



KSCHOOL

PREDICCIÓN DEL VALOR DE LAS ACCIONES EN EL ÍNDICE IBEX 35

TRABAJO FIN DE MÁSTER

Melisa Fernández Montero SEPTIEMBRE / 2020

ÍNDICE DE CONTENIDO

INTRODUCCIÓN	3
OBJETIVOS	3
ORGANIZACIÓN DEL PROYECTO.....	4
COMPRENSIÓN DE LOS DATOS	5
PREPARACIÓN DE LOS DATOS - ETLs	7
MODELOS UTILIZADOS EN EL PROYECTO	8
COMPARATIVA DE LOS MODELOS IMPLEMENTADOS	8
VISUALIZACIÓN	10
CONCLUSIONES	14

INTRODUCCIÓN

Hoy en día debido a una mayor cantidad y facilidad de acceso al contenido, un gran avance en la parte tecnológica y la disponibilidad de técnicas como Machine Learning se están creando herramientas predictivas cada vez más eficientes; esto aplicado a los mercados financieros puede tener grandes beneficios al intentar obtener la mayor rentabilidad posible asumiendo los menores riesgos.

Según José Luis Crespo Espert, Investigador del Instituto de Análisis Económico y Social (IAES), y Profesor titular de Economía Financiera de la Universidad de Alcalá (UAH) comenta «En la actualidad hay estudios que señalan que las cuatro quintas partes de la negociación en el mercado bursátil se realiza mediante trading automático o algorítmico, siendo los principales actores los inversores institucionales y grandes firmas de intermediación frente los pequeños inversores individuales.». También destaca sobre todo que en los próximos años «se espera un gran crecimiento de este tipo de trading», el aumento podría llegar a ser de una tasa superior al 10 por ciento. (Obtenido de <https://bigdatamagazine.es/trading-con-ia-una-forma-de-ganar-dinero-sin-saber-de-la-bolsa>)

El objetivo del presente trabajo es desarrollar y obtener el mejor algoritmo que permita predecir a corto plazo el comportamiento del IBEX 35 utilizando técnicas de aprendizaje supervisado. Para ello se ha utilizado el histórico del IBEX 35, los históricos de los 10 principales componentes que conforman el IBEX 35 (AENA, AMS, BBVA, CABK, FER, IBE, ITX, REP, SAN, TEF), además de otras métricas genéricas como son la deuda pública, IPC, tasa de paro y PIB.

El IBEX 35 al ser un índice ponderado por capitalización bursátil implica que no todos los valores que integran el índice influyen de igual manera en la variación final del selectivo español. La entrada o salida de valores del índice se realiza dos veces al año por el Comité Asesor Técnico (CAT); esto hace que los componentes pueden ir variando (la última modificación en los componentes fue el 7 de agosto de 2020). Por ambas razones, se decidió utilizar los 10 principales componentes ya que la totalidad de su ponderación supone un 70,70% del índice y aparte para disponer de histórico suficiente (lo marca la entrada de AENA en el índice el 12-02-2015).

Componentes

IBEX 35

Nombre valor	Código	% Pond	Nombre valor	Código	%Pond	Nombre valor	Código	% Pond
SANTANDER	SAN	12,67	GRIFOLS	GRF	2,69	INM. COLONIA	COL	0,68
IBERDROLA	IBE	11,76	ACS CONST.	ACS	2,17	ARCEL.MITTAL	MTS	0,66
INDITEX	ITX	11,73	ENDESA	ELE	2,10	ACCIONA	ANA	0,60
BBVA	BBVA	6,84	RED ELE.CORP	REE	1,96	ACERINOX	ACX	0,54
TELEFONICA	TEF	6,77	NATURGY ENER	NTGY	1,79	MASMOVIL	MAS	0,52
AMADEUS IT	AMS	6,21	SIEMENS GAM.	SGRE	1,28	VISCOFAN	VIS	0,46
REPSOL	REP	4,29	BANKINTER	BKT	1,21	CIE AUTOMOT.	CIE	0,44
FERROVIAL	FER	4,01	B. SABADELL	SAB	1,21	INDRA "A"	IDR	0,36
CAIXABANK	CABK	3,42	MERLIN PROP.	MRL	1,20	MELIA HOTELS	MEL	0,29
AENA	AENA	3,09	ENAGAS	ENG	1,10	MEDIASET ESP	TL5	0,29
INT.AIRL.GRP	IAG	2,94	MAPFRE	MAP	0,90	ENCE	ENC	0,18
CELLNEX	CLNX	2,91	BANKIA	BKIA	0,71			

OBJETIVOS

Los objetivos que se van a cubrir son:

- Mediante el uso de técnicas como web scraping / selenium se realizará la recopilación de la información.
- Creación de distintas ETLs para el procesamiento de dicha información.
- Utilización de los siguientes modelos: Regresión lineal múltiple, Series temporales y Support vector regression.

Este apartado se ha dividido en dos fases:

- Fase 0: se utilizaba como predictores la fecha de apertura y el volumen de la acción para ese día. Al final se decidió utilizar otros predictores porque intentar el precio de cierre en base a estos dos predictores no tenía mucho sentido. Se ha dejado el trabajo realizado pero no se tiene en cuenta en la elección final del mejor modelo.
- Fase 1: únicamente se han tenido en cuenta como predictores métricas genéricas.

- Fase 2: se han tenido en cuenta como predictores métricas genéricas y las métricas de los 10 principales componentes que conforman el índice.
- Mediante el uso de Streamlit se va a proporcionar una interface web que nos permitirá visualizar las predicciones para los próximos días.

ORGANIZACIÓN DEL PROYECTO

El código fuente se ha subido al siguiente repositorio:

https://github.com/yensaiFM/TFM_KSCHOOL_Data_Science

Como algunos modelos ocupan más de 100 Mb y no se pueden subir al repositorio git se pueden descargar desde la siguiente url:

<https://drive.google.com/file/d/1tyYCC0762zhul20jeyml7o4OCVjPNYwg/view?usp=sharing>

El presente proyecto está formado por los siguientes ficheros:

- **ETLs**
 - 01_Generate_historic_data.py
Este script se encarga de unificar los datos históricos con las métricas del índice IBEX 35 y los distintos componentes.
- **Recopilación de los datos**
 - 02_Export_historical_data_web_scraping.py
Utilizando la técnica de web scraping obtenemos los valores de cierre del índice IBEX 35 y los distintos componentes. Los datos se obtienen de la página web 'https://es.investing.com/'.
 - 03_Export_other_metrics_web_scraping.py
Utilizando las técnicas de web scraping y selenium se obtiene información de las métricas PER, BPA, Rentabilidad por dividendo, Precio/Cash flow, Precio/Valor contable, ebitda y beneficio neto del índice IBEX 35 y los distintos componentes. Los datos se obtienen de la página web 'https://www.expansion.com/mercados/cotizaciones/indices/ibex35_.IB.html'
 - 04_Export_default_other_metrics_web_scraping.py
Utilizando la técnica de web scraping se obtiene información de las métricas por defecto PER, Precio/Cash flow, Precio/Valor contable, ROE, dividendo neto por acción, BPA, Rentabilidad por dividendo, beneficio neto y ebitda de los diez componentes más importantes del IBEX 35. Los datos se obtienen de la siguiente página web 'https://www.expansion.com'.
- **Análisis de los modelos**
 - **FASE 0**
 - 06_Multiple_lineal_regression_data_filterby_20150212_apertura+vol.ipynb
Análisis desarrollado en python utilizando un modelo de regresión lineal múltiple utilizando únicamente los predictores apertura y volumen del IBEX 35.
 - 06_Multiple_lineal_regression_data_filterby_20150212_apertura+vol.Rmd
Análisis desarrollado en R utilizando un modelo de regresión lineal múltiple utilizando únicamente los predictores apertura y volumen del IBEX 35.
 - 07_Series_temporales_all_data_apertura+vol.ipynb
Análisis desarrollado en python utilizando los modelos ARIMA(p,d,q) sin variables exógenas y utilizando como variables exógenas apertura y volumen.
 - **FASE 1**
 - 06_01_Multiple_lineal_regression_all_data_deuda_publica+ipc+tasa_paro+pib.ipynb
Análisis desarrollado en python utilizando un modelo de regresión lineal múltiple utilizando los predictores deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset completo con los datos desde el 09-09-1991.
 - 06_01_Multiple_lineal_regression_all_data_deuda_publica+ipc+tasa_paro+pib.Rmd
Análisis desarrollado en R utilizando un modelo de regresión lineal múltiple utilizando los predictores deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset completo con los datos desde el 09-09-1991.

- [06_01_Multiple_lineal_regression_data_filterby_20150212_deuda_publica+ipc+tasa_paro+pib.ipynb](#)
Análisis desarrollado en python utilizando un modelo de regresión lineal múltiple utilizando los predictores deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset filtrado desde el 12-02-2015.
- [06_01_Multiple_lineal_regression_data_filterby_20150212_deuda_publica+ipc+tasa_paro+pib.Rmd](#)
Análisis desarrollado en R utilizando un modelo de regresión lineal múltiple utilizando los predictores deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset filtrado desde el 12-02-2015.
- [07_01_Series_temporales_all_data_deuda_publica+ipc+tasa+paro+pib.ipynb](#)
Análisis desarrollado en python utilizando los modelos ARIMA(p,d,q) sin variables exógenas y utilizando como variables exógenas deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset completo con los datos desde el 09-09-1991.
- [07_01_Series_temporales_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib.ipynb](#)
Análisis desarrollado en python utilizando los modelos ARIMA(p,d,q) y SARIMA(p,d,q)x(P,D,Q)s sin variables exógenas y utilizando como variables exógenas deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset filtrado desde el 12-02-2015.
- [08_01_Support_vector_regression_all_data_deuda_publica+ipc+tasa+paro+pib.ipynb](#)
Análisis desarrollado en python utilizando el modelo support vector regression utilizando los predictores deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset completo con los datos desde el 09-09-1991.
- [08_01_Support_vector_regression_data_filtrery_20150212_deuda_publica+ipc+tasa+paro+pib.ipynb](#)
Análisis desarrollado en python utilizando el modelo support vector regression utilizando los predictores deuda pública, IPC, tasa de paro y PIB únicamente para el índice IBEX 35. Se ha utilizado el dataset filtrado desde el 12-02-2015.

- FASE 2

- [06_02_Multiple_lineal_regression_data_filterby_20150212_deuda_publica+ipc+tasa_paro+pib+other_components.ipynb](#)
Análisis desarrollado en python utilizando un modelo de regresión lineal múltiple utilizando como predictores deuda pública, IPC, tasa de paro, PIB y los diez componentes más importantes que pertenecen a ese índice. Se ha utilizado el dataset filtrado desde el 12-02-2015.
- [07_02_Series_temporales_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib+other_components.ipynb](#)
Análisis desarrollado en python utilizando los modelos ARIMA(p,d,q) y SARIMA(p,d,q)x(P,D,Q)s utilizando como variables exógenas deuda pública, IPC, tasa de paro y PIB y los diez componentes más importantes que pertenecen a ese índice. Se ha utilizado el dataset filtrado desde el 12-02-2015.
- [08_02_Support_vector_regression_data_filtrery_20150212_deuda_publica+ipc+tasa+paro+pib+other_components.ipynb](#)
Análisis desarrollado en python utilizando el modelo support vector regression utilizando los predictores deuda pública, IPC, tasa de paro, PIB y los diez componentes más importantes que pertenecen a ese índice. Se ha utilizado el dataset filtrado desde el 12-02-2015.

- Visualización

- Todo lo correspondiente al directorio webpage.
Desde este website el usuario podrá obtener las predicciones para los próximos días de los modelos con mejor MSE obtenido en las distintas fases.

COMPRENSIÓN DE LOS DATOS

Esta fase se puede considerar una de las más importantes ya que sin un buen dato los resultados obtenidos en futuras fases no serán adecuadas. En definitiva, es importante entender que el análisis predictivo no es magia y que los algoritmos aprenden en base a los datos que les proporcionamos, por lo que el éxito del modelo depende principalmente de los datos de entrada.

Recopilación de los datos

El IBEX 35 al ser un índice ponderado por capitalización bursátil implica que no todos los valores que integran el índice influyen de igual manera en la variación final del selectivo español. La entrada o salida de valores del índice se realiza dos veces al año por el Comité Asesor Técnico (CAT); esto hace que los componentes pueden ir variando (la última modificación en los componentes fue el 7 de agosto de 2020). Por ambas razones, se decidió utilizar los 10 principales componentes ya que la totalidad de su ponderación supone un 70,70% del índice y aparte para disponer de histórico suficiente (lo marca la entrada de AENA en el índice el 12-02-2015).

Para la recopilación de los distintos tipos de información se han utilizado técnicas de web scraping y Selenium y también manuales.

Datos obtenidos de forma manual:

- Histórico del índice IBEX 35 y sus componentes: <https://es.investing.com/indices/spain-35-historical-data>
- Métrica tasa de paro: <https://www.epdata.es/evolucion-tasa-paro-espana/7c16eeb2-b201-450b-b56c-e00df71ee371/espana/106>
- Métrica deuda pública: <https://www.epdata.es/asi-aumentado-deuda-publica/f4dcc707-9e28-4d16-aadb-0010a33a4db5/espana/106>
- Métrica producto interior bruto (PIB): <https://www.epdata.es/evolucion-pib-espana/60ed4f95-ea9f-47ff-a4ae-c5c1b85b6aaa/espana/106>
- Métrica IPC general: <https://www.epdata.es/variacion-interanual-ipc-2003/c561378f-ab32-4681-8771-a1094d5657bc>

Datos obtenidos utilizando técnicas de web scraping y selenium:

- Valores de cierre del IBEX 35 y los diez componentes más importantes que pertenecen a ese índice: <https://es.investing.com>
- Desde la pestaña Ratios se han obtenido las métricas diarias PER, BPA, Rentabilidad por dividendo, Precio/Cash flow, Precio valor contable, EBITDA y B neto para los diez componentes más importantes que conforman el IBEX 35: https://www.expansion.com/mercados/cotizaciones/indices/ibex35_1.IB.html
- Desde la pestaña Análisis Fund. se han obtenido los históricos por año de las métricas comentadas anteriormente para los diez componentes más importantes que conforman el IBEX 35: https://www.expansion.com/mercados/bolsa/recomendaciones/analisis-fundamental/a/acciona_M.ANA.html (esta url sería para ACCIONA)

Estos scripts se ejecutan todos los días de Lunes a Viernes a partir de la hora de cierre de la Bolsa Española (17:35 horas).

Descripción de los datos

A continuación se muestra una breve descripción de las distintas variables predictoras:

Variable	Descripción
Fecha	Día de la muestra
Último	valor de cierre del índice IBEX 35 expresado en puntos para ese día (unidad puntos)
Deuda pública	conjunto de deudas que mantiene el Estado español frente a los particulares que pueden ser españoles o de otro país
IPC	indicador que mide la variación de los precios de una cesta de bienes y servicios en un lugar concreto durante un determinado periodo de tiempo
Tasa paro	mide el nivel de desocupación en relación con la población activa
PIB	suma de todos los bienes y servicios finales que produce un país o una economía, elaborados dentro del territorio nacional tanto por empresas nacionales como extranjeras, y que se registran en un periodo determinado
PER	(Price earning ratio o relación precio-beneficio). Mide la relación entre el precio en bolsa de una acción y los beneficios que obtiene año tras año.

Precio /Cash Flow	(P/FCF o "Price to Free Cash Flow"). Este ratio relaciona la capitalización bursátil de una compañía con sus flujos de caja libres.
Precio /Valor contable	(PVC o P/BV o "Price /Book Value"). Se obtiene de dividir el precio de una acción por su valor teórico contable, o lo que es equivalente, dividir la capitalización bursátil entre los fondos propios.
ROE	Rentabilidad sobre los recursos propios.
Dividendo neto por acción	
BPA	(Beneficio por acción). Cuánto de lo obtenido por la empresa como resultado de su actividad, correspondería a cada acción.
Rentabilidad por dividendo (últimos 12 meses)	Mide el porcentaje del precio de una acción que va a parar a los accionistas en forma de dividendo cada año.
Beneficio neto (millones)	
EBITDA	Ganancias antes de intereses, impuestos, depreciación y amortización.

En la FASE 2 del Análisis de los modelos al final se decidió utilizar únicamente las métricas PER, Precio valor contable, Rentabilidad por dividendo, beneficio por acción y EBITDA porque están enfocadas a ver el valor real de la compañía y todas se basan en datos de rentabilidad real y valor de los activos de la empresa.

PREPARACIÓN DE LOS DATOS - ETLs

El objetivo de esta fase es unir, transformar, manipular los datos de manera que podamos obtener un dataset de calidad. Los tratamientos que se han realizado son los siguientes:

- FASE 0
 - Se han dejado únicamente las variables predictoras apertura y volumen. Y como variable a predecir el valor último del índice IBEX 35.
- FASE 1
 - Se han dejado como variables predictoras únicamente las métricas generales deuda pública, IPC, tasa de paro y PIB. Y como variable a predecir el valor último del índice IBEX 35.
- FASE 2
 - Se han dejado como variables predictoras las métricas generales deuda pública, IPC, tasa de paro y PIB y como métricas específicas de cada uno de los diez componentes más importantes que componen el IBEX 35 último, PER, BPA, rentabilidad por dividendo, precio valor contable y EBITDA. Y como variable a predecir el valor último del índice IBEX 35.

En todas las fases ha sido necesario realizar las siguientes transformaciones:

- Cambiar el formato de miles y decimales
- En el campo volumen transformar los tipos de unidades (M millones) y B (billones)
- Quitar el % en el campo variación
- Rellenar los valores nulos utilizando el método "forward fill" de forma que los valores no nulos se copien hacia adelante siempre que se encuentren valores nulos
- Convertir los NaN a 0
- Convertir las columnas a float
- Añadir las fechas no presentes en el dataset y establecer el periodo B (business day frequency - de Lunes a Viernes) (requerido para el modelo de series temporales)

También se han realizado pruebas utilizando el dataset al completo con datos desde el 09-09-1991 y con datos filtrados desde el 12-02-2015; esto lo marca la entrada de AENA en el índice IBEX 35.

Para la unificación de todas estas variables predictores (proceden de distintos ficheros) se hace uso del script 01_Generate_historic_data.py; este permite definir mediante un flag si deseamos los datos filtrados o no. La salida será tantos csv como índice y componentes dispongamos.

MODELOS UTILIZADOS EN EL PROYECTO

Regresión lineal múltiple

Nos permite establecer la relación que se produce entre una variable dependiente Y y un conjunto de variables independientes (X1, X2,...Xn). Este análisis se aproxima más a situaciones de análisis real puesto que los fenómenos son complejos y deben ser explicados en la medida de lo posible por la serie de variables que, directa o indirectamente, participan en su concreción. En este tipo de modelos hay ciertos requisitos necesarios:

- Linealidad: la variable respuesta depende linealmente de las variables explicativas.
- Normalidad y equidistribución de los residuos: los residuos son las diferencias entre los valores predichos por el modelo y los realmente observados en la variable dependiente. Para tener un buen modelo de regresión no es suficiente con que los residuos sean pequeños; también se requiere que los mismos se distribuyan de modo normal y con la misma dispersión.

Series temporales

Estos modelos se basan en el estudio estadístico de muestras de variables recogidas secuencialmente a lo largo del tiempo. En este tipo de modelos hay tres elementos básicos a tener en cuenta: la tendencia, la estacionariedad y la aleatoriedad.

- Tendencia: mide si temporalmente los valores tienen una direccionalidad hacia arriba o hacia abajo; es decir, capta una pendiente general de los valores.
- Estacionariedad: mide la presencia de ciclos, de subidas y bajadas realizadas con una determinada regularidad.
- Aleatoriedad: mide desvíos respecto a los anteriores elementos.

Support vector regression

Este tipo de modelos permite resolver problemas de estimación de funciones multidimensionales. Ofrece muy buenos resultados ya que selecciona el hiperplano regresor que mejor se ajuste a nuestro conjunto de datos de entrenamiento. Este hiperplano viene acompañado por un rango (máximo margen), tanto del lado positivo como en el negativo, el cual tiene el mismo comportamiento o forma; todos los datos que se encuentren fuera de él son considerados errores por lo que es necesario calcular la distancia entre el mismo y los rangos.

COMPARATIVA DE LOS MODELOS IMPLEMENTADOS

En esta tabla comparativa de modelos únicamente se han tenido en cuenta los implementados en las FASE 1 y FASE 2. Los modelos implementados en la FASE 0 se decidieron desestimar ya que las variables predictoras que se estaban utilizando no tenían mucho sentido.

La métrica que se utiliza para comparar los modelos es el MSE (Error cuadrático medio). Es una medida de dispersión del error del pronóstico, que maximiza el error al elevar al cuadrado castigando aquellos periodos donde la diferencia fue más alta a comparación de otros.

Los valores de MSE Train que se muestran en la tabla son los obtenidos después de aplicar la técnica de Cross Validation. Para calcular el MSE Test se han utilizado 20 registros.

La siguiente tabla muestra los resultados de los modelos de la **FASE 1**:

Modelo	Datos filtrados	Predictores	MSE Train	MSE Test	R-cuadrado
Regresión lineal - 06_01_Multiple_lineal_regression_all_data_deuda_publica+ipc+tasa_paro+pib.ipynb	NO	SI	5785258,37	5505813.44	0.47
Regresión lineal - 06_01_Multiple_lineal_regression_data_filt	SI	SI	531877,46	341764,74	0,74

<u>erby_20150212_deuda_publica+ipc+tasa_paro+pib.ipynb</u>					
Serie temporal ARIMA - <u>07_01_Series_temporales_all_data_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	NO	NO	12825.81	26460.83	
Serie temporal ARIMA - <u>07_01_Series_temporales_all_data_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	NO	SI	12816.16	26451.99	
Serie temporal ARIMA - <u>07_01_Series_temporales_data_filterby_2_0150212_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	SI	NO	241875.08	35144.21	
Serie temporal ARIMA - <u>07_01_Series_temporales_data_filterby_2_0150212_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	SI	SI	281203.68	36145.24	
Serie temporal SARIMA - <u>07_01_Series_temporales_data_filterby_2_0150212_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	SI	NO	227050.16	40274.23	
Serie temporal SARIMA - <u>07_01_Series_temporales_data_filterby_2_0150212_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	SI	SI	413995.02	18366.52	
SVR - <u>08_01_Support_vector_regression_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	NO	SI	245448.57	591445.44	0.90
SVR - <u>08_01_Support_vector_regression_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib.ipynb</u>	SI	SI	61521.21	51435.96	0.95

La siguiente tabla muestra los resultados de los modelos de la **FASE 2**:

Modelo	Datos filtrados	Predictores	MSE Train	MSE Test	R-cuadrado
Regresión lineal - 06_02_Multiple_lineal_regression_data_filterby_20150212_deuda_publica+ipc+tasa_paro+pib+other_components.ipynb	SI	SI	116905.03	33303.04	0.95
Serie temporal ARIMA - 07_02_Series_temporales_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib+other_components.ipynb	SI	SI	520.51	30354.90	
Serie temporal SARIMA - 07_02_Series_temporales_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib+other_components.ipynb	SI	SI	967.20	55979.35	
SVR - 08_02_Support_vector_regression_data_filterby_20150212_deuda_publica+ipc+tasa+paro+pib+other_components.ipynb	SI	SI	73360.54	118924.09	0.92

Se ha marcado en amarillo los cinco mejores modelos teniendo en cuenta la métrica MSE sobre los datos de test.

Para más detalle sobre el análisis de los diferentes modelos acceder a cada uno de los NoteBooks.

VISUALIZACIÓN

Desde este web site el usuario podrá ver las predicciones para los próximos n días.

Tomando como base el precio de cierre del índice IBEX 35 de hoy (10/09/2020) con un valor de 6999,20 podemos compararlo con el primer día de las predicciones propuestas por los distintos modelos.

Predicción del valor de las acciones en el índice IBEX 35

Fases implementadas

Fase 01

Elige el modelo a analizar:

Regresión Lineal Múltiple

Número de días a predecir:

0 5 10

La predicción para los próximos 5 días:

0	
0	6,988.5940
1	6,988.5940
2	6,988.5940
3	6,988.5940
4	6,988.5940



Predicción del valor de las acciones en el índice IBEX 35

Fases implementadas

Fase 01

Elige el modelo a analizar:

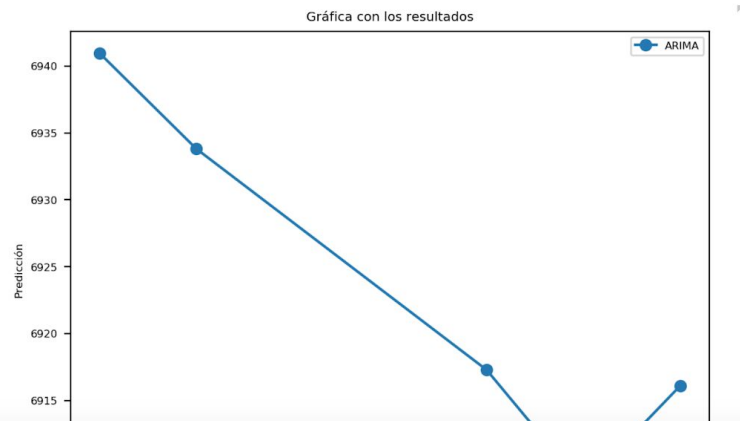
Serie temporal - ARIMA

Número de días a predecir:

0 5 10

La predicción para los próximos 5 días:

0	
Sep 10, 2020	6,940.9591
Sep 11, 2020	6,933.8117
Sep 14, 2020	6,917.2861
Sep 15, 2020	6,908.5850
Sep 16, 2020	6,916.0902



Fases implementadas

Fase 01

Elige el modelo a analizar:

Serie temporal - SARIMA

Número de días a predecir:

5

0 10

⑥

Sep 10, 2020	7,044.1568
--------------	------------

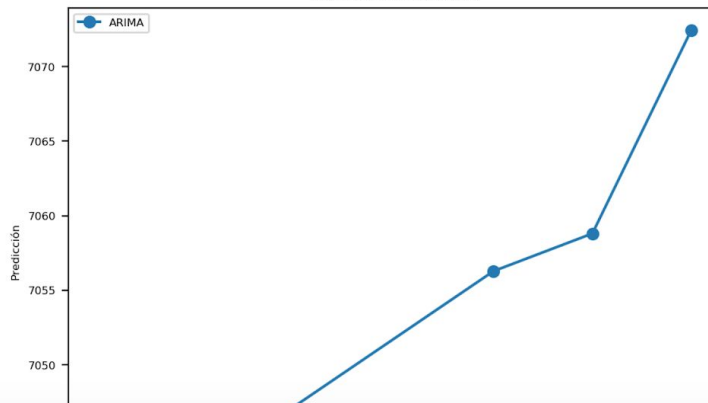
Sep 11, 2020	7,042.7760
--------------	------------

Sep 14, 2020	7,056.2875
--------------	------------

Sep 15, 2020	7,058.8116
--------------	------------

Sep 16, 2020	7,072.4444
--------------	------------

Gráfica con los resultados



Fases implementadas

Fase 01

Elige el modelo a analizar:

SVR

Número de días a predecir:

5

0 10

9

0 7,206.1001

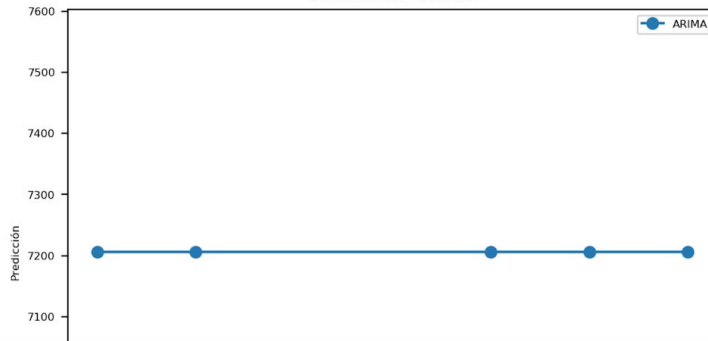
1 7,206.1001

2 7,206.1001

3 7,206.1001

4 7,206.1001

Gráfica con los resultados



Predicción del valor de las acciones en el índice IBEX 35

Fases implementadas

Fase 02

Elige el modelo a analizar:

Regresión Lineal Múltiple

Número de días a predecir:

0

5

10

La predicción para los próximos 5 días:

	0
0	6,213.0740
1	6,213.0740
2	6,213.0740
3	6,213.0740
4	6,213.0740



Predicción del valor de las acciones en el índice IBEX 35

Fases implementadas

Fase 02

Elige el modelo a analizar:

Serie temporal - ARIMA

Número de días a predecir:

0

5

10

La predicción para los próximos 5 días:

	0
Sep 10, 2020	3,826.1894
Sep 11, 2020	3,826.1894
Sep 14, 2020	3,826.1894
Sep 15, 2020	3,826.1894
Sep 16, 2020	3,826.1894



Predicción del valor de las acciones en el índice IBEX 35

Fases implementadas

Fase 02

Elige el modelo a analizar:

Serie temporal - SARIMA

Número de días a predecir:

0

5

10

La predicción para los próximos 5 días:

	0
Sep 10, 2020	2,615.4438
Sep 11, 2020	2,599.5638
Sep 14, 2020	2,636.0346
Sep 15, 2020	2,641.1825
Sep 16, 2020	2,660.9168



Predicción del valor de las acciones en el índice IBEX 35

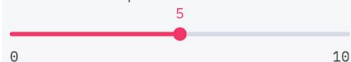
Fases implementadas

Fase 02

Elige el modelo a analizar:

SVR

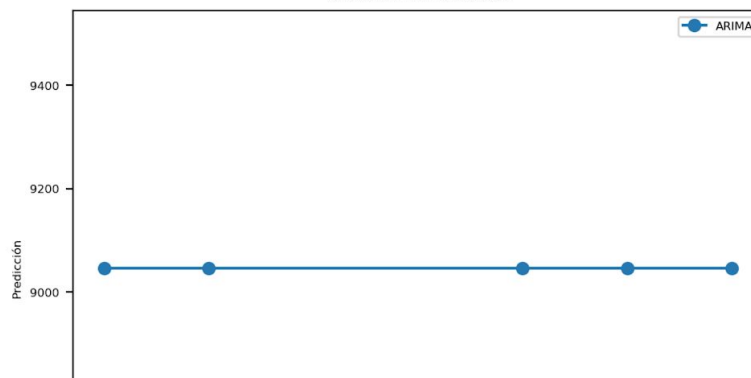
Número de días a predecir:



La predicción para los próximos 5 días:

0	
0	9,046.3413
1	9,046.3413
2	9,046.3413
3	9,046.3413
4	9,046.3413

Gráfica con los resultados



Como podemos observar, no todas las predicciones se aproximan tanto al precio de cierre real. Por lo general, los modelos de la FASE 1 responden mucho mejor que los modelos de la FASE 2. También hay que tener en cuenta que en la FASE 2 se está utilizando hasta un máximo de 61 variables predictoras y no se ha tenido demasiado tiempo para estudiar todas las variables.

CONCLUSIONES

El objetivo principal de este proyecto era “desarrollar y obtener el mejor algoritmo que permita predecir a corto plazo el comportamiento del IBEX 35 utilizando técnicas de aprendizaje supervisado”; según los modelos propuestos el que obtiene un menor MSE en Train sería la Serie temporal SARIMA de la FASE 1 y sin embargo la predicción propuesta para hoy 10/09/2020 es de 7044,15 existiendo una desviación con respecto a la real de “44,95” puntos. Esta desviación es bastante alta tratándose de la bolsa; en definitiva, los modelos obtenidos no son muy fiables para invertir uno su dinero.

Otro de los motivos por los que los modelos de la FASE 2 no están respondiendo tan bien es que al proporcionar las variables predictoras al modelo como no disponemos de dicha información estamos cogiendo los valores del último dato que disponemos y si para alguna de las variables no existe información se coge los valores por defecto; esto penaliza bastante las predicciones obtenidas.

Como mejoras a futuro se debería seguir trabajando en conseguir la estacionariedad de los datos en las series temporales, realizar un mayor análisis de las variables predictoras utilizadas en la FASE 2 e intentar probar con Redes neuronales.