapotzalco,
ing.romanmora@gmail.com. ORCID: 0000-0002-2112-7049

**** Investigador independiente
ruben_caudillo@hotmail.com. ORCID: 0000-0003-2814-3989

información necesaria y pertinente, para la caracterización del comportamiento de cada acción; posteriormente, la segunda fase de ajuste del modelo, involucra adaptar y resolver un modelo multiobjetivo para la predicción de precios de las acciones seleccionadas.

La base de datos empleada considera el comportamiento de doce acciones representativas en la Bolsa Mexicana de Valores en el periodo 2006 al 2016, el código fuente utilizado se encuentra disponible en "<http://bit.ly/396h3J1>"; los datos fueron obtenidos de una plataforma especializada sobre mercados financieros en Latinoamérica (Economatica, n.d.). Los resultados numéricos obtenidos muestran que la fase de filtrado es capaz de identificar un conjunto compacto de variables relevantes con alta influencia en el precio futuro de cada acción en particular. En la segunda fase, se emplearon los datos del 2016 como valores a predecir sobre el modelo multiobjetivo y, comparado con el modelo de regresión lineal múltiple, se observa una mejora considerable en la calidad de los datos pronosticados, haciendo que el modelo generado a partir de la segunda fase tenga una confiabilidad mayor al 95%.

Clasificación JEL: C32, C61, G12, G14.

Palabras clave: modelos multiobjetivo, modelo geométrico browniano, algoritmos genéticos, búsqueda armónica.

ABSTRACT

In this paper, a new methodology is presented to analyze and predict the behavior of stocks in the Mexican Stock Market based on the synergistic concatenation of non-parametric statistical strategies and multi-objective optimization models. This methodology involves two phases. The first (filtering) leverages an automated process for the analysis, evaluation, and selection of the necessary and relevant information for the characterization of the behavior of each action. The second (the model adjustment phase) involves adapting and solving a multi-objective model for the prediction of prices of the selected stocks.

The database used for this research includes the behavior of twelve significant stocks in the Mexican stock exchange during the 2006 to 2016 period, the source code used is available at "<http://bit.ly/396h3J1>"; the data was obtained from a specialized financial markets platform for Latin America. The numerical results show that the filtering phase can identify a compact set of relevant variables with a significant influence on the future price of each stock. In the second phase, the data from 2016 is used to predict the multi-objective model, which compared with the multiple linear regression model, provides a considerable improvement in the quality of the predicted observed data. The model generated from the second phase has a reliability greater than 95%.

JEL Classification: C32, C61, G12, G14

Keywords: Multi-objective Models, Brownian Geometric Model, Harmony Search, Genetic Algorithms.

Introducción

En este trabajo, se presenta una metodología para la organización, análisis y predicción de valores sobre el precio de cierre de acciones de la Bolsa Mexicana de Valores (BMV). Esta metodología se encuentra dividida en dos fases, en la primera se analiza, selecciona y filtra la información; en la segunda fase se emplea el conjunto de información reducido en la primera fase con el objeto de producir o generar un conjunto de coeficientes capaces de minimizar el error en los pronósticos, considerando la aleatoriedad de los movimientos. Las fases de la metodología propuesta tienen como finalidad resolver los problemas de selección de información y predicción de valores.

La calidad de los pronósticos realizados se puede mejorar con un modelo de regresión lineal múltiple a través de un filtrado adecuado de datos y del empleo de un modelo multiobjetivo. Cabe mencionar, que en la segunda fase, el modelo ajustado se empleó para pronosticar un año específico con el objeto de cuantificar y caracterizar el comportamiento de las acciones analizadas.

La hipótesis del trabajo sostiene que, se puede diseñar una metodología híbrida entre un enfoque multiobjetivo y la técnica del Movimiento Geométrico Browniano para la toma de decisiones en el mercado de acciones en México.

En términos generales, el problema abordado en este trabajo se define como, dado un conjunto de variables predictoras X , una variable respuesta (dependiente) Y , y \hat{Y} respuesta pronosticada, se desea encontrar un subconjunto X' tal que $X' \subset X$, el cual genere el $\arg \min(z)_{f(X') \rightarrow \hat{Y}} : \sum_{i=1}^n (Y - \hat{Y})^2$.

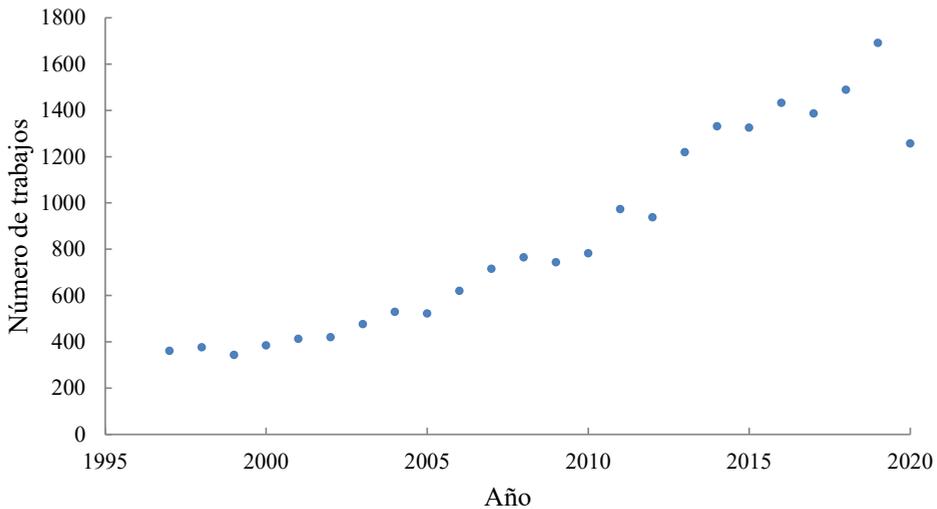
En la actualidad, la selección de la información pertinente para describir, caracterizar y analizar un fenómeno es una actividad relevante, ya que los avances tecnológicos han propiciado que las personas tengan acceso a una gran cantidad de información. Es decir, el desarrollo tecnológico ha incentivado una nueva forma de comunicación en red, asincrónica, rápida y efectiva sin depender de un lugar físico y, a veces, el lenguaje no es un impedimento (Díaz Gandasegui, 2011); sin embargo, la gran cantidad de información disponible lleva a que permita la toma de decisiones de manera estratégica.

Por otro lado, la predicción de precios en el mercado de valores es una actividad importante en el ámbito económico y bursátil. El sistema económico bursátil es complejo ya que intervienen múltiples actores (empresas, gobierno, sociedad, entre otros), y sus posibles interacciones. Desde ini-

cios del siglo pasado, la predicción de precios en el mercado de valores ha sido un campo de desarrollo científico, debido al gran impacto económico y social que conlleva esta actividad. Se debe mencionar, de acuerdo con la consulta realizada, que del año 2000 al 2019, se han publicado en la plataforma *ScienceDirect* (Elsevier, n.d.) 18,144 trabajos relacionados sobre este tema, sin embargo, en predicción de precios de la BMV, el promedio anual, es inferior al 3%.¹

En la Gráfica 1, se muestra el número de trabajos reportados en *ScienceDirect* sobre el pronóstico de acciones en las bolsas de valores en el mundo en el periodo del año 1997 al 2020 (con la consulta mencionada acotando los años). Por otro lado en la Gráfica 2, se muestra la comparación entre el porcentaje de investigaciones sobre pronósticos para las bolsas de Estados Unidos de América (EUA), Hong Kong y México, entre 1997 y 2020, en donde se observa que, durante la primera década de este siglo, alrededor del 50% del total de trabajos que han estudiado la Bolsa de Valores de EUA, han

Gráfica 1. Producción científica sobre pronóstico de acciones entre 1997 y 2020



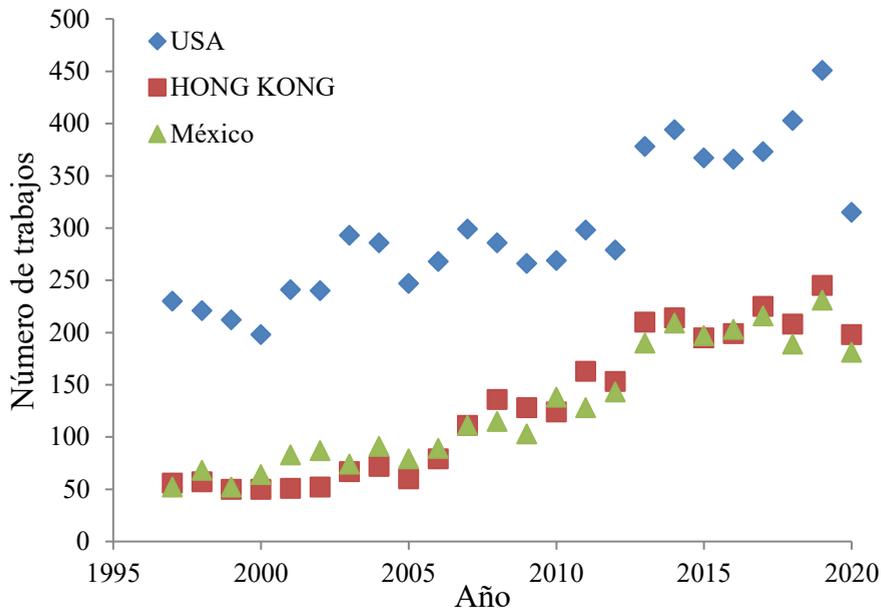
Fuente: elaboración propia a partir de los datos de ScienceDirect.

¹ Consulta realizada en el buscador de ScienceDirect con los términos “stock price prediction in stock exchanges” acotando los resultados en los años seleccionados.

ido disminuyendo en los años siguientes hasta representar sólo el 25% de los trabajos. Las investigaciones sobre las Bolsas de valores de México y de Hong Kong han representado entre el 10% y el 20% de las investigaciones.²

El objetivo de esta investigación, es presentar una metodología semiautomática, basada en estrategias de inteligencia artificial, para tratar y analizar la información de algunas acciones para realizar una predicción adecuada de los precios en la BMV. Cabe mencionar, que en la base de datos se incluye información de doce empresas mexicanas en el periodo comprendido de enero de 2006 a diciembre de 2016.

Gráfica 2. Comparación entre el porcentaje de investigaciones realizadas sobre bolsas de valores de EUA, México y Hong Kong.



Fuente: elaboración propia a partir de datos de ScienceDirect.

² Consultas realizadas en el sitio web de ScienceDirect con los mismos términos mencionados en la nota 1, y añadiendo las palabras "USA", "Hong Kong" y "México" respectivamente, obteniendo los resultados en los años mencionados

El presente documento se estructuró de la siguiente manera: en la sección 1 se muestra un conjunto de conceptos básicos relacionados con los temas de a) manejo y análisis de información, b) modelos para la predicciones de valores y c) métodos heurísticos en la solución de problemas de optimización; posteriormente en la sección 2 se aborda una revisión del estado del arte sobre pronóstico de acciones; en la sección 3 se presenta y describe la metodología propuesta así como los modelos de optimización implicados; en la sección 4 se aborda la metodología experimental empleada; en la sección 5 se ofrece un análisis de los resultados obtenidos; finalmente, en la última sección se plantean las conclusiones.

1. Conceptos básicos

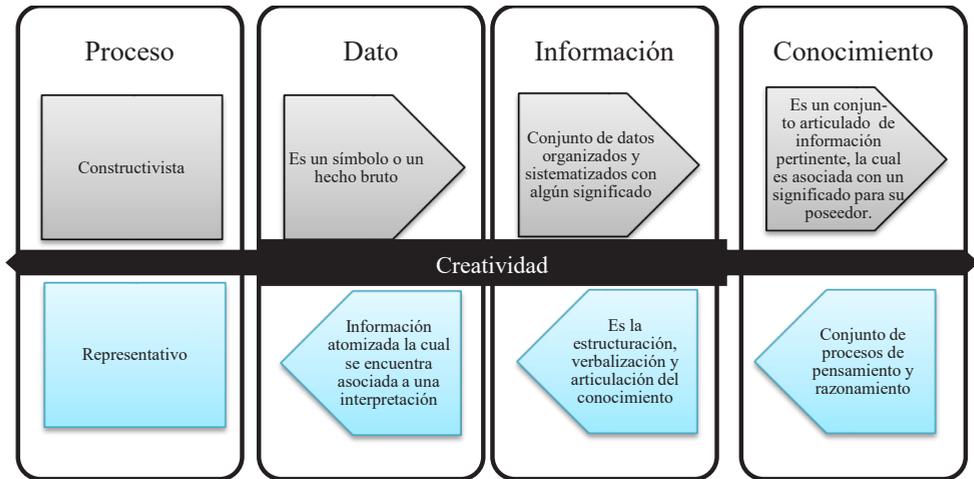
En esta sección se presentan los conceptos fundamentales en los que se basa esta investigación, su estructura es la siguiente: a) selección, análisis y manejo de información; b) pronóstico de valores y c) optimización como una herramienta para la toma de decisiones.

a) Selección, análisis y manejo de información

En la actualidad el ser humano se encuentra sumergido en un ecosistema hiperconectado y global lo que permite la rápida generación, difusión, análisis y debate de información, a esto se le ha denominado sociedad del conocimiento (Cerezo, 2016). En otras palabras, los miembros de la sociedad actual poseen la capacidad de obtener, generar y compartir información instantáneamente, desde cualquier lugar y en la forma que se prefiera. Lo que propicia innovaciones, modificaciones y desarrollos tecnológicos y científicos que impactan los medios económicos, sociales, culturales y políticos.

La posibilidad de acceder a una gran cantidad de información sobre cualquier tema conlleva el reto de transformar dicha información en conocimiento; como mencionan Juárez y Ponce (2019), en la sociedad actual el conocimiento es uno de los principales valores que poseen las personas. Para la construcción de conocimiento el ser humano emplea tanto su razonamiento (construcción *a priori* o proceso representativo), experiencia (construcción *a posteriori* o proceso constructivista), creatividad y su entorno sociocultural y tecnológico. El proceso descrito en la Gráfica 3 muestra una relación sinérgica entre los procesos constructivistas y representativos para el desarrollo de conocimiento.

Gráfica 3. Proceso para la construcción de conocimiento



Fuente: elaboración propia.

La producción y los procesos para la generación de conocimiento han sido objeto de amplios estudios (Montuschi, 2001; Rodríguez Gómez, 2006; Tobón *et al.*, 2015). Lo anterior ha llevado a lo que se ha denominado la cuarta revolución industrial.

La gran cantidad de información involucra la integración y empleo de estrategias y métodos para la generación de conocimiento como una aptitud primordial de los tomadores de decisiones; ya que de lo contrario se corre el riesgo de saturación, banalidad, irrelevancia, pérdida de tiempo, entre otros (Montuschi, 2001; Rodríguez Gómez, 2006; Tobón *et al.*, 2015).

Una estrategia empleada con objeto de manejar y analizar datos es el uso de sistemas de filtrado. En términos generales, se define el filtrado de información como una actividad que conlleva la selección y análisis con base en las necesidades. De acuerdo con Quiroga (2005), un sistema de filtrado es una herramienta inteligente, la cual tiene el propósito de valorar, seleccionar y entregar información adecuada, precisa y pertinente que corresponda a las necesidades de un usuario.

Una forma de procesar y reducir el tamaño de un conjunto de datos se conoce como Análisis de Componentes Principales (PCA, por sus siglas en inglés). Como menciona Jolliffe (1990), cuando se confronta con un conjunto de datos muy extenso, el instinto natural incita a intentar reducir su

tamaño, mientras se minimiza cualquier pérdida de información, para obtener mejor entendimiento e interpretar la estructura de los datos. Indica que, una estrategia alternativa, sería construir nuevas variables a partir de las originales, este enfoque parece menos intuitivo, pero tiene la ventaja que, por la misma cantidad de información perdida, es posible obtener mayores reducciones del conjunto de datos. PCA fue originalmente propuesto con un enfoque meramente estadístico por Pearson (F.R.S., 1901), la referencia más actual y recomendada se encuentra en la segunda edición del libro de Jolliffe (2002).

En el diseño y desarrollo de un sistema de filtrado se encuentra inmerso el siguiente problema de optimización: $\max(z) \sum_{i=1}^{|C|} \sum_{j=1}^{|S|} u_{ij} x_{ij}$; donde: x_{ij} es uno si el j -ésimo ítem es usado por el i -ésimo usuario y x_{ij} es cero en cualquier otro caso; u , una función de utilidad del j -ésimo ítem para el i -ésimo usuario tal que $u: C \times S \rightarrow \mathbb{R}^+$; C es un conjunto de usuarios y S un conjunto de posibles ítems. (Agichtein, *et al.*, 2006; Hu, Volinsky *et al.*, 2008; Rennie y Srebro, 2005; Yannakakis, 2009).

En el diseño de sistemas de filtrado se han empleado árboles de decisión, máquinas de soporte vectorial, redes neuronales, redes bayesianas, discriminantes lineales, regresión logística, entre otras.

b) Pronóstico de valores en la toma de decisiones

El diseño de herramientas confiables para pronosticar y predecir el comportamiento de fenómenos o sistemas, ha sido objeto de interés en las investigaciones desde hace varias décadas. En términos generales, pronosticar se define como la generación de información y conocimiento que permite prever, caracterizar y analizar el comportamiento futuro de situaciones con base en estrategias lógicas, racionales, científicas y de experiencia previa; se debe mencionar que, frecuentemente la información generada por el proceso de pronóstico es empleada para la toma de decisiones.

Rubio Guerrero (2017), menciona que pronosticar es una de las funciones más importantes dentro de las empresas, industrias y negocios; ya que a través de la información producida se toman decisiones. En la Tabla 1 se muestra la clasificación de los métodos de pronósticos comúnmente empleados.

Algunos autores diferencian los conceptos pronóstico y predicción, haciendo énfasis en que el pronóstico se basa en datos de naturaleza objetiva;

Tabla 1. Métodos comunes de pronóstico

Métodos de pronósticos	Cuantitativos	Se basan en el empleo de estrategias lógico-matemáticas para el análisis histórico de datos. Generalmente, se les denomina procedimientos objetivos.	<ul style="list-style-type: none"> • Series de tiempo • Modelos causales • Modelos de optimización • Modelos de simulación • Otros
	Cualitativos	Se basan en la experiencia e intuición; por lo tanto, son procedimientos con un grado de subjetividad.	<ul style="list-style-type: none"> • Método Delphi • Analogías con sistemas similares
	Mixtos	Se basan en la combinación sinérgica de estrategias cuantitativas y cualitativas.	Aquí se encuentran varias estrategias de IA y Machine Learning.

en contraste, predicción se basa en información subjetiva. En la actualidad, las metodologías de naturaleza mixta han crecido en su aplicación (Madero Suarez y Gómez López, 2013; Muñoz Poblete, 2013).

En el presente trabajo, se emplearon métodos cuantitativos que integran procedimientos causales, series de tiempo y de optimización en la generación del modelo predictivo. En las subsecciones siguientes se describen brevemente los modelos causales.

Un modelo es una representación o abstracción selectiva (cuantitativa o cualitativa) de las características de un sistema. De manera general, los modelos causales involucran que dado un conjunto X de variables descriptivas (también denominadas independientes) y una variable a explicar (llamada dependiente) encontrar una función $f(X, B)$ tal que $f(X, B) \rightarrow \hat{y}$; donde $y = \hat{y} + \varepsilon$, ε , es una medida de error entre y y \hat{y} , $B = \{\beta_0, \beta_1, \dots, \beta_n\}$, es un conjunto de coeficientes asociados a las variables X .

Los modelos de series de tiempo involucran que dado el comportamiento histórico de la variable y en el periodo $t - 1$ a $t - k$ encontrar una fun-

ción $f(y_{t-1,t-k}, \Psi)$ tal que $f(y_{t-1,t-k}, \Psi) \rightarrow \hat{y}_t$; donde, $y_t = \hat{y}_t + \mu$; y_t ; es el valor real de la variable y en el instante t ., \hat{y}_t es el valor calculado para la variable y en el instante t . μ es una cuantificación de ruido en función del tiempo, $\Psi = \{\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_k\}$ es un conjunto de coeficientes asociados a los valores históricos de y .

En la literatura consultada, el modelo Log-Normal basado en el Movimiento Geométrico Browniano (MGB), es empleado frecuentemente en los modelos de serie de tiempo. Algunos autores (Pérez Fernández *et al.*, 2015) describen al MGB como un modelo diferencial estocástico tipo $It\delta$, en el cual el valor de una variable a predecir (activo financiero), es modelada añadiendo una función de caminata aleatoria en el tiempo.

c) Optimización como una herramienta para la toma de decisiones

Los modelos de optimización se caracterizan por poseer: a) un conjunto de alternativas o variables de decisión (el valor de éstas afectan el desempeño del sistema), b) un conjunto de restricciones, las cuales representan relaciones o condiciones que el conjunto de variables de decisión están obligadas a satisfacer, y c) un conjunto de funciones objetivo, las que se emplean como medida cuantitativa sobre la calidad de las soluciones encontradas de un problema y son una función matemática de las variables de decisión.

Como se mencionó con anterioridad, la determinación de coeficientes en un problema de pronóstico puede implicar un sistema no lineal y por ende se involucra un problema de optimización NP-completo.

Si la función objetivo $f(\chi, \theta)$, espacio de búsqueda $S(y, f(\chi, \theta))$ o ambas son relaciones no lineales, entonces se tendrá que resolver una instancia del problema de optimización no lineal, el cual es NP-duro. Un conjunto de técnicas empleadas para resolver esta clase de problemas son los métodos metaheurísticos.

Una metaheurística es un procedimiento flexible e inteligente para generar buenas soluciones a un problema con un consumo razonable de recursos. Las soluciones generadas por estos procedimientos no son necesariamente óptimas. A continuación, se describen qué son las metaheurísticas de los algoritmos genéticos y la búsqueda armónica, las cuales serán empleadas en este trabajo.

Los algoritmos genéticos (GA) son un enfoque de optimización evolutiva basado en algoritmos de búsqueda aleatoria, (Holland, 1975). Es un método, muy popular debido a su robustez para la solución de problemas complejos y

no lineales. Tiene numerosas ventajas sobre otros métodos clásicos, ya que pueden converger fácilmente a una buena, o bien, a la mejor solución más rápido que otros. Se utilizan operadores genéticos, como selección, mutación y cruce para obtener soluciones. Los pasos básicos son: a) el algoritmo inicializa una población de posibles soluciones; b) se aplican los operadores genéticos; c) se evalúa la función para cada individuo; d) se eliminan, ocasionalmente las malas soluciones de la población; e) una nueva población es creada de nuevo utilizando los operadores, este proceso se repite hasta llegar a un criterio de paro.

Búsqueda armónica, propuesta en 2001 (Zong *et al.*, 2001), este algoritmo se adapta a los conceptos de creación musical en optimización. Las soluciones formuladas son llamadas “armonías” y cada valor de la misma se llama acorde. Se inicializa la memoria de armonías, seleccionando aleatoriamente los valores de las soluciones y se calcula su función objetivo, enseguida se improvisa una nueva solución tomando en consideración la memoria, si esta nueva solución “improvisada” es mejor que la peor en la memoria, se sustituye y se repite el proceso hasta el criterio de paro. Se debe recordar que se utilizan tres operaciones básicas las cuales son: recordar, adaptar e improvisar.³

En la siguiente sección, se ofrece una revisión bibliográfica con énfasis en los trabajos sobre pronósticos en bolsas de valores, que involucran el empleo de estrategias de inteligencia artificial.

2. Estado del arte

En la literatura existen múltiples trabajos que versan sobre el desarrollo de estrategias computacionales para predicción de acciones e índices. Dentro de éstos, se encuentran: el uso de redes neuronales artificiales (Kara *et al.*, 2011), Support Vector Machines (SVM) (Yeh, Huang, y Lee, 2011) y árboles de decisión (Nair, Mohandas, y Sakthivel, 2010) entre otros métodos de inteligencia artificial (IA). Algunos autores han optado por combinar técnicas de IA en sus modelos propuestos, para mejorar los resultados de predicción. Por ejemplo, en el estudio de Choudhry y Garg (2008), se realizó un modelo híbrido GA-SVM, implementando en los datos de entrada una primera fase con un Algoritmo Genético (GA), para posteriormente realizar, con los datos obtenidos, una segunda fase con SVM. Este trabajo contó con 35 valores

³ Para mayor información se puede consultar Zong *et al.*, (2001).

técnicos como entrada para el modelo, sin embargo, proponen variar este número, para intentar obtener un rendimiento mayor en la predicción de precios futuros. Trabajaron con tres conjuntos de valores (stocks) de los más prominentes de la bolsa de valores de India. Entre los resultados obtenidos de este estudio, se encuentra una mejora significativa en el rendimiento de la ejecución del modelo híbrido, en contraste con el modelo de una sola capa de SVM.

A finales del otoño de 2014, Patel *et al.*, (2015) implementaron algoritmos en dos capas para predecir los índices bursátiles. Se trabajó con los índices llamados “CNX Nifty” y “S&P Bombay Stock Exchange (BSE) Sensex” del mercado de valores de India. Estos investigadores realizaron predicciones a futuro de 1-10, 15 y 30 días, mostraron entre sus conclusiones, una mejor predicción cuando se utilizó un acercamiento de dos etapas, haciendo uso en la primera etapa de regresión de soporte vectorial (SVR), y en la segunda Redes Neuronales Artificiales (ANN), Random Forest (RF) y SVR. El resultado son modelos de predicción SVR-ANN, SVR-RF y SVR-SVR. Se compararon los resultados de estos modelos junto con los de una sola etapa, y se obtuvo un mejor resultado en los modelos de dos etapas, el modelo que mostró un mejor desempeño de éstos fue SVR-ANN. Este método final de implementación de dos capas es propuesto por los autores para ser aplicado en otros campos, como predicción del clima y consumo de energía, entre otros.

Como se mencionó con anterioridad, el empleo del movimiento geométrico Browniano (MGB) en tareas de pronóstico de precios es frecuente (Chakrabarti y Guhathakurata, 2019; Fonseca *et al.*, 2017; Mason y Wilmot, 2016), por ejemplo: algunos emplean al MGB para ciertas tareas dentro de algoritmos bioinspirados (Dehghani y Bogdanovic, 2018), y otros utilizan, también, la idea del movimiento browniano para implementar un modelo de predicción (Ntemi y Kotropoulos, 2020; Sikora *et al.*, 2019; Song *et al.*, 2019).

Mason y Wilmot (2016) realizaron predicciones de precio sobre el mercado emergente de combustibles renovables en Estados Unidos, el cual al manejar materias primas, se alimenta con nueva información que puede producir cambios bruscos o saltos; explican que puede tener un comportamiento más complejo que el MGB, así que, tomando como base el movimiento geométrico browniano, desarrollan modelos de predicción de precios, y obtuvieron resultados con una cantidad aceptable de error.

2.1. Modelo

En un modelo multiobjetivo, se plantean un conjunto de funciones objetivo (dos o más) en conflicto entre sí. Por ejemplo: $\text{Min } f_1(x), f_2(x), \dots, f_k(x)$ sujeto a: $g(x) \geq 0$ para todo $j = 1, \dots, p$, $h(x) = 0$ para todo $j = p - 1, \dots, m$, $x_i \geq 0$, $x \in \mathbb{R}^n, c \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n$

En optimización multiobjetivo, el término optimizar, se refiere a encontrar un ajuste entre las distintas funciones objetivo en lugar de una sola solución como en optimización global.

2.1.1. Modelo multiobjetivo para la determinación de predictoras

El modelo siguiente es empleado en la sección de filtrado con recocido simulado para la reducción de variables: $\min z_1 = \sum_{i=1}^N w_i$, $\min z_2 = \sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{N}$, $\min z_3 = \sum_{i=1}^N \sum_{j=1, j \neq i}^N \beta_{ij} * w_i * w_j$, $\min z_4 = -\sum_{i=1}^N \alpha_i$.

Sujeto a: $w_i = 1$, si la variable x_i es seleccionada como predictora; 0, si la variable x_i NO es seleccionada como predictora; $g(x, w) = \hat{y}_i$, $\beta_{ij} = 0$ si la variable x_i es estadísticamente similar a x_j , -1 , si la variable x_i NO es estadísticamente similar a x_j ; $\alpha_i = \text{absoluto (correlación } (y_i, x_i) \text{)}$, si $w_i = 1$, o 0, si $w_i = 0$, donde, $g(x, w)$ es una función de x y w que estiman a y .

En este trabajo, se considera que una variable es exactamente igual o una combinación lineal de otra variable, si la media entre ambas variables (x_i, x_j) es estadísticamente similar así como su varianza. También se considera que se desea encontrar una función $f(x) = \hat{y}$, tal que el error producido entre la y (real) y la \hat{y} (pronosticada), sea mínimo, tanto para el periodo de entrenamiento como para el periodo que se desea pronosticar. Cabe mencionar que determinar o encontrar un conjunto de funciones que cumplan estas características es un problema NP-completo, por lo cual no existe un método exacto que lo pueda resolver en tiempo polinomial.

Entonces, el problema de selección de parámetros en modelos de regresión se determina por una tupla $(\theta, x, y, f(x, \theta), S(\theta), \varepsilon, n, m)$, una solución de este problema involucra encontrar una configuración θ adecuada, tal que, sea un conjunto de argumentos mínimos.

Para la búsqueda de los parámetros para la implementación de un modelo ARMA (Auto-regresivo de medias móviles), se han utilizado GA (Ervural *et al.*, 2016), también han usado GA para la búsqueda de parámetros de vectores de soporte regresión (Chen, 2007). En este trabajo se implementa-

ron GA y Búsqueda Armónica (HS) (Zong *et al.*, 2001), las cuales se describen en la siguiente sección.

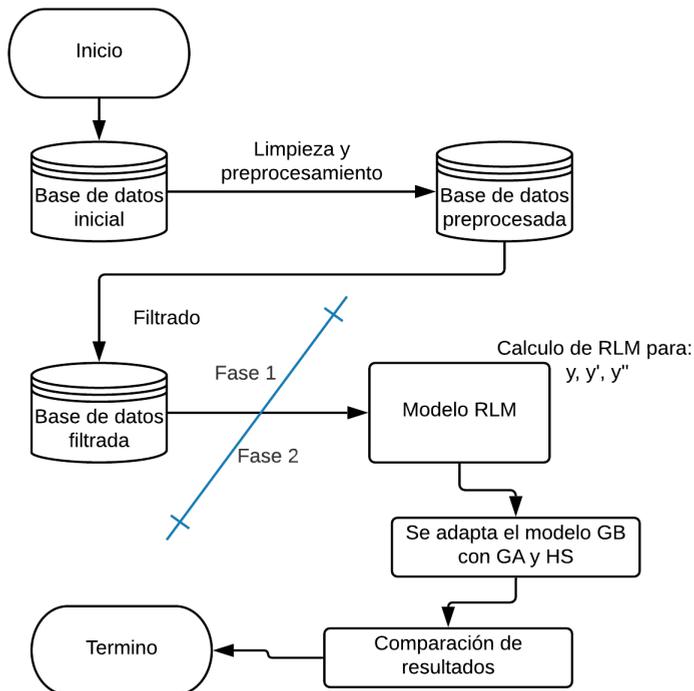
3. Metodología

En esta sección se expondrá el proceso metodológico realizado en este trabajo. La estructura de la presente sección es la siguiente: a) metodología experimental y b) diseño experimental. Se debe mencionar que se empleó la metodología propuesta por de-los-Cobos-Silva en (de-los-Cobos-Silva *et al.*, 2018).

a) Metodología experimental

El proceso completo que se utilizó para la realización de este trabajo se presenta en la Gráfica 4. Como se mencionó en la sección anterior, la parte de selección de variables es muy importante. Las bases de datos iniciales son

Gráfica 4. Diagrama del proceso completo



descargadas, cargadas en el sistema, analizadas y depuradas para mayor facilidad de uso y procesamiento (eliminación de valores nulos, cambio de tipo de datos, etcétera). La base de datos preprocesada se introduce en el proceso de filtrado de variables para reducir el número de entradas y minimizar el error en un modelo de regresión lineal múltiple (RLM). Una vez obtenida la base de datos filtrada, se formula el Modelo Geométrico Browniano y se adaptan las metaheurísticas para obtener valores óptimos de las variables del modelo. Cabe mencionar que para la primera fase se emplea recocido simulado mientras que para la segunda fase se emplea GA y HS. A continuación se presenta una explicación más detallada de cada parte del proceso completo.

Primera fase

La separación entre las variables predictoras y la que se va a predecir (valor de cierre) obedecen al interés mostrado en otros trabajos. En la Gráfica 5 se presenta el proceso del filtrado de datos. Debe mencionarse que sólo para las variables predictoras, se realizó lo siguiente:

La base de datos original, inicialmente de 11 variables, se amplía aplicando 12 funciones (ver Anexo 1) a cada una de ellas, generando así 132 variables que permitan determinar el valor de cierre de las acciones de un conjunto de 12 empresas mexicanas que cotizan en la BMV (Bolsa Mexicana de Valores).

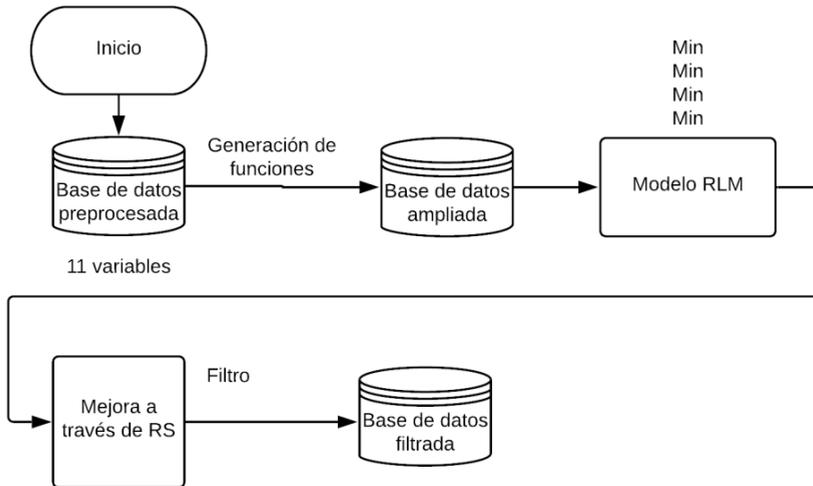
El objetivo de esta fase es minimizar “ α ”, el número de variables asociadas para describir el comportamiento de una variable dependiente, donde: $\alpha = \sum_{i=1}^k X_i$; $x_i = \{1, \text{ si se usa la } i \text{ variable, o } 0 \text{ en cualquier otro caso}\}$, $k = \text{número de total variables}$.

De esta manera se obtiene un problema multiobjetivo, el siguiente modelo es empleado para la determinación de coeficientes del modelo no lineal de pronóstico empleado en la segunda fase de este trabajo:

$$\text{Max } Z_1 = |\text{Corr}|, \text{Min } Z_2 = \frac{\text{SSE}_a}{N_a}, \text{Min } Z_3 = \frac{\text{SSE}_p}{N_p}, \text{Min } Z_4 = \alpha$$

Donde, $|\text{Corr}|$ es el valor absoluto del coeficiente de correlación lineal, SSE es el error cuadrático, N el número de datos, a es el conjunto de entrenamiento, p es el conjunto de prueba y α el conjunto de variables a usar.

Gráfica 5. Diagrama del filtrado de variables



Dado que es un modelo de regresión lineal múltiple, se considera la siguiente ecuación: $\hat{y}_{t+1} = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon$. El filtrado de variables se realiza de la siguiente manera, dadas k variables, para cada X_i , no se emplea en el modelo si y sólo si: el promedio de la variable X_i es igual al promedio de y' , $y: \delta_{X_i}^2 = \delta_{y'}^2$, la varianza de X_i es igual a la varianza de y' .

Se determina el conjunto de variables estadísticamente diferentes a y y cuyo coeficiente de correlación es diferente de cero, denotando el conjunto X_r . Posteriormente para cualquier par de variables en X_r se determina si i y j son estadísticamente similares, en caso de serlo se penalizará emplear simultáneamente a i y a j . Sea m_1 el modelo de regresión lineal múltiple $g(X_r) \rightarrow y$, se evalúa el modelo generado con base en el modelo multiobjetivo y se determina un conjunto θ de parámetros para la ejecución del recocido simulado multiobjetivo (sea p el número de puntos en el frente de Pareto, T_i temperatura inicial, T_f temperatura final, α criterio geométrico de descenso de temperatura, k número de ejecuciones antes del descenso de temperatura) para encontrar el conjunto mínimo de variables necesarias. Se genera un conjunto de $2p$ soluciones iniciales con base en la información de m_1 y se evalúan las soluciones encontradas, sea A el conjunto de p soluciones iniciales aquellas que se toman al considerar criterio de no dominancia y el criterio de suma ponderada.

A continuación, se presenta el pseudocódigo del algoritmo del recocido simulado para el filtrado:

Tabla 2. Algoritmo de recocido simulado para filtrado de variables.

Entrada Un conjunto expandido de variables predictoras

Ta = Ti

MIENTRAS Ta > Tf

PARA i =1: k

PARA j =1: p

S_i i-ésima solución en A

S_n sea una solución generada con búsqueda local

 Si S_n es mejor que S_i , entonces S_i en A es remplazado por S_n

 Si no se determina el criterio de metrópolis posteriormente se emplea éste decidir si S_i en A es remplazado por S_n .

B_j sea la solución desechada

FIN PARA

$T_a = \alpha * T_a$

$C = A \cup B$

 A son las p mejores soluciones de C

$B = []$

FIN PARA

SALIDA Subconjunto de X_T de variables predictoras

Fuente: elaboración propia

Segunda fase

La función del Modelo Geométrico Browniano para la predicción de \hat{y} (precio de la acción) es el siguiente:

$$\hat{y} = a_1 y + e^{a_2 y' + a_3 y'' + \mu_{error} + \sqrt{t} \times S_{error} \times \alpha}$$

Donde:

- a_i con $i = 1, 2, 3$ son constantes a encontrar,
- y es el resultado del cálculo con Regresión Lineal Múltiple obtenida con el filtrado ($f(x)$, en primera fase),
- y' es la primera derivada y calculada ($f(x')$),
- y'' es la segunda derivada de y calculada ($f(x'')$),

- μ_{error} es la media del error,
- t es el tiempo transcurrido,
- S_{error} es la varianza del error sobre y calculada,
- α es un aleatorio normalmente distribuido ($\alpha \sim N(0,1)$).

Una vez obtenida la base de datos filtrada (Gráfica 4), se inicializa el Modelo Geométrico Browniano, y se adaptan las metaheurísticas de GA y HS para la búsqueda de los mejores parámetros del modelo predictivo.

Algoritmos genéticos

Se aplicó el algoritmo genético básico, con una población de 10 individuos, cada cromosoma se representó por una cadena de números enteros compuesta por tres constantes a encontrar (a_1, a_2, a_3), la media del error (μ_{error}), la varianza del error sobre calculada (S_{error}), funciones objetivo 1 y 2 (O_1, O_2) como puede observarse en la Tabla 3. Estos individuos fueron almacenados en una matriz para su manipulación.

Tabla 3. Representación de un individuo en el algoritmo genético básico

a_1	a_2	a_3	μ_{error}	S_{error}	\bar{O}_1	O_2
-------	-------	-------	---------------	-------------	-------------	-------

Fuente: elaboración propia.

La creación de la población inicial es aleatoria, la selección de padres se realiza mediante torneo binario. Una vez que se obtiene la pareja de padres, se realiza la cruce para obtener un hijo, la cruce se realiza en dos puntos seleccionados de manera aleatoria del cromosoma, se transfiere la información genética a los hijos y se almacenan.

Con los nuevos hijos se genera la corrida de mutación, con una probabilidad de 15% con los hijos elegidos, se selecciona un “gen” a mutar. A continuación se unen la población anterior con los hijos y se ordenan por aptitud; se selecciona la mejor mitad y esta se convierte en la nueva generación. El proceso se repite hasta que se cumple el criterio de paro, en nuestro caso, al llegar a las dos mil generaciones procesadas.

Búsqueda Armónica

Se aplicó el algoritmo de Búsqueda Armónica (HS, por sus siglas en inglés), el cual contó con una memoria de armonías de diez elementos, una tasa de ajuste de tono de 10%, una tasa de consideración de armonía de 70%, un máximo número de iteraciones de dos mil y su información fue representada en cadenas de números enteros en la primera sección, representando la parte a buscar y el resto, variables calculadas con los primeros tres valores de la armonía, del Modelo Geométrico Browniano. Estas armonías se almacenaron en una matriz para su manipulación.

Cada armonía está igualmente compuesta que los individuos del AG. En la creación de la memoria de armonías aleatoriamente, se llena su información con los valores acotados en los que deben estar los parámetros, un número aleatorio de una distribución normal. Posteriormente se improvisa una nueva memoria de armonías considerando la actual, ajustes de acordes (constantes a buscar) y aleatoriedad. El algoritmo que se emplea para la creación de una nueva memoria de armonías se presenta en la Tabla 4.

Una vez llegado a este punto del proceso, se unen las memorias, la nueva improvisada y la anterior, se ordenan por mejores resultados de los objetivos ponderados, se selecciona la mejor mitad y ésta procede a ser almacenada como la memoria de armonías inicial para la siguiente iteración. Este

Tabla 4. Sección del algoritmo de Búsqueda Armónica para creación de una nueva memoria de armonías

```

Crea una memoria armónica inicial
MIENTRAS criterio de paro no satisfecho
PARA j=1: HM
PARA k=1: 3
    SI rand < (1-PAR)
        selection = una armonía aleatoria
        sol_nueva(j,k) = memoria_h(selection,k)
    SI rand < HCR
        s1 = raíz(varianza(memoria_h(:,k)
        SI s1==0
            s1=1
        FIN SI
        sol_nueva(j,k) = sol_nueva(j,k) + randn*(s1)

```

FIN SI

OTRO

apar1=mínimo(memoria_h(:,k)

apar2=máximo(memoria_h(:,k)

MIENTRAS apar2==apar1

apar2=apar2+randn

FIN MIENTRAS

sol_nueva(j,k) = apar1+randn*(apar2-apar1)

FIN SI

FIN PARA

sol_nueva(j,4) = Calcular media del error

sol_nueva(j,5) = Calcular varianza del error

sol_nueva(j,6) = Calcular Objetivo 1

sol_nueva(j,7) = Calcular Objetivo 2

FIN PARA

Determinar el conjunto de soluciones no dominadas

FIN MIENTRAS

Fuente: elaboración propia.

proceso se repite hasta llegar al criterio de paro, en este caso, hasta llegar a las dos mil generaciones procesadas. Se emplea el criterio de no dominancia y suma ponderada. Se evalúan ambos métodos usados por índices de calidad o desempeño, los cuales son: hipervolumen, distribución, distancia generacional invertida y cobertura de conjuntos.

b) Diseño experimental

Una vez obtenidas las bases de datos de la plataforma especializada, se guardaron en formato CSV (archivo separados por comas, por sus siglas en inglés), se realizaron cambios a formatos de número de los datos y se removieron los títulos de las columnas, se almacenaron en formato TXT (archivo de texto) para facilitar la manipulación en MATLAB.

Para realizar el filtrado de variables, se procesó cada caso en la base de datos en MATLAB,⁴ el análisis de una instancia implicó la ejecución de 20

⁴ Esto se ejecutó en la versión r2014b, en una computadora iMac Retina 5K, con un procesador Intel Core i5 3.2 GHz y memoria 8GB 1867 MHz DDR3, cada corrida del algoritmo se realizó en 8 horas cada una en promedio.

corridas, al final de cada ejecución, se obtuvieron las gráficas y tablas de información que pueden ser consultadas en “<http://bit.ly/396h3J1>”.

La siguiente parte del modelo también se ejecutó en MATLAB. La implementación de dos metaheurísticas para la obtención de los parámetros del MGB, ambas implementaciones se probaron 20 veces y se guardaron los resultados, tuvieron un tiempo de ejecución por cada acción de 2 a 3 horas por técnica heurística implementada. Con las soluciones generadas por cada método, se analizaron las medidas de desempeño y los frentes de Pareto de ambas técnicas heurísticas para obtener los resultados.

4. Datos

Para la realización de este trabajo se seleccionaron al azar, con el fin de no sesgar los datos, empresas grandes de distintos sectores que operan en la BMV, el cual es considerado un mercado financiero emergente (Comisión Nacional de Vivienda, 2016). Las bases de datos se obtuvieron de la plataforma de “Economatica” (Economatica, n.d.), posteriormente fueron depuradas para eliminar los valores faltantes por días sin operación (días festivos) y para el filtrado de variables. La base de datos depurada, antes de la aplicación de funciones a las mismas, consta de 12 variables, las cuales son, valor de cierre, apertura, valor mínimo, valor máximo, volumen, número de operaciones, promedio, RSI, AccRSI, MACD, Momento a 10 días y Bollinger. El periodo de análisis de la muestra comienza en enero de 2006 y termina en diciembre de 2016, fue elegido debido a los eventos recientes ocurridos en México, entre otros: elecciones presidenciales, aumento en índices de violencia, “guerra contra el narcotráfico”, crecimiento de redes sociales, y en el mundo, desde la recesión en 2008, la crisis de deuda de la eurozona en 2011, el Brexit y las elecciones en USA en 2016.

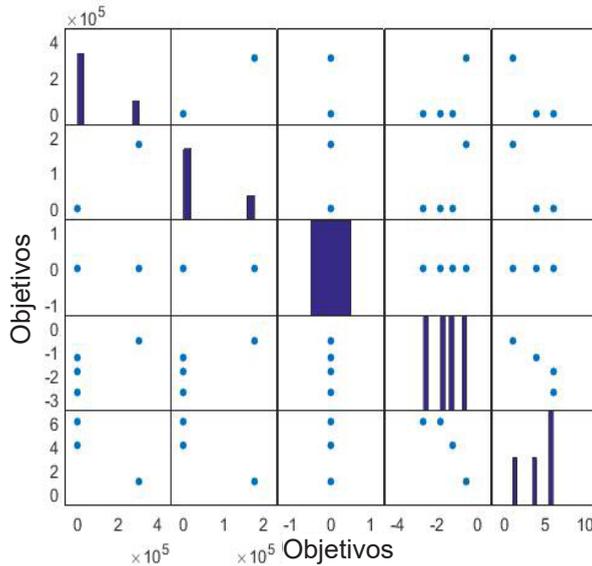
5. Análisis de resultados

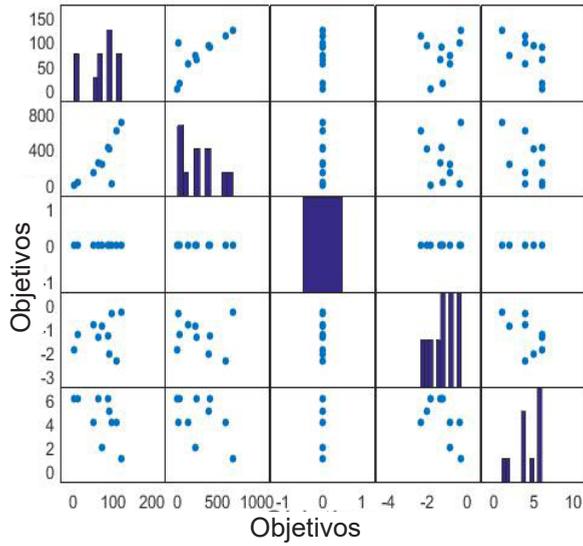
Como fue mencionado en la segunda sección, el filtrado es una herramienta que ayuda a reducir el número de variables necesarias y mejora la calidad de los resultados obtenidos, en la Gráfica 10, se puede observar la representación de los frentes de Pareto de los cinco objetivos del modelo, descritos en la sección 3, para el pronóstico de la acción de “Bimbo”. Se puede apreciar una dispersión considerable en los puntos, y no es posible identificar una curva bien definida en la mayoría de los casos.

Los resultados de las demás acciones, tanto gráficas como tablas, obtenidos en este trabajo pueden ser consultados en “<http://bit.ly/396h3J1>”, por motivos de espacio, sólo se presentan los resultados de la acción de “Bimbo”. Después de realizado el filtrado de variables, se observa en la Gráfica 6, que las curvas de los frentes comienzan a dibujarse más claramente en cada comparación. La forma de generación de las soluciones permitió evitar emplear variables colineales, por ende, el objetivo 3, del modelo presentado siempre toma el valor de cero.

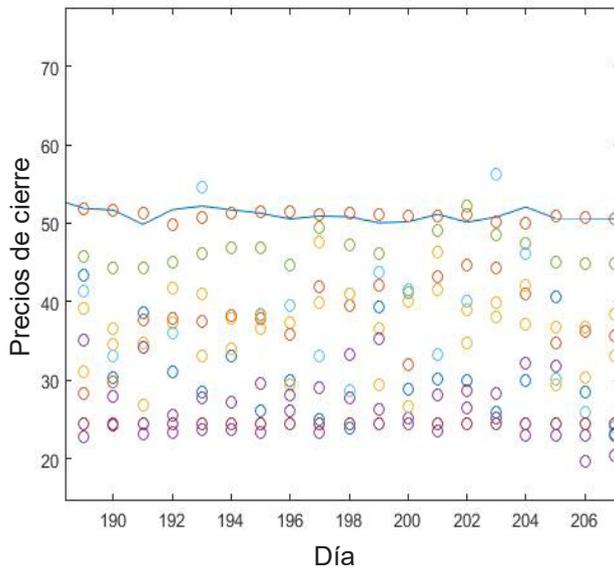
Es posible observar esta mejora en el modelo, al comparar la Gráfica 7, se muestran los pronósticos del valor de la acción antes del filtrado de datos y los pronósticos después del mismo, en ambas gráficas, el eje “y” corresponde al valor en pesos (\$), el eje “x” es el número de días del periodo de prueba, la línea continua es el valor real de la acción y los círculos pequeños representan los pronósticos de los modelos generados; se puede ver un incremento en los modelos generados que reducen la distancia al valor real de la acción a pronosticar.

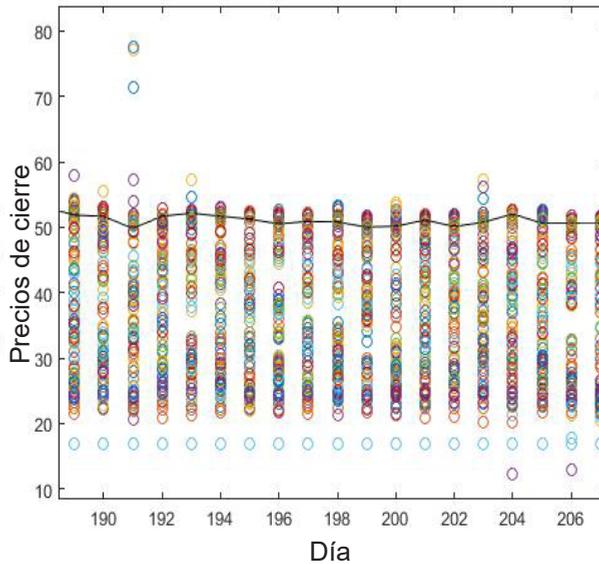
Gráfica 7. Frentes de Pareto de los objetivos del modelo antes (izquierda) y después (derecha) del filtrado de datos





Gráfica 8. Frente de Pareto del modelo con GA (izquierda) y HS (derecha)





Por último, se concentran en la Tabla 5, los resultados del filtrado de variables de las 12 acciones con las que se trabajó, de un total de 132 variables en cada caso, se redujeron a estos números, de izquierda a derecha, en la primera columna se presenta el número máximo de variables usadas, del total de variables de la base de datos ampliada, en los modelos creados para cada acción. En la segunda columna, el número mínimo de variables usadas, en la tercera y cuarta columna, el promedio y mediana, respectivamente, de variables usadas en los modelos generados. Se nota una gran disminución del número de variables a emplear, cerca de 12 variables en el peor de los casos y alrededor de 2 a 3 variables en los mejores modelos.

En la Tabla 6, se muestran los resultados obtenidos al evaluar las técnicas empleadas para la búsqueda de mejores variables del MGB, de la acción de “Bimbo”, la primer métrica (Distribución), se observa una cercanía al Pareto óptimo (valor cercano a 1), mayor en el GA, y en las siguientes métricas, donde se espera obtener un valor cercano a 0, un mejor desempeño por parte del GA, exceptuando el hipervolumen, estos resultados del desempeño pueden ser mejorados en un trabajo posterior, ajustando el algoritmo de búsqueda usado.

Tabla 5. Comparación del número de variables utilizadas en cada acción (stock)

1	10	2	6	7
2	9	2	5	5
3	7	1	4	4
4	7	2	4	4
5	9	2	6	6
6	13	2	5	5
7	7	1	4	4
8	5	1	3	3
9	9	4	6	5
10	9	2	6	6
11	9	2	5	4
12	11	2	6	7

En la Gráfica 8 se muestran los frentes de Pareto obtenidos del modelo multiobjetivo utilizando GA y HS respectivamente, en las cuales es posible observar los puntos próximos a la curva del frente de Pareto óptimo. A pesar de obtener una capacidad de predicción satisfactoria, y un error aceptable, las técnicas heurísticas usadas pueden ser mejoradas, para así, llegar a obtener valores de predicción más acertados y lograr un mejor resultado en sus métricas de desempeño. En la Tabla 7, se exponen los resultados resumidos de las medidas del desempeño de ambos algoritmos, para la búsqueda de parámetros del modelo generado para cada acción, 1 representa un buen desempeño, 0 menor que el obtenido y, -1 un mal desempeño.

6. Conclusiones

La metodología propuesta en este trabajo ha mostrado una alta flexibilidad para el filtrado y pronóstico de datos a partir de la concatenación sinérgica de dos modelos de optimización. Destacando que la primera fase permite identificar el conjunto de variables predictoras importantes reduciendo a 5% las variables de la base de datos ampliada. Por otro lado, el proceso de

Tabla 6. Tabla de medidas del desempeño de algoritmos multiobjetivo

Métrica	Algoritmo	Máximo	Mínimo	Promedio	Mediana	Varianza	Desviación estándar
Distribución	HS	0.89331447	0.73627844	0.81479645	0.81479645	1.23E-02	0.11104125
	GA	0.9375684	0.92663061	0.9320995	0.9320995	5.98E-05	0.00773418
Distancia Generacional Inversa	HS	0.27314163	0.10430499	0.18872331	0.18872331	0.01425291	0.11938554
	GA	0.1548839	0.07730731	0.1160956	0.1160956	0.00300906	0.05485493
Space	HS	0.34250562	0.296266	0.31938581	0.31938581	0.00106905	0.03269635
	GA	0.2185164	0.03603323	0.12727482	0.12727482	1.67E-02	0.12903508
Hipervolumen	HS	0.39362027	0.31168131	0.35265079	0.35265079	0.003357	0.05793959
	GA	0.47255828	0.33761021	0.40508424	0.40508424	9.11E-03	0.0954227

Gráfica 8. Pronóstico del valor/stock (Bimbo) antes (izquierda) y después (derecha) del filtrado de variables

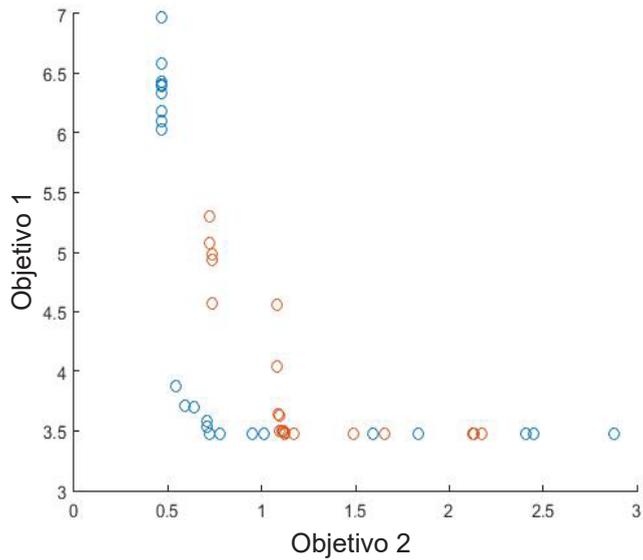
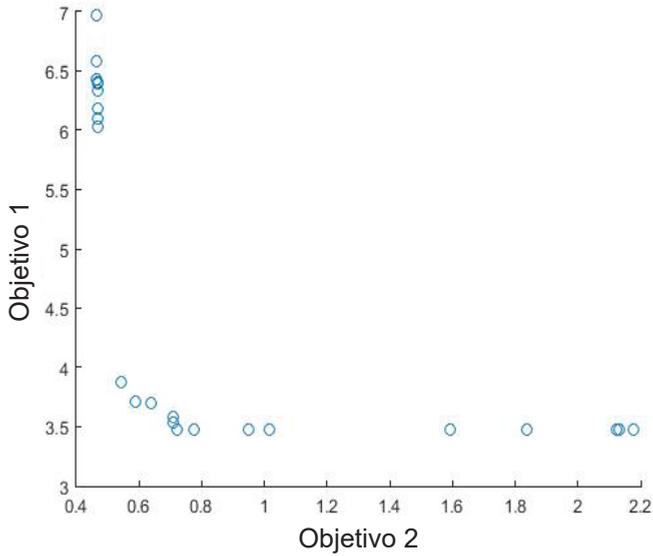


Tabla 7. Resultados del desempeño de ambos algoritmos

Acción	Alsea	Amxl	Bimbo	Cemex	Cocacola femsakofi	femsaubd	Gbanorte	Gcarso	Ginbursa	Gmexico	TelevisaCPO	Walmex
BH	0	1	0	-1	-1	-1	-1	0	-1	-1	0	-1
AG	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1

calibración de los modelos mejora el comportamiento de las funciones de pronóstico generando datos ajustados y cercanos a los datos reales.

Se desarrolló un modelo multiobjetivo para la predicción de los precios de las acciones, donde se logró minimizar el número de variables de entrada, maximizar la correlación lineal y minimizar el error sobre el conjunto de datos de entrenamiento y de prueba, de los cuales se obtuvo un buen resultado. En este trabajo se aporta una base de datos del histórico de acciones representativas de la BMV con indicadores técnicos, del periodo comprendido de 2006 a 2016, ésta se limpió de valores faltantes y se procesó para poder ser manipulada con mayor facilidad. La base de datos se amplió utilizando funciones, contando al final con 132 variables, las cuales se redujeron a través del proceso de recocido simulado entre 3 a 9 variables en la mayoría de los casos.

De esta manera se determinaron las variables con mayor influencia para la predicción del precio en cada acción, en el periodo comprendido. Con esta nueva base de datos de variables seleccionadas, se implementó el MGB, para la predicción de precios, con dos metaheurísticas, GA y HS, las cuales tuvieron un buen desempeño en la predicción, de estas dos técnicas, Búsqueda Armónica obtuvo mejores resultados para la construcción del MGB, mejorando los resultados en la predicción. Se generaron modelos *ad hoc* para cada una de las acciones analizadas, los resultados de selección para cada una de las variables están disponibles en “<http://bit.ly/396h3J1>”. Como se mostró en las tablas de hipervolumen y distancia generacional invertida, ambas técnicas resultan eficientes para resolver el problema multiobjetivo. Sin embargo, cabe destacar que el mejor comportamiento lo presentan los resultados obtenidos bajo la metodología de Búsqueda Armónica. Considerando lo anterior, las áreas de oportunidad para futuros trabajos involucran

adaptar los procedimientos de restricciones ϵ con los presentados en este trabajo para mejorar la calidad de los resultados y mejorar el frente de Pareto obtenido. Con base en la interpretación de los resultados obtenidos, se concluye que la hipótesis fue aceptada.

Referencias bibliográficas

- Agichtein, E., Brill, E., Dumais, S., y Ragno, R. (2006). "Learning User Interaction Models for Predicting Web Search Result Preferences". *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, vol. 2006, pp. 3–10.
- Cerezo, P. (2016, December). La generación Z y la información. Los auténticos nativos digitales: ¿estamos preparados para la Generación Z? *Revista de Estudios de Juventud*. pp. 95–109. Madrid. Recuperado de <http://www.injuve.es/sites/default/files/2017/29/publicaciones/revistainjuve114.pdf>
- Chakrabarti, P., y Guhathakurata, K. (2019). "Which is the Right Option for Indian market: Gaussian, Normal Inverse Gaussian, or Tsallis? *IIMB Management Review*, vol. 31, núm. 3, pp. 238–249. Elsevier Ltd.
- Chen, K.-Y. (2007). "Forecasting Systems Reliability based on Support Vector Regression with Genetic Algorithms". *Reliability Engineering and System Safety*, vol. 92, núm. 4, pp. 423–432. Elsevier. Recuperado en octubre 1, 2019, de <https://www.sciencedirect.com/science/article/pii/S0951832006000147>
- Choudhry, R., y K. Garg. (2008). "A Hybrid Machine Learning System for Stock Market Forecasting". *World Academy of Science, Engineering and Technology*, vol. 2, núm. 15, pp. 315–318.
- Comisión Nacional de Vivienda. (2016, October 25). "La BMV ofrece una gran oportunidad de crecimiento para empresas del sector vivienda. Ciudad de México". Retrieved from <https://www.gob.mx/conavi/prensa/la-bmv-ofrece-una-gran-oportunidad-de-crecimiento-para-empresas-del-sector-vivienda?idiom=es>
- de-los-Cobos-Silva, S. G., Mora-Gutiérrez, R. A., Gutiérrez-Andrade, M. A., Rincón-García, E. A., Ponsich, A., y Lara-Velázquez, P. (2018). "Development of Seven Hybrid Methods Based on Collective Intelligence for Solving Nonlinear Constrained Optimization Problems". *Artificial Intelligence Review*, vol. 49, núm. 2, pp. 245–279. Springer Netherlands.
- Dehghani, H., y Bogdanovic, D. (2018). "Copper Price Estimation Using Bat Algorithm". *Resources Policy*, núm. 55, pp. 55–61. Elsevier Ltd.

- Díaz Gandasegui, V. (2011). "Mitos y realidades de las redes sociales. Información y comunicación en la Sociedad de la Información". *Prisma Social: revista de ciencias sociales*, núm. 6, pp. 340–366. Recuperado en febrero 26, 2020, de <https://dialnet.unirioja.es/servlet/articulo?codigo=3686439>
- Economática. (n.d.). *Economática.com*. Recuperado en octubre 23, 2019, de <https://economática.com/>
- Elsevier. (n.d.). *ScienceDirect*. Recuperado en marzo 3, 2020, de <https://www.sciencedirect.com/>
- Ervural, B. C., Beyca, O. F., y Zaim, S. (2016). "Model Estimation of ARMA Using Genetic Algorithms: A Case Study of Forecasting Natural Gas Consumption". *Procedia - Social and Behavioral Sciences*, núm. 235, pp. 537–545. Elsevier. Recuperado en octubre 1, 2019, de <https://www.sciencedirect.com/science/article/pii/S1877042816316007>
- Fonseca, M. N., Pamplona, E. de O., Valerio, V. E. de M., Aquila, G., Rocha, L. C. S., y Rotela Junior, P. (2017). "Oil Price Volatility: A Real Option Valuation Approach in an African Oil Field". *Journal of Petroleum Science and Engineering*, núm. 150, pp. 297–304. Elsevier B.V.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial systems: An Introductory Analysis with Applications to Biology, Control, and Artificial intelligence*. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Oxford, England: U Michigan Press.
- Hu, Y., Volinsky, C., y Koren, Y. (2008). "Collaborative Filtering for Implicit Feedback Datasets". *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 263–272. Recuperado en marzo 9, 2020, de <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.167.5120>
- Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer-Verlag New York. Recuperado en marzo 5, 2020, de <https://www.springer.com/gp/book/9780387954424#otherversion=9780387224404>
- Jolliffe, I. T. (1990). "Principal Componente Analysis: A Beginner's Guide — I". Introduction and Application. *Weather*, vol. 45, núm. 10, pp. 375–382. John Wiley & Sons, Ltd. Recuperado en marzo 5, 2020, de <http://doi.wiley.com/10.1002/j.1477-8696.1990.tb05558.x>
- Juárez Hernández, L. G., y Ponce Lopez, I. (2019). "Construcción y validación de un instrumento para evaluar el abordaje de la sociedad del conocimiento en docentes". *Apuntes Universitarios*, vol. 10, núm. 1, pp. 40–66. Recuperado en febrero 26, 2020, de <https://dialnet.unirioja.es/servlet/articulo?codigo=7237431>

- Kara, Y., Acar Boyacioglu, M., y Baykan, Ö. K. (2011). "Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange". *Expert Systems with Applications*, vol. 38, núm. 5, pp. 5311–5319.
- Madero Suarez, I. P., y Gómez López, L. F. (2013). "El proceso de comprensión lectora en alumnos de tercero de secundaria". *Revista Mexicana de Investigación Educativa*, vol. 18, núm. 56, pp. 113–139. Recuperado en marzo 26, 2020, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-66662013000100006
- Mason, C. F., y Wilmot, N. A. (2016). "Price discontinuities in the market for RINs". *Journal of Economic Behavior and Organization*, núm. 132, pp. 79–97. Elsevier B.V.
- Montuschi, L. (2001). *Datos, información y conocimiento. De la información a la sociedad del conocimiento*. Recuperada de <https://econpapers.repec.org/RePEc:cem:doctra:192>
- Muñoz Poblete, C. (2013). "Métodos mixtos: una aproximación a sus ventajas y limitaciones en la investigación de sistemas y servicios de salud". *Revista Chilena de Salud Pública*, vol. 17, núm. 3, pp. 218–223. Universidad de Chile.
- Nair, B. B., Mohandas, V. P., y Sakthivel, N. R. (2010). "A Decision Tree- Rough Aet Hybrid System for Stock Market Trend Prediction". *International Journal of Computer Applications*, vol. 6, núm. 9, pp. 1–6. Recuperado en octubre 1, 2019, de <http://www.ijcaonline.org/volume6/number9/pxc3871449.pdf>
- Ntemi, M., y Kotropoulos, C. (2020). "A dynamic dyadic particle filter for price prediction". *Signal Processing*, núm. 167. Elsevier B.V.
- Patel, J., Shah, S., Thakkar, P., y Kotecha, K. (2015). "Predicting Stock Market Index Using Fusion of Machine Learning Techniques". *Expert Systems with Applications*, vol. 42, núm. 4, pp. 2162–2172. Recuperado de <http://www.sciencedirect.com/science/article/pii/S0957417414006551>
- Pérez Fernández, D., Cortés López, J. C., y Villanueva Micó, R. (2015). *Cálculo estocástico en finanzas: Aplicación del modelo browniano activo subyacente FCC.MC en el IBEX-35*. Universidad Politécnica de Valencia. Recuperado en noviembre 27, 2019, de <http://cotizaccion.imm.upv.es/Dani>
- Quiroga, L. M. (2005). "Sistemas de filtrado: un puente tecnológico entre la oferta y demanda de información en línea al servicio de la toma de decisiones". *Las nuevas tecnologías de la información y de la comunicación (TIC) y la institucionalidad social: hacia una gestión basada en el conocimiento-LC/L*, pp. 49–57. Recuperado en febrero 26, 2020, de <https://repositorio.cepal.org/handle/11362/6109>

- Rennie, J. D. M., y Srebro, N. (2005). "Fast Maximum Margin Matrix Factorization for Collaborative Prediction". *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning* (pp. 713–720).
- Rodríguez Gómez, D. (2006). *Modelos para la creación y gestión del conocimiento: una aproximación teórica*. Educar, vol. 37).
- Rubio Guerrero, G. (2017). "Perspectiva multivariante de los pronósticos en las PYMES industriales de Ibagué (Colombia)". *Revista Facultad de Ciencias Económicas*, núm. 2, pp. 25–40. Recuperado en febrero 26, 2020, de <https://dialnet.unirioja.es/servlet/articulo?codigo=6128520>
- Sikora, G., Michalak, A., Bielak, Ł., Miśta, P., y Wyłomańska, A. (2019). "Stochastic Modeling of Currency Exchange Rates with Novel Validation Techniques". *Physica A: Statistical Mechanics and its Applications*, núm. 523, pp. 1202–1215. Elsevier B.V.
- Song, Y., Liu, T., Liang, D., Li, Y., y Song, X. (2019). "A Fuzzy Stochastic Model for Carbon Price Prediction Under the Effect of Demand-related Policy in China's Carbon Market". *Ecological Economics*, núm. 157, pp. 253–265. Elsevier B.V.
- Tobón, S., Guzmán, C. E., Silvano Hernández, J., y Cardona, S. (2015). "Sociedad del conocimiento: Estudio documental desde una perspectiva humanista y compleja". *Paradigma*, vol. 36, núm. 2, pp. 7–36. Recuperado en febrero 26, 2020, de http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1011-22512015000200002
- Yannakakis, G. N. (2009). "Preference Learning for Affective Modeling". *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*. Recuperado en marzo 9, 2020, de <https://www.researchgate.net/publication/224088130>
- Yeh, C. Y., Huang, C. W., y Lee, S. J. (2011). "A Multiple-Kernel Support Vector Regression Approach for Stock Market Price Forecasting". *Expert Systems with Applications*, vol. 38, núm. 3, pp. 2177–2186.
- Zong Woo Geem, Z. W., Joong Hoon Kim, J. H., y Loganathan, G. V. (2001). "A New Heuristic Optimization Algorithm: Harmony Search". *SIMULATION*, vol. 76, núm. 2, pp. 60–68. Sage PublicationsSage CA: Thousand Oaks, CA. Recuperada en octubre 1, 2019, de <http://journals.sagepub.com/doi/10.1177/003754970107600201>

ANEXO 1

En este trabajo se emplearon estas funciones para ampliar la base de datos inicial. $\sin X$

X^2	El valor de X elevado al cuadrado
X^3	El valor de X elevado al cubo
$\begin{cases} \log(X), X \neq 0 \\ 0, X = 0 \end{cases}$	Logaritmo del valor absoluto de X , para X distinto de cero
X^{-1}	El valor de X elevado a la menos 1
X^{-2}	El valor de X elevado a la menos 2
X^{-3}	El valor de X elevado a la menos 3
$X_i X_j \{i \neq j\}$	El producto de , siempre que i no sea igual a j
	$\sin X$. El seno del valor de X
	$\cos X$. El coseno del valor de X
	$ \sqrt{X} $. El valor absoluto de la raíz cuadrada de X
$\begin{cases} \frac{1}{X}, X \neq 0 \\ 0, X = 0 \end{cases}$	El cociente de 1 sobre el valor de X , para X distinto de cero
$\begin{cases} \frac{1}{\log X }, \log X \neq 0 \\ 0, X = 0 \end{cases}$	El cociente de 1 sobre el valor del logaritmo del valor absoluto de X , siempre que el logaritmo del valor absoluto de X sea distinto de cero
$\begin{cases} \frac{1}{1 - e^{- X }}, 1 - e^{- X } \neq 0 \\ 0, 1 - e^{- X } X = 0 \end{cases}$	El cociente de 1 sobre 1 menos la constante e elevado a la menos el valor absoluto de X , siempre que 1 menos la constante e elevado a la menos el valor absoluto de X sea distinto de cero

