

Utilización de Técnicas de Soft Computing para la estimación de comportamientos de Valores del IBEX 35

Arturo Peralta¹, Ricardo Rejas², Francisco P. Romero¹, José A. Olivas¹,
Jesús Serrano-Guerrero¹

¹Departamento de Tecnologías y Sistemas de la Información,
Universidad de Castilla La Mancha, España

²Facultad de Derecho y Economía,
Universidad Camilo José Cela, España

Abstract. La predicción de comportamientos bursátiles es uno de los principales retos de todo inversor de renta variable. Por este motivo, en la actualidad existen multitud de métodos que tratan de ayudar en esta labor. Sin embargo, la ocurrencia de sucesos imprevistos hace inválida cualquier estimación inicial basada únicamente en comportamientos previos. En el presente trabajo se propone un método basado en el uso de técnicas de Soft Computing de ayuda para prever el comportamiento de un Valor, considerando no sólo históricos de comportamientos previos, sino también la opinión de posibles inversores o expertos analistas del sector.

Keywords. Predicción bursátil, Extracción de conocimiento, Soft Computing.

1. INTRODUCCIÓN

Como bien es sabido, en el ámbito de los mercados financieros, el objetivo de todo inversor es predecir el comportamiento bursátil. Sin embargo, puesto que los movimientos de la Bolsa se ven influenciados por multitud de factores, predecir su comportamiento exacto conllevaría conocer la psicología individual de cada inversor, así como el modo en que ha invertido. Por este motivo, en la actualidad existen multitud de modelos que tratan de ayudar a detectar tendencias. Sin embargo, la mayoría de ellos se basan en estudios de tipo puramente técnicos o fundamentales, dejando en un segundo plano aspectos como la psicología de masas, las opiniones de expertos, o la aparición de sucesos imprevistos capaces de generar inestabilidad.

Resolver estas debilidades es precisamente el objetivo de este estudio. Para ello, son considerados patrones de comportamiento aprendidos del pasado, y un conjunto de indicadores económicos y variables externas capaces de representar el sentimiento social y el denominado “factor humano”, para así proporcionar un marco de referencia completo de ayuda para la toma de decisiones. Con este objetivo, en este estudio se propone el uso de técnicas de Soft Computing, para la extracción de conocimiento oculto en datos y en comentarios bursátiles proporcionados por la masa de posibles inversores o expertos del sector, de utilidad para la toma de decisiones.

2. TRABAJOS RELACIONADOS

La estimación de comportamientos bursátiles supone un aspecto clave para cualquier inversor. Por este motivo existen numerosos trabajos [1] que tratan, mediante el uso de técnicas basadas en análisis de tipo técnico o fundamental, de dar solución a este problema. Sin embargo, la incertidumbre motivada por el denominado “factor humano” hace interesante investigar soluciones que conjuguen técnicas tradicionales con otras basadas en Soft Computing como la lógica borrosa o algoritmos de clustering. Sin embargo, son escasos los estudios existentes que apliquen estas técnicas para la estimación de comportamientos bursátiles, pudiendo ser englobados en tres grandes grupos dependiendo del tipo de herramientas usadas para resolver la problemática de la estimación de tendencia bursátil. Todos ellos excluyen de sus cálculos los posibles factores externos influyentes.

Dentro del primero de los grupos son englobados los estudios más destacados están centrados en el uso exclusivo de redes neuronales son los siguientes. En [2] se realiza un estudio sobre la aplicación de redes neuronales y el cálculo de algunos indicadores representativos, como VAMA (Volume Adjusted Moving Average) y EMV (Ease of Movement) para su uso en predicciones bursátiles de un conjunto de compañías seleccionadas pertenecientes al índice S&P (Standard and Poor's). Por su parte, en [3] se presenta un estudio basado en el uso de redes neuronales y el análisis de series temporales, para la predicción de comportamientos a muy corto plazo. Por último, destaca [4] proponiendo un modelo borroso basado en redes neuronales para clasificar comportamientos con el objetivo de predecir tendencias a corto plazo.

El segundo grupo de trabajos corresponde a aquellos estudios que presentan modelos híbridos, al incluir, además del uso de redes neuronales, algoritmos genéticos. Así pues, en [5] se propone un modelo que, tomando como entrada un amplio conjunto de datos relativos a análisis técnicos y conocimiento experto, haciendo uso de redes neuronales y algoritmos genéticos, trata de predecir el precio de cierre de un valor. Con este mismo objetivo, en [6] se estudia el uso combinado de ambas técnicas y la extracción de reglas para sugerir el momento de compra o venta, mientras que en [7] el análisis es más centrado en las oscilaciones diarias con un objetivo puramente especulador.

Finalmente, el tercer grupo de trabajos identificados son aquellos que comprenden el uso de alguna técnica de Soft Computing. De entre estos son relevantes estudios como [8] en el que se describe un sistema que, tomando como entrada un conjunto de indicadores financieros, haciendo uso de redes neuronales y métodos estadísticos, trata de calcular una serie de ratios para predecir el comportamiento bursátil de un conjunto de compañías del S&P.

3. MARCO PARA LA PREDICCIÓN DE VALORES BURSÁTILES

La predicción de comportamientos bursátiles supone un proceso de difícil solución. El conjunto formado por los factores externos (datos que pudieran afectar negativamente: porcentaje de paro, datos macroeconómicos, consumo, IPC, PIB, etc.) y el inevitable “factor humano”, hacen que, en la mayoría de los casos, las estimaciones basadas sólo en comportamiento anteriores, puedan distar demasiado respecto a la realidad.

En este contexto surge la investigación de uso de técnicas de Soft Computing a partir de datos históricos y de comentarios bursátiles como una posible solución para mejorar estas previsiones, centrándose en el estudio de situaciones no ideales, afectadas por factores externos que pudieran distorsionarla. A continuación se detallan cada uno de los elementos y fases que componen el proceso propuesto.

3.1 Información Bursátil

La información bursátil procede de repositorios de libre acceso, como Visual Chart¹ o Yahoo! Finanzas², con el comportamiento de cada uno de los valores del IBEX 35 durante años. Estos registros contienen datos como: Máximo, Mínimo, % Variación, Volumen de negociación, BPA, PER, EV/Ebitda, Máximo y Mínimo de semana anterior, etc. Además, se consideran comentarios bursátiles de expertos analistas, y comentarios en forma textual publicados por lectores de carácter general y obtenidos mediante el desarrollo de un procedimiento informático especializado que permite extraer los comentarios que se encuentran en artículos en las webs usadas. Por otro lado, para recuperar microtextos desde Twitter, se ha desarrollado una interface que utiliza el Search API proporcionado por Twitter. Para la recuperación de microtextos desde Facebook se lleva a cabo aprovechando el Graph API³ de Facebook.

3.2 Factores Externos

Los factores externos corresponden a aquellos elementos que pueden afectar en la previsión obtenida de modo ideal, tales como: conocimiento de nuevos datos macroeconómicos, porcentaje de paro, IPC, etc. Adicionalmente, la ocurrencia de grandes sucesos imprevistos, causantes de inestabilidad en mercados, como pueden ser los producidos por guerras o atentados terroristas, la publicación de medidas económicas gubernamentales como la privatización de empresas públicas, la concesión de ventajitas fiscales a los inversores bursátiles o a sectores empresariales, etc., suponen una fuente de factores imprevistos, que dificultan la estimación de tendencias y toma de decisiones. Así pues, en este trabajo, es tomada en cuenta la opinión de expertos del sector y de los posibles inversores con el objetivo de conocer y valorar la impresión de la sociedad más allá de la extraíble únicamente a partir de datos económicos.

¹ <https://www.visualchart.com>

² <https://es.finance.yahoo.com>

3.3 Extracción de conocimiento

Este proceso se compone de las siguientes fases:

Limpieza de datos: En esta fase se eliminan los datos no relevantes, corrección de erratas y detección de posibles faltas de datos. Los diferentes procesos que deben realizarse para la limpieza de la información a procesar son:

- *Detección de sesiones de datos erróneos.* Se detectan datos incongruentes: fechas imposibles, datos ilógicos, valores faltantes, etc. Se usan técnicas para detección de valores atípicos, como el Análisis de Valores Atípicos de Mahalanobis como propone Matsumoto [9].
- *Preproceso lingüístico:* Eliminación de *stop words*, *stemming* y corrección ortográfica de los comentarios textuales asociados a cada valor.
- *Análisis de sentimiento:* En esta fase se realiza un proceso de análisis de sentimiento sobre los comentarios textuales, con el objetivo de agruparlos según el nivel de positividad que presentan sobre la evolución de un valor. Para ello se usa un algoritmo de clustering basado en el uso de funciones de similitud que analizan la distancia, a nivel de palabras y del conjunto total del texto, agrupando aquellas con una baja distancia y alta similitud.

Transformación de los datos: En esta fase se calculan diferentes indicadores que caracterizan el comportamiento de los valores en cuanto a dimensiones temporales como días y semanas.

- *Caracterización de los días:* Para la caracterización de los días es necesario un análisis de los datos proporcionados por los repositorios de históricos de cotizaciones de cada valor, como su precio de cierre, máximo, mínimo, volumen o capitalización. A partir de estos datos, se establece un conjunto de indicadores globales diarios:

Tabla 1. Indicadores característicos de los días

Indicador	Descripción
VA_P: Variación Precio.	Porcentaje de variación del precio de cierre respecto del obtenido en la sesión anterior, es decir, $VA_P = ((PC_i - PC_{i-1}) / PC_i) * 100$
VA_V y VA_C: Variación Volumen y Capitalización	Porcentaje de variación respecto del obtenido en la sesión anterior, es decir, VA_V y $VA_C = (VOL_i - VOL_{i-1}) / VOL_i * 100$
PER: Relación entre Precio y Beneficios.	Relación existente entre el precio y el beneficio de la acción, es decir, $PER = (PC / Beneficios)$
BPA: Beneficio por Acción.	Beneficio obtenido por cada acción, es decir, $BPA = Beneficio / N^{\circ} Acciones$
RDIV: Rentabilidad por Dividendo	Relación entre los dividendos y el precio de la acción, es decir, $RDIV = Dividendo \text{ por acción} / PC$

- *Caracterización de las semanas:* Una semana es un conjunto ordenado de días. Por tanto, una semana es caracterizada por la secuencia ordenada de los indicadores anteriores calculados para sus días, teniendo en cuenta la naturaleza del día dentro de la semana (día laborable, festivo, víspera, etc.). Además puede contar con los valores agregados de estos indicadores.
- *Caracterización de los Valores:* Haciendo uso de los indicadores anteriormente especificados, es posible el cálculo de indicadores globales para los valores capaces de representar y caracterizar sus comportamientos.

Tabla 2. Indicadores característicos de los valores

Indicador	Descripción
T_C: Tendencia a corto plazo.	Previsión de tendencia considerando los datos de cotización de la última semana.
T_M: Tendencia a medio plazo.	Previsión de tendencia considerando los datos de cotización del último mes.
T_L: Tendencia a largo plazo.	Previsión de tendencia según datos de cotización de seis últimos meses.
E: Previsión según expertos.	Previsión para el valor, obtenida a partir del procesamiento de descripciones textuales de fuentes de opiniones de expertos establecidas.
S: Factor Social.	Representa la impresión por parte de la masa de posibles inversores sobre un Valor, calculable a partir de procesamiento de sus comentarios.

Selección de los indicadores más decisivos. Anteriormente se han descrito indicadores capaces de representar días y semanas de cotización, así como los propios valores. Sin embargo, en caso de analizar los indicadores correspondientes a una posible cartera con un número elevado de valores, puede resultar complejo seleccionar de manera certera cual o cuales son aquellos que verdaderamente resultarán más ventajosos. Para ello, se propone la elaboración de un árbol de decisión tal y como se verá posteriormente en el caso de estudio.

3.4 Explotación del conocimiento

A partir del conocimiento extraído y mediante la aplicación de algoritmos de clustering y árboles de decisión, se almacenan las reglas, los indicadores calculados y los prototipos obtenidos, como base de apoyo al proceso de estimación y predicción bursátil. De esta forma, el Inversor, a partir de la base de conocimiento extraída, obtiene una recomendación acerca de qué valor o valores resultarán más beneficiosos. Mediante este conocimiento extraído se pretende contar con una información valiosa, capaz de servir de ayuda para la planificación de una cartera de valores.

4. CASO DE ESTUDIO

En el presente caso de estudio se describe la experiencia obtenida de la aplicación de la propuesta a todos los valores del IBEX 35 para la extracción del conocimiento necesario para seleccionar aquellos valores para los que se estima que resultaría más beneficiosa una inversión a corto plazo, tomando como base la predicción del comportamiento calculada para los siguientes días.

Concretamente, la propuesta es aplicada en tres escenarios distintos: durante una semana de cotización estable, durante una semana considerada inestable por la existencia de factores externos que pudieran causar inestabilidad (el reciente atentado terrorista del pasado 13 de noviembre en París), y durante una semana con un comportamiento bursátil negativo.

4.1 Información Bursátil

Para la aplicación de la propuesta al presente caso de estudio han sido seleccionados los datos bursátiles relativos a cotizaciones del total de los 35 valores pertenecientes al IBEX 35 durante los 2 meses previos a la semana para la cual se desea estimar su tendencia. La siguiente tabla muestra la división del conjunto de datos usados para cada uno de los tres periodos establecidos y la semana para la que se pretende estimar la tendencia bursátil:

Tabla 3. Información Bursátil utilizada

	Información bursátiles usada	Tendencia a estimar
Escenario 1: semana estable	16/08/2015-16/10/2015	19/10/2015-23/10/2015
Escenario 2: semana inestable	13/09/2015-13/11/2015	16/11/2015-20/11/2015
Escenario 3: semana negativa	27/09/2015-27/11/2015	30/11/2015-04/12/2015

Los datos bursátiles son los relativos al precio de cierre de cada valor, al máximo y mínimo diario, al volumen de negociación, la capitalización y el valor de los dividendos. Además, se cuenta con un conjunto de opiniones bursátiles ofrecidas por expertos y público general, que una vez agrupados dan lugar a una lista cronológica de unos 200 comentarios por cada valor tal y como se describió en la sección 3.

4.2 Limpieza de Datos

El proceso de limpieza comenzó con la revisión de los datos y comentarios de las 75 jornadas de cotización usadas. Además, para los comentarios bursátiles se realizó un proceso de limpieza de textos, y posteriormente se representaron en modo vectorial, para su clasificación en 4 grupos según su nivel de positividad aplicando un algoritmo de clustering como se describió en la sección 3. Tras el proceso de limpieza, un total de 1660 comentarios fueron descartados, considerándose como válidos cerca de 5500.

4.3 Cálculo de Indicadores

Una vez calculados los indicadores capaces de representar una semana de cotización, sus valores se han transformado a una escala entre 1y 5 mediante un algoritmo de clustering divisivo junto con el criterio del experto para mantener de ese modo un escalado lógico en cuanto al significado del indicador.

Los indicadores usados son: *Com* (completitud de la semana), *Max* (máximo subida en la semana anterior), *Min* (máxima bajada en la semana anterior), *Inf* (Previsión de publicación de informes económicos), *Ant* (Comportamiento del en la semana previa), *TC* (tendencia según datos a corto plazo), *E* (Previsión según expertos) y *S* (Factor social):

Tabla 4. Indicadores extraídos para valores del IBEX35

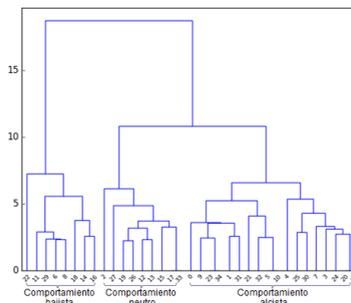
Valor	Com	Max	Min	Inf	Ant	TC	E	S
Abertis	5	4	4	1	3	4	4	5
Acciona	5	5	4	1	4	4	4	4
Acerinox	5	4	4	3	2	3	3	3
Acs	5	5	3	1	3	4	3	4
Aena	5	4	2	1	3	4	3	4
Amadeus	5	4	4	1	5	5	4	4
Arcelormital	5	2	2	3	2	2	2	3
B. Popular	5	4	3	1	4	4	3	4
B. Sabadell	5	2	2	3	2	2	2	2
Bankia	5	5	4	1	3	4	4	5
Bankinter	5	4	4	1	5	5	4	4
Bbva	5	2	2	3	1	3	2	2
Caixabank	5	3	3	1	3	3	4	3
Dia	5	3	3	1	4	3	3	3
Enagas	5	2	2	3	1	2	1	2
Endesa	5	3	2	1	3	3	3	3
Fcc	5	2	2	3	2	2	1	2
Ferrovial	5	3	3	2	3	3	3	3
Gamesa	5	2	2	3	1	2	1	1
G. Natural	5	3	3	1	3	4	3	3
Grifols	5	4	4	1	3	4	3	4
Iag	5	5	4	1	5	4	4	5
Iberdrola	5	3	1	1	2	3	2	2
Indítex	5	5	4	1	3	5	4	5
Indra	5	4	4	1	3	5	3	4
Mapfre	5	4	3	1	4	5	4	4
Mediaset	5	3	3	2	3	4	3	3
Ohl	5	3	2	2	2	4	3	3
R. Eléctrica	5	2	2	3	2	3	2	2
Repsol	5	4	4	1	3	4	3	4
Sacyr	5	4	4	1	4	4	4	4
Santander	5	5	4	1	4	4	4	5
T. Reunidas	5	3	3	2	3	3	3	3
Telefónica	5	5	4	1	3	5	4	5

4.4 Selección de los valores con mejor tendencia.

El análisis de la anterior tabla mediante un algoritmo de clustering jerárquico permite obtener un dendrograma que nos permita observar cómo están distribuidos los diferentes valores según los indicadores establecidos.

La función de distancia utilizada ha sido la distancia Manhattan y como política de recálculo de distancia entre clusters “el vecino más lejano” con el fin conseguir grupos homogéneos de valores. El resultado arroja tres grupos de elementos (Figura 1).

Fig. 1. Dendrograma



Una vez obtenidos los distintos grupos, el objetivo es tratar de determinar cuáles de los indicadores resultan más influyentes. Con esta intención, se realiza un análisis basado en árboles de decisión para tratar de determinar las variables más decisivas para seleccionar los valores que resultarán más rentables a corto plazo. La siguiente figura muestra el árbol de decisión obtenido a partir de la clasificación obtenida con el dendrograma. Como se puede observar en la Figura 2, los indicadores MAX y T_C resultan ser los más influyentes. De modo similar, los indicadores ANT y E resultarían también bastante relevantes. La Tabla 5 muestra resaltados aquellos valores con mejor puntuación para los indicadores T_C y MAX. Esos valores son los que mejor comportamiento deben ofrecer a corto plazo, resultando por tanto más beneficioso realizar una inversión. Una vez seleccionados los valores que deberían ofrecer un mejor comportamiento durante la semana 19/10/2015-23/10/2015, a continuación se muestra la evolución real de dicho conjunto de valores seleccionados. Se trata de una semana bursátil estable, sin importantes acontecimientos socio-económicos que pudieran interferir demasiado en el comportamiento previsto. La Tabla 6 muestra los precios de cierre de los tres valores seleccionados como más interesantes a corto plazo (Indítex y Telefónica), en comparación con propio índice del IBEX 35 durante la semana analizada.

Tabla 5. Valores seleccionados como más rentables

Valor	Com	Max	Min	Inf	Ant	TC	E	S
Acciona	5	5	4	1	4	4	4	4
Acs	5	5	3	1	3	4	3	4
Inditex	5	5	4	1	3	5	4	5
Santander	5	5	4	1	4	4	4	5
Telefónica	5	5	4	1	3	5	4	5

Fig. 2. Árbol de decisión con entropía 0.1

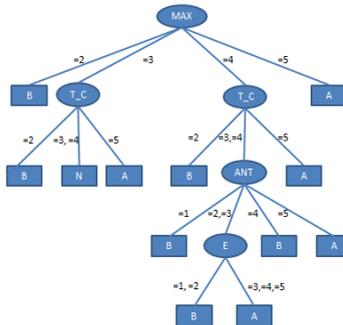


Tabla 6. Precios de cotización para los valores seleccionados en una semana estable

Valor	19/10	20/10	21/10	22/10	23/10	Variación Semanal (%)
Inditex	31,14	31,2	31,51	32,5	33,5	8,84
Telef.	11,01	10,79	10,82	11,38	11,68	5,51
IBEX35	10207,3	10100,6	10157,6	10365,4	10476,3	2,39

El resultado de ambos valores es muy superior a la media del IBEX 35, por lo que su elección para realizar una inversión a corto plazo habría resultado muy beneficiosa. Además, con el fin de validar la propuesta, se ha tratado de predecir comportamientos bursátiles sobre otros dos tipos de semanas: sobre una semana inestable (16/11-20/11) influida por factores externos y sobre una semana complicada (30/11-04/12) con fuertes bajadas. Tal y como muestran los resultados (Tabla 7), todos los valores seleccionados ofrecieron un buen comportamiento a corto plazo, por lo que puede concluirse que la aplicación del modelo propuesto ha funcionado correctamente, permitiendo seleccionar los valores más rentables en semanas no estables.

5. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha presentado una serie de técnicas que permiten el modelado del comportamiento bursátil de un conjunto de valores. Con ello se han conseguido definir comportamientos en cuanto a tipología de días o semanas de cotización, o características propias de los valores analizados, mediante el uso de técnicas de Sof Computing. Una de las posibles líneas de trabajo futuro es la reducción del nivel de error obtenido tras la estimación. Para ello es posible explorar la incorporación de otras variables que puedan afectar global o localmente a la tendencia bursátil.

Tabla 7. Precios de cotización para los valores seleccionados en una semana no estables

Valor	16/11	17/11	18/11	19/11	20/11	Variación Semanal (%)
Sacyr	2,20	2,28	2,35	2,38	2,32	4,98
Repsol	10,85	11,39	11,59	11,72	11,45	7,41
IBEX35	10124,5	10363,8	10261,1	10354,7	10290,3	1,77
	30/11	01/12	02/12	03/12	04/12	
Bankinter	6,80	6,94	6,87	6,84	6,89	2,84
Gamesa	16,57	16,28	16,36	16,11	16,34	-0,31
IBEX 35	10386,9	10379,2	10342	10092,9	10078,7	-2,25

REFERENCIAS

- García M. C., Jalal A. M., López J. M. (2013). "Métodos para predecir índices bursátiles". *Ecos de Economía*, Vol. 37; 51-82.
- Chavarnakul T., Enke D. (2008). "Intelligent technical analysis based equivolume charting for stock trading using neural networks". *Expert Syst. with Appl.*, Vol. 34; 1004-1017.
- Karen E. (2007). "Automatic extraction and identification of chart patterns towards financial forecast". *Applied Soft Computing*, Vol. 7; 1197-1208.
- Svalina I., Galzina V., et al. (2013): "An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices". *Expert Systems with Applications*, nº 40, pp. 6055-6063.
- Armando G. (2005). "A hybrid genetic-neural architecture for stock indexes forecasting". *Information Sciences*, Vol. 170; 3-33.
- Mabua S., Hirasawab K., Obayashia M., Kuremotoa T. (2013). "Enhanced decision making mechanism of rule-based genetic network programming for creating stock trading signals". *Expert Syst. with Appl*, Vol. 40; 6311-6320.
- Evansa C., Pappasa K., Xhafab F. (2013): "Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation". *Mathematical and Computer Modelling*, Vol. 58; 1249-1266.
- Lam M. (2004): "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis". *Dec. Support Systems*, Vol. 37; 567-581.
- Matsumoto, S., Kamei, Y., Monden, A., Matsumoto, K. (2007). "Comparison of Outlier Detection Methods in Faultproneness Models". *ESEM. IEEE Computer Society*; 461-463.