Evaluación del Módulo 2 – Parte 2

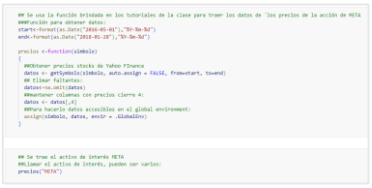
Contexto

Este reporte presenta un resumen detallado de los pasos llevados a cabo para aplicar y comparar varios modelos de pronóstico, (MARS, LSTM, RF y SSA) sobre la serie temporal de precios de cierre del activo META. El período de análisis comprende desde el 1 de mayo de 2016 hasta el 28 de enero de 2018. A continuación, se describen los procedimientos seguidos para cada uno de los modelos mencionados.

MARS

Para el modelo MARS se proceden a realizar los siguientes pasos.

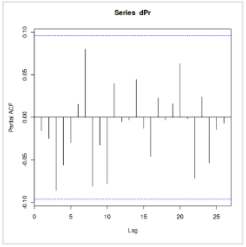
1. A través del uso de la herramienta de R Studio se procede a cargar la información.



2. Una vez cargada la información se procede a realizar el preprocesamiento de los datos: Se transforman los datos para que queden indexados como fecha, se realiza la partición en datos de entrenamiento y de prueba con estos últimos comprendiendo el 5% de los datos iniciales y se verifica si la tendencia es significativa y si existen rezagos significativos.

A tibble: 1 × 16

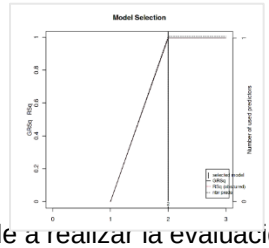
frequency	nperiods	seasonal_period	trend	spike	linearity	curvature	e_acf1	e_acf10
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	1	0.9901344	1.737658e-09	19.22551	4.38829	0.7349023	1.041272



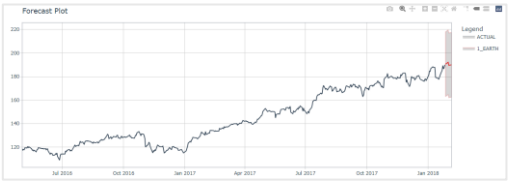
En el caso de estos datos se identificó que la tendencia si es significativa (el valor en tsfeatures fue cercano a 1) pero no se identificaron rezagos significativos por lo que se generan los primeros 5 rezagos para correr el modelo.

3. Se corre el modelo con los 5 rezagos y la tendencia. Y se verifica el resultado.

```
## Se corre el modelo MARS
#2) MARS
mars_mod <- earth(
  META ~ ., data=Pr_all_train, pmethod="backward" )
```



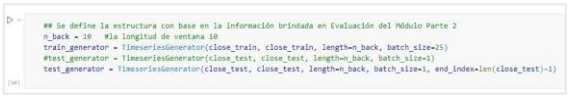
4. Posteriormente, se procede a realizar la evaluación del modelo mediante el cálculo de la Mean Squared Error (MSE) y Root Mean Squared Error (RMSE). Estas métricas son empleadas para determinar la eficacia del modelo y facilitar su comparación con otros en términos de desempeño. Este cálculo se realiza con la predicción con los datos de prueba.
5. Se realiza el pronóstico de 10 días con modeltime.



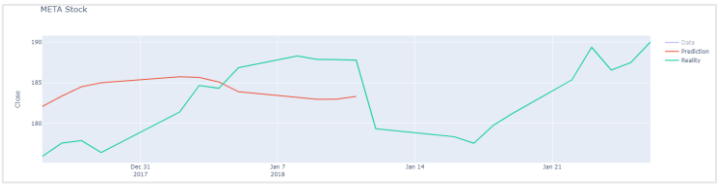
LSTM

1. A través del uso de Python se procede a cargar la información.

2. Una vez cargada se procede a modificar la información para que quede un dataframe como un objeto de tiempo y los precios de cierre queden como un array en numpy. Hecho esto se procede a realizar la partición en datos de entrenamiento y prueba y se procede a definir la estructura con una longitud de ventana de 10.



3. Se procede a correr el modelo con las características definidas en la evaluación y se evalúa el modelo con el MSE y RMSE como se realizó en el modelo de MARS.



4. Se realiza el pronóstico de 10 días.

Random Forest

1. Se carga la información de la misma forma que se realizó en LSTM.
2. Una vez cargada la información se procede a ajustar el formato de los datos a fecha con periodicidad diaria, llenar los espacios vacíos con el valor del día anterior y a dividir los datos en entrenamiento y prueba.

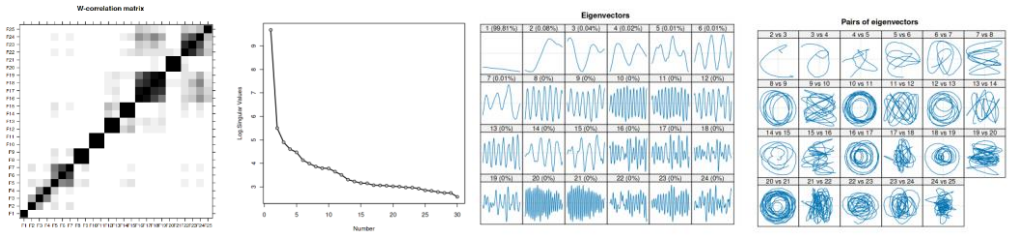
3. Se realiza el modelamiento con forecaster y 15 lags y se calcula la MSE y RMSE.
4. Se usa gridsearch para determinar los valores óptimos de los hiperparamateros.
5. Se corre el modelo óptimo.
6. Se realiza el pronóstico de 10 días.

***Singular Spectrum Analysis* (SSA)**

1. Con el uso de R Studio se carga la información que se usara en el modelo.

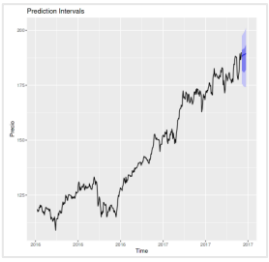
2. Se transforman los datos de formato xts a ts y se parten en datos de entrenamiento y prueba.

   Se aplica la función SSA. Se realiza la interpretación de las cuatro gráficas



   Con base en estas se determinó que los vectores 1,2,3,4 y 5 eran candidatos de vectores de tendencia y los pares 8 y 9, 10 y 11 representaban la estacionalidad por lo que se tomaron en cuenta durante el proceso de reconstrucción.

3. Se realiza la reconstrucción con estos parámetros. Se calculan las métricas MSE y RMSE con los datos de prueba.
4. Se realiza el pronóstico de 10 días con un SSA corrido con los datos completos.



Análisis de los Resultados

Se realiza un análisis pequeño de las ventajas y desventajas de cada modelo:

Modelo	Ventajas	Desventajas
MARS	- Capacidad para seleccionar variables de manera automática. - Buena dicotomía de sesgo-varianza.	- No es posible calcular un intervalo de confianza. - No generan tan buen ajuste como los modelos Boosted.
LSTM	- Capacidad para modelar dependencias a largo plazo en datos secuenciales. - Capacidad de dar resultados muy precisos.	- Mayor complejidad de implementación y ajuste de hiperparámetros.
Random Forest	- Manejo de relaciones no lineales	- Dificultad para interpretar los resultados.
SSA	- Descompone la serie temporal en componentes interpretables (tendencia, estacionalidad, ruido) - Robusto frente a la presencia de ruido y datos faltantes.	- Requiere un ajuste cuidadoso de los parámetros.

Con todos los modelos aplicados se procede a realizar un análisis de las métricas obtenidas para determinar qué modelo tuvo el mejor desempeño:

Métricas de error	MARS	LSTM	RF	SSA
RMSE	5,9969888	5,1693267822265625	6,054678958784857	4,78038400698083
MSE	35,96387467	29,558761596679688	36,65913729395208	22,8520712541981
MAE	5,237771			4,26223510796316

Como se puede observar en la tabla el modelo con el mejor desempeño fue el SSA. Esto tiene sentido ya que el modelamiento por SSA es recomendado para series de datos que presentan comportamientos no paramétricos y esto se suele presentar en datos de series de tiempo financieros.

Anexos

Para el desarrollo de este documento se realizó la aplicación de los 4 modelos definidos (MARS, LSTM, Random Forest y SSA) de acuerdo con lo que se solicitó. Este desarrollo junto con la explicación de cada paso realizado de forma más completa se puede encontrar en el siguiente repositorio de github:

Enlace: https://github.com/jmanjarresm/Pron-sticos---Evaluaci-n-Modulo-2

En este repositorio se pueden encontrar las carpetas cada una con el notebook correspondiente para el desarrollo del modelo.

Ejemplo: