



Minería de datos en series de tiempo

Data's Minery in Series of Time

Alejandra Coria Olivares¹ Doricela Gutiérrez Cruz² Carmen Liliana Rodríguez Páez³

Para citar este artículo: A. Coria; D. Gutiérrez, y C. L. Rodríguez (2015). Minería de datos en series de tiempo. *Revista Vínculos*, 12(2), 148-156.

Recibido: 12-08-2015 / **Modificado:** 22-08-2015 / **Aprobado:** 20-09-2015

Resumen

En este trabajo se presenta el análisis en series de tiempo del precio del maíz (*Zea Mays L.*). Este cultivo es considerado de gran importancia económica a nivel mundial. Se seguirá el proceso clásico de la minería de datos determinando el coeficiente de Hurst (H) obteniendo valores entre $0 < H < 1$. Considerando lo anterior, resulta necesario entender las fluctuaciones en cuanto al precio del maíz que permitan identificar su comportamiento futuro en el mercado, como una enseñanza didáctica de la minería de datos mediante la prueba estadística rango reescalado (R/S), teniendo como resultado en el estado de Sinaloa $H=0,494$, lo que permite identificar una inestabilidad futura en el mercado con base en el costo del cereal; Jalisco con $H=0,580$ y Michoacán con $H=0,527$, lo que muestra estabilidad futura en el costo del maíz.

Palabras clave: coeficiente de Hurst, desviación estándar, rango reescalado.

Abstract

This job represents the analysis in series of time of the corn price (*Zea Mays L.*) Considered in great economy importance in word level, we can follow the classic proses in the minery data the hurts coefficient. Getting values between $0 < H < 1$. Considering the last we have said is necessary to understand the fluctuations about the corn price letting to identify the behavior in the future market, like a didactic teaching of the minery data through. the statistic range prove (S/R), having the result of $H=0,494$ letting to identify the future instability in the market in base to the cereal coast, Jalisco with $H=0,580$ and Michoacán with $H=0,527$ showing future stability in the coast of the corn.

Keywords: Hurst coeficiente, standard desviation, rescaled range.

- 1 Estudiante, Ingeniería en Sistemas Inteligentes, Unidad Académica Profesional Nezahualcóyotl, Nezahualcóyotl Edo de México, México. Correo: coria_922@hotmail.com.
- 2 Profesor investigador, Dra. en Ingeniería de Sistemas, Unidad Académica Profesional Nezahualcóyotl, Nezahualcóyotl Edo de México, México. Correo: gutierrezcruzdo@yahoo.com.mx.
- 3 Profesor investigador, Dra. en Ingeniería de Sistemas, Unidad Académica Profesional Nezahualcóyotl, Nezahualcóyotl Edo de México, México. Correo: carmenlilianapa@hotmail.com.

1. INTRODUCCIÓN

Los métodos estadísticos son una herramienta impredecible en la generación y aplicación del conocimiento [1-2], como es el caso de la minería de datos (MD), la cual consiste en el descubrimiento de estructuras interesantes, inesperadas o valiosas en grandes conjuntos de datos, proporcionando herramientas para extraer información [3-4]. Durante este proceso se descubren nuevas y significativas relaciones, patrones y tendencias [5], las cuales normalmente no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados de ellos [6]. Por esta razón, las técnicas más usuales se orientan a la predicción automática de tendencias, comportamientos y al descubrimiento de patrones desconocidos existentes en bases de datos [7]. Las técnicas predictivas, como su nombre lo indica, predicen el valor de un atributo de un conjunto de datos a partir de otros valores ya conocidos, induciendo una relación entre ellos. Para Riquelme [8], es considerada como una disciplina de importancia estratégica para la planeación y la toma de decisiones, basándose en la rapidez para identificar, procesar y extraer la información que realmente es importante, descubriendo conocimiento y patrones en bases de datos [9].

La MD es utilizada como apoyo en técnicas relacionadas como: la estadística, el reconocimiento de patrones, algoritmos de aprendizaje y comportamientos sistémicos que pongan de manifiesto interrelaciones entre los datos que sirvan para predecir comportamientos futuros en la administración de series de tiempo [10]. Lo anterior la convierte en un área de investigación importante que consta principalmente de cuatro fases: recopilación, limpieza, explotación y transformación y evaluación e interpretación de los datos [11]. Es una disciplina de importancia estratégica para la planeación y toma de decisiones basándose en la rapidez para identificar, procesar y extraer información significativa a partir de datos complejos [12]. En series de tiempo proporciona métodos que integran las técnicas en

una herramienta computacional para demostrar por métodos experimentales que dichas técnicas funcionan correctamente [13]. Considerando lo anterior como un sistema complejo cuyos elementos interactúan de manera no lineal [14] bajo un enfoque fractal que mide el grado de detalle estructural que se va agregando conforme va incrementando la escala, entonces, la dimensión fractal sirve como cuantificador de patrones caóticos complejos [15]. Particularmente se espera que el comportamiento espacio-temporal de sistemas complejos permita predecir y prevenir adecuadamente fallas posibles en estructuras ingenieriles [16].

Dentro de los métodos que pueden mencionarse se encuentra el análisis del recorrido estandarizado, también llamado análisis de rango reescalado, o análisis R/S, es una prueba estadística utilizada para cuantificar la dinámica de una serie temporal y determinar la existencia de características fractales en un sistema [17]. El valor de esta prueba es su sensibilidad para distinguir correlacionales que se presentan como consecuencia de la tendencia que presentan las observaciones a desviarse del valor medio durante un tiempo más o menos prolongado, mide el rango de las desviaciones de las sumas parciales de una serie temporal respecto a su media [18]. Consiste en obtener el coeficiente de escalamiento, también denominado, coeficiente de Hurst que puede tomar cualquier valor entre 0 y 1 [19].

2. COEFICIENTE DE HURTS Y DIMENSIÓN FRACTAL

El análisis R/S fue desarrollado por Harold Hurst [20], sin embargo, no fue sino hasta 1968 [21], cuando Benoit B. Mandelbrot, Van Ness [22], Mandelbrot y Wallis lo introdujeron como herramienta de análisis estadístico. En 1957, Mandelbrot demostró que el valor del exponente de Hurts se encuentra acotado entre 0 y 1, explicando que este fenómeno ocurría debido a un síntoma de invariancia al cambio de escala, que es característica clave de las series de tiempo fractales.

El coeficiente de Hurst puede ser determinado a partir de la dimensión fractal o viceversa [23], utilizando la fórmula:

$$D = 2 - H \quad (1)$$

Dado lo anterior, si $H=0,5$ entonces $D=1,5$, lo que significa que los dos valores coinciden con un proceso aleatorio e independiente (ruido blanco). Para los valores que se encuentren entre $0,5 < H \leq 1$, es decir, series de tiempo persistentes o correlacionado, la dimensión fractal se acercará a una línea con un aspecto de suavidad y menos picos que un proceso aleatorio (ruido negro). Para un serie que tome valores entre $0 < H < 0,5$, es decir, una serie anti-persistente o anti-correlacional, la dimensión fractal arrojará un valor mayor que el anterior y tomará un aspecto más puntiagudo e irregular que un proceso aleatorio, con tendencia a acercarse cada vez más a una superficie (ruido rosa).

Para el presente trabajo, se aplicó el proceso general del KDD (*Knowledge Discovery in Databases*) [24], con el objetivo de adquirir como resultado las fluctuaciones en cuanto al precio del maíz a través de técnicas que permitan identificar su comportamiento futuro, obteniendo el coeficiente de Hurst que puede tomar valores entre 0 y 1 [19], como una enseñanza didáctica de la minería de datos.

3. METODOLOGÍA

Las técnicas usadas en el minado de datos son muy variadas, pues no todas son aplicables en cualquier conjunto de datos. El empleo de métodos estadísticos de uso frecuente son los siguientes: descripción uni- y multivalente de datos, diversos contrastes de hipótesis, modelos de regresión y de regresión logística, análisis discriminante, técnicas de reducción de dimensión como el análisis en componentes o en coordenadas principales, o el análisis vectorial, series temporales, árboles de decisión, algoritmos genéticos y otras técnicas estadísticas de visualización y representación de datos [25].

La MD, como proceso, intenta cubrir todos los enfoques o puntos de vista. El proceso de descubrimiento es una secuencia iterativa que comprende los pasos siguientes: este trabajo fue analizado con una herramienta basada en software licenciado denominado Benoit®, software de análisis fractal que permite medir la dimensión fractal y/o exponente de Hurst de conjuntos de datos utilizando su elección del método(s) para el análisis de los patrones auto-similares. Aunque el proceso de la MD en cada caso concreto puede ser radicalmente distinto, el proceso común en todos ellos se suele componer de cinco etapas principales, las cuales se muestran en la Figura 1.

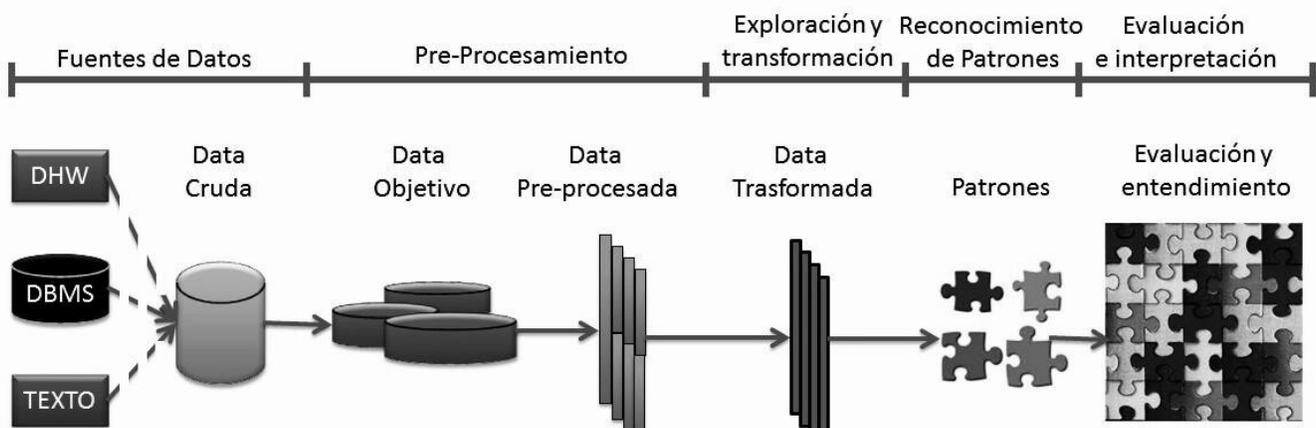


Figura 1. Proceso de la minería de datos.

Fuente: elaboración propia.

3.1 Fase 1: recopilación e integración (fuentes de datos-extracción)

Integración de los datos. Múltiples fuentes de datos pueden ser combinadas. La integración de datos se define como el proceso de combinar datos que residen en diferentes fuentes y permitirle al usuario final tener una vista unificada de todos [26]. De forma general, los datos son la materia prima bruta, en el momento que el usuario les atribuye algún significado especial pasan a convertirse en información. En este trabajo se presenta el análisis en series de tiempo del maíz (*Zea mays L.*) considerado de gran importancia económica a nivel mundial [27] por la gran variedad de productos que se obtienen de él, como fuente de un gran número de productos industriales, para ganado o bien para consumo humano. Con una superficie cultivada en más de 140 millones de hectáreas, es considerado como el cereal de mayor demanda a nivel mundial con una producción de 817,1 millones de toneladas, con un consumo de 16,76 kg año⁻¹ [28].

México es el tercer importador de maíz de Estados Unidos [29]. Se produjeron 20,1 millones de toneladas [30] con el 18% del valor de producción del sector agrícola (88 mil mdp en 2012 y 78 mil en 2013) y concentra el 33% de la superficie sembrada en el territorio nacional, de lo cual el estado de Sinaloa es el principal productor al concentrar el 16,5% del total, le sigue en importancia Jalisco con el 14,7% y Michoacán con el 8,2% [31]. Poco más del 50% de la producción nacional se destina para consumo humano, lo cual representa una ingesta superior a los 128 kg de maíz/año correspondiente a Kcal/persona/día.

Considerando lo anterior, resulta necesario entender las fluctuaciones en cuanto al precio de este a través de técnicas que permitan identificar su comportamiento, como una enseñanza didáctica de la minería de datos.

Los datos correspondientes al precio del maíz fueron adquiridos del Sistema Nacional de Información e Integración de Mercados, que vienen desglosados por año y mes, tomando como referencia al estado

de Sinaloa, Jalisco y Michoacán los cuales son los tres estados de la república mexicana con mayor producción del maíz, teniendo como reseña la primera semana del mes de enero de 1998 hasta el mes de mayo del 2015. Se obtienen tres series de tiempo de una cantidad de 908 datos cada una, teniendo un total de 2,724 datos por lo que se hizo uso de Excel® para su ordenamiento y dejarlos listos para su posterior tratamiento.

3.2 Fase 2: limpieza, selección y transformación (preprocesamiento)

Procediendo a la limpieza en Excel®, en la columna donde se encuentra cada estado de la República (Sinaloa, Jalisco y Michoacán) se homologa un mismo formato para todos los datos, en el precio se cambia la coma por un punto para estandarizar el *precio* y se eliminan los símbolos con que se escribió la cantidad en costo excluyendo así el ruido que podrían estar causando, dejando solo la cantidad en su forma más básica de número.

Los datos originalmente venían dados por mes, por lo que para fines de estudio se alteraron esos meses en semanas ordenándolos en dos columnas, una para las fechas y otra para la cantidad en precio, reguladas en forma ascendente y de forma continua. Generalmente, es necesario realizar alguna transformación a los datos para obtener materia prima adecuada. El éxito de un proceso de MD depende de tener datos íntegros, completos y consistentes para poder eliminar información incorrecta o inconsistente [32].

- Limpieza de los datos. Se removieron los datos inconsistentes.
- Selección de datos. Se recuperan los datos relevantes para su análisis posterior.
- Transformación de datos. Los datos son transformados y consolidados en formas apropiadas para las operaciones de resumen y agregación.

La limpieza, selección y transformación de datos pueden detectar y solucionar problemas de datos

no resueltos durante la integración evitando problemas como datos faltantes, valores duplicados y datos incorrectos [33]. En esta etapa se juntan todos los datos en un solo archivo .txt por cada serie de tiempo, posteriormente se modifica a extensión .ts, verificando que no exista alguna inconsistencia o datos faltantes para así poder trabajarlos posteriormente con el Software Benoit®.

3.3 Fase 3: explotación y transformación

Los resultados se obtienen al analizar las series de tiempo con el software Benoit®, así mismo permite medir la dimensión fractal y/o el coeficiente de Hurst de un conjunto de datos usando algunos métodos que tiene disponibles para el análisis de patrones auto-similares o auto-fines e incluye el método R/S que es el objeto de estudio para el minado de datos.

- Minería de datos. Proceso esencial donde métodos inteligentes son aplicados para extraer patrones o modelos entre los datos.

Para mayor precisión y resultados más fiables se trabaja la serie de tiempo con el Software Benoit®, el cual realiza un procesamiento mucho más extenso de la serie de tiempo por medio del método R/S.

- Evaluación de los patrones o modelos.

Ya identificados los patrones o modelos realmente interesantes que representen conocimiento basados en medidas de interés, se procede a compilar en el Software Benoit®.

Se introduce la serie de tiempo, se selecciona el algoritmo R/S y al ejecutarlo nos muestra el coeficiente de Hurst (H), la desviación estándar (SD) y la dimensión fractal (DF).

- Presentación del conocimiento.

Utilizando el método Rango re-escalado en Benoit® para analizar la información obtenida en cada estado de los cuales se hizo este trabajo, los resultados se muestran en la tabla 1, que muestra los valores obtenidos para el coeficiente de Hurst (H), dimensión fractal (FD) y desviación estándar (SD) de cada estado:

3.4 Fase 4: evaluación e interpretación de los resultados

Tomando los valores obtenidos del método R/S de cada uno de los estados (tabla 3), se hace la descripción de los resultados. Es importante aclarar que R/S es un estadístico con media cero, expresado en términos de desviación estándar y cuyo análisis es de tipo no paramétrico, ya que no se requiere una distribución específica. Donde H es el número asociado a una serie temporal [34], FD es el número que refleja la media topológica de un conjunto fractal a escalas distintas (irregularidad en los datos) [35] y SD es un índice numérico de la dispersión de un conjunto de datos [36].

Si $0 \leq H < 0,5$ como en el caso del estado de Sinaloa es no correlacionado por lo que la dimensión fractal aumentara, por lo cual la representación gráfica se acercará a una superficie [37]. A esto se le llama ruido rosa. Se dice entonces que es una

Tabla 1. Tabla de valores de H, FD y SD de los estados: Jalisco, Michoacán y Sinaloa.

Estado	Jalisco			Michoacán			Sinaloa		
	Coeficiente de Hurst	Dimensión fractal	Desviación estándar	Coeficiente de Hurst	Dimensión fractal	Desviación estándar	Coeficiente de Hurst	Dimensión fractal	Desviación estándar
Resultados	0,580	1,492	0,0183422	0,527	1,473	0,0480625	0,494	1,506	0,0807737

Fuente: elaboración propia.

serie antipersistente, con frecuencia se les denomina con reversion a la media. Es decir, la serie ha estado arriba de un determinado valor que hace las veces de la media de largo plazo en el periodo anterior, entonces es más probable que esté abajo en el periodo siguiente y viceversa. Mientras mayor es la SD, mayor es la dispersión de los datos. Por tanto, se puede apreciar que las fluctuaciones en cuanto al precio del maíz no permiten identificar el comportamiento futuro en el mercado ya que el costo del maíz se presentara de manera inestable. Si $0,5 < H \leq 1$, como es en el caso del estado de Jalisco y Michoacán, se dice que tiene correlación positiva o que la serie es persistente (fuerza de dependencia). Es decir, si la serie esta arriba (o abajo) de su media de largo plazo en el periodo anterior,

lo más probable es que continúe arriba (o abajo) en el periodo siguiente (periodos estables), entre más se acerque H a 1 la serie será determinística, en este intervalo el color de ruido es negro. Por tanto se puede apreciar que las fluctuaciones en cuanto al precio del maíz permiten identificar el comportamiento futuro en el mercado ya que el costo del maíz se presentara de manera estable. Mientras mayor es la SD, mayor es la dispersión de los datos teniendo un incremento de la DF. Podemos decir que las fluctuaciones en cuanto al precio del maíz a través del método R/S permitieron identificar el comportamiento futuro de los mercados de los estados de Sinaloa, Jalisco y Michoacán. Otro resultado arrojado por Benoit® son las gráficas de persistencia que se muestran en la Figura 2.

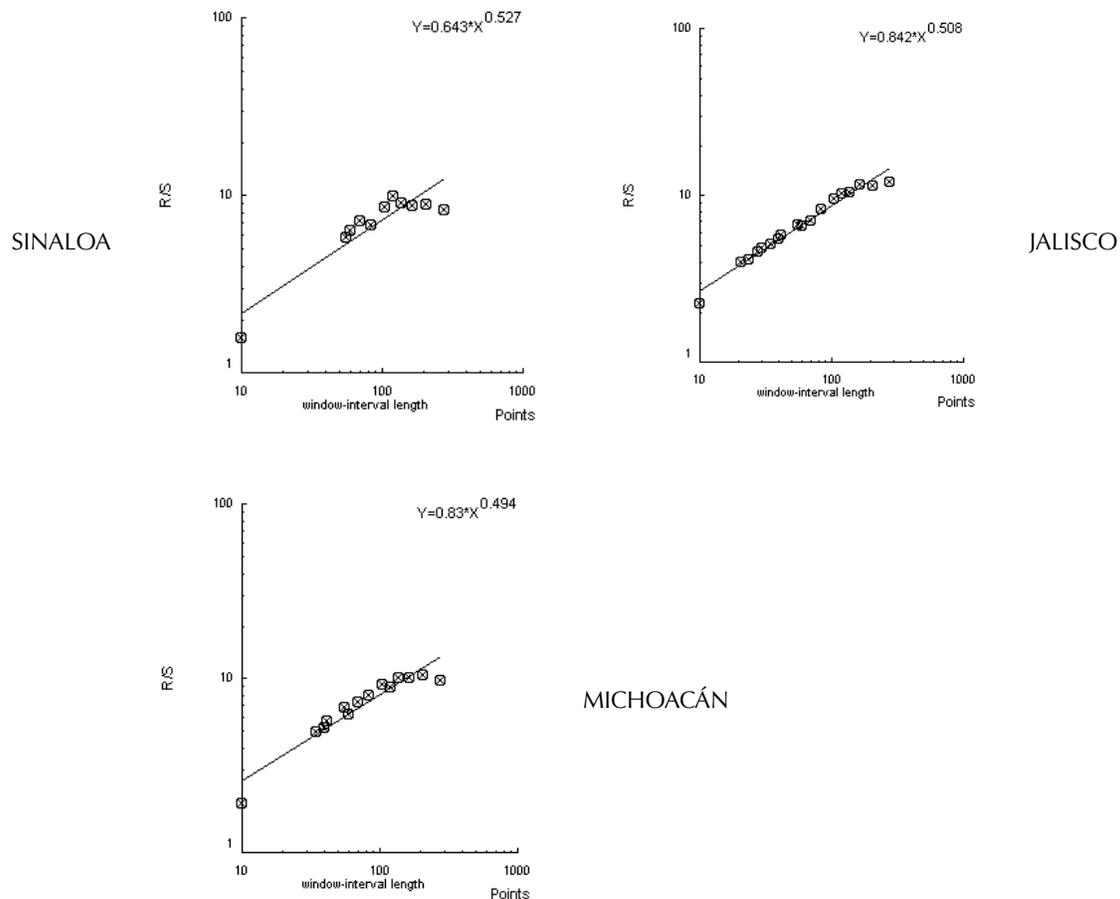


Figura 2. Graficas de persistencia.

Fuente: elaboración propia.

Si $0 \leq H < 0,5$, como en el caso del estado de Sinaloa, se dice entonces que es una serie antipersistente ya que los datos se encuentran dispersos. Teniendo una representación gráfica de manera no lineal, por tanto se puede apreciar que las fluctuaciones en cuanto al precio del maíz no permiten identificar el comportamiento futuro en el mercado ya que el costo de este se presentará de manera inestable.

Si $0,5 < H \leq 1$, como es el caso de los estados de Jalisco y Michoacán, se dice entonces que son series persistentes ya que los datos se encuentran menos dispersos, teniendo una representación gráfica de manera más lineal, por tanto se puede apreciar que las fluctuaciones en cuanto al precio del maíz permiten identificar el comportamiento futuro en el mercado ya que el costo del maíz se presentará de manera estable.

4. CONCLUSIONES

A pesar de las grandes diferencias en el precio del maíz entre los tres estados analizados existen regularidades estadísticas básicas, lo que corrobora que el análisis R/S es una herramienta de análisis estadístico para evaluar la ocurrencia de eventos poco comunes y la complejidad de la minería de datos en el estudio de las series de tiempo estructural del precio en diferentes mercados nacionales.

El análisis R/S permitió demostrar que las series de tiempo están caracterizadas por el fenómeno de memoria o dependencia a largo alcance. La estructura de memoria varía según el precio de la serie de tiempo. Las estructuras de memoria de las secuencias fueron específicas para cada estado de la república mexicana, las cuales estadísticamente presentaron un comportamiento persistente o antipersistente. Estos resultados fueron consistentes en base a la complejidad del precio. La persistencia está asociada a estructuras estables con alta probabilidad de cumplir funciones específicas, mientras que la antipersistencia se relaciona con estructuras inestables que buscan funcionalidad.

Los parámetros R/S permiten definir matemáticamente conjuntos de secuencias entre los precios

del maíz, lo que indica que el procedimiento del análisis R/S potencialmente podría utilizarse como una herramienta que complemente los métodos clásicos de comparación de series de tiempo, o podría implementarse en los algoritmos de uso común de las herramientas de minería de datos para mejorar la capacidad de predicción; también podría utilizarse para influir sustancialmente en las estadísticas de las puntuaciones del alineamiento de secuencias y complementar los análisis de comparación de secuencias que sean complejas.

Por lo que podemos decir que las fluctuaciones en cuanto al precio del maíz a través del método R/S permitieron identificar el comportamiento futuro en los mercados en cuanto al costo del cereal en los estados de Sinaloa, Jalisco y Michoacán, obteniendo una enseñanza didáctica de la minería de datos.

Otros ejemplos de la aplicación del método R/S son: predicción de genes de genoma vegetal mediante un procedimiento estadístico con base al análisis R/S y el coeficiente de Hurts para el desarrollo de caracterizar, predecir genes y los componentes estructurales de exones e intrones en los genomas eucariotas completos de *Arabidopsis thaliana*, *Oryza sativa* y *Mus musculus* [34].

El estimando del exponente de Hurts y la dimensión fractal de una superficie topográfica a través de la extracción de perfiles que nace como una alternativa para interpretar y representar de una manera más precisa las diferentes formas presentes en la naturaleza de las simulaciones de campos aleatorios, tipo movimiento browniano fraccionado, aplicadas al estudio de observables espaciales en geociencias [35], entre otros estudios se encuentran: que mediante invarianza temporal al escalado de series de tiempo con información pluviométrica diaria determinando el grado de invarianza al escalado de una serie de tiempo de precipitación pluvial permitiéndole determinar en forma consistente y a diferentes escalas de tiempo la precipitación pluvial [36].

El coeficiente de Hurts y el parámetro α -estable para el análisis de series financieras buscan suponer que las series de tiempo presentan una distribución normal y que los rendimientos sucesivos

son independientes. Asimismo, se analiza el caso del tipo de cambio Fix peso-dólar en México, obteniendo como resultado el análisis de persistencia y antipersistencia de una serie financiera en periodos de alta volatilidad [37].

5. AGRADECIMIENTOS

Los autores agradecen a la Universidad Autónoma del Estado de México (UAEM) y al PROMEP por el financiamiento otorgado.

REFERENCIAS

- [1] J. M. Caridad & J. M. Ocerín, "Econometría: modelos econométricos y series de temporales", Barcelona: Reverté, 1998.
- [2] J. Hand & M. Kamber, "Data Mining. Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
- [3] J. L. Gordillo, E. Martínez & C. R. Stephens, "Develando Estrategias de Mercado: Minería de Datos aplicada al Análisis de Mercados Financieros". *Scielo*, México, vol. 16 no 2, pp. 221-231, 2012.
- [4] S. Sumathi, & S. Sivanandam, "Introduction to Data Mining and its Applications", Berlin: Springer. 2006.
- [5] J. Hernández; M. J. Ramírez & C. Ferri, "Introducción a la Minería de Datos", Madrid: Pearson. 2004.
- [6] Microsoft Corporation, "Conceptos de Minería de Datos. SQL Server 2012", 2015, [En línea] Disponible en: <http://msdn.microsoft.com/es-es/library/ms174949.aspx>
- [7] Y. J. Marcano & R. Talavera, "Minería de Datos como soporte a la toma de decisiones empresariales" *Scielo*, Opción, Maracaibo, vol. 23, no 52, 2007.
- [8] J. C. Riquelme & K. Gilbert, "Minería de datos: conceptos y tendencias", *Redalyc*, vol. 10, no 29, pp. 11-18, 2006.
- [9] F. Hernández & Y. Sánchez, "Minería de datos técnicos aplicada al diagnóstico de entidades clínicas", *Scielo*, vol.4, no 2, pp. 174-183, 2012.
- [10] C. A. Espinoza & J. P. Mcphee, "Estudio de la geometría fractal en roca fracturada y series de tiempo", Tesis M. SC., Universidad de Chile, 2008.
- [11] G. Cáceres & J. E. Rodríguez., "Agrupamiento de datos de series de tiempo. Estado del arte", *Vínculos*, vol. 8, no 1, 2011.
- [12] T. I. Contreras, "Análisis fractal de un sistema complejo: epilepsia". Tesis M. SC., Instituto Politécnico Nacional, 2007.
- [13] J. M. Rodríguez, "Clasificación de series de tiempo por minería de datos". Tesis M. SC., Instituto Politécnico Nacional, México D.F.: 2006.
- [14] O. Matamoros; A. Balankin; L. M. Hernández, "Metodología de predicción de precios del petróleo basada en dinámica fractal", *Redalyc*, vol. 9, no 1, pp. 3-11, 2005.
- [15] R. García, "Sistemas complejos", Barcelona: Gedisa, p. 202, 2006.
- [16] A. Balankin, K. Oleshko, A. Ortega, O. Susarrey, D. Morales, I. Campos, J. Martínez, J. Marquez, A. García, L. Morales & J. García, "Comportamiento fractal de sistemas complejos", *Científica*, 2000.
- [17] M. I. Almaza, K. López & C. E. Téllez, "Aplicación del análisis de ranfo reescalado R/S para la predicción de genes en el genoma vegetal", *Acta Agron.* vol.59, no 4, Palmira, 2010.
- [18] H.E. Hurts,, "Long-term storage capacity of reservoirs", *Trans. Am. Soc. Civil Engineers*, 116, pp. 770-808, 1951.
- [19] O. Sotolongo, "Experiencia en la enseñanza de la física de los sistemas complejos en la cátedra Henri Poincaré". *Revista Cubana de Física*, vol. 23, no 2, pp.127-134, 2006.
- [20] H. E. Hurts C.M.G., M.A., D.Sc. F. Inst. P., "The problem of long-term storage in reservoirs", *International Association of Scientific Hydrology*, boletín 1(3), pp. 13-27. DOI: 10.1080/02626665609493644. 2010.
- [21] B. B. Mandelbrot, "The Fractal Geometry of Nature". Oxford: International Business Machines Thomas J. Watson Research Center, 1968.

- [22] B. Mandelbrot & J.W. Van Ness, "Fractional noises and applications". *SIAM Review*, vol. 10, pp.422-437, 1968.
- [23] G. Sierra, "Procesos de Hurts y movimientos brownianos fraccionales en mercados fractales", *Revista de Administración, Finanzas y Economía*, vol. 1, no 1, pp. 1-21, 2007.
- [24] D. Luengas, E. Ardila & J. F. Moreno, "Metodología en interpretación del coeficiente de Hurts", *Redalyc*, n.º 5, pp. 265-290, 2010.
- [25] P. Adriaans & D. Zantinge, *Data mining*. Addison Wesley Pub. Co., 1998.
- [26] P. Cabena, P. Hadjinian, R. Stadler, J. Verhess & A. Zanasi, *Discovering Data Mining from Concept to Implementation*. Upper Saddle NJ, Prentice Hall, 1997.
- [27] TLCAN (Tratado de Libre Comercio en América Latina). 2015, [En línea] Disponible en: <http://www.tlcan.com.mx/>.
- [28] FAO (Food and Agriculture Organization of the United Nations): FAOSTAT (FAO Statistical Databases) Agriculture, Fisheries, Forestry, Nutrition. Roma: 2015. [En línea] Disponible en: <http://faostat.fao.org/default.aspx/>.
- [29] M. A. Asturias, "Maíz de alimento sagrado a negocio del hambre", Ecuador, Ed. Acción Ecológica, pp. 105. 2004.
- [30] SIAP (Servicio de Información Agroalimentaria y Pesquera), (2015), [En línea] Disponible en: <http://www.campomexicano.gob.mx/campo/index.php>.
- [31] SHCP (Secretaría de Hacienda y Crédito Público 2015, [En línea] Disponible en: <http://www.shcp.gob.mx/LASHCP/MarcoJuridico/contabilidadGubernamental/SCG2014/Paginas/2014.aspx>.
- [32] J. A. Berry & L. Gordon, *Mastering Data Mining*. John Wiley, 2000.
- [33] U. Fayyad, G. Piatetsky-Shapiro & P. Smyth, "From Data Mining to Knowledge Discovery in Databases". *AI Magazine*, vol. 12, n.º 3, 1996.
- [34] M. I. Almanza, K. López & C. E. Téllez, "Aplicación del análisis de rango reescalado R/S para la predicción de genes en el genoma vegetal", *Redalyc, Acta Agronómica*, vol. 59, n.º4, 2010.
- [35] O. Y. Quintero & J. Ruíz, "Estimación del exponente de Hurst y la dimensión fractal de una superficie topográfica a través de la extracción de perfiles". *Revista Geomática*, n.º5, pp. 84-91, 2011.
- [36] M. A. Velásquez, J. J. Velásquez, J. A. Muñoz, M. R. Martínez, K. Oleschko, I. Sánchez & B. Figueroa, "Invarianza temporal al escalado de series de tiempo con información pluviométrica diaria". *Rev. Méx. Cienc. Agríc.*, vol. 1, n.º 4, 2010.
- [37] R. Rodríguez, "El coeficiente de Hurst y el parámetro α -estable para el análisis de series financieras aplicación al mercado cambiario mexicano", *Scielo*, vol. 59, n.º 1, 2012.

