

Estimación de tendencia bursátil a corto plazo mediante técnicas de Soft-Computing. Aplicación práctica sobre valores del IBEX 35

Estimation of short-term market trend by Soft-Computing techniques. Practical application over IBEX 35' values

A. Peralta¹, R. Rejas²

Recibido: 21 de marzo de 2016

Aprobado: 30 de mayo de 2016

Resumen

La predicción de comportamientos bursátiles es sin duda el principal problema al que suele enfrentarse todo inversor del mercado de valores. Sin embargo, prever el comportamiento de un mercado tan fluctuante como el determinado por el conjunto de Valores que cotizan en el IBEX 35, donde multitud de factores pueden influir en él, supone todo un reto.

Actualmente son multitud los métodos, modelos y herramientas que tratan de ayudar a conseguirlo. No obstante, la mayoría de ellos realizan sus estimaciones únicamente en base a datos de comportamientos previos, por lo que la aparición de sucesos imprevistos, suelen hacer inservible cualquier estimación inicial al poco tiempo.

Ante este tipo de problemas, el uso de técnicas de Soft-Computing, debido a su capacidad para trabajar con datos imprecisos e incompletos, supone una herramienta polivalente en multitud de entornos. Por este motivo, en el presente trabajo se propone un modelo basado en su uso, como herramienta de ayuda para la previsión de comportamientos bursátiles.

Palabras clave: Predicción bursátil, Detección de tendencia, Estimación de comportamientos, Reconocimiento de patrones, Clustering.

1 Departamento de Tecnologías y Sistemas de la Información, Universidad de Castilla La Mancha, España (arturo.peralta@alu.uclm.es)

2 Facultad de Derecho y Economía, Universidad Camilo José Cela, España (rrejas@ucjc.edu)

Abstract

Predicting stock market behavior is one of the major challenges for any investor of variable income. However, predict the behavior of such a fluctuating market as the determined by the set of values of the IBEX 35 index, where many factors can influence on it, it is a problem with difficult solution.

Currently there are a wide variety of available methods to help investors perform this task. However, most of them make their estimations based solely on their previous behaviors, so the occurrence of unforeseen external factors, can rendered invalid any initial estimation.

Faced with such problems, the use of Soft-Computing techniques, due to its ability to work with inaccurate and incomplete data, results a versatile tool in many environments. Therefore, in this paper is proposed a method based on the use of soft-computing techniques to help predict the stock market behavior.

Keywords: Stock market prediction, Trend detection, Estimation of behaviors, Pattern recognition, Clustering

1. Introducción

Sin duda, en el entorno de los mercados financieros, el objetivo de todo inversor es estimar el comportamiento bursátil. Sin embargo, dado que habitualmente las fluctuaciones de la Bolsa se ven influenciadas por numerosos factores, predecir su comportamiento exacto implicaría conocer la psicología individual de cada posible inversor, así como el modo por el que ha invertido. No obstante, esto no hace imposible obtener ganancias aprovechando los movimientos de la Bolsa.

Por esta razón, actualmente existen numerosos métodos y modelos que intentan ayudar a detectar tendencias. Sin embargo, la mayoría de ellos se basan en estudios de tipo puramente técnico o fundamental, relegando a un segundo plano la estimación del impacto que pudiera tener la aparición de sucesos imprevistos capaces de generar inestabilidad, cuestiones relativas a la psicología de masas, o la publicación de opiniones de expertos.

Dar solución a estos puntos débiles es el objetivo principal de este trabajo. Con este fin, en el modelo propuesto se utilizan diversos criterios para detectar el inicio y fin de un movimiento tendencial, para conseguir que los inversores puedan obtener una estimación de la evolución futura del precio de cierre de valores del IBEX 35. Para conseguirlo, son considerados patrones de comportamiento derivables de comportamientos pasados, y cálculo de un conjunto de indicadores y variables externas capaces de dar representación al sentimiento social, para de ese modo proporcionar un marco de referencia de ayuda para la toma decisiones bursátiles.

No obstante, trabajar con datos con un alto nivel de incertidumbre, incompletos, o hasta contradictorios no resulta sencillo, motivo por el cual las técnicas más tradicionales de tratamiento de datos no resultan suficientes y óptimas. Además, generar un modelo capaz de extraer conocimiento acerca del sentimiento social, es decir, del conjunto de posibles inversores interesados en un valor del IBEX35, o conjunto de ellos, tomando como entrada opiniones personales en forma de comentarios bursátiles vía foros, tweets, u otras plataformas web, resulta igualmente complejo de abordar con métodos clásico. Por este motivo, en este trabajo se propone la utilización de un conjunto de técnicas provenientes del Soft-Computing como el KDD (Knowledge Discovery in Databases) ^[22] y el uso de Categorías Prototípicas Borrosas ^{[20] [29]} como herramientas de ayuda para la extracción de conocimiento imperceptible en datos y en los comentarios bursátiles proporcionados por expertos del sector o por el conjunto de posibles inversores, de ayuda para tomar decisiones bursátiles.

Sin embargo, actualmente resultan escasos los trabajos existentes que apliquen estas técnicas para la estimación de comportamientos bursátiles, pudiendo ser englobados en tres grandes grupos. Concretamente, aquellos basados en el uso exclusivo de redes neuronales, los que utilizan modelos híbridos combinando algoritmos genéticos con redes neuronales y, por último, los centrados en el uso de técnicas matemáticas aunque mejorados con alguna herramienta de Soft-Computing para permitir la representación de conocimiento de modo borroso.

Dentro del primero de los grupos, los estudios más destacados centrados en el uso exclusivo de redes neuronales son los siguientes. En ^[5] se realiza un estudio sobre la aplicación de redes neuronales y el cálculo de algunos indicadores representativo, 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 S&P. Por su parte, en ^[21] y ^[11] se presentan dos estudios basados en el uso de redes neuronales y el análisis de series temporales, para la predicción de comportamientos a muy corto plazo. Asimismo, destaca ^[27] proponiendo un modelo borroso basado en redes neuronales para clasificar comportamientos con el objetivo de predecir tendencias a corto plazo. Por último, resulta destacable ^[19], en el cual se propone el uso de un Perceptrón multicapa, optimizado mediante técnicas extraídas del análisis de comportamientos durante reacciones quí-

micas, que utilizando datos bursátiles recopilados durante más de 15 años, es capaz de realizar modelados y predicciones del mercado de valores.

Continuando con anteriormente clasificación, en el segundo de los grupos de trabajos relacionados se encuentran aquellos que presentan modelos híbridos, al incluir, además del uso de de redes neuronales, algoritmos genéticos. Así pues, en [4] se propone un modelo que, tomando como entrada 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 [16] se estudia el uso combinado de ambas técnicas, aunque centrado en la extracción de reglas capaces de sugerir el momento de compra o venta, algo similar a lo descrito en [8] aunque en este caso más enfocado en las oscilaciones diarias con un objetivo especulador. Finalmente, resulta relevante un modelo centrado en la aplicación de redes bayesianas como elemento capaz de extraer las relaciones e influencias existentes en un gran conjunto de factores macroeconómicos, con el objetivo de analizar y prever comportamientos del mercado de valores americano y chino. [28]

Dentro del tercero de los grupos identificados, aquellos basados en el uso de métodos matemáticos/estadísticos junto con la aplicación de alguna técnica proveniente del Soft-Computing, son relevantes estudios como [24] . En él se presenta un modelo basado en redes neuronales y funciones estadísticas complejas, capaces de calcular el indicador RIS (Relative Strength Index), mostrando su aplicación sobre el índice general del IBEX 35, y sobre un valor concreto. De modo similar, en [14] 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 Standar and Poor's. Igualmente es destacable [30] un modelo basado en series geométricas y funciones matemáticas donde los indicadores de volatilidad, el riesgo, y el precio son representados como números borrosos. De modo relacionado, en [2] se propone un sistema basado en la aplicación de ciertos cálculos estadísticos o matemáticos, como medias móviles, medias móviles exponenciales, o factores de divergencia, que unidos con factores económicos mesurables, sirven de entrada para un modelo basado en el uso de técnicas Fuzzy para el modelado de comportamientos bursátiles del mercado de Bombay. Finalmente, es destacable [3] un modelo híbrido basado en el uso

conjunto de redes neuronales y un potente sistema estadístico de reducción de indicadores económicos, que mediante técnicas de simulación permite crear modelos predictivos de comportamientos bursátiles.

El principales inconveniente de los anteriores trabajos es que, en general, la mayoría de ellos suelen tomar como fuente de entrada únicamente datos relativos a comportamientos previos, o el cálculo de indicadores derivables a partir de ellos, dejando en segundo plano las posibles inclemencias y sucesos no considerados en origen que pudieran afectar a la tendencia esperada.

El presente trabajo se estructura del siguiente modo: en el apartado número 2 se presenta el modelo desarrollado para la extracción de conocimiento, que parte de la selección y preproceso de las fuentes de datos seleccionados hasta la caracterización y definición de patrones de comportamiento de los Valores. En el apartado 3 se muestra la aplicación de la propuesta sobre un caso práctico real concreto. En el apartado 4 se discuten los resultados obtenidos en el caso de estudio analizado y en otros dos adicionales realizados siguiendo el mismo procedimiento, con el objetivo de validar el sistema propuesto. Finalmente, en el punto 5 se detallan las conclusiones alcanzadas y las posibles líneas de trabajo futuro.

2. Sistema de estimación de comportamientos bursátiles mediante Soft-Computing

A continuación se muestra un esquema general y se describen de las distintas fases y elementos que intervienen en el modelo propuesto (Fig. 1).

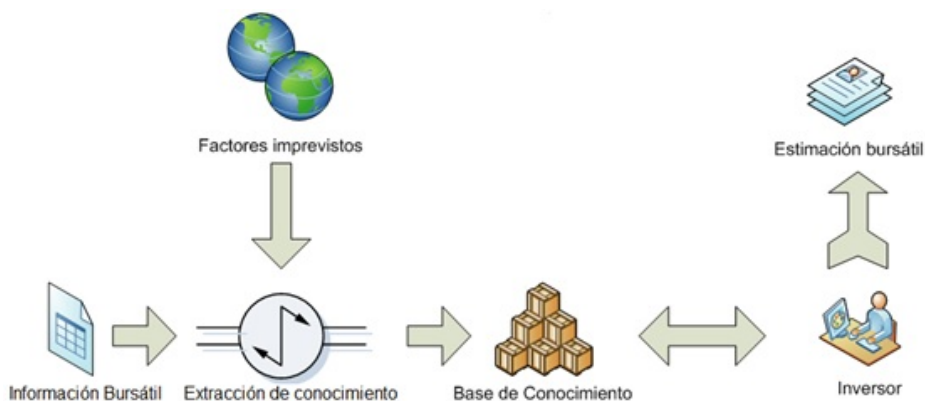


FIGURA 1. MODELO DEL SISTEMA DE EXTRACCIÓN DE CONOCIMIENTO

Las fases llevadas a cabo y los elementos involucrados en este proceso son los siguientes:

1. Información bursátil: Representa el conjunto de información considerada para la extracción de conocimiento. Concretamente, son usados datos provenientes de repositorios de libre acceso, relativos al comportamiento de cada uno de los valores del IBEX 35 durante los últimos meses. Estos registros contienen datos como: Máximo, Mínimo, % Variación, Volumen de negociación, Rentabilidad/Dividendo, PER, BPA EV/Ebitda, fecha, Máximo de semana anterior, Mínimo de semana anterior, etc. Además, son utilizados comentarios bursátiles provenientes de expertos analistas del sector, y comentarios textuales publicados por posibles inversores interesados.

A continuación se muestran algunos ejemplos de comentarios bursátiles usados (Tabla 1 y Tabla 2):

TABLA 1. RECOMENDACIONES BURSÁTILES EXPERTAS

Fuente	Expansión (www.expansion.com)	
Sección	Recomendaciones de expertos para Banco Santander	
Firma	Comentario	Día
Societe Generale	Mantener posiciones	25/09/2015
RBC	Igual que el mercado	25/09/2015
Natixis	Neutral	25/09/2015

TABLA 2. COMENTARIOS BURSÁTILES PROVENIENTES DE FOROS DE OPINIÓN

Fuente	Rankia (www.rankia.es)	
Sección	Foro sobre cotizaciones del Banco Santander	
Usuario	Comentario	Día/Hora
Sergio	Creo que se está estabilizando; posiblemente en pocos días vernos de nuevo el valor en 5.5 euros.	25/09/2015 11:09
Anónimo	Estamos superando el bache de los últimos días.	25/09/2015 11:58
Andrés R.	Prefiero no arriesgar demasiado y mantener mis posiciones.	25/09/2015 12:07

2. Factores externos: Corresponde a aquellos elementos externos que pueden afectar en una previsión obtenida de modo ideal. Existen multitud

de indicadores y factores que estudian y reflejan la situación específica de las distintas economías de cada zona bursátil, aunque algunos de los más relevantes y considerados en este trabajo son [17] :

- Crecimiento del PIB: Este indicador refleja la riqueza, el potencial económico de un país y según su gráfica histórica de mayor o menor crecimiento, se podrá operar en su mercado adaptando las inversiones según su ciclo bursátil específico.
 - Nivel de inflación: La alerta negativa de este indicador a la hora de invertir saltará cuando el nivel de inflación empiece a subir por encima del 2% y se aproximen subidas de tipos, acción que suele repercutir de forma negativa en los mercados.
 - Precio Bonos del Estado: La variación en sus precios marcará la posibilidad de subidas en los tipos de interés debiendo saber que si el precio del Bono baja excesivamente, es posible encontrar en la antesala de una subida por el BCE.
 - Datos macro económicos de Estados Unidos: todos los datos internos y externos europeos son importantes, pero índices como el Dow Jones, Nasdaq o S&P500, son para el inversor global el mejor termómetro de la actividad económica y fieles reflejos en momentos de alta volatilidad, para observar cambios de tendencias y situaciones de riesgo inversor ante futuras caídas en las bolsas de valores mundiales.
3. Extracción de Conocimiento: Haciendo uso de los datos anteriores, es llevado a cabo el proceso de extracción de conocimiento, mediante las siguientes fases:
- 3.1 Limpieza de Datos, eliminación de datos considerados no relevantes, corrección de posibles erratas y detección de posibles faltas de datos durante ciertos periodos.
 - 3.2 Transformación de datos, en registros procesables y cálculo de indicadores para la caracterización de días de cotización, semanas y los propios Valores analizados.
 - 3.3 Extracción de prototipos, ciclos, y patrones de comportamiento.
4. Base de conocimiento: Haciendo uso del conocimiento extraído y utilizando técnicas de lógica borrosa, son almacenadas los indicadores cal-

culados y los prototipos obtenidos, como base de apoyo para el proceso de estimación y predicción bursátil.

Un ejemplo de los indicadores calculados son: el porcentaje máximo de la subida o de la bajada experimentada en la semana anterior, la previsión de publicación de informes económicos que pudieran afectar a la previsión considerable en condiciones estables, etc.

5. Inversor: Encargado de realizar las acciones bursátiles correspondientes, tomando en cuenta la base de conocimiento extraído para la toma de decisiones.
6. Estimación Bursátil: El Inversor, a partir del conocimiento relativo al valor o conjunto de valores de interés, considerando las características de los días y semanas calculadas, y tomando en cuenta los patrones de comportamiento extraídos, es capaz de obtener una estimación del comportamiento bursátil. En este proceso es usado todo el conocimiento representado anteriormente, para inferir el comportamiento futuro del conjunto de valores analizados y pertenecientes al IBEX35.

De este modo, el proceso de extracción de conocimiento, a partir de datos de comportamientos bursátiles históricos, para la conformación de una base de conocimiento útil, puede ser modelado como muestra la siguiente figura (Fig. 2).



FIGURA 1. PROCESO DE EXTRACCIÓN DE CONOCIMIENTO

En este trabajo se han abordado todas las fases del proceso descrito, con el objetivo de conseguir un mecanismo automatizado mediante el cual, a partir de datos históricos de cotizaciones bursátiles lograr caracterizaciones de comportamientos, de los días y semanas de cotización, con el objetivo de

conformar una base de conocimiento útil para un inversor del mercado de valores.

Mediante este conocimiento extraído se pretende contar con una información valiosa, capaz de servir de ayuda para la planificación de una cartera de valores, de acuerdo a nuevas necesidades y a los posibles factores externos hagan necesario modificar los objetivos y adaptar la planificación inicial.

3. 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 próximos días. Concretamente, se ha pretendido utilizar el método propuesto para seleccionar los valores del IBEX35 para los que resultaría realizar una inversión durante la semana 25/01/2016-29/01/2016.

El estudio no solamente se centra en la extracción de caracterizaciones, patrones de comportamiento o cálculo de indicadores, sino en su uso de modo conjunto y complementario con el objetivo de permitir seleccionar los Valores que deberían resultar más rentables para realizar una inversión.

Para conseguir esto, se utilizan como fuente de entrada datos relativos de comportamientos bursátiles previos y comentarios textuales provenientes de analistas expertos o público general. Además, debido a la multitud de factores considerados, se hace uso de técnicas de decisión para determinar cuáles de los parámetros obtenidos resultarán más importantes, de modo comparativo, permitir la elección del conjunto de valores que deberían resultar más interesantes a corto plazo.

3.1. Fuente de datos

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 6 meses previos a la semana para la cual se desea estimar su tendencia, es decir, durante el periodo: 22/07/2015 – 22/01/2016.

Los datos bursátiles considerados para el cálculo de indicadores y el posterior proceso de extracción de conocimiento, 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.

Adicionalmente, se cuenta con un conjunto de opiniones bursátiles ofrecidas por expertos y público general, que una vez agrupados dan lugar a un fichero de tipo Excel, con 35 hojas (una por valor), en la que se listan de modo cronológico una media de unos 200 comentarios por cada valor. Para conseguir estos comentarios bursátiles se desarrolló un software que, haciendo uso de APIs de dominio público, como la Search API proporcionado por Twitter, o la Graph API³ ofrecida por Facebook. Para la recopilación de comentarios expertos se utilizaron las ediciones on-line de prensa especializada del sector.

3.2. Limpieza de datos

El proceso de limpieza comienza con la revisión de los datos y comentarios de las 130 jornadas de cotización usadas. Además, para los comentarios bursátiles se realiza un proceso de limpieza de textos. Las principales técnicas usadas fueron:

- Detección de datos bursátiles erróneos. Son detectados datos incongruentes: fechas inválidas, datos ilógicos, valores perdidos, etc. Para ellos se utilizan técnicas para detección de valores atípicos, como la Prueba de Grubbs ^[9], la Prueba de Dixon como propone Li y Edwards ^[15], el Análisis de Valores Atípicos de Mahalanobis como describe Matsu-moto et al ^[18] o el modelo de Regresión Lineal Simple proporcionado por Rousseeuw y Leroy ^[26].
- Detección y supresión de Stop Words. En esta fase es eliminada toda la información implícitamente sobrante, carente de significancia, que pudiera estar contenida en los comentarios textuales utilizados. Para realizar esta tarea se utilizan listas de “stop words” (palabras vacías) que automáticamente son eliminadas los textos procesados, como se propone en Korfhage ^[13].
- Corrección ortográfica. Para ello se utilizan diccionarios de corrección automática estandarizados. Además, estos diccionarios pueden ser en-

riquecidos mediante el uso de diccionarios adicionales y especializados para contextos determinados

- Agrupación de comentarios bursátiles según su nivel de positividad o negatividad. En esta fase se realiza un proceso de extracción del sentimiento desprendido de los comentarios textuales utilizados. El objetivo es conseguir una agrupación de los mismos según el nivel de positividad o negatividad que presentan sobre evolución de un valor. Para ello se utiliza un algoritmo de clustering como el expuesto por Romero et al ^[25] basado en la representación vectorial de los comentarios textuales, y el uso de funciones de similitud y vecindad, que analizan la distancia, a nivel de palabras y del conjunto total del texto, agrupando aquellas con alta similitud y baja distancia. Adicionalmente, se utilizan técnicas de stemming como las propuestas por Korenius et al ^[12] para la extracción de raíces léxicas para la agrupación de conceptos.

Tras el proceso de limpieza, un total de 1425 comentarios bursátiles fueron descartados, y un total de 23 datos económicos relativos a cotizaciones fueron descartados o corregidos.

3.3. Cálculo de indicadores

Una vez recolectados los datos y los comentarios bursátiles, se ha llevado a cabo su limpieza, y se han calculado un conjunto de indicadores (o calculables a partir de datos de comportamientos pasados) establecidos como aquellos que mejor describen a una semana de cotización.

- Completitud (COM): Representa, de modo numérico, si la semana bursátil es completa o contiene algún festivo.
- Máxima subida en semana anterior (MAX): Porcentaje máximo de subida conseguido durante la semana anterior.
- Máxima bajada en semana anterior (MIN): Porcentaje máximo de bajada obtenido en la semana anterior.
- Previsión de informes (INF): Previsión de publicación de informes económicos de la empresa durante la semana
- Comportamiento en semana anterior (ANT): Comportamiento del Valor en la semana previa (alcista, bajista o neutro).

- T_C (tendencia según datos a corto plazo): Tendencia esperada a corto plazo según el análisis realizado a partir de los datos recopilados de comportamientos previos.
- E (Previsión según expertos): Predicción de comportamiento según expertos, obtenido a partir del estudio de sus análisis y comentarios bursátiles en formato textual.
- S (Factor social): Sentimiento de posibles inversores, obtenido mediante el estudio de sus comentarios en formato textual, en foros especializados.

Cada una de estas características puede ser considerada como bipolar y dicotómica, ya que cada elemento tiene nivel o grado, pudiendo servir para distinguir unos elementos de otros o verificar su semejanza. Es por ello que se pueden considerar como construcciones de un emparrillado ^[1].

Las construcciones contemplan la asignación de valores graduales entre sus polos. La escala puede ser simbólica o numérica y no puede variar el número de escalones (1 a 5) en toda la malla aunque los nombres utilizados para cada símbolo sean diferentes.

Por ello, los valores de los indicadores primero se han normalizado y posteriormente se han transformado a este tipo de escala. Para dividir una serie de números en cinco grupos se ha utilizado un algoritmo de clustering divisivo, como el propuesto en [6] sobre el conjunto de datos disponibles y el criterio del experto para mantener de ese modo un escalado lógico en cuanto al significado del indicador.

Los resultados obtenidos se observan en la siguiente tabla (Tabla 3).

TABLA 3. INDICADORES PARA VALORES DEL IBEX 35

Valor	Com	Max	Min	Inf	Ant	TC	E	S
Abertis	5	5	3	1	3	3	4	4
Acciona	5	4	4	1	4	4	4	4
Acerinox	5	3	3	3	2	3	3	3
Acs	5	4	4	1	3	4	4	4
Aena	5	2	3	1	3	3	2	3
Amadeus	5	3	3	1	4	3	3	2
Arcelormital	5	2	3	2	2	2	2	3

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

3.4. Selección de los valores con mejor tendencia

El proceso análisis de la anterior tabla mediante clusters o grupos consiste en detectar grupos homogéneos de construcciones o elementos. Para realizar este análisis se ha de definir una medida de similitud mediante el uso de

una función capaz asociar un valor numérico a cualquier par de vectores para calcular su nivel de semejanza o similitud. Mediante la función de similitud se puede construir una matriz de proximidad, cuadrada, donde cada elemento de ella mide la similitud entre el elemento fila y columna que se presenta. Dicha matriz de proximidad es la fuente de entrada al análisis de clusters.

Para realizar el cálculo de la distancia o similitud pueden ser usadas distintas expresiones, aunque para el presente caso de estudio se ha optado por utilizada la distancia Manhattan entre dos vectores.

La siguiente imagen muestra la matriz de similitud resultante. En ella, los tonos de color gris oscuro representar elementos con una gran similitud, mientras que los tonos de color gris claro representan elementos con una mayor diferencia (Fig. 3).

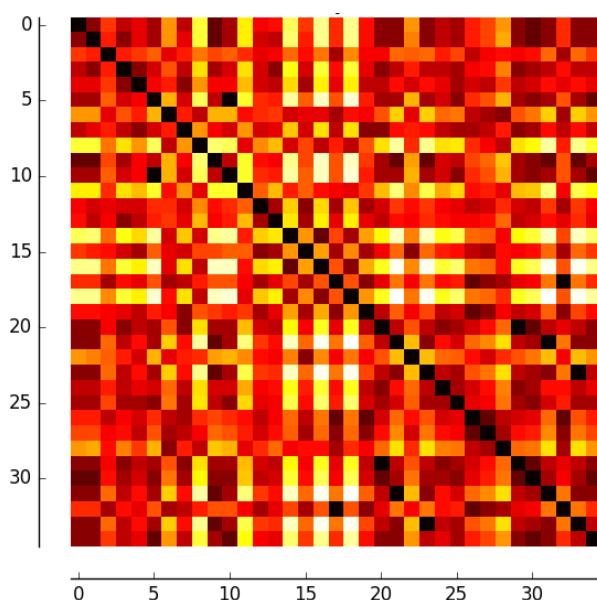


FIGURA 3. MATRIZ DE SIMILARIDAD

Tras la aplicación de un algoritmo de clustering jerárquico a la anterior matriz de similitud es posible obtener un dendrograma como el que se muestra en la siguiente figura (Fig. 4).

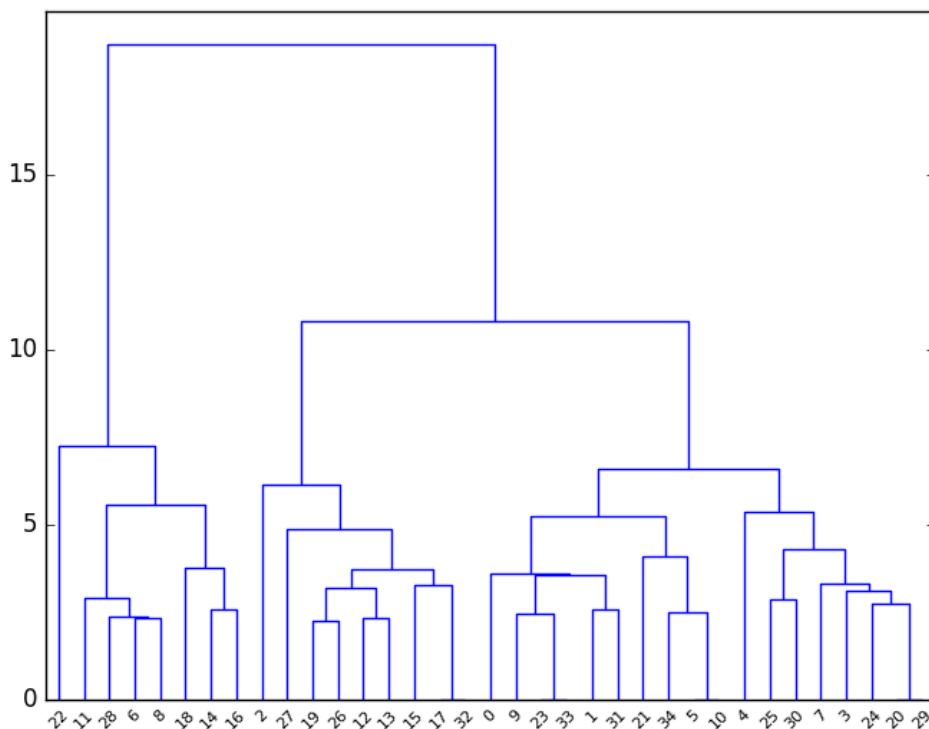


FIGURA 4. DENDROGRAMA

Para su cálculo se ha utilizado el método del enlace completo (el vecino más lejano) como política de cálculo de distancia entre clusters con el objetivo de que el algoritmo sea menos sensible a valores extremos, consiguiendo grupos de valores más homogéneos. En este caso y a partir del estudio del dendrograma se puede inferir que existen tres grupos de elementos bastante homogéneos.

Tras la obtención de los tres grupos anteriores, el siguiente objetivo es intentar determinar qué indicadores calculados anteriormente resultan más influyentes a la hora de estimar la tendencia de un Valor. Con este objetivo, se realiza sobre los resultados obtenidos un análisis basado en árboles de decisión. Mediante él se tratará de obtener una representación gráfica capaz de determinar sin duda qué variables resultan más útiles para realizar la selección de valores para los cuales debe resultar más rentable realizar una inversión en un corto periodo de tiempo.

A continuación se muestra el árbol de decisión obtenido a partir de la clasificación obtenida con el dendrograma (Fig. 5). Para su cálculo se ha aplicado un algoritmo de árboles de decisión de tipo ID3^[23], estableciendo un nivel de entropía máximo de 0.1 para cada hoja

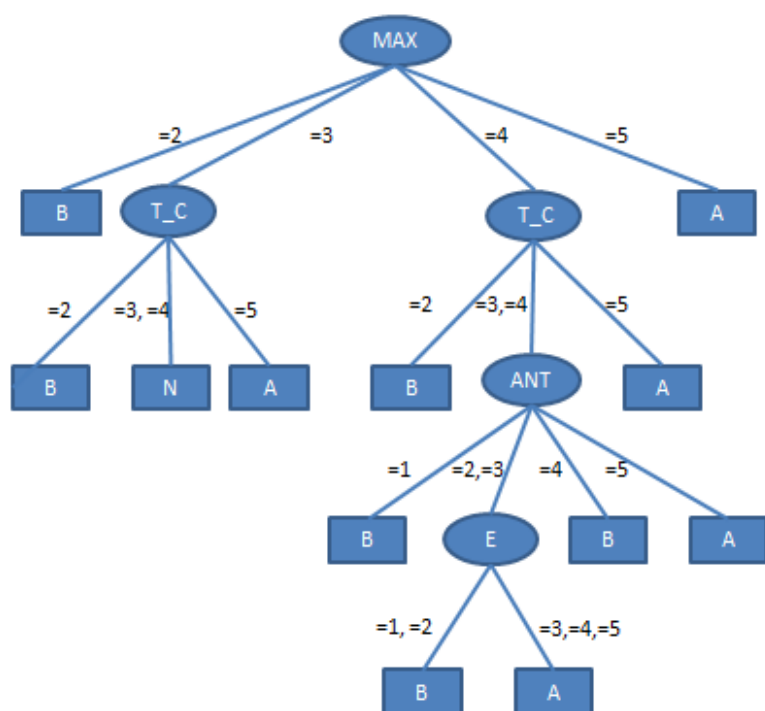


FIGURA 5. ÁRBOL DE DECISIÓN CON ENTROPÍA 0.1

Sin embargo, con un nivel de entropía relativamente alto, como es 0.1, algunos de los valores son erróneamente clasificados. Por el contrario, con un nivel de entropía menor, como podría ser 0.01, el árbol resultante se muestra demasiado complejo e inmanejable como para permitir concluir qué indicadores de los utilizados resultan más influyentes.

Por esta razón, y dado que realmente lo que se trata de conseguir es determinar qué indicadores son los que resultan ser más influyentes, con el objetivo de poder seleccionar el conjunto de valores que deberían resultar más rentables a corto plazo, se considera más interesante representar la informa-

ción anterior mediante diagramas de influencia. La siguiente figura muestra el diagrama de influencia obtenido (Fig. 6).

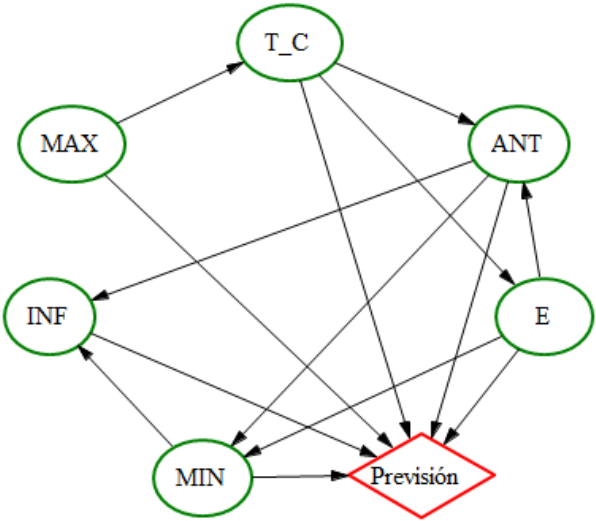


FIGURA 6. DIAGRAMA DE INFLUENCIA

Con esta representación se observa claramente que los indicadores correspondientes a T_C (Tendencia a corto plazo) y MAX (Máximo de semana anterior), resultan ser los más influyentes. De modo similar, en un siguiente nivel de influencia, los indicadores E (Opinión de los expertos) y ANT (Comportamiento en la semana anterior) resultarían también bastante relevantes. De ese modo, podría concluirse que aquellos valores que consigan una mejor puntuación en estos indicadores deberían ser aquellos con un comportamiento bursátil más favorable en sus más inmediatas jornadas.

La tabla siguiente (Tabla 4) muestra, resaltados en gris oscuro aquellos valores con mejor puntuación para los indicadores MAX y T_C. Esos valores son los que mejor comportamiento deben ofrecer a corto plazo, resultando por tanto más beneficioso realizar una inversión. De modo análogo, los valores resaltados en gris claro, serían aquellos valores que ocuparían un segundo lugar en nivel de interés, al haber obtenido unos resultados algo peores (aunque altos) en T_C, MAX, ANT y E.

TABLA 4. VALORES SELECCIONADOS COMO MÁS RENTABLES

Valor	Com	Max	Min	Inf	Ant	TC	E	S
Endesa	5	5	4	1	4	5	5	5
IAG	5	5	4	1	4	4	4	5
Ferrovial	5	5	4	2	3	5	5	5
G.Natural	5	5	4	1	4	5	3	5
Abertis	5	5	4	1	4	4	4	4

4. Resultados y discusión

Una vez seleccionados los tres valores que deberían ofrecer un mejor comportamiento, a continuación se muestra la evolución real de dicho conjunto de valores durante la siguiente semana al cálculo de los indicadores. Se trata de una semana bursátil estable, sin importantes acontecimientos socio-económicos que pudieran interferir demasiado a los comportamientos previstos.

Tal y como se indicó, para realizar el cálculo de los indicadores, fueron usados datos y comentarios bursátiles recopilados durante 6 meses hasta el 22/01/2016. Por ese motivo, para comprobar si efectivamente la elección de estos valores resulta acertada, se estudia su comportamiento en la siguiente semana, es decir, durante el periodo 25/01/2016-29/01/2016.

La siguiente tabla (Tabla 5) se muestra los precios de cierre de los tres valores seleccionados como más interesantes a corto plazo (Endesa, Ferrovial y Gas Natural), en comparación con propio índice del IBEX 35 durante la semana analizada.

TABLA 5. PRECIOS DE COTIZACIÓN DE VALORES SELECCIONADOS - SEMANA 1

	25/01	26/01	27/01	28/01	29/01	Variación Semanal (%)
Ferrovial	19,57	19,54	19,59	19,66	20,14	2,70
Gas Natural	17,60	17,75	17,74	17,83	18,02	1,69
Endesa	16,82	17,26	17,54	17,49	17,77	3,13
IBEX35	8567,7	8692,5	8741,0	8590,0	8815,8	1,06

La siguiente gráfica muestra la variación, en modo porcentual, de los tres valores (Fig. 7).

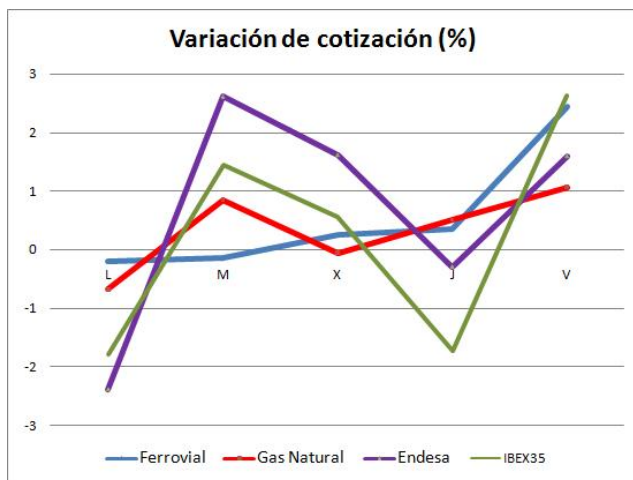


FIGURA 7. EVOLUCIÓN SEMANAL DE VALORES SELECCIONADOS

Adicionalmente, con el objetivo de valorar el modelo propuesto, el experimento es realizado de modo íntegro, siguiendo los mismos pasos indicados, con el objetivo de predecir comportamientos bursátiles sobre otras dos semanas posteriores. La primera de ellas, la comprendida entre los días 01/02/2016-05/02/2016; una semana complicada para el IBEX35 con fuertes bajadas. Y la segunda, la correspondiente a los días 15/02/2016-19/02/2016, una semana con un comportamiento alcista. El objetivo es comprobar el funcionamiento del modelo en situaciones estables, alcistas y bajistas.

Tras la aplicación de la propuesta sobre la semana 01/02/2016-05/02/2016, los valores más recomendados según el método propuesto fueron: Aceronix, Ohl y BBVA. El comportamiento de dicha selección de valores en comparación con la media del IBEX35 fue el siguiente (Tabla 6):

TABLA 6. PRECIOS DE COTIZACIÓN DE VALORES SELECCIONADOS - SEMANA 2

	01/02	02/02	03/02	04/02	05/02	Variación Semanal (%)
Aceronix	8,18	7,93	8,29	9,04	9,18	11,00
Ohl	4,93	4,57	4,31	4,67	4,91	0,20
BBVA	5,8	5,6	5,48	5,77	5,9	0
IBEX35	8788,5	8528,7	9314,5	8468,1	8499,5	-3,59

De modo similar, tras la aplicación de la propuesta sobre la semana 15/02/2016-19/02/2016 los valores seleccionados fueron: Telefónica, Indra y Sacyr. Su comportamiento respecto del IBEX fue el mostrado en la siguiente tabla (Tabla 7).

TABLA 7. PRECIOS DE COTIZACIÓN DE VALORES SELECCIONADOS - SEMANA 3

	15/02	16/02	17/02	18/02	19/02	Variación Semanal (%)
Telefo.	4,45	4,4	4,51	4,53	4,58	6,02
Indra	8,06	8,2	8,12	8,38	8,47	7,89
Sacyr	1,43	1,45	1,49	1,45	1,42	5,97
IBEX35	8179,2	8137,6	8364,9	8295,4	8194,2	3,45

Como puede comprobarse, en los tres experimentos, todos los valores seleccionados ofrecieron un buen comportamiento a corto plazo, en todos los casos bastante superior a la media del IBEX. Considerando que la media de las variaciones de los valores seleccionados es muy positiva, y bastante superior a la media del IBEX 35, puede concluirse que haber seleccionado dichos conjuntos de valores óptimos para una inversión a corto plazo habría resultado acertado y rentable.

5. Conclusiones y trabajo futuro

En el presente trabajo se ha definido un modelo enmarcado de dentro del entorno del descubrimiento de conocimiento prototípico. En él se han usado un conjunto de técnicas de Soft-Computing, como los algoritmos de clustering, la lógica borrosa o las Categorías Protototípicas Borrosas, que permiten el modelado del comportamiento, en lo que se refiere a la distribución temporal, de un conjunto de valores bursátiles.

Con el modelo propuesto se han conseguido definir comportamientos o características propias de los valores analizados. A partir de estas caracterizaciones, y siguiendo un modelo predictivo del comportamiento ha sido posible realizar estimaciones de tendencia bursátiles. Además, la propuesta ha sido probada con un caso real de estudio, pudiendo verificar su precisión

tomando en cuenta datos de comportamientos reales anteriores y posteriores al experimento.

A pesar de los buenos resultados obtenidos, una línea de trabajo futuro es valorar la posibilidad de definir nuevos indicadores, o hacer uso de variables externas adicionales. El objetivo sería mejorar la capacidad predictiva y adaptativa, en especial en situaciones en complejas en las que pudieran aparecer sucesos imprevistos.

De modo similar, podrían definirse un conjunto variables adicionales, capaces de representar el perfil del inversor según el nivel de resigo que esté dispuesto a asumir.

Finalmente, de modo complementario, otra posible línea de trabajo futuro sería el desarrollo de un sistema borroso capaz de auto-generar resúmenes lingüísticos a partir del procesamiento de los datos de registros y de los comentarios bursátiles textuales. En este contexto, actualmente existen diversas aproximaciones borrosas para la generación de resúmenes lingüísticos, que habitualmente se construyen mediante proposiciones cuantificadas para diferentes aplicaciones, como los propuestos en [7] [10] que fácilmente podrían ser adaptados al contexto de este trabajo. Mediante este tipo de resúmenes sería posible obtener una representación, en lenguaje natural, que aúne las impresiones generalizadas de expertos analistas del sector o de la sociedad.

4. Referencias

- [1] P. M. Alexander, J. J. Van Loggerenber, : “The repertory grid: discovering a 50-year-old research technique”. Proceedings of SAICSIT, pp 192–199, 2005.
- [2] T. Anbalagan, Uma S. «Classification and Prediction of Stock Market Index Based on Fuzzy Metagraph». Procedia Computer Science, nº 47, pp 214-221, 2015.
- [3] Anish C. M., Babita M. «Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis». Journal of the Korean Statistical Society, nº 45, pp 64-76, 2015.
- [4] G. Armando. !A hybrid genetic-neural architecture for stock indexes forecasting”. Information Sciences, Vol. 170. Pg. 3-33, 2005.

- [5] Chavarnakul T., Enke D. "Intelligent technical analysis based equivolume charting for stock trading using neural networks". *Expert Systems with Applications*, n°34, pp. 1004-1017, 2008.
- [6] Chavent M., Lechevallier Y., Olivier B. "DIVCLUS-T: A monothetic divisive hierarchical clustering method". *Comput. Stat. Data Anal.*, pp 687-701. DOI=10.1016/j.csda.2007.03.01, 2007.
- [7] Díaz-Hermida F., Bugarín A. "Semi-fuzzy Quantifiers as a Tool for Building Linguistic Summaries of Data Patterns". *IEEE Symposium on Foundations of Computational Intelligence*, pp 45-52, 2011.
- [8] Evansa C., Pappasa K., Xhafab F. "Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation". *Mathematical and Computer Modelling*, n°58, pp. 1249-1266, 2013.
- [9] Grubbs, F. "Procedures for Detecting Outlying Observations in Samples". *Technometrics*, V.11, N.1, pp 1-21, 1969.
- [10] Kacprzyk J. "Computing With Words Is An Implementable Paradigm: Fuzzy Queries, Linguistic Data Summaries, and Natural-Language Generation". *IEEE Trans. Fuzzy Systems*, n° 18, pp. 451-472, 2010.
- [11] Karen E. "Automatic extraction and identification of chart patterns towards financial forecast". *Applied Soft Computing*, n° 7, pp. 1197-1208, 2007.
- [12] Korenius T, Laurikkala J., Järvelin K., and Juhola M. "Stemming and lemmatization in the clustering of finish text documents" in *CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management*. New York, NY, USA: ACM, pp. 625-633, 2004.
- [13] Korfhage R. R. "Information storage and retrieval". John Wiley & Sons, Inc.
- [14] Lam M. (2004): "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis". *Decision Support Systems*, n° 37, pp. 567-581, 1997.
- [15] Li D., Edwards E. "Automatic Estimation of Dixon's Test for Extreme Values Using a SAS Macro Driven Program". *PharmaSug*, Toronto, 2001.
- [16] Mabua S., Hirasawab K., Obayashia M., Kuremotoa T. "Enhanced decision making mechanism of rule-based genetic network programming for creating stock trading signals". *Expert Systems with Applications*, n° 40, pp. 6311-6320, 2013.
- [17] Madrigal J. A. "Invertir en bolsa con sentido común". *Anaya Multimedia*. ISBN: 9788441537194, 2015.

- [18] Matsumoto, S., Kamei, Y., Monden, A., Matsumoto, K. "Comparison of Outlier Detection Methods in Faultproneness Models". Proceedings of the First international Symposium on Empirical Software Engineering and Measurement ESEM. IEEE Computer Society, Washington, USA, 2007.
- [19] Nayal S. C., Misra B. B., Behera H. S. «Artificial chemical reaction optimization of neural networks for efficient prediction of stock market indices». Ain Shams Engineering Journal, in press, 2015.
- [20] Olivas J.A. "Contribución al Estudio Experimental de la Predicción basada en Categorías Deformables Borrosas", Universidad de Castilla La Mancha, España. Tesis Doctoral, 2000.
- [21] Oliveira F. A. "Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4". Expert Systems with Applications, nº 40, pp. 7596–7606, 2014.
- [22] Peralta A., Romero F. P., Polo M., Olivas J. A. "Knowledge extraction of the behaviour of software developers by the analysis of time recording logs". FUZZIEEE. pp. 1-8, 2010.
- [23] Quilan, J.R. "Induction of decision trees". Machine Learning 1, nº 81-106, 1986.
- [24] Rodríguez-González A., García-Crespo A, et al. "CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator". Expert Systems with Applications. Elsevier, 2011.
- [25] Romero F. P., Peralta A., Soto A., Olivas J. A., Serrano-Guerrero J. "Fuzzy optimized self-organizing maps and their application to document clustering", Soft- Computing. DOI 10.1007/s00500-009-0468-3, Springer, 2009.
- [26] Rousseeuw P., Leroy A.: "Robust Regression and Outlier Detection". John Wiley & Sons. New York, USA, 1996.
- [27] Svalina I., Galzina V., et al. "An adaptive network-based fuzzy inference system (AN-FIS) for the forecasting: The case of close price indices". Expert Systems with Applications, nº 40, pp. 6055–6063, 2013.
- [28] Wang L., Wang Z., Zhao S., Tan S. «Stock market trend prediction using dynamical Bayesian factor graph». Expert Systems with Applications, nº 42, pp 6267-6275, 2015.
- [29] Zadeh, L. A. "A note on prototype set theory and fuzzy sets". Cognition, nº 12, pp. 291- 297, 1982.
- [30] Zhanga W-G., Xiaob W-L., Konga W-T, Zhang Y. "Fuzzy pricing of geometric Asian options and its algorithm". Applied Soft Computing, nº 28, pp. 360–367, 2015.