

“Modelo de precio de las acciones del índice Dow Jones por medio de series de tiempo en modelo ARIMA y aplicación de modelo red neuronal.”

“Price model for the shares of the Dow Jones index through time series in the ARIMA model and application of the red neural model.”

Oscar Cortes Chala

ocortesc@libertadores.edu.co

Carlos Eduardo Wilches Moreno

[cewilchesm@libertadores.edu.co](mailto:cwilchesm@libertadores.edu.co)

Trabajo de grado para optar al título de especialista en Estadística Aplicada.

RESUMEN:

En el presente trabajo se plantea un objetivo de análisis de bajo un modelo estadístico que permita estimar los precios de las acciones del índice Dow Jones (DJI); índice que pertenece a bolsa de New York compuesto por 30 de las acciones más significativas de todas las industrias, salvo transporte y servicios públicos que cotizan en la bolsa. Se realiza la estimación y el correspondiente análisis estadístico basándonos en información histórica desde el periodo 1 de enero del 2019 al 1 de enero del 2022; utilizando inferencia estadística desde el punto de vista de series de tiempo con herramientas tales como ARIMA y aplicación del modelo red neuronal para posteriormente comparar dichas propuestas y así seleccionar el modelo más acertado y su análisis descriptivo de la variable objetiva DJI.

Palabras clave: trading, volatilidad, estimación, portafolio de inversión, estacionariedad en activos bursátiles, DJI.

ABSTRACT

In the present work, an analysis objective is proposed under a statistical model that allows estimating the prices of the shares of the Dow Jones index (DJI); New York Stock Exchange-owned index comprised of 30 of the most significant stocks in all industries except publicly traded transportation and utilities. The estimate and the corresponding statistical analysis are made based on historical information from the period January 1, 2019 to January 1, 2022; using statistical inference from the point of view of time series with tools such as ARIMA and application of the neural network model to subsequently compare these proposals and thus select the most accurate model and its descriptive analysis of the DJI target variable.

Keywords: trading, volatility, estimation, investment portfolio, stationarity in stock assets, DJI.

INTRODUCCION

La economía mundial es sin duda alguna un tema fundamental para los economistas, sobre todo para aquellos que se dedican en el área financiera, pues los acontecimientos resultan ser variables importantes para el comportamiento de un activo financiero, por ejemplo el descubrimiento de una mina de plata en el África puede ser una oportunidad o una amenaza en Wall Street; las declaraciones de un mandatario asiático pueden significar toda una fluctuación y sensibilidad en el mercado occidental o el incremento de la violencia en un país sudamericano se puede traducir en una fuga de capitales e inversionistas y en un desplome

de la bolsa nacional. Más que un tema de interés, la economía mundial significa un conjunto de economías entrelazadas entre sí y en algunos casos co-dependientes entre ellas. La economía mundial es un diagnóstico de cómo se están comportando las economías de los diversos países y sirve como guía para comprender hechos históricos como shocks, recesiones, “burbujas” y depresiones.

Una vez entendida la importancia de la economía mundial es importante cuestionarse la influencia de esta sobre nuestros portafolios y toma de decisiones.

Las bolsas de valores facilitan financiación a las empresas para que puedan hacer sus proyectos de inversión y por otra parte a los inversionistas dan su dinero para que haya un retorno de inversión generando rentabilidad. Una de las ventajas de dichas instituciones es que dan liquidez a los instrumentos financieros y con respecto al presente estudio nos enfocaremos en el índice Dow Jones (DJI) en la bolsa de New York.

La bolsa de New York es el más grande mercado bursátil del planeta con respecto a volumen transado de valores y muchas empresas agregadas a dicha entidad. Sus índices bursátiles son S&P500, Nasdaq, Dow Jones; con respecto a este último nos vamos a basar en nuestro estudio.

PLANTEAMIENTO DEL PROBLEMA:

En el presente trabajo se pretende establecer un modelo de predicción fundamentados en las series de tiempo y modelación ARIMA y aplicación de red neuronal para poder estimar el comportamiento del índice Dow Jones (DJI) en la bolsa de New York; Se entiende que en el mercado de capitales todo es volátil y más en los mercados de renta variable, por el principal interés de estimar un modelo optimo e idóneo para hacer los respectivos pronósticos.

OBJETIVO GENERAL:

- Pronosticar el índice Dow Jones de la bolsa de New York usando las series de tiempo y aplicación de modelo red neuronal.

OBJETIVOS ESPECIFICOS:

1. Usar el modelo ARIMA para definir las series temporales y estudiar las observaciones continuas para encontrar patrones y hacer proyecciones de los precios de las acciones.
2. Aplicar la modelación red neuronal para estudiar los valores de salida esperados en función de los datos entrantes para establecer predicción de sucesos y simulaciones.
3. Comparar los modelos y análisis descriptivo de la variable DJI.

JUSTIFICACION:

Es importante realizar este estudio ya que se considera que los rendimientos de las compañías muchas veces no solo dependen del desempeño de esta sino de factores exógenos que pueden marcar shocks o perturbaciones en el comportamiento de los precios de sus acciones, si nos referimos a compañías que coticen en bolsa.

Una de las causas principales de la presente investigación es el interés que despierta el conocimiento sobre el mercado bursátil y más en el diario vivir de sus operaciones en términos de oferta y demanda y su estimación en función de los precios de las acciones; esto nos permite en materia de toma decisiones en función de inversión utilizar herramientas estadísticas que nos ayudan a predecir y más cuando conocemos que dicho mercado es muy volátil y por este motivo resulta interesante aplicar las herramientas estadísticas y bajo la metodología aquí usada nos introduce al campo teórico práctico y se espera que con este trabajo de estudio se apruebe aplicar los conocimientos adquiridos.

MARCO TEORICO

Para una mayor comprensión del tema y una visión más amplia de las variables de uso en esta investigación, además de procesos y pruebas claves en el desarrollo del modelo y el análisis de series de tiempo proporcionaremos al lector las definiciones necesarias.

Índice Dow Jones:

El índice bursátil industrial Dow Jones muestra el progreso de 30 empresas industriales con una gran capitalización bursátil que cotizan en el mercado de valores de la bolsa de New York.

Su utilidad es de valiosa contribución a los inversionistas ya que permite dar información sobre las inversiones de las empresas inscritas en dicho índice. Las acciones que componen el Dow Jones son las siguientes: sector tecnológico, empresas petroleras, entidades financieras, acciones en telecomunicaciones, etc.

El cálculo del índice bursátil industrial Dow Jones se hace de la siguiente forma: se debe conocer antes que todos los precios de las 30 acciones que están en el índice y el divisor de Dow, se suman los precios de todas las empresas y se divide entre el divisor de Dow y el resultado es el índice Dow Jones.

Series de tiempo:

Otro concepto fundamental a tener en cuenta es el de series de tiempo el profesor José Alberto Mauricio la define como “Una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes modelos”. (Mauricio, 2007, pág. 1) de aquí nacen otros dos términos y condiciones que consideramos importantes, series de tiempo estacionarias y series de tiempo no estacionarias,

la primera nos indica que “un proceso estocástico es estacionario cuando las propiedades estadísticas de cualquier secuencia finita $y_{t1}, y_{t2}, \dots, y_{tn}$ ($n \geq 1$) de componente (y_t) son semejantes a las de la secuencia $y_{t1+h}, y_{t2+h}, \dots, y_{t+h}$ para cualquier número entero $h = \pm 1, \pm 2, \dots$ ” (Mauricio, 2007), es decir, en palabras más técnicas “Un proceso de series de tiempo estacionario es aquel en el que sus distribuciones de probabilidad se mantienen estables con el paso del tiempo en el siguiente sentido: si se toma cualquier colección de variables aleatorias de la secuencia y se las desplaza h periodos, la distribución de probabilidad conjunta debe permanecer inalterada.” (Wooldridge, 2008, pág. 378) Por otro lado las series de tiempo no estacionarias hacen referencia a “Un proceso estocástico (y_t) no estacionario cuando las propiedades estadísticas de al menos una secuencia finita $y_{t1}, y_{t2}, \dots, y_{tn}$ ($n \geq 1$) de componente (y_t) , son diferentes de la secuencia $y_{t1+h}, y_{t2+h}, \dots, y_{t+h}$ para al menos un número entero $h > 0$ ”. (Mauricio, 2007)

De acuerdo a lo expresado, se puede decir que “una condición suficiente para que un proceso sea no estacionario es que la esperanza condicional de algunos de sus componentes sea distinta de la de los otros”. (Mauricio, 2007)

Modelo ARIMA:

El modelo autorregresivo integrado de media móvil expresa la metodología Box & Jenkins para fabricar series de tiempo, cuyo proceso esté compuesto por cuatro pasos: identificación, estimación, comprobación y pronóstico. En el primer paso de identificación se usan datos históricos para proponer inicialmente un modelo Box & Jenkins. En el segundo paso de estimación se utilizan la misma información histórica para estimar los criterios de modelo propuesto inicialmente. En el tercer paso de comprobación se estructuran muchos parámetros

para verificar si el modelo es apropiado o si es imperativo hacer un modelo mucho mejor. El cuarto paso de pronóstico se debe emplear para hacer proyecciones de las series de tiempo.

En términos generales los modelos de series de tiempo de Box & Jenkins son usados para hacer pronósticos en series de temporales tipo estacionarias. Una serie de tiempo es estacionaria cuando su media y su varianza son constantes. Si la serie no estacionaria se debe transformar la información hasta tener una serie temporal estacionaria, una técnica para hacer dicha transformación es hacer la diferenciación y se puede hacer varias veces hasta que haya estacionariedad esto quiere decir construir series en la que cada dato sea la diferencia de dos datos seguidos de la serie inicial para anular la influencia de elementos que no tienen nada que ver con los ciclos o la tendencia.

Este modelo se aplica en series de tiempo estacionarias que en termino resumidos es que las medias y las varianzas son invariantes a lo largo del intervalo propuesto de tiempo. El modelo ARIMA se define de la siguiente forma:

$$Y_t = -(\Delta^d Y_t - Y_t) + \phi_0 + \sum_{i=1}^p \phi_i \Delta^d Y_{t-1} - \sum_{i=1}^q \theta_i \epsilon_{t-1} + \epsilon_t$$

En donde:

d = diferencias que son imperativas para transformar la serie original en estacionaria en los casos que se presente, muy seguramente se tendrá que hacer este proceso porque los mercados bursátiles en algunos momentos son volátiles.

$\phi_1 \dots \dots \dots, \phi_p$ = pertenece al componente autorregresivo del modelo.

$\theta_1 \dots \dots \dots, \theta_q$ = dichas observaciones son de las medias móviles del modelo.

ϕ_0 = es constante.

ε_t = margen de error.

Red Neuronal:

Las redes de neuronas artificiales permiten establecer a partir de una metodología de aprendizaje la supervisión de modelos con número elevados de parámetros que configuran a determinar e identificar problemas con estructuras muy complejas y que generalmente son subyacentes de dichos parámetros.

Una red neuronal básica va a realizar dos transformaciones de los datos, y por tanto es un modelo con tres capas: una capa de entrada (input layer) consistente en las variables originales $X=(X_1, X_2, \dots, X_p)$, otra capa oculta (hidden layer) con M nodos, y la capa de salida (output layer) con la predicción (o predicciones) final $m(X)$.

METODOLOGIA

Tipo de estudio:

En el presente trabajo, se utiliza la metodología de modelo ARIMA para establecer la correlación y pronóstico del comportamiento de la serie de datos establecida para el índice bursátil Dow Jones (DJI) y a través de red neuronal determinar una segunda ejecución de pronóstico para lograr buscar bajo un estadístico el patrón más indicativo del comportamiento de la serie.

Área de estudio y recolección de datos:

El estudio involucra distintas ramas del área económica y financiera, en la generación de la toma de decisiones los inversionistas tienen distintas herramientas que les permiten analizar el comportamiento del mercado bursátil, en la necesidad de realizar modelos que permitan pronosticar nuestra área de estudio involucra el comportamiento de un índice en donde se categorizan bajo una ponderación a la capitalización bursátil y el cual corresponde a 30 empresas que cotizan en la bolsa Dow Jones y que permiten establecer el comportamiento que depende de la oferta y demanda (fuerza de mercado). El índice con el cual se realiza el estudio es el índice Dow Jones (DJI) y bajo el cual se descarga la serie de datos del comportamiento de índice desde el periodo 1 de enero del 2019 al 1 de enero del 2022, se utiliza inferencia estadística de la serie y se realiza la descarga de los datos del rango de tiempo establecido y que se realiza el estudio bajo la metodología de series de tiempo univariada, la recolección de los datos se establece a partir de la plataforma Yahoo! Finance que permiten descargar el comportamiento bajo el rango de tiempo establecido para el área de estudio.

Universo y muestra:

El universo corresponde al comportamiento diario del índice Dow Jones (DJI), el cual depende diariamente de volatilidades que se determinan de acuerdo con la fuerza de mercado que conlleve a la toma de decisiones por parte de los inversionistas y por ende el factor explícito del índice y su comportamiento diario. Se establece que la serie descargada conlleva un análisis estadístico que permita visualizar el comportamiento de la varianza que depende de la volatilidad o los cambios diarios asociados a la serie.

Método:

El método utilizado para análisis estadístico y pronóstico del índice se utiliza la serie de tiempo univariada que permita realizar el proceso para determinar el comportamiento de la serie bajo el supuesto determinístico que permita bajo el pronóstico realizar el proceso de análisis con los siguientes pasos:

- 1) Identificación del molde de acuerdo con la naturaleza de los datos.
- 2) Estimación del parámetro implícito del modelo.
- 3) Verificación de los supuestos y que los residuos no estén autocorrelacionados.
- 4) Uso del modelo para los fines que motivaron la construcción que generalmente son los de pronóstico de la serie.

RESULTADOS

El índice bursátil Dow Jones (DJI) muestra el progreso de 30 empresas industriales con una gran capitalización bursátil que cotizan en el mercado de valores de la bolsa de New York, los datos que se analizan son el periodo comprendido entre 01 de enero de 2019 al 01 de enero de 2022.

Para hacer el respectivo estudio se utiliza Rstudio y como primer paso se convierte la serie de tiempo con el comando TS y posteriormente se grafica como se evidencia en la imagen:

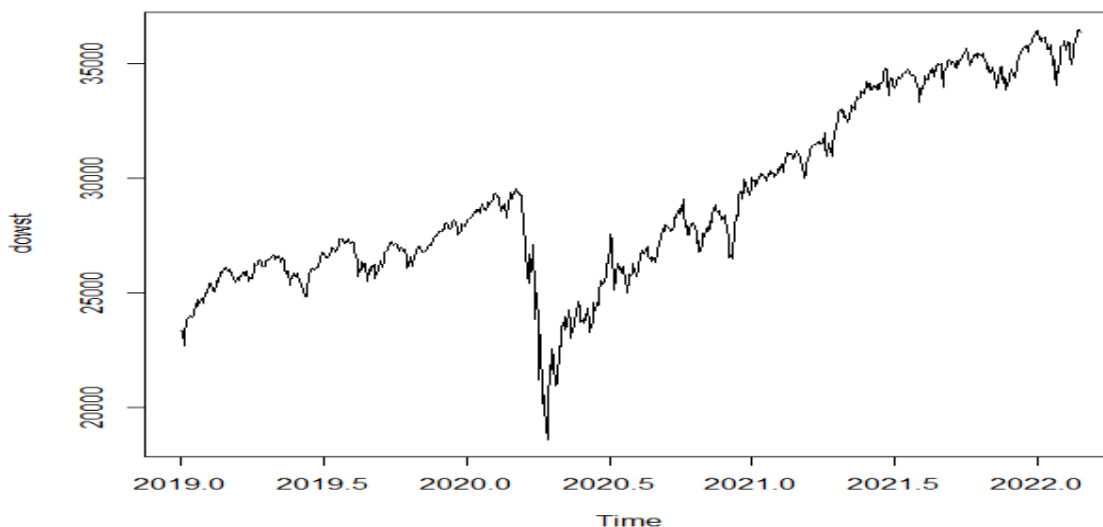
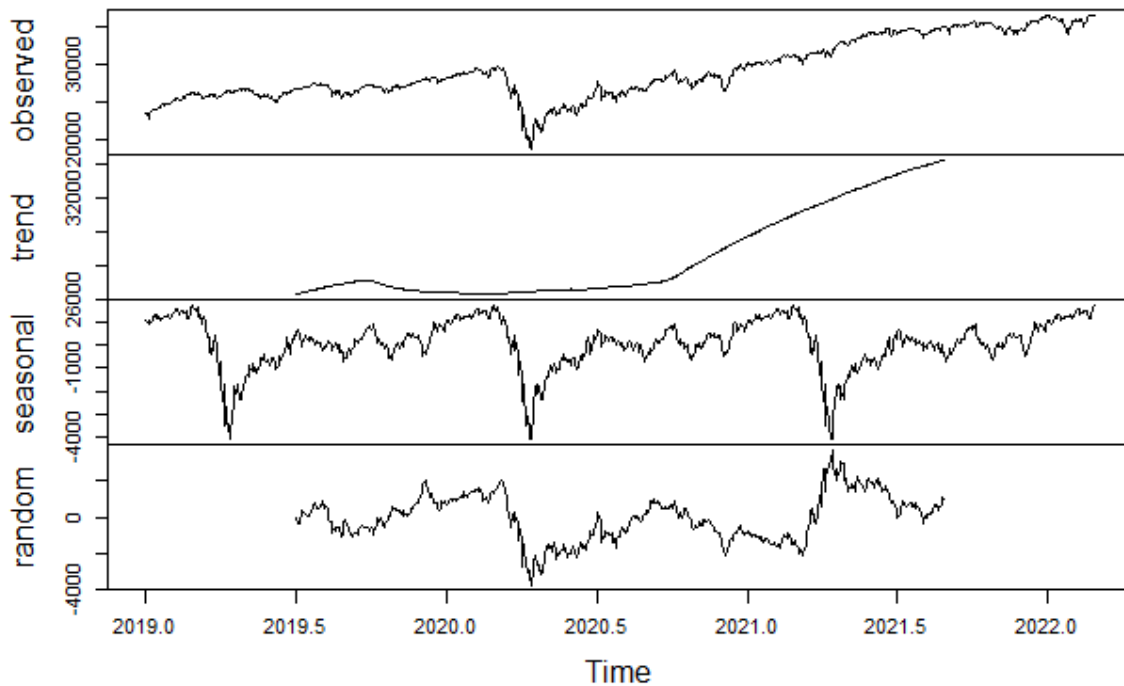



Grafico 1: (DJI) serie de tiempo entre los tiempos comprendidos de estudio.

En esta primera grafica se puede observar que en el periodo entre 2019 al 2022 los precios en el año 2019 tuvieron una tendencia alcista con pequeños descensos para luego tener una caída vertiginosa en el primer semestre del año 2020 por la emergencia sanitaria COVID-19, en la segunda parte de ese año empezó aumentar los precios de las acciones siguiendo esta evolución en el año 2021. En las series de tiempo se deben tener en cuenta tres componentes, a saber: nivel de tendencia, nivel de estacionariedad y nivel aleatorio para hacer proyecciones, para evidenciar estos tres niveles se procede a graficar y tener un análisis previo de las series:

Decomposition of additive time series



 Figura 2: (DJI) descomposición de la serie.

En la gráfica de descomposición se observan cuatro graficas de las cuales se clasifican de la siguiente manera:

1. Grafica de la serie de tiempo del índice DOW JONES desde el 01 de enero de 2019 al 01 de enero de 2022 que anteriormente se hizo una descripción preliminar.
2. Grafica de tendencia de la serie de tiempo en la cual nos soporta la conclusión que se dio de que empezando el periodo en estudio hay cifras ascendentes, posterior al año 2020 esas cantidades descienden producto de la pandemia y los precios alcistas a partir del segundo semestre del año 2020 siguiendo esa evolución en el año 2021.
3. Grafica estacional de la serie de tiempo está determinando que puede no es cíclico y visualmente tanto la media como la varianza no son constantes, pero se podría trabajar con estos datos para hacer proyecciones haciendo uso de la diferenciación para volver la serie en estacionaria.
4. Grafica del nivel aleatorio de la serie de tiempo, con respecto a este componente es un valor residual que se traduce en las desviaciones de los factores reales de la serie de tiempo de los valores esperados en los efectos de la serie de tendencia y estacional. Este nivel aleatorio lo causan momentos de corto plazo, son imprevistos y no ocurren a menudo y que pueden influenciar a las series de tiempo por este motivo no son predecibles y no se puede pronosticar datos.

Test de Dickey Fuller:

Haciendo la prueba DF proyectada anualmente se evidencia que hay un P VALOR de 0.43 que es mayor a 0.05 y esto quiere decir que la serie con los datos originales no es estacionaria, para volver la series de tiempo del índice Dow jones estacionaria se debe hacer la diferenciación, haciendo este proceso se evidencia que solo se debe hacer una vez y haciendo el estadístico de prueba nos corroboran estos resultados como se observa a continuación:

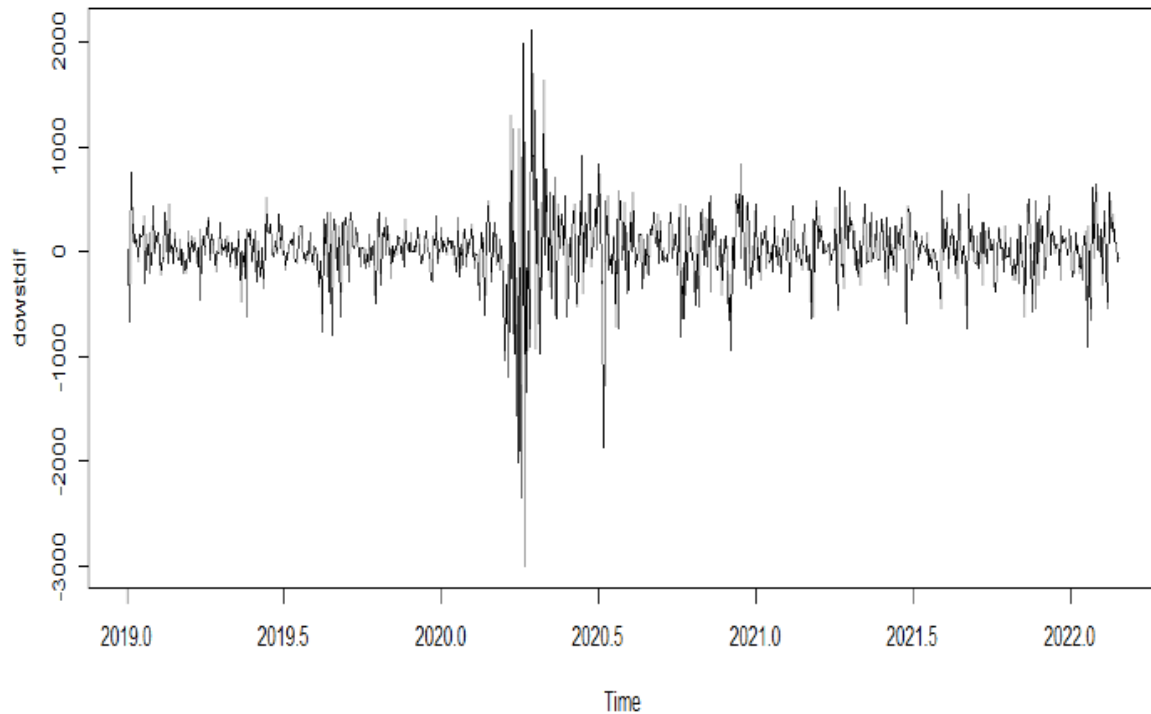


Gráfico 3: (DJI) Test Dickey Fuller

Al hacer el test de Dickey Fuller de la primera diferencia nos da un P VALOR de 0.01 esto quiere decir que nos da debajo de 0.05 y por ende la serie de tiempo del índice Dow Jones ya es estacionaria y podemos seguir trabajando en nuestros modelos.

Graficas de autocorrelación y autocorrelación parcial:

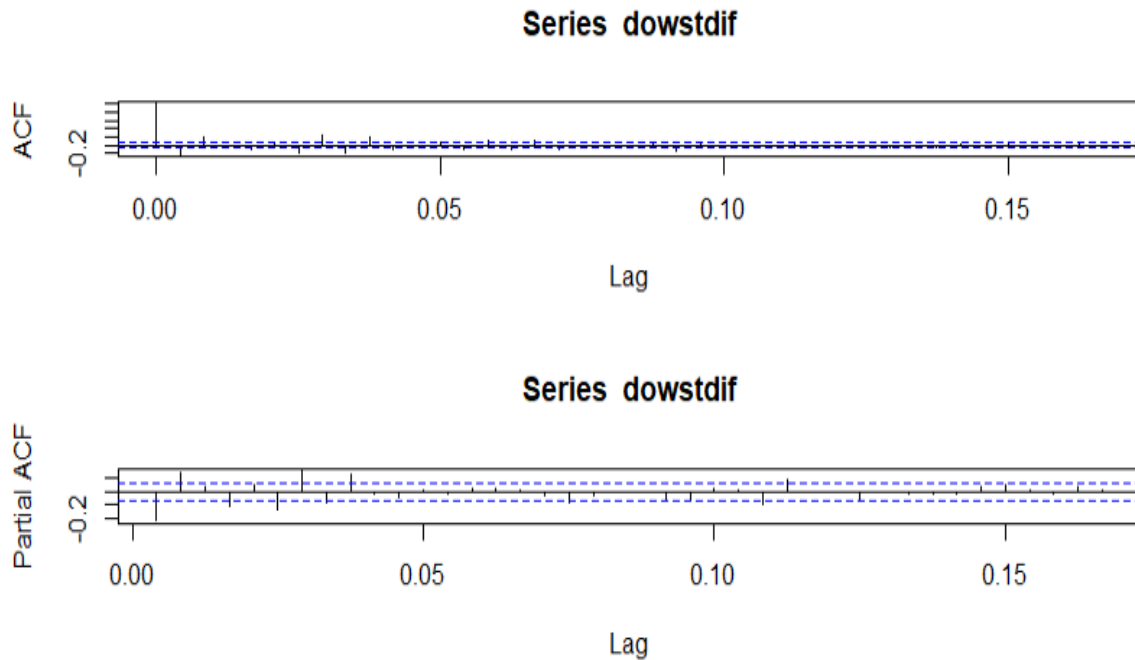


Gráfico 4: (DJI) autocorrelación y autocorrelación parcial.

- Grafica ACF: se observa que hay varios lags autorregresivos que se sobresalen del horizonte, hasta 9 AR.
- Grafica PACF: se evidencia que hay las de medias móviles que salen del intervalo como ejemplo 10 MA.

Modelo ARIMA:

De acuerdo a las gráficas de ACF y PACF identificamos dos modelos Arima (2, 1,0) y (2, 1,1) y se obtuvieron los siguientes resultados:

- Arima (2, 1,0): en dicho modelo arrojé unos coeficientes en los cuales al hacerles la significancia los valores obtenidos fueron $2.369056e-07$ y $8.495605e-06$ de los autorregresivos por ende son datos menores a 0.05 es decir no hay que hacer ajuste a esta propuesta.
- Arima (2, 1,1): en este modelo a los coeficientes se les hizo la respectiva significancia y dio como resultado $3.103162e-01$, $5.736746e-05$ y $2.373974e-01$, esto quiere decir que tanto el autorregresivo 1 y la media móvil 1 son mayores que 0.05 y solo el autorregresivo 2 es significativo ya que es menor que 0.05 y con estos datos se debe ajustar la propuesta.

Residuales:

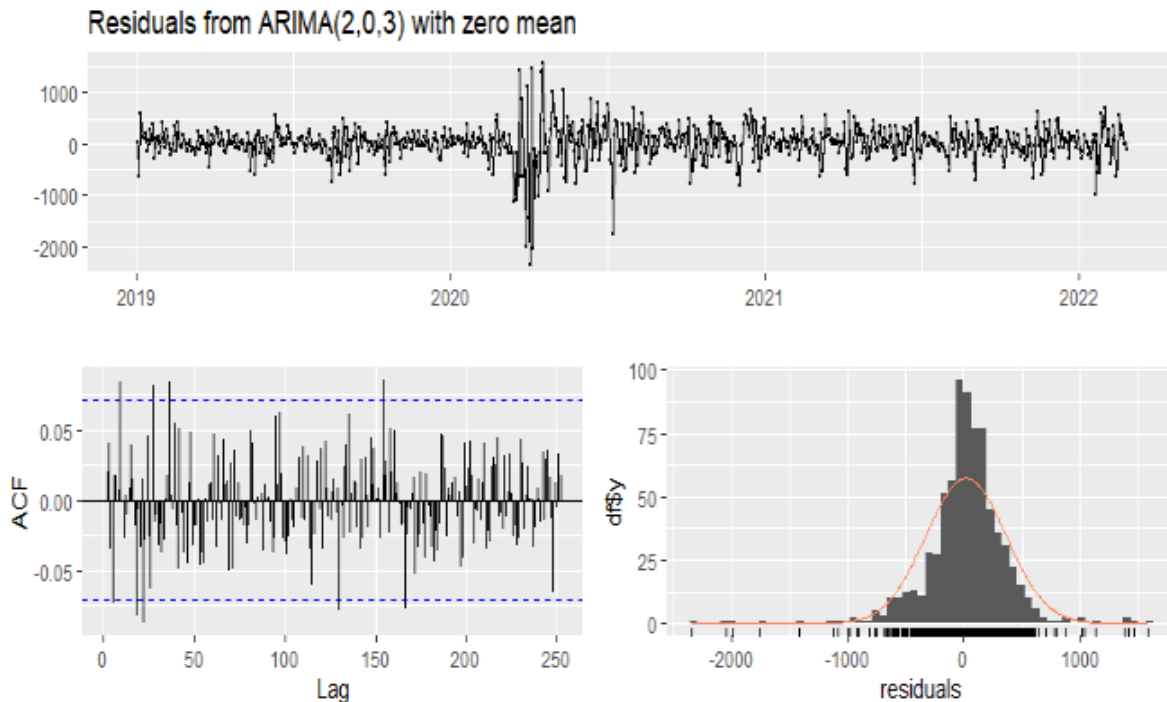


Gráfico 5: (DJI) Residuales Modelo ARIMA.

- Se observa los valores residuales están dentro del intervalo de la gráfica ACF.
- En la gráfica de los residuales los valores tienden a cero y tiene una distribución normal.

Test de Ljung Box:

La prueba de Ljung Box en ambos modelos se calculó con respecto a los residuales de cada uno y se debe definir no rechazar la hipótesis nula de que si hay ruido blanco y se evidencio lo siguiente:

- Modelo 1 Arima (2, 1,0): el p valor 0.8034 es una cifra mayor que el 0.05 es decir no se rechaza la hipótesis nula de que hay ruido blanco.
- Modelo 2 Arima (2, 1,1): el p valor 3.49e-07 es una cifra menor que el 0.05 es decir se rechaza la hipótesis nula de que hay ruido blanco.

Test de Jarque Bera:

La prueba Jarque Bera en ambos modelos se calcula con respecto a los residuales de cada una de ellas y se debe definir si no se rechaza la hipótesis que la distribución es normal y se concluyo lo siguiente:

- Modelo 1 Arima (2, 1,0): el p valor $2.2e-16$ es una cifra menor que el 0.05 es decir se rechaza la hipótesis nula de que la distribución es normal.
- Modelo 2 Arima (2, 1,1): el p valor $2.2e-16$ es una cifra menor que el 0.05 es decir se rechaza la hipótesis nula de que la distribución es normal.

Pronostico del modelo y gráfica:

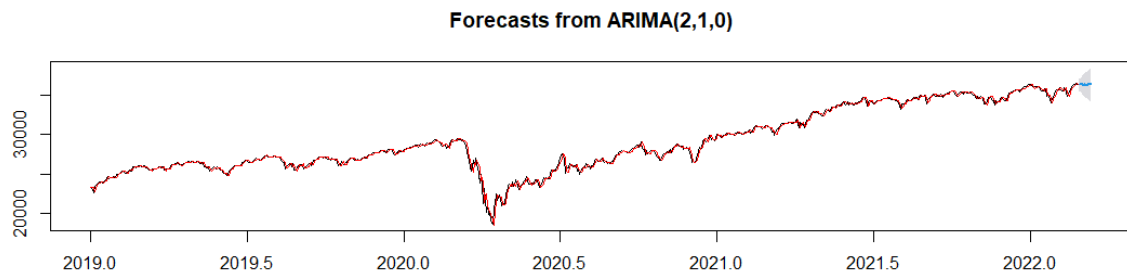


Gráfico 6: (DJI) pronostico modelo ARIMA modelo 1.

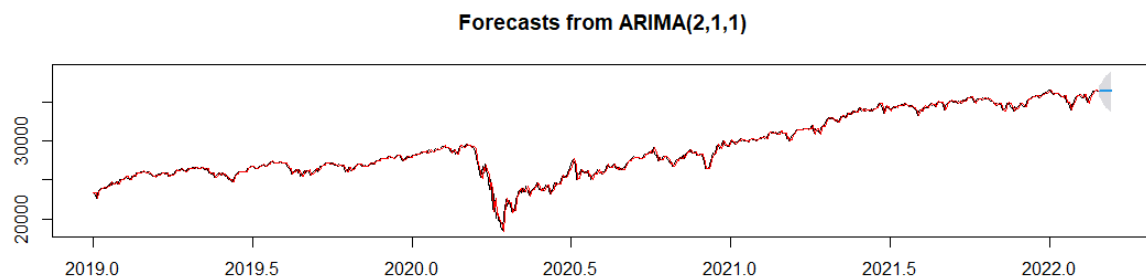


Gráfico 7: (DJI) pronostico modelo ARIMA modelo 2.

- Se observa que hay una clara tendencia del pronóstico a la baja y se hizo la proyección a 10 datos.

- Se observa en la comparación que no es mucha la diferencia con los valores reales y los valores esperados según las dos graficas esto quiere decir que nuestro modelo ARIMA se acerca mucho a la realidad y tendría un impacto tanto en la oferta como en la demanda de acciones en el índice Dow Jones.

Modelo red neuronal de retroalimentación:

Este modelo incluye redes de una sola capa y valores retrasados para pronosticar series de tiempo de una sola variable, en este caso el índice Dow Jones y al convertir las series de tiempo en una red nos arrojó los siguientes resultados: Model: NNAR (10, 1,6) [240] con un sigma al cuadrado de 80625.

Grafica de los residuales:

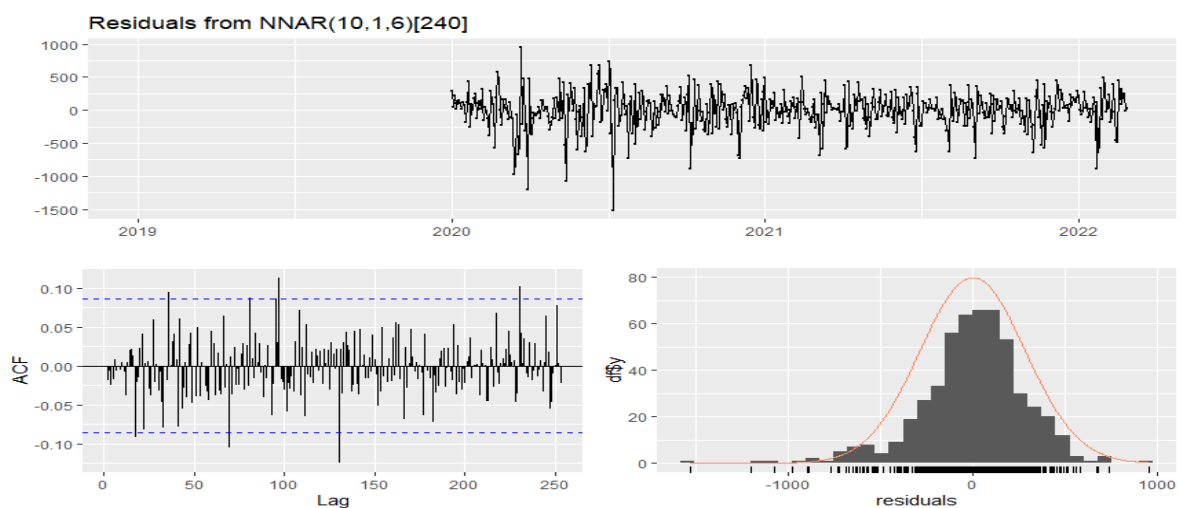


Gráfico 7: (DJI) residuales modelo red neuronal.

- Se observa que la mayor de los valores residuales está dentro del intervalo de la gráfica ACF.
- En la gráfica de los residuales los valores tienden a cero y tiene una distribución normal.

Pronostico del modelo y gráfica:

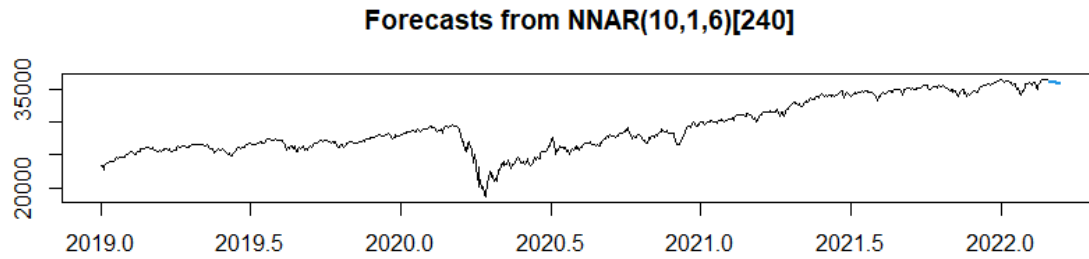


Gráfico 8: (DJI) pronostico del modelo neuronal.

- Se observa que hay una clara tendencia del pronóstico a la baja y se hizo la proyección a 10 datos.
- Se observa que no es mucha la diferencia con los valores reales y los valores esperados esto quiere decir que nuestro modelo red neuronal de retroalimentación se acerca mucho a la realidad y tendría un impacto tanto en la oferta como en la demanda de acciones en el índice Dow Jones.

Gráficas de comparación y cálculo de RMSE y MAPE:

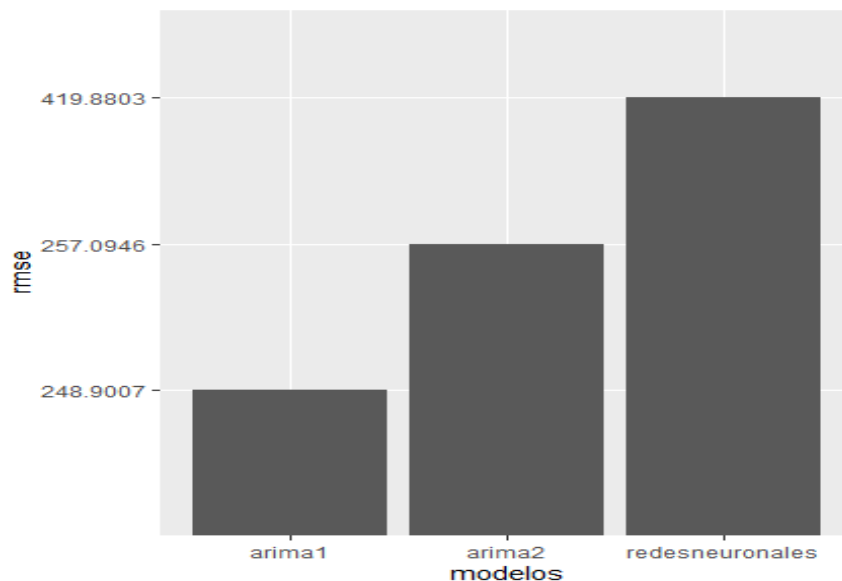


Gráfico 9: (DJI) comparativo modelos y RMSE.

De acuerdo a la gráfica anterior del comparativo entre los modelos y el valor calculado del RMSE se evidencia que el menor dato es del modelo ARIMA número 1 lo cual nos indica que tiene un margen de error bajo y nos ayuda a una buena proyección en los precios de las acciones en el índice Dow Jones.

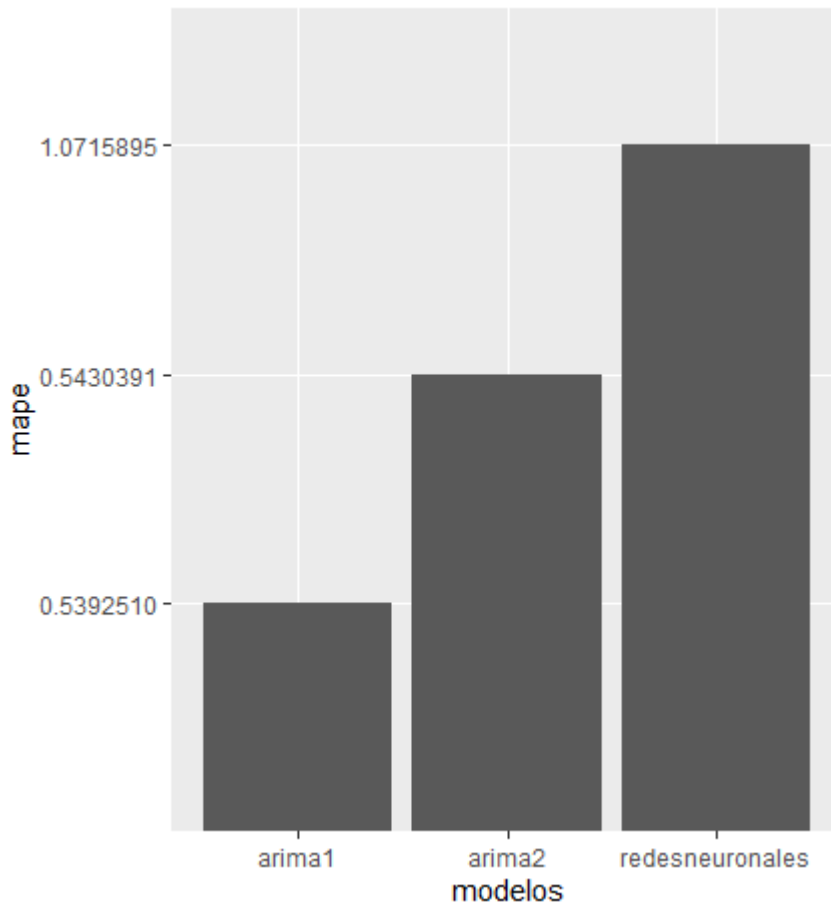


Gráfico 10: (DJI) comparativo modelos y MAPE.

De acuerdo a la gráfica anterior del comparativo entre los modelos y el valor calculado del MAPE se evidencia que el menor dato es del modelo ARIMA número 1 lo cual nos indica que tiene un margen de error bajo y nos ayuda a una buena proyección en los precios de las acciones en el índice Dow Jones.

Conclusiones:

De acuerdo a lo estudiado y analizado en el presente documento sobre el pronóstico de los precios de las acciones en el índice Dow Jones se evidencia que los modelos de series de tiempo pueden ayudar hacer una buena proyección ya que permite mediante varias propuestas escoger la más cercana y poder tomar decisiones de oferta y demanda en el mercado bursátil. Al escoger una sola variable pudimos ejecutar series temporales univaridas

en este caso ARIMA y redes neuronales, la metodología Box Jenkins definitivamente es la más idónea y completa que aporta un buen proceso a la hora de estudiar las series de tiempo y de hecho fue la más acertada comparando los valores reales con los valores esperados y con un margen de error muy bajo con un intervalo de confianza al 95%. Se espera en el presente trabajo pueda servir y se puede aplicar que es lo que se busca con el presente trabajo.

Bibliografía:

- Canal Lic. Lourdes Cuellar. Arima en Rstudio [Archivo de Vídeo]. YouTube. https://www.youtube.com/watch?v=PBcdIK5oKtQ&list=PL2VzkGOpbFyVhZIXOZJ_okDWGDdhqACpg&index=2
- Calviño, J. (26 de agosto de 2020). Pronostico del IPC: arima, suavizado exponencial, redes neuronales. Rpubs. <https://rpubs.com/Jacalti/PronosticosIPC>
- Alonso, A. (19 de abril de 2019). Gráficos en R. Rpubs. <https://rpubs.com/aafernandez1976/graficos>
- Promedio industrial Dow Jones (^DJI). Yahoo! Finance. <https://finance.yahoo.com/quote/%5EDJI/history?p=%5EDJI>. (24 de junio de 2022).
- Gujarati, D. y Porter, D. (2010). Econometría. Mc Graw Hill.