



UNIVERSIDAD TECNOLÓGICA DE PANAMÁ
VICERRECTORÍA DE INVESTIGACIÓN, POSTGRADO Y EXTENSIÓN
DIRECCIÓN DEL SISTEMA DE ESTUDIOS DE POSTGRADO
FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES

MAESTRÍA EN ANALÍTICA DE DATOS

MODELOS PREDICTIVOS
PROYECTO FINAL
ANÁLISIS PREDICTIVOS DE SERIES TEMPORALES PARA EL ÍNDICE NYA
COMPOSITE DE LA BOLSA DE VALORES DE NEW YORK EXCHANGE

TRABAJO ELABORADO POR:
FRANCISCO D. TUDISCO S.

8-910-131

PROFESOR:
Ph. D. JUAN CASTILLO

2024

Indice

Introducción	3
Componentes Tecnológicos Utilizados	4
Recolección de los Datos	4
Tratamiento Y Procesado De Los Datos	5
Minado de los Datos con la herramienta de Weka	5
Análisis Exploratorio de Datos	6
Histograma de Frecuencia.....	7
Correlación.....	7
Modelos Estadísticos de Pronósticos Series Temporales	8
Promedio Móviles.....	8
Suavizamiento Exponencial Simple	9
Suavizamiento Exponencial Holt-Winter.....	10
Modelo SMOreg	11
Modelo RandomTree	12
Modelo MultiLayer Preceptron.....	13
Tabla de Resultados	14
Conclusiones	15
Referencias Bibliográficas	16

Introducción

El NYA Composite Index, es un índice compuesto de la Bolsa de Valores que mide el desempeño de diferentes acciones y activos, incluidas acciones extranjeras, American Depositary Receipts, fideicomisos de inversión inmobiliaria y acciones de seguimiento, todos aparecen en New York Exchange.

Los datos de NYA fueron obtenidos a través de la plataforma de Yahoo Finance, que abarcan a partir del año (enero 2023) a la fecha, estos datos son agrupados de forma semanal y son la fuente del análisis estadístico realizado en la presentación como proyecto final de la materia Modelos Predictivos.

Los modelos de pronósticos de Series temporales, son modelos estadísticos inferencial enfocado en el análisis basado en la suposición de que la historia de la demanda pasada es un buen indicador de la demanda futura. Estos modelos son herramientas que nos aportan información sobre la precisión y consistencias de la predicción realizada.

Los análisis estadísticos presentado tiene un enfoque de estadística inferencial, con la premisa de realizar predicción de Serie Temporal sobre ¿cómo va a cerrar el índice NYA al final de las siguientes semanas?

Componentes Tecnológicos Utilizados

Recolección de los Datos

Yahoo Finance

NYSE - Delayed Quote - USD

NYSE COMPOSITE (DJ) (^NYA)

☆ Follow

18,663.14 -214.86 (-1.14%)

At close: September 6 at 4:08 PM EDT



Ilustración 1: Portal Financiero de Yahoo! Finance

A través del portal <https://finance.yahoo.com/>, se recolectaron los datos históricos a partir de enero 2023 a la fecha, un año de datos agrupados semanalmente. Los datos obtenidos se descargaron en formato CSV, y el dataset tiene 7 columnas, que son Date, Open, High, Low, Close, Adj Close.

- El campo de Date, indica la fecha de cada lunes que inicia la semana en la Bolsa.
- Campo Open, indica el valor monetario en dólares que abre la transacción en la Bolsa.
- Campo High, indica el punto más alto en valor monetario en dólares de la transacción en la Bolsa.
- Campo Low, indica el punto más bajo en valor monetario en dólares de la transacción en la Bolsa.
- Campo Close, indica el valor monetario en dólares que cierra la transacción en la Bolsa.
- Campo Adj Close, indica el valor monetario en dólares que cierra ajustado de la transacción en la Bolsa.

Tratamiento Y Procesado De Los Datos

Posterior de la obtención de lo datos, es requerido emplear diferentes técnicas y métodos para procesar tratar los datos. Para llevar a cabo la tarea fue requiero la utilización del lenguaje de programación de Python, y librerías específicas, como:

- Pandas, para la manipulación y tratamiento de los datos.
- Numpy, librería específica para funciones matemáticas y estadísticas.
- Matplotlib, librería específica representar visualmente y graficar los datos.
- Sweetviz, librería específica para generar automáticamente Análisis Exploratorio de Datos.
- ydata_profiling.ProfileReport, librería específica para generar automáticamente Análisis Exploratorio de Datos.

La herramienta de Microsoft Excel, se utilizó como método final de almacenamiento de los datos, en otras palabras, cumplió una función de base de datos, y desarrollo de los modelos estadísticos de pronósticos.

Minado de los Datos con la herramienta de Weka

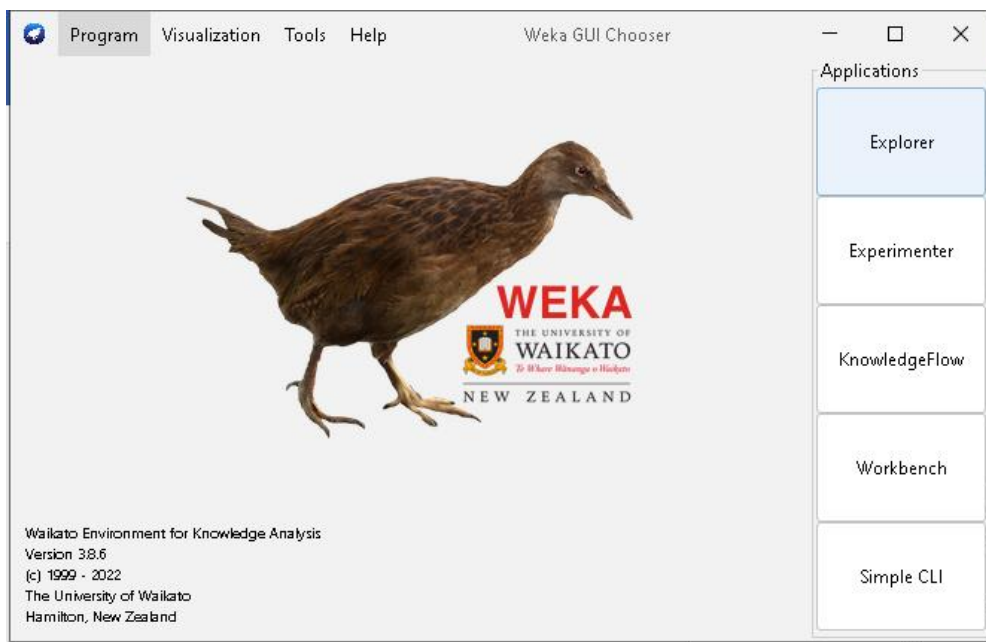


Ilustración 2: Weka, Herramienta de Minería de Datos.

Los datos procesados y limpiados, se utilizaron en la herramienta de Weka, en donde con la gran cantidad de modelos de Minería de datos se llevaron a cabo, múltiples ejecuciones

de modelos, con el objetivo de identificar que modelo era el más preciso y consistentes con el dataset de NYA Composite Index.

Análisis Exploratorio de Datos

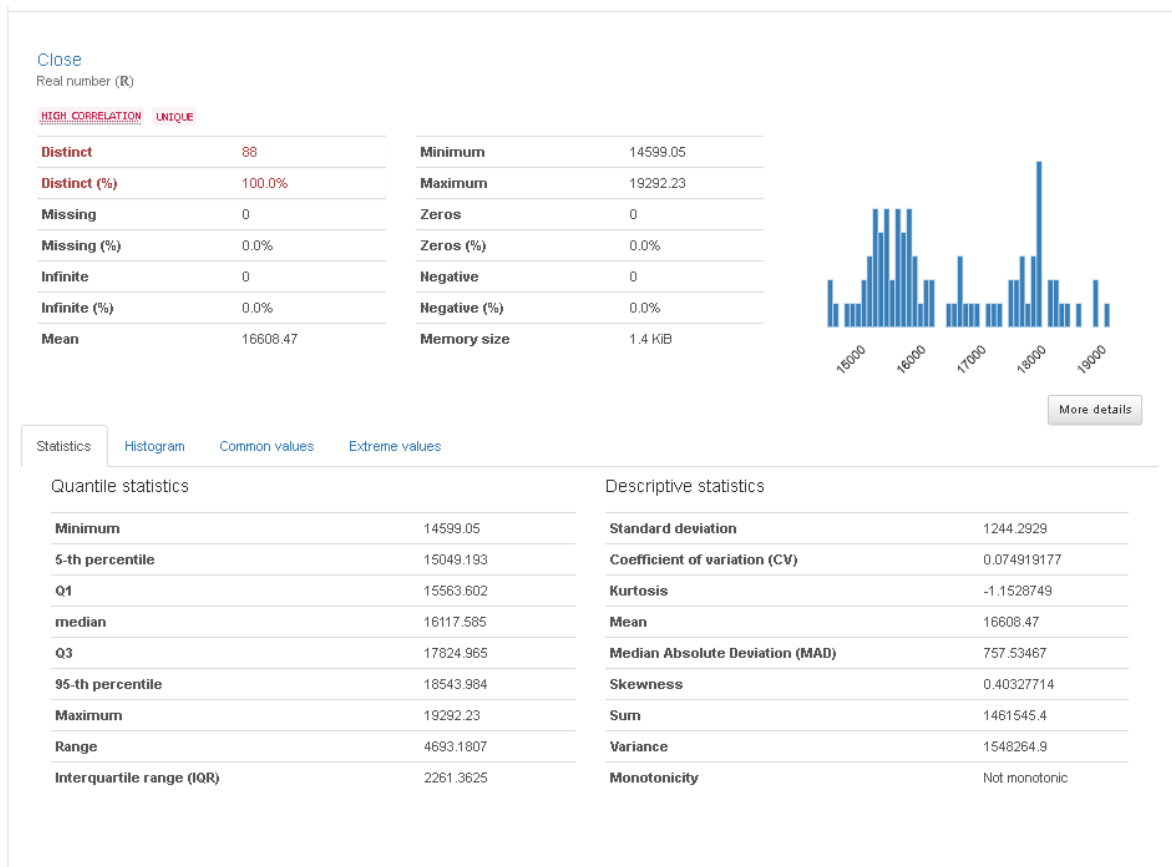


Ilustración 3: Análisis Exploratorio de Datos: Enfoque en el Atributo Close

Al dataset obtenidos se realizó un análisis exploratorio de los datos, con el objetivo de conocer detalladamente las características y descripciones de los datos, con este análisis se pudo obtener información de forma rápida como los puntos máximos, mínimos, desviación estándar, coeficiente de variación y otros valores de importancia para los análisis estadísticos de pronósticos de serie temporal.

En el análisis Exploratorio de Datos, indica que el atributo Close del dataset, presenta 88 valores distintos, siendo los valores máximos de 19,292.23, y valores mínimos de 14,599.05, con una media de 16,117.58, desviación estándar de 1244.29, y coeficiente de variación de 7.49%.

Histograma de Frecuencia

El análisis exploratorio de datos, genera un histograma de Frecuencia, en donde nos indica la distribución no tan uniforme del atributo Close, y hay ligera presencia de asimetría en los datos.

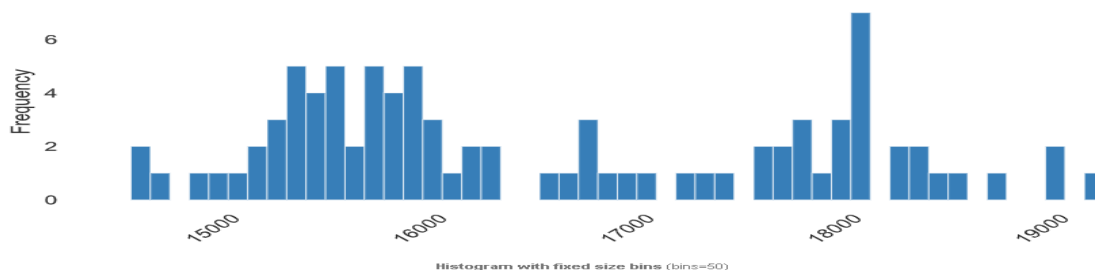


Ilustración 4: Histograma de Frecuencia

Correlación

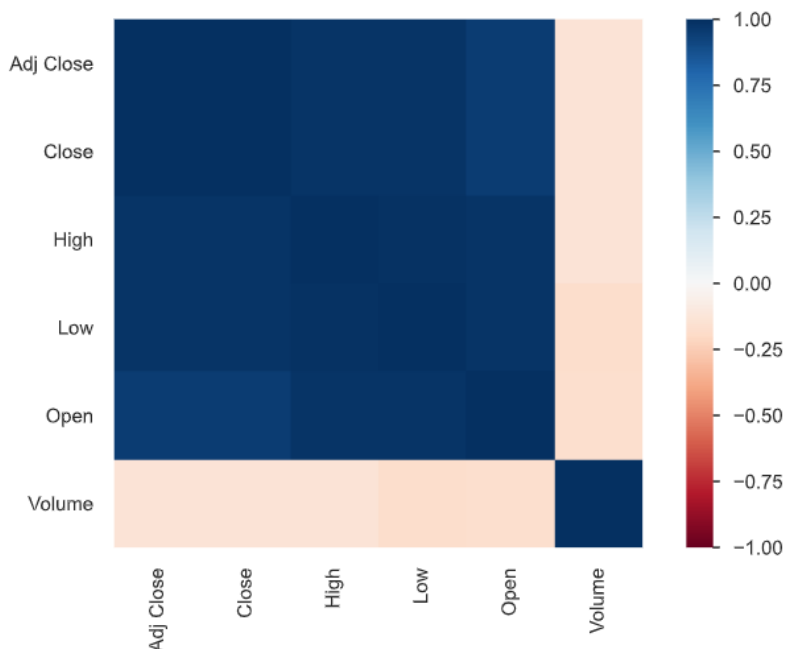


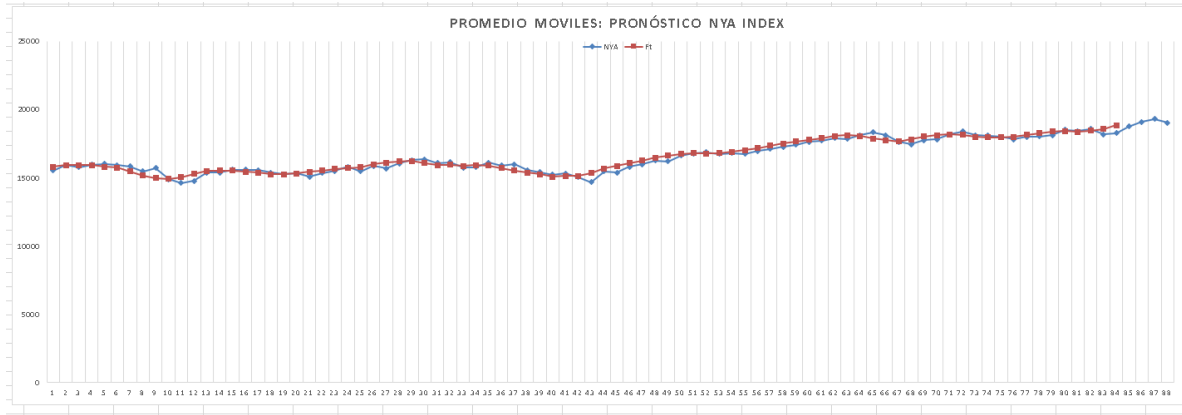
Ilustración 5: Análisis Exploratorio de Datos: Correlación

En la gráfica de Heatmap, se muestra como los diferentes atributos del dataset presentan alta correlación entre sí. Sugiriendo una tendencia entre los atributos de Open, Low, High, Close y AdjClose., en otras palabras, cuando el valor aumenta para una de los atributos, los demás tendrán un comportamiento similar. Y están asociados con el color azul.

Mientras que el atributo de Volumen presenta menos relación con el resto de las variables. Sugiere que el atributo de volumen es prácticamente independiente.

Modelos Estadísticos de Pronósticos Series Temporales

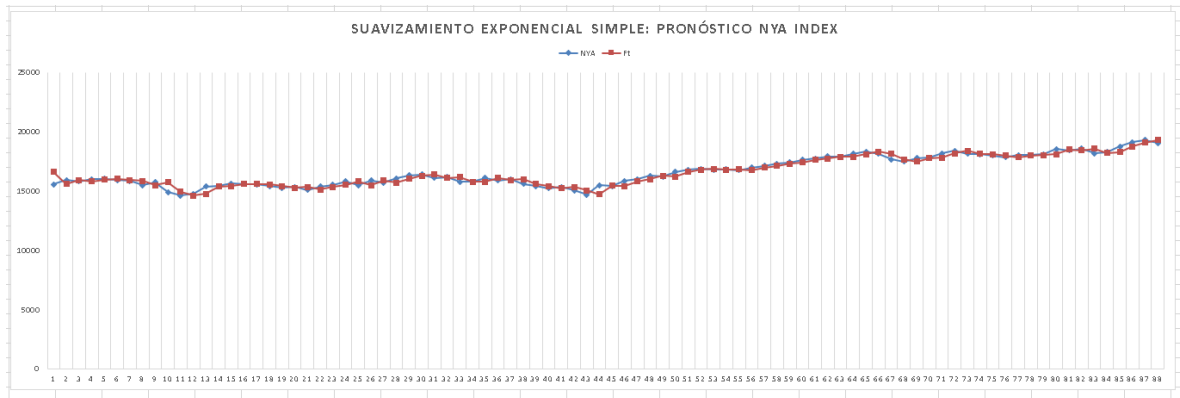
Promedio Móviles



Este modelo de pronósticos da como resultado:

- Mean Absolute Desviation (MAD): 313.35
- Mean Absolute Percentage Error (MAPE): 1.90
- Rango Ts Superior: 3.03
- Rango Ts Inferior: -2.29
- Desviación Estándar: 391.69

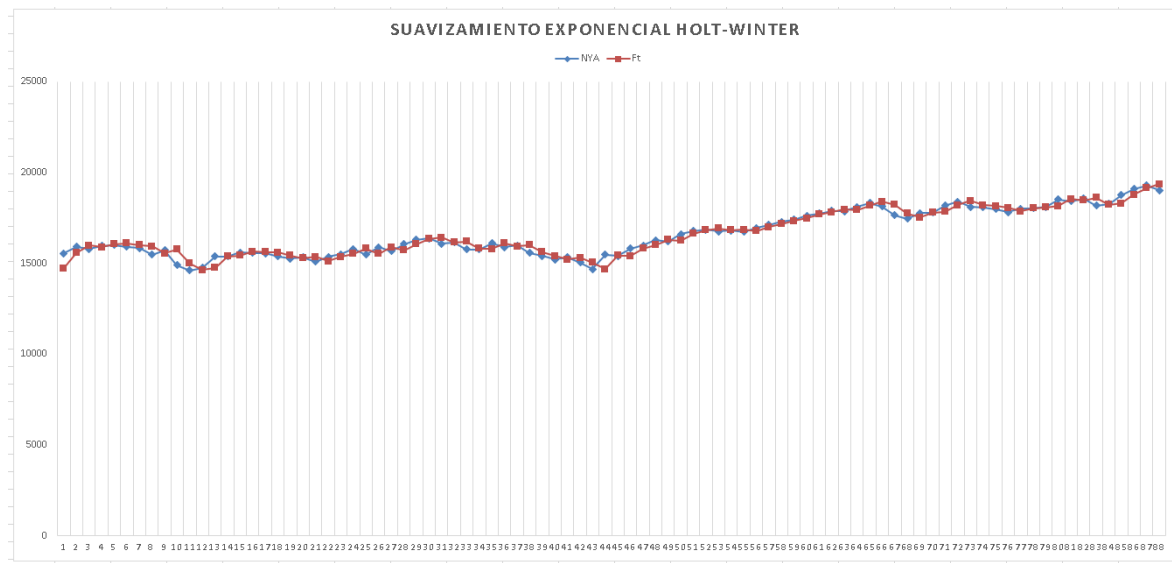
Suavizamiento Exponencial Simple



Este modelo de pronósticos da como resultado:

- Mean Absolute Desviation (MAD): 228.29
- Mean Absolute Percentage Error (MAPE): 1.39
- Rango Ts Superior: 8.13
- Rango Ts Inferior: --12.26
- Desviación Estándar: 285.36

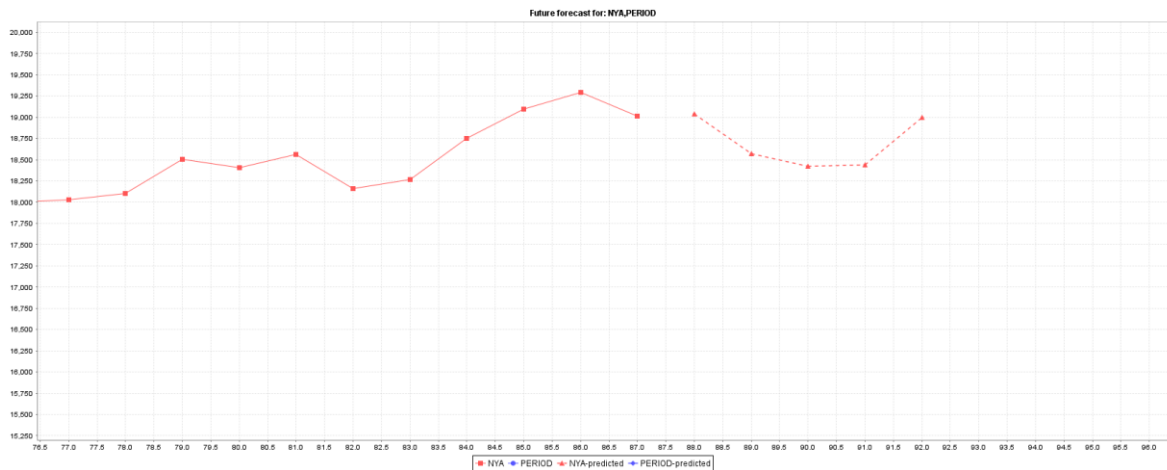
Suavizamiento Exponencial Holt-Winter



Este modelo de pronósticos da como resultado:

- Mean Absolute Deviation (MAD): 223.38
- Mean Absolute Percentage Error (MAPE): 2000.58
- Rango Ts Superior: 6.92
- Rango Ts Inferior: -4.64
- Desviación Estándar: 279.22

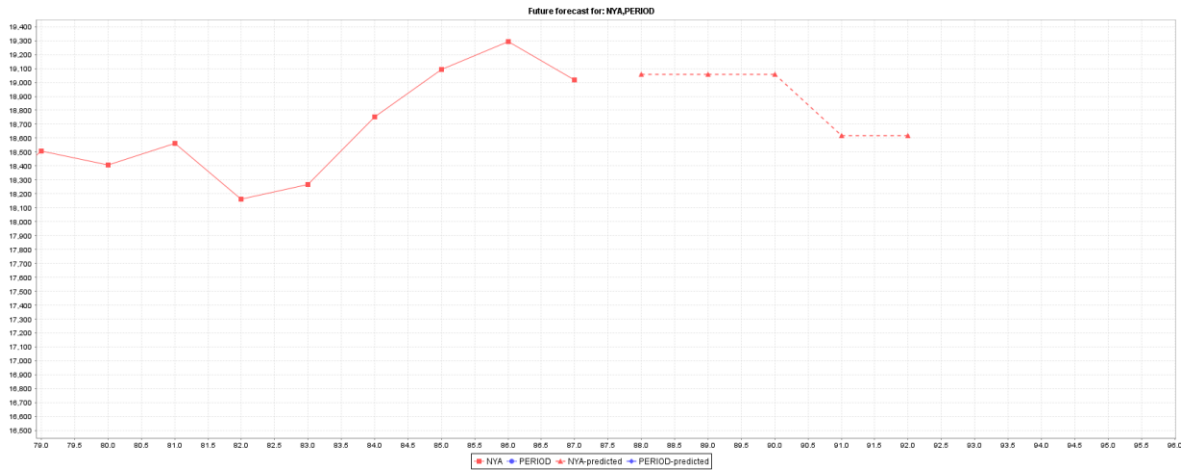
Modelo SMOreg



Este modelo de pronósticos da como resultado:

- Mean Absolute Desviation (MAD): 24.71
- Mean Absolute Percentage Error (MAPE): 13.79
- Rango Ts Superior: -
- Rango Ts Inferior: -
- Desviación Estándar: 30.89

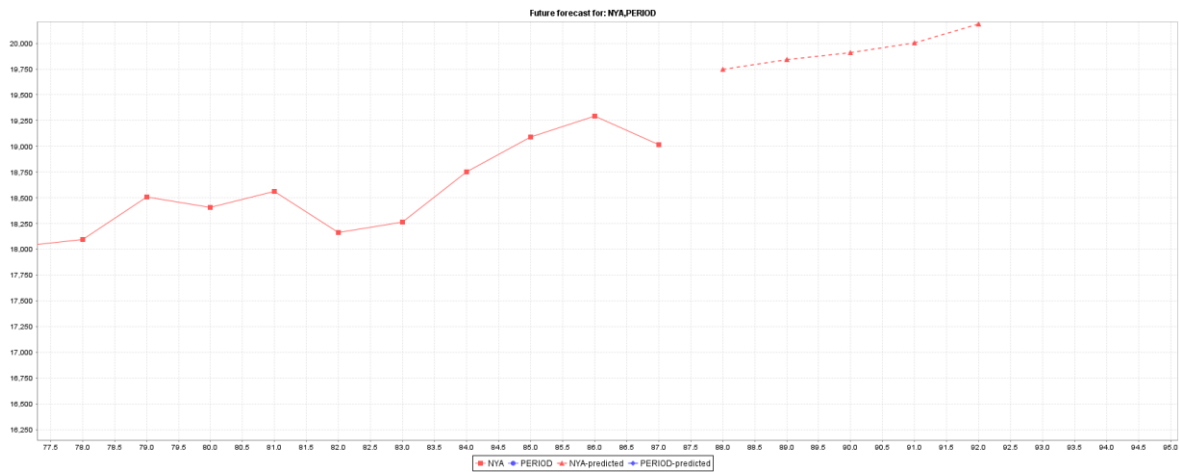
Modelo RandomTree



Este modelo de pronósticos da como resultado:

- Mean Absolute Deviation (MAD): 28.01
- Mean Absolute Percentage Error (MAPE): 15.6
- Rango Ts Superior: -
- Rango Ts Inferior: -
- Desviación Estándar: 35.01

Modelo MultiLayer Preceptron



Este modelo de pronósticos da como resultado:

- Mean Absolute Desviation (MAD): 135.99
- Mean Absolute Percentage Error (MAPE): 76.51
- Rango Ts Superior: -
- Rango Ts Inferior: -
- Desviación Estándar: 169.99

Tabla de Resultados

Método	MAD	MAPE	Rango TS inf	Rango TS sup	Desv. Est.
Promedio Movil	313.35	1.90	-2.29	3.03	391.69
Suavización Expo	228.29	1.39	-12.26	8.13	285.36
Holt	223.38	2000.58	-4.64	6.92	279.22
MultiLayer Preceptron	135.99	76.51	-	-	169.99
SMOreg	24.71	13.79	-	-	30.89
RandomTree	28.01	15.6	-	-	35.01

Promedio Móvil tiene un MAD de 313.35 y una desviación estándar de 391.69, lo que sugiere errores más altos y mayor variabilidad en comparación con otros métodos.

Suavización Exponencial tiene una menor MAD (228.29) y MAPE (1.39) que el Promedio Móvil, pero su rango es bastante amplio (-12.26 a 8.13).

Holt muestra un error MAPE extremadamente alto (1958.01), lo que indica que este método tiene un rendimiento muy deficiente en este caso.

SMOreg y RandomTree tienen los errores más bajos en términos de MAD y MAPE, con una desviación estándar muy baja, lo que sugiere que son los métodos más precisos y consistentes.

Multilayer Perceptron tiene un error MAPE alto (76.51), lo que indica que no es tan eficiente para este caso en particular.

El siguiente enlace es la ubicación del repositorio de Github creado para el proyecto:

https://github.com/francusd/Modelos_Predictivos

Conclusiones

En este análisis estadístico, hemos comparado detalladamente los resultados de diferentes modelos de pronósticos de series temporales, incluyendo el modelo de Promedio Móviles, Suavizamiento Exponencial Simple, Suavizamiento Exponencial Holt-Winter, Modelo SMOreg, Modelo RandomTree, Modelo MultiLayer Preceptron. A través de diversas métricas de evaluación, como la Desviación Media Absoluta (MAD), error cuadrático medio (MSE) y el error porcentaje absoluto medio (MAPE), desviación Estándar, se ha determinado que cada modelo presenta ventajas y desventajas específicas dependiendo del contexto y la naturaleza de los datos.

El modelo ARIMA mostró un rendimiento robusto en series temporales con patrones de tendencias lineales claras. Por otro lado, el modelo de suavizamiento exponencial simple fue más efectivo en datos con fluctuaciones menores.

Los modelos ejecutados con la herramienta WEKA fueron más complejas y con mejor optimización de variable, y demandantes en términos de recursos computacionales, demostraron una capacidad superior para capturar patrones y dependencias en los datos.

En resumen, la elección del modelo de pronóstico más adecuado está basado en las características específicas de la serie temporal, por lo que podemos concluir que los modelos que ofrecen el mejor equilibrio con mayor precisión y consistencia de Pronóstico en el contexto de nuestro análisis sería:

- SMOreg

- RandomTree

Referencias Bibliográficas

- 1) Chopra, S., & Meindl, P. (2013). *Administración de la cadena de suministro: Estrategia, planeación y operación* (5th ed.). Pearson Educación de México.
- 2) Wikipedia contributors. (n.d.). *Series de tiempo para pronósticos*. In Wikipedia, The Free Encyclopedia. Retrieved September 8, 2024, from https://es.wikipedia.org/wiki/Series_de_tiempo_para_pron%C3%B3sticos
- 3) Investing.com. (n.d.). *NYSE Composite Index (NYA)*. Retrieved September 8, 2024, from <https://www.investing.com/indices/nyse-composite>
- 4) Yahoo Finance. (n.d.). *NYSE Composite (^NYA) Interactive Stock Chart*. Retrieved September 8, 2024, from <https://finance.yahoo.com/quote/%5ENYA/chart/>
- 5) Investopedia. (n.d.). *NYSE Composite Index*. Retrieved September 8, 2024, from <https://www.investopedia.com/terms/n/nysecompositeindex.asp>
- 6) Python Software Foundation. (n.d.). Welcome to Python.org. Retrieved September 8, 2024, from <https://www.python.org/>
- 7) Pandas development team. (n.d.). pandas: Python Data Analysis Library. Retrieved September 8, 2024, from <https://pandas.pydata.org/>
- 8) Matplotlib development team. (n.d.). API Reference — Matplotlib 3.9.2 documentation. Retrieved September 8, 2024, from <https://matplotlib.org/stable/api/index.html>
- 9) Stojiljković, M. (n.d.). Linear Regression in Python. Real Python. Retrieved September 8, 2024, from <https://realpython.com/linear-regression-in-python/>
- 10) Aprende Machine Learning. (2019, February 26). Pronóstico de series temporales con redes neuronales en Python. Retrieved September 8, 2024, from <https://www.aprendemachinelearning.com/pronostico-de-series-temporales-con-redes-neuronales-en-python/>
- 11) DataCamp. (2017, March). Python Exploratory Data Analysis Tutorial. Retrieved September 8, 2024, from <https://www.datacamp.com/tutorial/exploratory-data-analysis-python>