

ANÁLISIS Y CLASIFICACIÓN DE INDICADORES TÉCNICOS MEDIANTE SUPPORT VECTOR MACHINE

ANALYSIS AND CLASSIFICATION OF TECHNICAL ANALYSIS INDICATORS BY SUPPORT VECTOR MACHINES

Javier OLIVER

Universidad Politécnica de Valencia, Facultad de Administración y Dirección de
Empresas. Spain.

Email: jaolmun@ade.upv.es

Resumen:

La búsqueda de modelos para la predicción de la tendencia de los índices bursátiles se ha desarrollado en las últimas décadas. Los indicadores y osciladores técnicos son los inputs más utilizados en todos los modelos. Éstos se basan fundamentalmente en los precios y dirección del propio índice. Esto puede provocar ciertos problemas en las estimaciones y procesos de aprendizajes de los diferentes modelos, como multicolinealidad y autocorrelación para el caso de modelos lineales y problemas de sobreoptimización y ruido en otros casos como en las redes neuronales.

Se plantea filtrar los diferentes indicadores y osciladores técnicos a utilizar en los diferentes modelos. Para ellos, se va a analizar el impacto que tienen éstos en el proceso de predicción de la tendencia de un índice bursátil. El modelo utilizado es la support vector machine que permite encontrar las características tanto de los inputs (indicadores y osciladores) como del output (la tendencia del índice). Este mapeo de la relación de los indicadores y la tendencia ofrece información relevante sobre si dicha contribución a su predicción es estable en el tiempo. Por tanto, se seleccionarán aquellos inputs cuyas características estabilicen las predicciones en los modelos. Así pues, se deben descartar aquellos indicadores irregulares, aunque puntualmente

puedan alcanzar ratios de acierto algo más elevadas que los más estables. Este proceso provocará obtener predicciones de la tendencia más consistentes.

Palabras clave: Support Vector Machines, tendencia, índice bursátil Dow Jones Industrial, indicadores técnicos.

Abstract:

The search for models which can accurately forecast the market trend has developed over the past decades. Technical indicators and oscillators are the most usually employed inputs in the prediction models. These inputs basically rely on prices and the evolution of the index itself, which may cause some problems like multicollinearity and autocorrelation, in the case of linear models, or overoptimization and noise, in the case of neural networks. This paper proposes filtering the inputs to be employed in the models. To this end, their impact on the forecast will be analysed. A support vector machine will be used to this end, in order to characterize both inputs (indicators and oscillators) and output (market trend). Doing this, it can be assessed whether the relationship between the different inputs and the market trend offers relevant information regarding the contribution of the inputs in the prediction process and whether this contribution remains constant over time. Those inputs will be selected, which obtain more stable forecasts in order to obtain more consistent predictions.

Keywords: Support Vector Machines, trend, stock index, Dow Jones Industrial Average, technical indicators

JEL: G110, G140, G170

1. INTRODUCCIÓN

El análisis y estudio de la tendencia de los índices bursátiles se ha realizado mediante el uso de diferentes tipos de modelos. Por un lado destacan modelos de regresión logística, que se aplican cuando se supone una relación lineal entre los inputs y output. Este modelo es muy utilizado para el caso que la variable respuesta sea dicotómica. El modelo presenta una función que describe la probabilidad del suceso 1 o 0, para el caso de tener tendencia alcista o bajista. Puede verse un ejemplo de la utilización de este modelo en (Zaidi. 2016) que analiza la tendencia índice Ksa. También en el análisis de la tendencia del índice bursátil de la India (Dutta et al. 2015).

La support vector machine es otro tipo de modelo de clasificación muy utilizado en la predicción de la tendencia de los índices bursátiles. Este modelo separa el problema en un hiperplano y mediante un problema de optimización minimiza las distancias generadas entre las características de los inputs y las características de los outputs. En Huang (2005) se analiza la tendencia semanal del índice Nikkei 225. Oztekin et al. (2016) utilizan la SVM en la tendencia diaria de los índices bursátiles emergentes, como el índice de Estambul. El análisis de la tendencia diaria mediante indicadores técnicos es muy frecuente. En Kim (2003) se analiza la tendencia del índice coreano mediante doce indicadores y osciladores técnicos. Kara et al. (2011) realiza una comparación entre la SVM y las redes neuronales para la predicción de la tendencia del índice de Estambul utilizando nueve indicadores y osciladores. Patel et al. (2015) utiliza los mismos indicadores aplicándolo a los índices Cnx Nifty y S&P de Bombay. En Ballings et al. (2015) puede verse una recopilación de las técnicas de clasificación más empleadas en la predicción de las tendencias bursátiles.

El otro tipo de modelos muy extendido para la predicción de la tendencia bursátil son las redes neuronales y algoritmos genéticos. Por ejemplo, en Moghaddam et al. (2016) se estudia la tendencia diaria del índice Nasdaq, si bien utilizando como inputs los precios de cierre retardados. En Oliver (2016) utiliza la red neuronal backpropagation con el índice Ibex-35. George et al. (2009) recopila los principales modelos de redes neuronales.

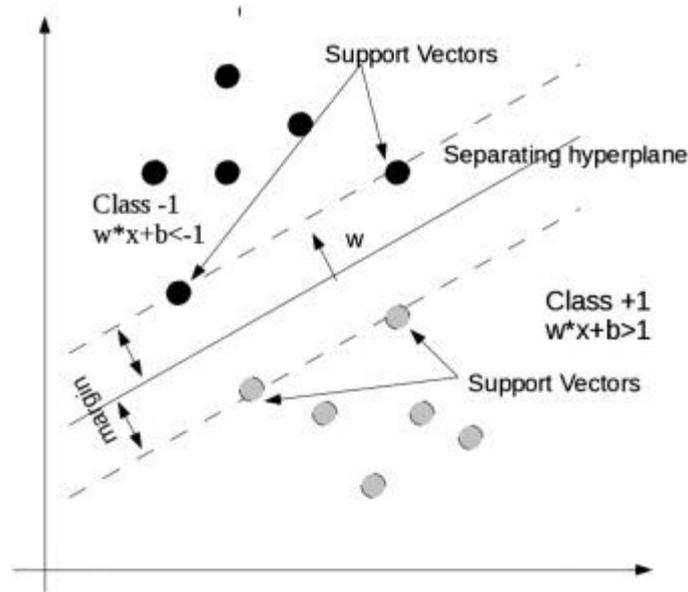
A partir de estos modelos surgen otros como combinación de diferentes técnicas y modelos. En Qiu et al. (2016) añaden un algoritmo genético a una red neuronal para la predicción del índice Nikkei 225. Otros autores combinan la SVM con otras metodologías. Así, en Huang et al. (2008) realizan un análisis individual del impacto de las características de cada input en los outputs del modelo. En Kumar et al. (2016) plantean diferentes modelos híbridos de la SVM como el Proximal Support Vector Machine o Random Forest Support Vector Machine, entre otros. En García et al. (2017) se utiliza algoritmos genéticos y heurísticos como el tabú search.

En estos estudios sobre la tendencia de los índices bursátiles se han empleado indicadores y osciladores técnicos cuyo cálculo se basa en los precios del propio índice. El objetivo de este trabajo es analizar la conveniencia del uso de algunos de estos indicadores como inputs de los diferentes modelos. Mediante la Support Vector Machine, se observará el comportamiento de cada uno de ellos frente a la variable respuesta, es decir, frente a la tendencia. Con esto se pretende obtener qué indicadores presentan una estabilidad en su contribución a la predicción de la tendencia. Aquellos indicadores más estables deberían formar parte de los diferentes modelos a construir, eliminando aquellos más inestables que pueden provocar distorsiones en el proceso de aprendizaje.

2. DESCRIPCIÓN DE LA METODOLOGÍA

La Support Vector Machine fue desarrollada por Vapnik y otros (Cortes & Vapnik.1995, Vapnik et al.1997) como una alternativa de machine learning de las redes neuronales y el perceptron (Rummelhart et al. 1986). El modelo se basa en incorporar el vector de inputs en un espacio de características multidimensional Z mediante un mapeo no lineal. Por tanto, se trata de separar el problema en diferentes espacios dimensionales. La support vector machine define el margen de la máxima distancia entre las diferentes clases del hiperplano (fig.1).

Figura 1. Separación del problema en dos dimensiones. (Zaroggiani et al.2013)



El problema fundamental reside en encontrar una separación del hiperplano que generalice correctamente para dar una solución satisfactoria al problema. Dado que las dimensiones del mapa de características son elevadas, los datos para el entrenamiento del algoritmo no siempre generalizan correctamente en todos los hiperplanos. Por tanto, la support vector machine trata de generar un modelo, basado en los datos de entrenamiento, que prediga los valores objetivo, teniendo en cuenta sus características.

El modelo resuelve un problema de optimización, partiendo de los set de datos de aprendizaje compuesto por las diferentes pares de instancias y etiquetas (x_i, y_i) con $i = 1 \dots n$, donde x_i es un número real e $y_i \in \{-1, 1\}^n$. (Chih-Wei et al,2010).

$$\min_{w,b,\varepsilon} \frac{1}{2} w^T w + C \sum_{i=1}^n \varepsilon_i$$

$$\text{sujeto a } y_i(w^T \phi(x_i) + b) \geq 1 - \varepsilon_i$$

$$\varepsilon_i \geq 0$$

Los vectores para el entrenamiento x_i , son mapeados en el espacio con n dimensiones mediante la función ϕ . El modelo añade un parámetro de penalización, $C > 0$ para el

término del error. Además añade una función interna (kernel) $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ (Smola et al. 1998b,2003). Las principales funciones kernel utilizadas en la support vector machine son (Jian et al. 2007):

- Lineal $K(x_i, x_j) = x_i^T x_j$
- Polinomial $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$
- Función radial base (RBF) $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$
- Sigmoidal $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Donde γ, r y d son los parámetros kernel de la función.

3. RESULTADOS

Para el análisis de la elección conveniente de indicadores técnicos para la predicción de la tendencia de un índice bursátil, se ha seleccionado los datos diarios del índice Dow Jones Industrial. El periodo analizado comprende del 2 enero de 1920 al 13 de febrero del 2018.

En la tabla 1 puede verse la descripción de los indicadores utilizados como inputs para su estudio como son el CCI, William's percentage range, ATR, ADX y PSY.

La muestra ha sido dividida en un 80% para el entrenamiento y el resto para el testeo. Como output del modelo se ha indicado la tendencia del índice bursátil comparando el cierre de la sesión anterior con la sesión actual, siendo 1 cuando la tendencia es alcista y 0 cuando se trata de tendencia bajista.

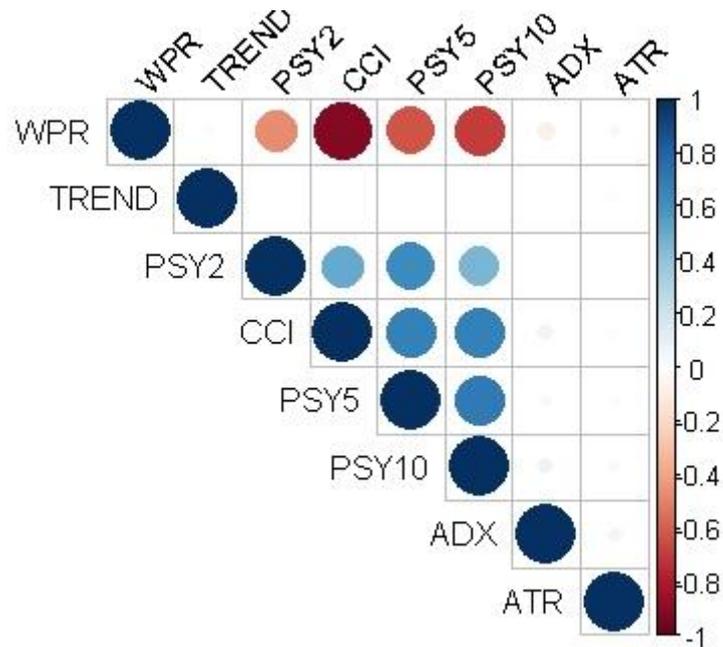
El modelo empleado para el estudio del comportamiento de los diferentes indicadores técnicos en la predicción de la tendencia, ha sido la support vector machine con la función radial base como función kernel.

Tabla 1. Indicadores técnicos analizados

Indicador	Descripción	Ecuación
$CCI(n)_t$	Commodity channel index ($n = 20$)	$(M_t - SM(n)_t)/(0.015 * D(n)_t)$
$PSY(n)_t$	Ratio of rising days ($n = 2, 5, 10$)	$A(n)_t/n$
$LW\%R(n)_t$	William's percentage range ($n = 14$)	$100 \times (H_n - C_t)/(H_n - L_n)$
ATR	Average True Range	$[(ATR_{t-1} * 13) + ATR_t]/14$
ADX	Average Directional Index	$[ADX_{t-1} * 13 + DX_t]/14$

En primer lugar se analiza las correlaciones de los indicadores técnicos con la tendencia del índice bursátil (figura 2). Como puede observarse, existe una correlación muy baja entre éstos y la dirección del índice. Esto se debe fundamentalmente a la definición de la tendencia como variable dicotómica. Si se tomase la correlación con los precios de cierre diarios del índice, las correlaciones aumentarían en la mayoría de casos. Este hecho no indica que los indicadores no sean buenos predictores de la tendencia. Para determinar su conveniencia debe analizarse mediante un modelo de clasificación como la support vector machine. Se destaca la alta correlación existente entre el indicador CCI y William's. Esta información puede ser relevante para su incorporación como inputs en modelos lineales, dado los problemas que pueden provocar como la multicolinealidad. Sin embargo, pueden ser necesarios en modelos más complejos como las redes neuronales.

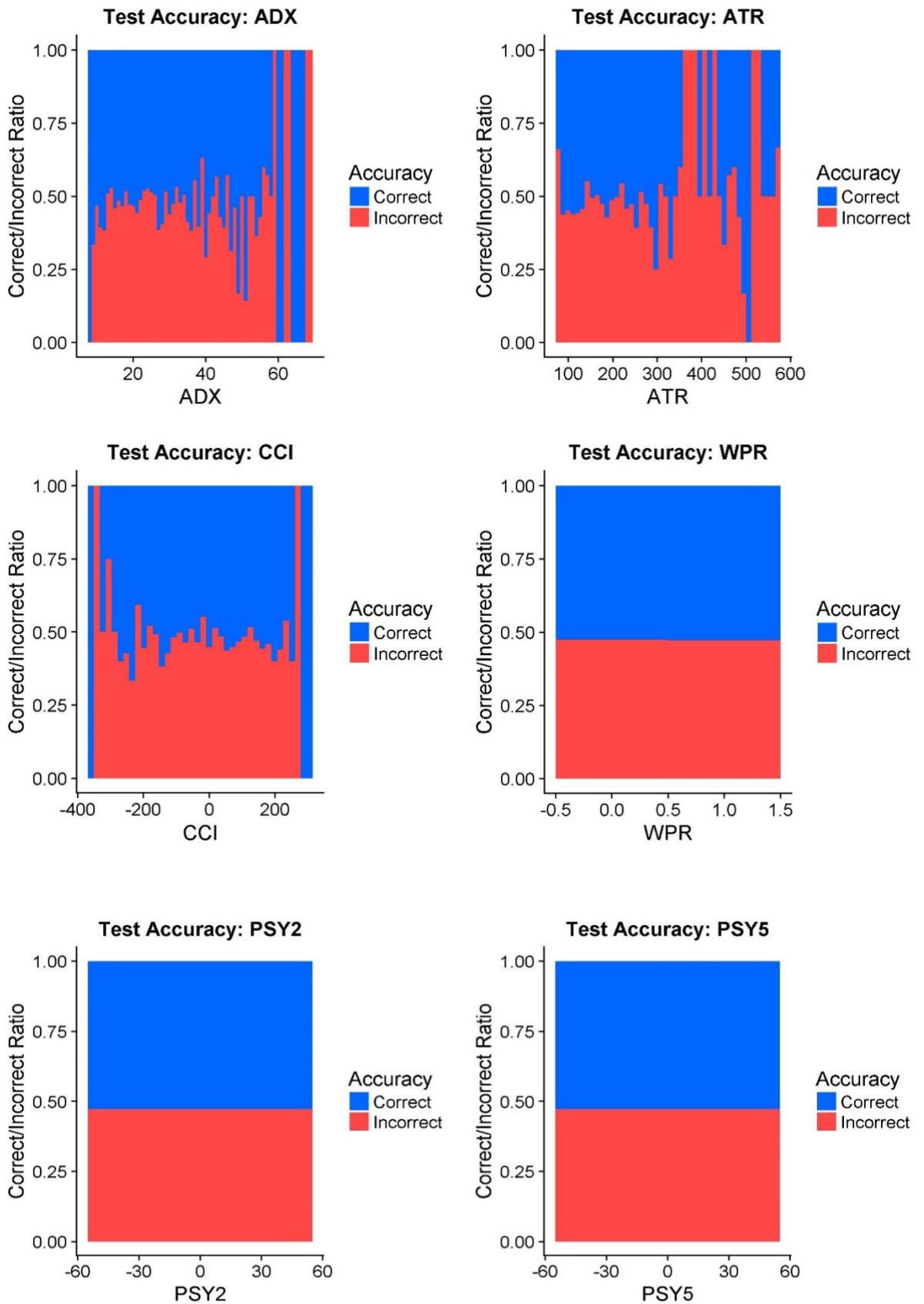
Figura 2. Correlaciones cruzadas de los indicadores

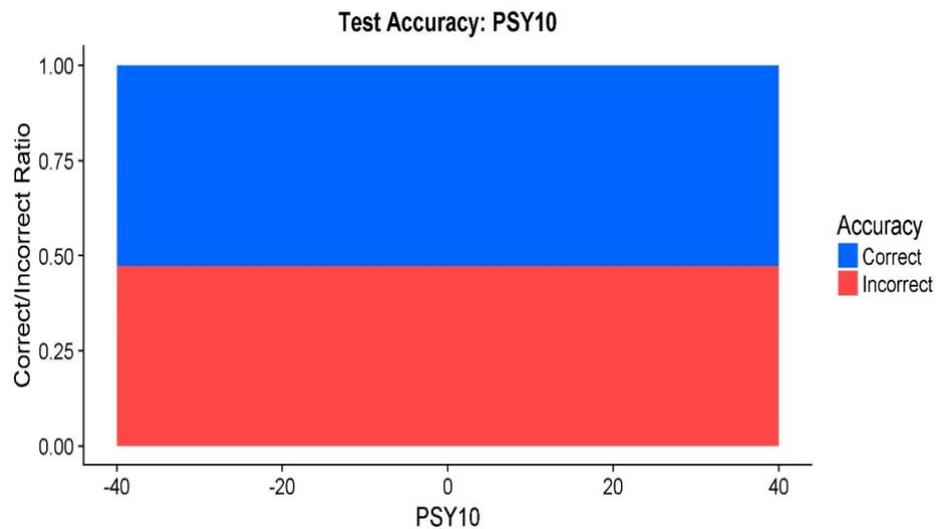


En la figura 3 puede verse los principales resultados de la utilización de los indicadores técnicos seleccionados en el modelo SVM. Dado que se trata de un modelo de clasificación, permite comprender la conveniencia de unos indicadores frente a otros. Para cada uno de ellos, se determina el ratio de errores y aciertos en la predicción de la tendencia del índice. Los indicadores ADX, ATR y CCI presentan irregularidades en el ratio de acierto de la tendencia mientras que el indicador William's y el indicador PSY para cada uno de los periodos seleccionados son más estables en su relación a la tasa de aciertos.

A pesar que en algún caso los primeros indicadores puedan presentar en algún momento tasas de acierto muy elevadas, dicha irregularidad los convierte en inputs inestables para el modelo. Por tanto, serían recomendables aquellos que presenten una estabilidad en su tasa de aciertos, incluso en el caso de tasas algo más reducidas. Los indicadores cuya tasa de acierto de la predicción es irregular puede provocar problemas en las estimaciones y procesos de aprendizaje de los diferentes modelos.

Figura 3. Análisis de los indicadores en el modelo SVM





4. CONCLUSIONES

El modelo de clasificación Support Vector Machine se ha utilizado, en este caso, para el estudio de diferentes indicadores técnicos para la predicción de la tendencia de un índice bursátil. El objetivo del modelo es determinar qué indicadores son los más adecuados para su uso como inputs en un modelo posterior, bien en un modelo SVM u otro como una red neuronal. Por tanto, sirve como clasificación de los mismos. La conveniencia de la elección de indicadores que presenten tasas de acierto estables evidencia la utilización de este tipo de modelos.

En estudios futuros puede analizarse cómo la preselección de indicadores técnicos mediante la utilización de modelos de clasificación puede mejorar la predicción de la tendencia de los indicadores técnicos. Para ello, pueden incorporarse estos indicadores estables como inputs para otros modelos.

5. REFERENCIAS

- Ballings, M.; Van den Poel, D.; Hespeels, N.; Gryp, R. (2015). Evaluating multiple classifiers for stock Price direction prediction. *Exper Systems with Applications*, 42, 7046-7056.
- Chih-Wei, H.; Chih-Chung, C., Chih-Jen, L. (2010). A Practical Guide to Support Vector Classification. Department of Computer Science. National Taiwan University. <http://www.csie.ntu.edu.tw/~cjlin>
- Cortes, C.; Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297
- Duta, A.; Bandopadhyay, G.; Sengupta, S. (2015). Predicting of Stock Performance in Indian Stock Market using Logistic Regression. *International Journal of Business and Information*, 7, 105-136
- García, F.; Guijarro, F.; Oliver, J. (2017). Index tracking optimization with cardinality constraint: a performance comparison of genetic algorithms and tabu search heuristics. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-017-2882-2>
- George, S.; Valavanis, K. (2009). Surveying stock market forecasting techniques-Part II: Soft computing methods. *Expert Systems with Applications*, vol.3, 3, part2, 5932-5941.
- Huang, C.; Yang, D., Chuang, Y. (2008). Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications*, 34, 2870-2878.
- Huang, W.; Nakamori, Y., Shou-Yabg, W. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32, 2513-2522.
- Jiang, T.; Wang, S., Wei, R. (2007). Support Vector Machine with Composite Kernels for Time Series Prediction. *LNCS 4493, Part. III*, 350-356.
- Kara, Y.; Boyacioglu, M.A., Baykan, O.M. (2011). Predicting direction of stock Price index movement using artificial neural networks and support vector machines: The simple of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38, 5311-5319.

- Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 307-319.
- Kumar, D.; Meghwani, S.S.; Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science*, 17, 1-13
- Moghaddam, A.H.; Moghaddam, M.H.; Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, 21, 89-93.
- Oliver, J. (2016). Modelling conditional volatility in the spanish Ibex-35 stock index using high frequency data. A comparison of the egarch model and the backpropagation neural network. *Finance, Markets and Valuation*, vol. 2, 2, 21-37.
- Oztekin, A.; Kizilaslan, R.; Freund, S., Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal Operational Research*, 253, 697-710.
- Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. (2015). Predicting stock and stock Price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, 42, 259-268.
- Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. (2015). Predicting stock market index using fusión of machine learning techniques. *Expert Systems with Applications*, 42, 2162-2172.
- Qui, M.; Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *Plos Ones*, 11 (5), 1-11
- Rumelhart, D.E.; Hinton, G.E., Williams, R.J.(1986). Learning internal representations by backpropagating errors. *Nature*, 323, 533-536.
- Smola, A.J.; Schölkopf, B., Müller, K.R. (1998b) The connection between regularization operators and support vector kernels. *Neural Networks*, 11, 637-649
- Smola, A.J., Schölkopf, B. (2003). A Tutorial on Support Vector Regression. *NeuroCOLT Technical Report*, 98-030.
- Vapnik, V.; Golowich, S., Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information*, 9, 281-287

Zaidi, M. (2016). Forecasting stock market trends by logistic regression and neural networks: Evidence from Ksa stock market. *International Journal of Economics, Commerce and Management*, IV, 6, 220-234.

Zarogianni, E.; Moorhead, T.W.; Lawrie, S. (2013). Towards the identification of imaging using multivariate pattern classification. *NeuroImage: Clinical*, 3, 279-289.