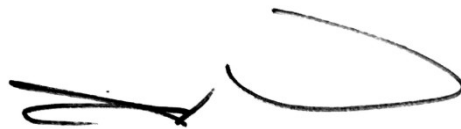


Nota de Aceptación:

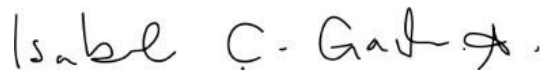
Proyecto de Diseño Aprobado. en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana Cali para optar el título de Ingeniero Industrial.



HERNÁN CAMILO ROCHA NIÑO
Decano de la Facultad de Ingeniería y Ciencias



JORGE ENRIQUE ÁLVAREZ PATIÑO
Director Carrera Ingeniería Industrial



ISABEL CRISTINA GARCÍA ARBOLEDA
Director(a) Proyecto de Diseño



LUZ ELENA VINASCO ISAZA
Jurado 1



KATHLEEN SALAZAR SERNA
Jurado 2



Pontificia Universidad
JAVERIANA
Cali

Santiago de Cali, 22 de noviembre de 2021.

Profesor
Álvaro Figueroa Cabrera
Curso PD2 - Ingeniería Industrial-
Pontificia Universidad Javeriana Cali.

Asunto: Aval unificado proyectos de diseño II

Cordial saludo.

Por medio de la presente, y con base en lo revisado en las últimas versiones, confirmo que los proyectos de diseño identificados con los números 102 y 106, están listos para ser sustentados. Sin más, agradezco sus comentarios en las entregas anteriores y quedo atenta de brindar cualquier información que considere pertinente.

Atentamente,

Isabel C. Garcia

Isabel Cristina Garcia Arboleda



Modelo estadístico para medir el impacto económico del covid-19 a través del índice COLCAP

Paula Andrea Bravo Aponte^{a,c}, María Camila Cardona González^{a,c}, Carlos Andrés Hernández Giraldo^{a,c}, María José Martínez Herrera^{a,c}

Isabel Cristina García^{b,c}

^aEstudiante de Ingeniería Industrial

^bProfesora, Directora del Proyecto de Grado, Departamento de Ciencias Naturales y Matemáticas

^cPontificia Universidad Javeriana, Cali, Colombia

Resumen

La situación de confinamiento ocasionada por la llegada del virus SARS-COV-2, generó un impacto en el sector financiero, en especial en la Bolsa de Valores de Colombia (BVC), donde para muchas empresas se tomó como una oportunidad de crecimiento y para otras fue la causa de grandes pérdidas económicas. El presente proyecto pretende diseñar un modelo estadístico, que integre el confinamiento ocasionado por la aparición del virus, con el impacto que se produce en la Bolsa de Valores de Colombia (BVC), a través de la evaluación del índice COLCAP. El objetivo es realizar un análisis desde el ámbito financiero para diseñar un modelo estadístico, que permita que las empresas tengan una base de conocimiento cuando se presenten situaciones similares al confinamiento. Para este proyecto inicialmente se realizó la contextualización y justificación del proyecto, definiendo la situación presentada a nivel nacional, además de los grupos de interés en el proyecto y sus requerimientos. Posteriormente, se realizó la recolección de datos de un periodo de dos años, para definir las variables que influyen en el índice COLCAP y para realizar un estudio pre-confinamiento, confinamiento y post-confinamiento. Después del estudio se eligió el modelo de serie de tiempo (ARIMA) dado que es el que mejor se adecúa a la situación comprendida. Finalmente, debido a que la intención del proyecto no es solo medir el impacto económico, sino también ser aplicable y adaptable en las situaciones similares, se ajustó el modelo al índice IMAE.

Palabras claves: Bolsa de Valores de Colombia (BVC), confinamiento, índice COLCAP, series de tiempo.

Abstract

The lockdown situation resulting from the virus SARS-COV-2, impacted the financial sector, especially in the “Bolsa de Valores de Colombia” (BVC), where many companies took it as an opportunity of growth and for others, it was the cause of a lot of economic losses. The present project aims to design a statistic model, that integrates the lockdown forced by the virus, with the impact produced in the “Bolsa de Valores de Colombia” (BVC), through the index evaluation COLCAP. The objective is to do an analysis from the financial side to design a statistic model that lets the companies have basic knowledge when they have similar situations to the lockdown. Initially for this project we did a contextualization and justification of the project, defining the situation presented at the national level, also the groups of interest in the project and their requirements. Subsequently, we did a data recollection for a period of two years, to figure out variables that influence the index COLCAP and to do research(studies) in pre lockdown, lockdown and post lockdown situations. After the study we choose the time series model (ARIMA) which is the one fits better the situation understood. Finally, due to the project intentions, it is not only due to the size of the economic impact otherwise, it is also to be applicable and adaptable in similar situations. We adjust the model to the index IMAE.

Key Words: Index COLCAP, lockdown, time series.

Tabla de contenido

I. PROJECT CHARTER	4
II. DEFINIR	5
A. Contexto y Justificación	5
B. Grupos de interés.....	7
C. Requerimientos.....	9
III. MEDIR	10
A. Plan de recolección de datos.....	10
B. Exploración del mercado.....	13
IV. ANALIZAR	17
A. Análisis de Oportunidad	17
B. Revisión de literatura.....	21
C. Exploración de ideas y selección de alternativa	22
D. Objetivos	24
E. Plan de trabajo (PdT).....	25
V. DISEÑAR	25
A. Desarrollo del diseño de la solución	25
B. Validación del diseño propuesto.....	30
VI. VERIFICAR	35
A. Medición de los impactos y resultados.....	35
B. Estandarización de la solución	39
C. Conclusiones	40
D. Recomendaciones.....	40
VII. GLOSARIO	41
VIII. REFERENCIAS	42
IX. ANEXOS	44

Índice de Tablas

TABLA I. ESCALA DE CALIFICACIÓN SELECCIONADA PARA LOS GRUPOS DE INTERÉS	7
TABLA II. GRUPOS DE INTERÉS	7
TABLA III. REQUERIMIENTOS DE LOS GRUPOS DE INTERÉS.....	9
TABLA IV. INDICADORES DE DESEMPEÑO A MEDIR DE VARIABLES DE ENTRADA	10
TABLA V. RESULTADOS DE INDICADORES DE DESEMPEÑO EN VARIABLES DE SALIDA	13
TABLA VI. ANÁLISIS PESTEL DEL PROYECTO	13
TABLA VII. RESULTADOS DE LOS SUPUESTOS DE ANOVA EN MODELOS DE REGRESIÓN	20
TABLA VIII. REVISIÓN DE LITERATURA.....	21
TABLA IX. ANÁLISIS DE LAS ALTERNATIVAS	22
TABLA X. MÉTODO VISUAL PARA LOS CRITERIOS	23
TABLA XI. PORCENTAJE RELATIVO PARA CADA CRITERIO	23
TABLA XII. ESCALA DE PUNTUACIÓN PARA LOS CRITERIOS DE EVALUACIÓN.....	23

TABLA XIII. RESULTADOS DE LA METODOLOGÍA AHP.....	24
TABLA XIV. RESULTADOS DE LA COMPARACIÓN DE MODELOS POR MEDIO DEL MSE.....	24
TABLA XV. PLAN DE TRABAJO.....	25
TABLA XVI. RESULTADOS DE LA COMPARACIÓN DE TODAS LAS ALTERNATIVAS DE MODELOS POR MEDIO DEL MSE.....	30
TABLA XVII. INTERVALOS DE CONFIANZA DEL ÍNDICE COLCAP EN R-STUDIO.....	34
TABLA XVIII. INTERVALOS DE CONFIANZA DEL IMAE VALLE EN R-STUDIO.....	38
TABLA XIX. COSTO DE IMPLEMENTAR O PROGRAMAR UN SHINY PARA R-STUDIO.....	38
TABLA XX. TABLA DE ANEXOS.....	44

Índice de Figuras

Fig. 1. Caída de la capitalización bursátil en el año 2020 [15].....	5
Fig. 2. Información de operaciones de compraventa de la acción Bancolombia [4].....	6
Fig. 3. Matriz de Interés vs. Poder.....	8
Fig. 4. Serie de tiempo para TRM.....	11
Fig. 5. Serie de tiempo para DTF.....	12
Fig. 6. Serie de tiempo para el Precio del Petróleo.....	12
Fig. 7. Serie de tiempo para Tasa de Desempleo en Estados Unidos.....	12
Fig. 8. Serie de tiempo para Tasa Prime.....	12
Fig. 9. Matriz DOFA del proyecto.....	14
Fig. 10. Gráfico del modelo de diseño experimental/aleatorización [16].....	15
Fig. 11. Ejemplo gráfico del modelo de discontinuidad en la regresión [16].....	15
Fig. 12. Gráfico del modelo de dobles diferencias [16].....	16
Fig. 13. Gráfico del modelo de pareamiento [16].....	16
Fig. 14. Diagrama de correlación entre indicadores de entrada y el índice COLCAP.....	17
Fig. 15. Coeficiente de determinación (R^2 ajustado).....	18
Fig. 16. Gráfico 4 en 1 para evaluación de supuestos.....	18
Fig. 17. Prueba de normalidad para los residuos.....	18
Fig. 18. Prueba de igualdad de varianzas.....	19
Fig. 19. Datos para encontrar el valor d_l y d_u en tablas.....	19
Fig. 20. Prueba de aleatoriedad e independencia para los residuos (Minitab).....	19
Fig. 21. Gráfico de serie de tiempo del índice COLCAP en R-Studio.....	27
Fig. 22. Componentes de la serie temporal del índice COLCAP.....	27
Fig. 23. Gráfico de diferencia en diferencia del valor del índice COLCAP en R-Studio.....	28
Fig. 24. Gráfico de rezagos en la serie de valores diarios del índice COLCAP.....	28
Fig. 25. Función de autocorrelación simple de una serie temporal.....	29
Fig. 26. Función de autocorrelación parcial.....	29
Fig. 27. Prueba para validar el supuesto de distribución normal.....	31
Fig. 28. Gráfica de puntos de cambio en R-Studio.....	32
Fig. 29. Validación de los puntos de cambio.....	33
Fig. 30. Predicción del modelo para el índice COLCAP.....	33
Fig. 31. Comparación entre el pronóstico y los datos reales.....	34
Fig. 32. Pronóstico de la serie de tiempo ARIMA.....	35
Fig. 33. Gráfico de serie de tiempo del IMAE VALLE en R-Studio.....	35
Fig. 34. Prueba para validar el supuesto de distribución normal del IMAE Valle.....	36
Fig. 35. Gráfica de puntos de cambio en R-Studio del IMAE Valle.....	37
Fig. 36. Validación de los puntos de cambio del IMAE Valle.....	37
Fig. 37. Predicción del modelo para el IMAE Valle.....	38
Fig. 38. Diagrama de flujo del proceso a seguir para correr el modelo en R-Studio.....	39

I.PROJECT CHARTER

Descripción (<i>Business case</i>)				Planteamiento del problema (<i>Problem statement</i>)		
La situación de confinamiento ocasionada por la llegada de la pandemia por el COVID-19 ocasionó un gran impacto en el sector financiero, que se convirtió en una oportunidad de proyecto al no encontrar ningún modelo que integre la aparición del virus SARS-COV-2 con el impacto que se produce en la Bolsa de Valores de Colombia (BVC). Debido a lo mencionado, se decide diseñar un modelo estadístico que realice la integración del confinamiento al análisis financiero y logre estimar el comportamiento del índice COLCAP. El modelo también se podrá adaptar a posibles situaciones que tengan similitudes a la vivida con la aparición del virus o se podrá utilizar en caso de una continuación prolongada de la pandemia.				La situación por tratar en el proyecto es la inclusión del confinamiento causado por el virus SARS-COV-2 en el diseño de un modelo estadístico de serie de tiempo que realice un análisis financiero de la Bolsa de Valores de Colombia.		
Impacto de los actores (<i>Stakeholder's business needs</i>)				Restricciones	Especificaciones	Marco legal
Los actores más influenciados son las entidades financieras y el gobierno nacional.				-Información disponible limitada. -Acceso restringido a la base de datos de la bolsa de valores en la universidad por tema pandemia, a pesar de contar con bases de datos online.	-Mejor aprovechamiento de los recursos.	-Ley 510 de 1999. Se dictan disposiciones en relación con el sistema financiero y asegurador, el mercado público de valores, las superintendencias Bancaria y de Valores.
Indicadores de Desempeño (<i>KPI's</i>)						
Variable		Actualidad			Meta	
Calidad del modelo		N.A.			Alcanzar un porcentaje de satisfacción del 70%.	
Visualización financiera		N.A.			Mostrar la representación gráfica de forma adaptable y accesible de los 5 indicadores financieros.	
Confiabilidad del modelo		N.A.			Obtener un error de máximo el 10%.	
Objetivo general (<i>Goal statement</i>)						
Diseñar un modelo estadístico que considere como variable de entrada el confinamiento debido al virus SARS-CoV-2, que permita predecir el comportamiento del índice COLCAP para medir el impacto económico en la Bolsa de Valores de Colombia (BVC).						
Objetivos específicos (<i>Project scope</i>)						
<ul style="list-style-type: none"> Analizar la influencia de las variables de entrada en el índice COLCAP durante el periodo marzo 2019-marzo 2021. Mostrar la infografía de la influencia de las variables de entrada en el modelo haciendo uso de un tablero de mando (dashboard) en el periodo marzo 2019-marzo 2021. Construir un modelo adaptable a situaciones extremas como la pandemia COVID-19, donde la variable confinamiento sea evaluada para realizar la predicción del índice COLCAP y sea ajustable a este modelo. 						
Plan de Trabajo (<i>Project Plan</i>)				Equipo de trabajo (<i>Team members</i>)		
Actividad	Fecha Inicio	Fecha Fin	Área IISE	Nombre	Rol	
Evaluación del comportamiento de los datos de entrada a través de estadística descriptiva.	27/04/2021	29/04/2021	5. Quality & Reliability Engineering	Carlos Andrés Hernández	Information giver	
Análisis de los índices de las variables de entrada a través del modelo de regresión múltiple.	06/05/2021	07/05/2021	5. Quality & Reliability Engineering	Paula Andrea Bravo	Elaborator	
Investigación del método.	09/08/2021	16/08/2021	11. Information engineering.	Todos	Information seeker	
Elaboración de un dashboard para facilitar el interactuar del usuario con el modelo.	17/08/2021	25/08/2021	11. Information engineering.	María Camila Cardona y Paula Andrea Bravo	Initiator/Coordinator	
Definición del modelo que ha sido seleccionado para realizar la predicción del índice (modelo base).	26/08/2021	09/09/2021	11. Information engineering.	Todos	Information seeker	

Construcción del modelo con las variables de entrada definidas en el proyecto.	10/09/2021	25/10/2021	12.1 Product Design & Development	Todos	Elaborator/Orientor
Diseño de una encuesta de satisfacción dirigida a los clientes.	26/10/2021	12/11/2021	5. Quality & Reliability Engineering	María Camila Cardona	Elaborator
Evaluación del error relativo porcentual de los resultados arrojados por el modelo.	01/11/2021	05/11/2021	5. Quality & Reliability Engineering	Paula Andrea Bravo	Elaborator
Conclusiones del modelo a partir de los resultados entregados.	07/11/2021	16/11/2021	11. Information engineering.	Carlos Andrés Hernández	Opinion giver/Initiator

II.DEFINIR

A. Contexto y Justificación

Desde el año 2020 se dio un suceso que ningún sector económico, social y cultural esperaba. El surgimiento del virus catalogado con el nombre de SARS-CoV-2 trajo consigo cambios que impactaron de forma positiva y negativa en el mundo, realizando una transformación de todas sus dimensiones. De las organizaciones del sector financiero que fueron ampliamente impactadas por la situación, son las bolsas de valores, como por ejemplo la de Nueva York, la de Tokio, y específicamente la Bolsa de Valores de Colombia (BVC), la cual sufrió un golpe que para muchas empresas se tomó como una oportunidad de crecimiento y para otras fue la causa de grandes pérdidas económicas y estructurales.

El COLCAP es uno de los índices bursátiles de la Bolsa de Valores de Colombia (BVC), el cual indica las variaciones de los precios de las 20 acciones más líquidas [1]. A partir de este, se ve reflejado el impacto de la pandemia denominada COVID-19, en donde los primeros meses del año 2020 las acciones se venían debilitando, dando un claro reflejo de la situación global y de la caída de la economía en muchos sectores a nivel mundial como consecuencia de las estrictas cuarentenas impuestas en cada país por el rápido avance del número de contagiados y muertos a causa del virus. Por esta razón, el COLCAP retrocedió 36,2% solo en tres jornadas, siendo así, el nivel más bajo del 2020 con 894,03 unidades, que a su vez era una pequeña muestra de lo que iba a ocurrir a lo largo del año. En consecuencia, el valor de las empresas en la bolsa, definido también, como la capitalización bursátil, cayó cerca del 40%, lo cual representó una pérdida de más de 48 613 millones de pesos [2] como se muestra en la Fig.1.

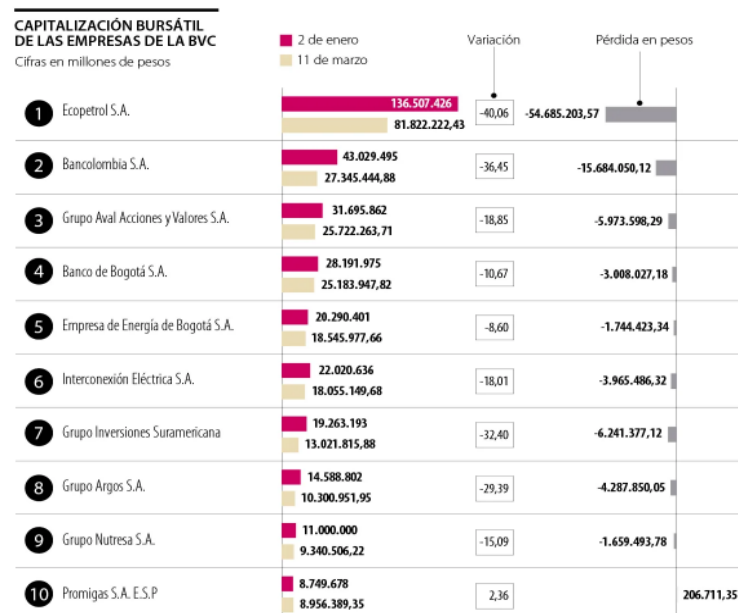


Fig. 1. Caída de la capitalización bursátil en el año 2020 [15]

Posterior a este suceso, el panorama para el año 2021 era un poco más alentador, debido a que, al generar el levantamiento de algunas restricciones y la implementación de pruebas piloto de apertura con normas de bioseguridad, como estrategia para impulsar la reactivación sectorial del comercio, la industria y el turismo; empresas como BANCOLOMBIA, ECOPETROL y

BANCO DE BOGOTÁ generaron un crecimiento positivo con respecto al año anterior del -7%, por lo tanto el COLCAP repuntó cerca del 21,9% [2].

La cuantificación del impacto económico requiere un análisis específico caso por caso, teniendo en cuenta que existen cambios institucionales para la regulación y buen manejo en la Bolsa de Valores de Colombia (BVC) por consiguiente, no sólo es importante la evaluación de los cambios en la economía nacional, también es necesario realizar un análisis de lo que sucede con los índices accionarios a nivel mundial, lo que permitirá profundizar en casos particulares, aplicable y viable, donde la importancia de las variables o los factores macroeconómicos en la bolsa enmarcan la trayectoria del impacto a nivel local e internacional.

Una de las empresas que conforman las 20 acciones más líquidas es Bancolombia, quien no fue ajena a la crisis mundial vivida, donde a pesar de que el valor de las acciones llegaba a ser histórico para su negociación en la Bolsa Colombiana antes de la pandemia, tuvo un gran impacto, a tal punto que dejó pérdidas insuperables, que no se veían desde el año 2008, lo cual se puede observar en la Fig. 2.



Fig. 2. Información de operaciones de compraventa de la acción Bancolombia [4]

El panorama anterior, traza la importancia de estudiar el problema planteado, proporcionando un análisis de la realidad desde las dimensiones económicas, sociales, culturales y políticas, las cuales traen consigo una visión local y a su vez global. Por lo tanto, entender que empresas como Bancolombia quienes representan la mayor parte de la población colombiana y que a su vez no se limita solo a ser una entidad bancaria, tiene como consecuencia que crisis como estas impacten a diferentes sectores y se vea reflejado no solo en sistemas económicos como la bolsa sino también en la vida cotidiana de cada persona.

Actualmente no existen modelos que integren la aparición del virus SARS-COV-2 con el impacto que produce en la Bolsa de Valores de Colombia (BVC), por lo tanto, es de gran importancia hacer énfasis en este hecho para utilizarlo como una oportunidad.

Debido a la situación identificada anteriormente, se considera necesario profundizar en el tema para el diseño de un modelo estadístico, que pueda predecir el comportamiento del índice COLCAP y medir el impacto económico de la bolsa en situaciones como la pandemia debida al virus SARS-CoV-2. El modelo se podrá adaptar a posibles situaciones que tengan similitudes a la vivida con la aparición del virus o también se podrá utilizar en caso de una continuación prolongada de la pandemia.

El proyecto trae como beneficio la realización de análisis de forma objetiva desde el ámbito financiero, debido a que por medio de esté no se espera tener un modelo que muestre cómo se comporta el virus, sino que se enfoca en evaluar el indicador COLCAP que regula gran parte de la economía del país, permitiendo que empresas nacionales e internacionales tengan bases para validar propuestas de negocio y tomar decisiones cuando se presentan situaciones o fenómenos que hagan que las ciudades se cierren en un confinamiento y se vea afectada la economía del país.

Es importante resaltar que la participación de empresas colombianas en la Bolsa de Valores de Colombia (BVC) como Bancolombia, son vitales para el desarrollo del sector económico, es como tener un sello de calidad ante la comunidad internacional. Por consiguiente, es necesario reconocer que la cooperación del país en organizaciones como la OCDE es clave, para generar confianza a posibles inversores. Esta organización está encaminada a proponer a sus países miembros, políticas que ayuden a construir economías sólidas tanto de sus países desarrollados como en vía de desarrollo. Contar con modelos, análisis e indicadores permiten prever conjeturas económicas futuras, las cuales impulsan el desarrollo social, económico, ecológico y ambiental entre otros [5].

B. Grupos de interés

Teniendo en cuenta la propuesta de un modelo estadístico, se establecen los siguientes grupos de interés para el proyecto:

- Sector financiero y bancario
- Sector educativo
- Gobierno Nacional (Banco de la República)
- Equipo de proyecto
- Indicador Mensual de Actividad Económica (IMAE).

De acuerdo con eso, en la TABLA I, se explica cuál fue el criterio utilizado para evaluar el interés y el poder que tiene cada uno de los grupos en el proyecto de diseño.

*TABLA I.
ESCALA DE CALIFICACIÓN SELECCIONADA PARA LOS GRUPOS DE INTERÉS*

Escala	Interés	Poder
1	No hay ningún interés en el proyecto	No tiene ningún poder sobre el proyecto
2	Tiene poco interés	Tiene poco poder
3	Tiene interés moderado	Tiene poder moderado
4	Tiene interés en el proyecto	Tiene poder sobre el proyecto
5	Tiene alto interés en el proyecto	Tiene alto poder sobre el proyecto

Donde:

- 1: No tiene ningún interés o poder en los resultados del proyecto, ya que, no implica un impacto en las actividades que realizan.
- 2: Tiene poco interés en el proyecto debido a que se puede tener una influencia baja en las decisiones que tome a partir del modelo.
- 3: Se interesa de forma parcial en el proyecto, ya que los resultados que se obtienen pueden tener un poco de influencia en algunos aspectos de interés.
- 4: Hay interés en los resultados del proyecto, porque pueden influenciar en las decisiones que se tomen.
- 5: Hay un alto interés como consecuencia de que se tendrá la oportunidad de generar grandes cambios en el sector al que pertenece.

Con ayuda de la escala y las especificaciones planteadas, en la TABLA II se definen dichos grupos de interés y se explica cómo van a verse afectados por esto, dando mayor claridad del nivel de atención que se debe tener para cada uno.

*TABLA II.
GRUPOS DE INTERÉS*

Código	Partes Interesadas	Descripción	¿Qué interés tiene con el proyecto?	Interés	Poder
●	Equipo de proyecto	El equipo de trabajo se compone de cuatro estudiantes de la Pontificia Universidad Javeriana que cuentan con diferentes personalidades, se encargan de realizar el proyecto y de hacer la investigación.	El equipo de proyecto está interesado en realizar el proyecto para aprobar las materias proyecto de diseño 1 y 2, para cumplir los requisitos de grado y para contribuir a la universidad con sus procesos de investigación.	5	4

●	Indicador Mensual de Actividad Económica (IMAE)	Institución de educación superior de alta calidad que ejerce la investigación a nivel regional y global.	El Indicador Mensual de Actividad Económica (IMAE) tiene interés en el proyecto con la intención de continuar con procesos de investigación y enriquecimiento de las bases de datos.	5	3
●	Sector financiero y bancario	Son entidades financieras que tienen por función principal la captación de recursos para realizar operaciones activas de crédito [3].	Los establecimientos bancarios tienen interés en el proyecto para utilizarlo como información de referencia del comportamiento económico de las bolsas en el momento de aparición del COVID-19.	4	1
●	Sector educativo	Instituciones de educación media y superior como colegios, escuelas técnicas y universidades.	Las entidades educativas tienen interés en el proyecto para fines de investigación, creación de nuevos indicadores con la información obtenida y procesos de aprendizaje de economía e historia.	3	1
●	Gobierno Nacional (Banco de la República)	Ministerio y Superintendencia financiera de Colombia, Banco de la República, Ministerio de Comercio.	El Gobierno nacional de Colombia tiene interés en el proyecto con el fin de obtener información de la bolsa de valores en la situación de la pandemia, debido a que muchas entidades públicas y privadas cotizan en ella y les será útil para entender la situación económica pasada, actual y para las futuras negociaciones, específicamente en el Banco de la República.	5	1

Al tener una descripción detallada y una clasificación del nivel de impacto y poder que pueden llegar a asumir cada uno de los interesados en el proyecto, se procede a realizar la matriz de Interés vs. Poder, Fig. 3.

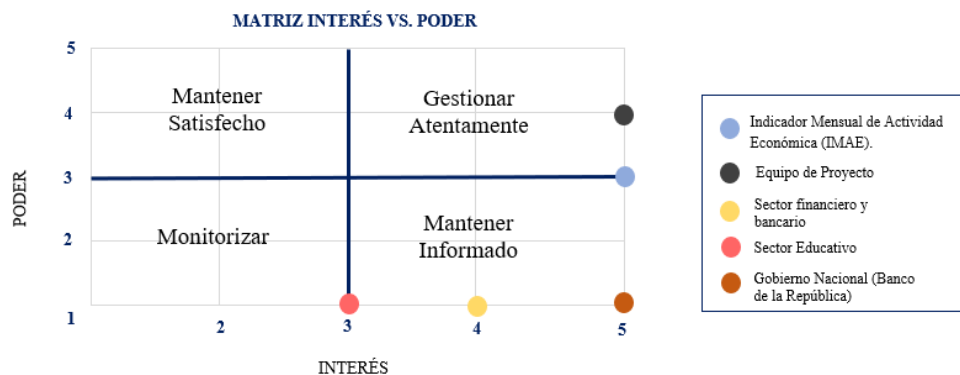


Fig. 3. Matriz de Interés vs. Poder

A partir de la matriz de la Fig. 3 se puede concluir que gran parte de los interesados se encuentran en el cuadrante “Mantener Informados” lo que indica un interés alto sobre los resultados que se generen a partir del proyecto, pero no cuentan con poder de influencia para su realización. En este caso, el equipo de trabajo debe mantener una constante actualización de la información para generar una visión más clara del tipo de consecuencias que tendrá el proyecto en cada una de ellas, con el fin de lograr un trabajo colaborativo.

Para los interesados que hacen parte del cuadrante “Gestionar atentamente”, se buscará hacer un mayor esfuerzo para maximizar la participación de cada uno de ellos al interior del proyecto, teniendo en cuenta las necesidades, requerimientos y su satisfacción con los resultados.

Finalmente, se puede observar que en este caso ninguno de los grupos de interés se encuentra dentro del cuadrante “Monitorizar” y “Mantener Satisfecho”.

C. Requerimientos

Los requerimientos son parte fundamental para el proyecto, con el fin de satisfacer las necesidades, expectativas y beneficios que puedan esperar u obtener las partes interesadas con los resultados finales. A partir de esto, es necesario identificar adecuadamente cada uno de los requerimientos, lo cual permitirá un correcto desarrollo del proyecto, teniendo en cuenta que, cada uno de ellos presenta restricciones y especificaciones de diseño, así como también las leyes, normas y estándares. Dicha información se encuentra en la TABLA III.

TABLE III.
REQUERIMIENTOS DE LOS GRUPOS DE INTERÉS

GRUPOS DE INTERÉS	VoC (REQUISITOS GRUPOS DE INTERÉS)	RESTRICCIONES DE DISEÑO	ESPECIFICACIONES DE DISEÑO	LEYES, NORMAS Y ESTÁNDARES	
				Legislación y Requisitos Aplicables	Importancia o Efecto
Equipo de proyecto	-Cumplimiento de los requisitos de los grupos de interés, aplicables dependiendo de las restricciones. -Fortalecimiento de habilidades estadísticas y de análisis de datos.	Los integrantes del equipo tienen poco conocimiento de modelación de eventos estadísticos.	-Generación valor a los grupos de interés con los resultados del proyecto. -Proporcionamiento de información confiable. - Cumplimiento de las especificaciones del material brindado para el curso de Proyecto de diseño.	-Acuerdo No.567 del Reglamento de estudiantes [6].	Derechos, deberes y consideraciones de las relaciones entre los estudiantes.
Indicador Mensual de Actividad Económica (IMAE)	-Que el modelo sea confiable en la exactitud de la información. -Que el modelo diseñado en el proyecto sea adaptable para implementar en el indicador mensual de actividad económica (IMAE).	Acceso restringido a la base de datos de la bolsa de valores en la universidad por tema pandemia, a pesar de contar con bases de datos online.	-Toma de decisiones en situaciones extremas como la pandemia COVID-19 haciendo uso de otro tipo de indicadores.	-Normas de la universidad. -Normas de bioseguridad en la universidad. -Normas del laboratorio contabilidad y finanzas.	-Uso de las instalaciones, laboratorios, equipos y programas. -Ingreso a la universidad bajo normas de bioseguridad.
Sector educativo	Que el modelo diseñado sirva como referencia para otros estudiantes o instituciones.	Educación reducida en el sector financiero, inversiones y manejo de bolsa de valores.	Permite que la comunidad tenga mayor extensión de búsqueda en bases de datos.	-Ley 715 de 2001 [9]. - Ley 819 de 2003 [10].	-Normas orgánicas en materia de recursos y se dictan otras disposiciones para organizar la prestación de los servicios de educación. -Normas orgánicas del presupuesto.
Sector financiero y bancario	Que el modelo diseñado genere información en tiempo real	La información que se puede obtener puede no ser totalmente confiable.	-Mejor aprovechamiento de sus recursos. -Toma de decisiones a partir de los resultados planteados.	-Ley 510 de 1999 [7]. -Ley 45 de 1990 [8].	-Se dictan disposiciones en relación con el sistema financiero y asegurador, el mercado público de valores, las superintendencias Bancaria y de Valores. -Se expiden normas en materia de intermediación financiera.

Gobierno Nacional (Banco de la República)	-Que el modelo suministre datos reales para obtener indicadores de comportamiento e impacto. -Contar con un proyecto aplicable para la organización.	-Mala gestión de la economía nacional. -Intereses privados e individualistas. -Información disponible limitada.	-Visión clara del impacto que generan las decisiones tomadas para el sector económico. -Definición de mejores estrategias de regulación.	-Ley 35 de 1993 [11]. -Ley 1314 de 2009 [12].	-Normas generales a las cuales debe sujetarse el Gobierno Nacional para regular las actividades financiera, bursátil y aseguradora y cualquier otra relacionada con el manejo, aprovechamiento e inversión de recursos captados del público. -Se regulan los principios y normas de contabilidad e información financiera y de aseguramiento de información aceptados en Colombia.
---	---	---	---	--	---

III.MEDIR

A. Plan de recolección de datos

Para continuar con el desarrollo del proyecto, se debe definir de forma clara el plan de recolección de datos (PRD), el cual es fundamental para una buena toma de información, pues en él se pueden definir las variables que influyen y afectan actualmente el índice COLCAP. Al ser un proyecto de investigación aplicada, no se realiza una toma de datos primaria, sino que los mismos serán recolectados de fuentes secundarias oficiales y bases de datos abiertas que son actualizadas constantemente, sin desconocer que son verídicas. En este apartado se realizó una investigación y profundización con el fin de establecer la base de datos y variables asociadas a la oportunidad que ha sido identificada en la propuesta del proyecto, en donde apoyados de las diferentes plataformas que impactan directamente a la Bolsa de Valores de Colombia (BVC) y su entorno, se definieron las siguientes variables de entrada en el modelo a través de la información encontrada en el histórico de las escuelas de economía para la recolección de datos, las cuales se encuentran en la TABLA IV.

Es así como el plan se desarrollará desde la base de datos proporcionada por el Banco de la República y otras fuentes de información, en las que se exponen los gráficos y datos necesarios para el proyecto, la cual estará delimitada desde el 25 de marzo del 2019 con un corte al 25 de marzo del 2020 en donde se da inicio al confinamiento y finalmente se tendrá el cierre al 25 de marzo del 2021, resultando así un estudio de dos años con una mirada pre-confinamiento, confinamiento y post-confinamiento. A su vez, la información proporcionada tiene un respaldo de confiabilidad y es actualizada diariamente dependiendo de la variable o información requerida, la recolección de datos se dará de manera organizada y concreta, en donde cada una de las variables tomará su lugar de manera que se relacione finalmente en el modelo a ejecutar.

Por lo tanto, la búsqueda de cada variable se delimitará según las fechas descritas con anterioridad, posterior a ello se depurará y se hará una limpieza de la información que no es requerida por medio de la minería de datos, con el fin de obtener y ordenar la información necesaria para generar una base de datos interna para el desarrollo del proyecto.

El plan de recolección de datos (PRD) que se ha mencionado anteriormente se puede encontrar en el ANEXO 1, y el entorno de los indicadores se encontrará en el ANEXO 2. Con el fin de tener más claridad de los indicadores y variables de entrada del modelo se realiza la TABLA IV.

TABLA IV.
INDICADORES DE DESEMPEÑO A MEDIR DE VARIABLES DE ENTRADA

Variable	Objetivo	Descripción	Indicador
Promedio de las tasas de operaciones de compra y venta de dólares	Se requiere, ya que es uno de los factores para observar el comportamiento del índice COLCAP.	Media de las tasas de operaciones de compra y venta de dólares al día [17].	$Tasa\ representativa\ del\ mercado\ (TRM)$ $= \frac{\sum_{i=1}^n [(MTC_i \times TTC_i) + (MTV_i \times TTV_i)] + \sum_{k=1}^t (MBV_k \times TBV_k)}{\sum_{i=1}^n [(MTC_i + MTV_i)] + \sum_{k=1}^t MBV_k}$ [17]

Interés a ahorradores por CDT	Se usa para dar una idea aproximada del costo de los recursos en la economía y cómo eso afecta al índice COLCAP.	Tasa semanal promedio ponderada por monto de las captaciones que se paga a los ahorradores por los CDT a 90 días [18].	$DTF = \frac{\sum Tasa\ de\ interés_i\ CDT\ 90\ días \times Cant.\ de\ recursos_i\ CDT\ 90\ días}{Total\ de\ recursos\ del\ CDT\ a\ 90\ días\ en\ todas\ las\ entidades}$ [18]
Precio del Petróleo	Se requiere como una variable de la economía que podría tener impacto en el índice COLCAP.	Se da de acuerdo con la oferta y la demanda del recurso, también cambia debido a otros factores externos [19].	Precio diario del petróleo Brent en dólares/barril
Desempleo de Estados Unidos	Se tiene en cuenta debido a la influencia que tiene Estados Unidos en el comportamiento de las diferentes bolsas de valores alrededor del mundo.	Se calcula dividiendo el número de desempleados entre el total de personas que forman la fuerza laboral [20].	$Tasa\ de\ desempleo = \frac{N^\circ\ de\ desempleados}{Población\ activa\ (PA)} \times 100$ [20]
Interés mínimo en Estados Unidos	Se requiere como una variable de la economía que podría tener impacto en el índice COLCAP.	Tasa de interés mínima de los Bancos en Estados Unidos. Se calcula un promedio entre los principales bancos comerciales de Estados Unidos [21].	$Tasa\ Prime = \frac{\sum Tasa\ de\ interés_i}{25}$ Donde: $i = Cantidad\ de\ bancos\ más\ grandes\ en\ Estados\ Unidos$ [21]
Tiempo de cuarentena	Presenta las temporadas (tiempo) en las que se presentó confinamiento, para poder evaluar los cambios en la bolsa en un determinado momento.	Variante binaria, siendo 1 día en estado de confinamiento y 0 día sin estado de confinamiento.	Índice de confinamiento: $T(i) = \begin{cases} 1, & Si\ i = Estado\ de\ Confinamiento \\ 0, & Si\ i = Sin\ Estado\ de\ Confinamiento \end{cases}$

Se debe tener en cuenta que las variables de entrada que han sido nombradas en la TABLA IV no cuentan con una actualidad y una meta, ya que ellas no buscan mejorar el índice COLCAP, sino que dan como resultado una serie de datos o insumos de vital importancia que serán procesados por el modelo.

Los datos recolectados de las variables mencionadas anteriormente se encuentran en el ANEXO 3. A continuación, se dispondrá el análisis de los datos de forma gráfica en el periodo de marzo 2019 hasta marzo 2021, donde se puede observar la variación de los indicadores debido a las afectaciones en el año 2020 por el confinamiento estricto a raíz de la pandemia generada por el virus SARS-CoV-2.

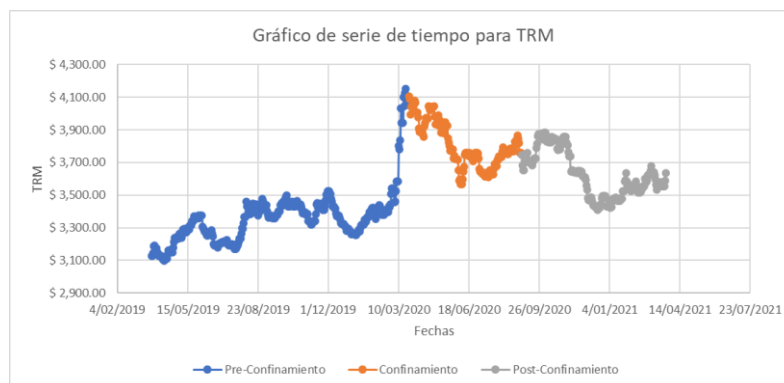


Fig. 4. Serie de tiempo para TRM

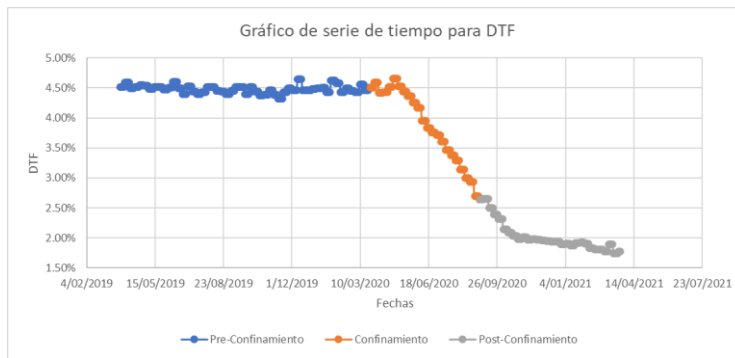


Fig. 5. Serie de tiempo para DTF

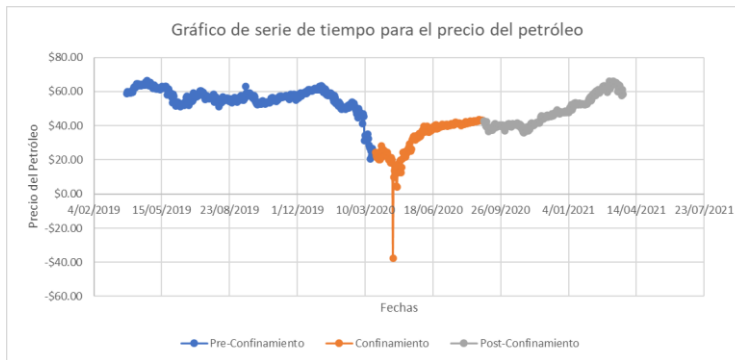


Fig. 6. Serie de tiempo para el Precio del Petróleo



Fig. 7. Serie de tiempo para Tasa de Desempleo en Estados Unidos

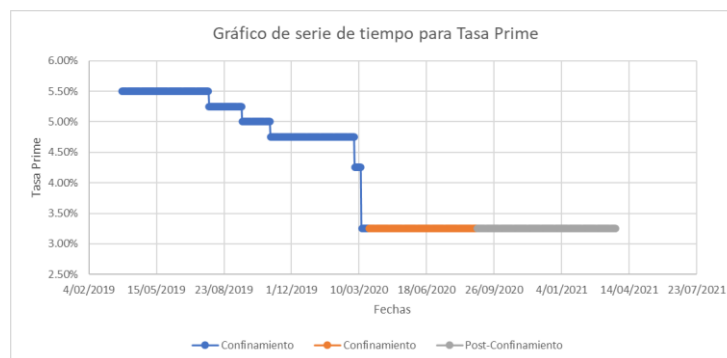


Fig. 8. Serie de tiempo para Tasa Prime

Para obtener más información del comportamiento de los indicadores anteriormente mostrados, se realizó un análisis estadístico descriptivo que se expone en el ANEXO 4.

Por otra parte, se observa como en la Fig. 6 el precio del petróleo tiene un dato atípico por causa de una caída en la demanda debido a la pandemia del coronavirus que llevó al colapso del crudo el 20 de abril de 2020 [32].

B. Exploración del mercado

Después de conocer las variables de entrada, se presentará en la TABLA V los resultados de actualidad y metas relacionados a unas variables de salida, ya que es importante verificar que el modelo realmente funciona frente a unos términos planteados que indican a dónde se quiere llegar con lo que está siendo propuesto en el proyecto. En este apartado también se presentará un análisis PESTEL y DOFA con el propósito de conocer los factores externos del entorno del problema y los factores internos del proyecto. Finalmente, se realizará una breve explicación de algunos modelos de medición de impacto planteados por el Banco Interamericano de Desarrollo, con el fin de conocer e identificar cómo se reportan este tipo de problemas en el mercado actual.

TABLA V.
RESULTADOS DE INDICADORES DE DESEMPEÑO EN VARIABLES DE SALIDA

Variable	Objetivo	Descripción	Indicador	Actualidad	Meta
Calidad del modelo	Ofrecer un modelo apto para el propósito que ha sido planteado y que cumplirá con todos los criterios requeridos para la medición de impactos.	Mide la calidad del proyecto a través de la satisfacción, en la cual se realiza una encuesta con una escala de calificación de 1 a 5.	$\%Calidad = \frac{\left(\frac{\sum \text{Puntos obtenidos en encuesta de satisfacción}}{\text{Total encuestas de satisfacción}}\right)}{5} \times 100$	No se encuentra información en el momento.	Alcanzar un porcentaje de satisfacción del 70%.
Visualización financiera	Dar una visualización como estrategia para la comprensión de la influencia de las variables financieras en la construcción del índice COLCAP.	Se realizará un dashboard influenciado por una infografía.	-Tasa representativa del mercado (TRM) -Depósito a término fijo (DTF) -Brent -La tasa de desempleo de Estados Unidos -Tasa Prime	No se encuentra información en el momento.	Mostrar la representación gráfica de forma adaptable y accesible de los 5 indicadores financieros.
Confiabilidad del modelo	Brindar información confiable de fuentes secundarias de alta calidad generando un valor del índice COLCAP más preciso.	Se mide la capacidad de predicción por medio de un intervalo de confianza al 90% en los distintos períodos analizados. En el rango marzo 2019 - marzo 2021.	$\text{Error relativo porcentual} = \frac{ V_{Real} - V_{Aproximado} }{Valor Real} \times 100$	No se encuentra información en el momento.	Obtener un error de máximo el 10%.

En la TABLA VI se encuentra el análisis PESTEL, el cual ayudó a identificar los factores externos e internos que afectan el entorno de la situación expuesta por el proyecto.

TABLA VI.
ANÁLISIS PESTEL DEL PROYECTO

P	E	S	T	E	L
POLÍTICO	ECONÓMICO	SOCIOCULTURAL	TECNOLÓGICO	ECOLÓGICO	LEGAL
Favorable					
Políticas de regulación del mercado de valores que protegen a los inversionistas.	Subsidios y proyectos para reactivación económica.	Normas estrictas de bioseguridad para el cuidado de la vida.	-Desarrollo de constantes innovaciones tecnológicas que permiten tener velocidad, consistencia, oportunidad, seguridad y confianza en la realización de transacciones.	Disminución de la contaminación del aire y del agua por confinamiento de la población (menos carros en las vías y menos personas en las playas).	Creación de decretos para ministerios, entidades, presidencia, instituciones, superintendencias, entre otros; para mejorar la situación de la población debido al virus SARS-COV-2.

			-Actualización constante en bases de datos abiertas encontradas en la web.		
Desfavorable					
Mala gestión de los recursos públicos.	Crisis económicas por medidas de confinamiento.	Confinamientos estrictos para la población.	-Bajo porcentaje de uso debido a los altos costos de implementación. -Mala programación de plataformas tecnológicas, lo que conlleva a tener información errónea.	Aumento del desperdicio de la energía eléctrica por teletrabajo y clases virtuales debido al confinamiento.	Reformas tributarias y nuevos impuestos.

En la Fig. 9 se encuentra la matriz DOFA, elaborada con el propósito de analizar de forma práctica y estratégica el proyecto a partir de la evaluación de factores internos (fortalezas y debilidades) y factores externos (oportunidad y amenaza) que han sido identificados al realizar el análisis PESTEL.

Matriz DOFA	Fortalezas (F)	Debilidades (D)
	Oportunidad (O)	Estrategia FO
O1: Innovaciones tecnológicas en el mercado financiero que permiten tener velocidad, seguridad y confianza. O2: Información verídica y actualizada constantemente en bases de datos abiertas.	F1: Conocimiento de las diferentes bases de ingeniería para el estudio del proyecto. F2: Interpretación de indicadores y bases de datos en línea F2O2: Uso y apoyo de las bases de datos que brinda el Banco de la República, Yahoo! Finanzas y el índice COLCAP para obtener datos verídicos y óptimos en el proyecto.	D1: Actualmente no se cuenta con modelos de impacto económicos en situaciones similares. D1O1: Apoyándose en modelos de impacto actuales en el mercado y aprovechando la innovación tecnológica en el ámbito financieros se desarrollará el modelo propuesto.
Amenaza (A)	Estrategia FA	Estrategia DA
A1: Mala programación de plataformas tecnológicas. A2: Crisis económicas por medidas de confinamiento.	A1: Realizar controles estrictos en el proceso de programación para mitigar el riesgo de error en las diferentes plataformas económicas.	D1A2: Hacer uso de los modelos de impacto actuales para adaptarlos a crisis económicas o confinamientos similares

Fig. 9. Matriz DOFA del proyecto

Después de investigar los diferentes indicadores o modelos de medición de impacto en los que se tiene una relación directa o indirecta con el proyecto propuesto, se logran identificar algunos modelos estadísticos brindados por el Banco Inter-Americano de Desarrollo de los cuales logramos extraer su aplicabilidad e importancia en problemas como el que se está estudiando, por lo tanto, como punto de partida se evaluó que los modelos además de medir el impacto, tuvieran un efecto de causalidad en donde finalmente logre relacionar las variables propuestas y sus interdependencias entre sí. De esta manera, se recopilieron modelos como: la aleatorización o diseño experimental, discontinuidad en la regresión, dobles diferencias (diferencia en diferencias) y pareamiento, estos se elegirán teniendo en cuenta los diferentes motivos, razones, relaciones y que recojan a su vez las diferentes restricciones que se presentan. A continuación, serán descritos y relacionados con el proyecto:

- Diseño experimental/aleatorización:

Es el modelo de mejor rendimiento disponible en la evaluación de impacto, el cual da a cada unidad la misma oportunidad de recibir tratamiento, a su vez genera dos grupos estadísticamente idénticos, y es factible en evaluaciones prospectivas cuando la demanda excede la oferta del beneficio. En contraste con el estudio que se realizará tanto al mercado COLCAP y las diferentes variables como la TRM, DTF, el valor del precio del petróleo, entre otras, el modelo logra proporcionar desde una justificación estadística y matemática la medición del impacto, que en este caso es dado por el acontecimiento de la pandemia, logrando así obtener resultados cuantitativos y a su vez explicativos para el fenómeno estudiado. Adicionalmente, es necesario tener en cuenta que el modelo desarrolla tanto una validez interna como externa, la cual asegura que los resultados en la muestra serán representativos de la población a un nivel determinado de error muestral (validez externa) y por otro lado asegura que el efecto observado sobre la variable dependiente se deba al tratamiento y no a otros factores (validez interna) [16].

Las ventajas de uso de este modelo desde una visión cuantitativa producen el mejor contrafactual posible, siendo a su vez intuitivo y fácil de comunicar.



Fig. 10. Gráfico del modelo de diseño experimental/aleatorización [16]

- Discontinuidad en la regresión:

El modelo está basado en un índice o parámetro claramente definido o también como un corte conocido para la elegibilidad, este modelo es posible cuando las unidades pueden ordenarse a lo largo de una dimensión cuantificable sistemáticamente relacionada con la asignación del tratamiento, por lo tanto, el efecto se mide en la discontinuidad, en otras palabras, el impacto es estimado alrededor del corte y no se puede generalizar a toda la población [16].

Un ejemplo gráfico de este modelo se presenta en la Fig. 11.

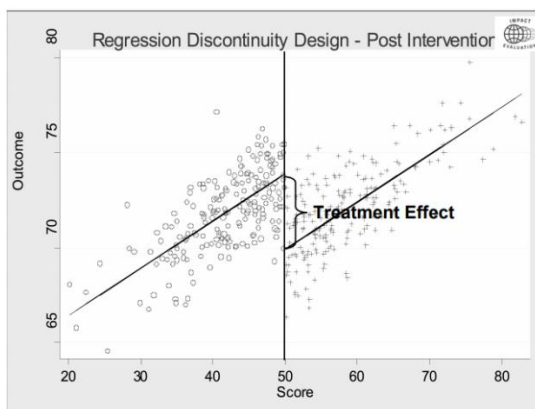


Fig. 11. Ejemplo gráfico del modelo de discontinuidad en la regresión [16]

En relación al proyecto propuesto, al estudiar gráficos proporcionados por la Bolsa de Valores de Colombia (BVC) en donde se muestra los diferentes precios a través del tiempo de las diferentes acciones que participan en ella, se puede entender como la crisis ocasionada por la pandemia genera diferentes rupturas en los precios de grandes empresas como Bancolombia, donde la caída generó un cambio totalmente marcado en el precio de sus acciones, así entender que este tipo de fenómenos traen consigo la oportunidad de estudio y aplicación frente a modelos como el presentado en donde haciendo uso de hechos históricos se puede generar un análisis y representación.

Finalmente, este modelo tiene como ventaja la producción de una estimación insesgada del efecto del tratamiento en la discontinuidad, además no es necesario “excluir” a un grupo de hogares/individuos del tratamiento.

- Dobles diferencias (diferencia en diferencias)

Este modelo compara el cambio en los resultados entre tratamientos y no tratamientos, el impacto es medido en la diferencia del cambio en los resultados, siendo así explicado como: $\text{Impacto} = (Y_{t1} - Y_{t0}) - (Y_{c1} - Y_{c0})$ [16].

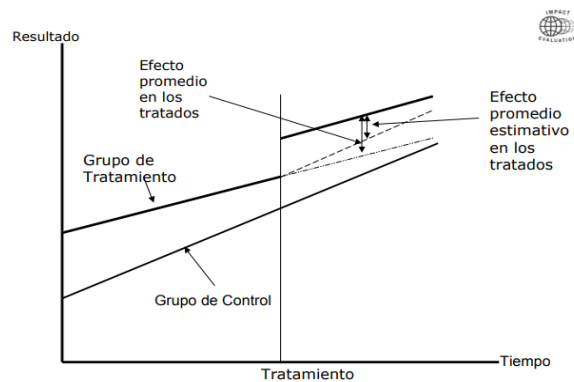


Fig. 12. Gráfico del modelo de dobles diferencias [16]

En el modelo se tiene como supuesto principal que las tendencias (pendientes) son las mismas en los tratamientos y los controles, es así como se necesita un mínimo de 3 puntos en el tiempo para verificar esto y estimar el tratamiento.

Al realizar el contraste entre el modelo expuesto y el proyecto, se observa que haciendo uso de él se buscó finalmente hacer uso de las variables y de los datos recolectados para que a través de esto se pueda medir el impacto desde la diferencia entre ellos, en donde la información obtenida estará alrededor de un tiempo determinado teniendo en cuenta el surgimiento de la crisis que dará como resultado el impacto generado, al igual que las implicaciones que se tienen y la aplicabilidad de este para situaciones o panoramas similares.

- Pareamiento:

Este modelo se define tomando la comparación ideal que pareo el grupo de tratamiento de uno más grande o de la población estudiada, los pares son seleccionados en base a similitudes en las características observadas, por lo tanto, supone que no hay sesgo de selección basado en características no observables [16].

Se alcanza el mismo resultado pareando sobre la probabilidad de participación, esto supone que, dada X, la participación es independiente de los resultados, en la Fig. 13 se visualiza lo descrito anteriormente.

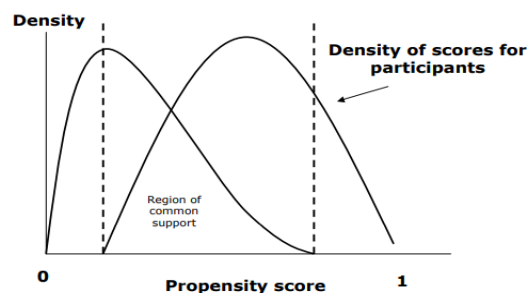


Fig. 13. Gráfico del modelo de pareamiento [16]

Como modelo es aplicable en el sentido de que proporciona un estudio fundado en grupos más pequeños y representativos, los cuales logran a través de su comportamiento trazar la tendencia o comportamiento para cualquier otro grupo de la población. Específicamente hablando de Bancolombia en donde se explicó y justificó anteriormente su comportamiento e impacto en las diferentes áreas y dimensiones de Colombia, hace que a través de la aplicación de modelos como el propuesto se exponga un comportamiento generalizado y aplicable en diferentes áreas o situaciones críticas como la presentada.

En conclusión, para el uso de alguno de los modelos expuestos se analizará el impacto en cada una de las variables y la situación comprendida. También se diseñará un nuevo modelo, que a su vez recoja y satisfaga las diferentes restricciones expuestas y por su parte desarrolle un resultado satisfactorio, con el fin de no solo medir el impacto sino de ser aplicable en las diferentes situaciones que se presenten.

IV. ANALIZAR

A. Análisis de Oportunidad

Al concluir la etapa medir, en donde se logró precisar y abrir el panorama desde el enfoque práctico, en el cual el desarrollo investigativo de la situación problema, los diferentes modelos de impacto y la construcción de la base de datos, dan resultados que se deben analizar, con el fin de llegar a conclusiones y respuestas a los diferentes interrogantes o problemas planteados, que su vez logran demarcar lo que se busca en el proyecto, el alcance, el impacto y la oportunidad que se tiene. Siendo la última, la propuesta de valor que se ofrece frente al problema o situación planteada inicialmente, en donde implica estudiar detenidamente las características del proyecto, para descubrir oportunidades, nichos y necesidades que con los recursos que cuenta o podría contar el proyecto se puedan cubrir.

Para conocer en mayor profundidad la oportunidad, partiendo del hecho que el proyecto tiene como fin el desarrollo de un modelo estadístico que logre precisar, medir y evaluar de una mejor manera el mercado de la bolsa de valores en situaciones o crisis como lo es la pandemia por el virus SARS-CoV-2 o eventos similares, se ofrece un modelo adaptativo y aplicativo a diferentes sectores. Surgen diferentes interrogantes tales como: ¿Cómo ofrecer al mercado la confiabilidad y seguridad de un modelo real, preciso y aplicable?, ¿El modelo que se desarrolla es una propuesta de valor suficiente para ser usado y aplicado a diferentes sectores?, y ¿Qué se le garantiza al cliente a la hora de aplicar el modelo en la situación o mercado que se presente?

En la etapa medir, se realizó la minería, recolección y construcción de la base de datos correspondiente, en la cual, se pudo estudiar el comportamiento de los datos obtenidos por medio del uso de resúmenes estadísticos, gráficos de cajas o bigotes y el análisis de la relación de cada variable cuantitativa por medio de gráficos de correlación, por lo tanto se analizó de forma individual cada una de las variables frente al índice COLCAP, dando como resultado relaciones altas que pueden ser directas o inversas frente a dicho índice como se muestra en la Fig. 14 y es por esto que se procede a analizar la relación de todas las variables en conjunto y se decide aplicar el modelo de regresión múltiple, ya que este trata de ajustar modelos lineales o linealizables entre una variable dependiente y más de una variable independiente, de modo que se analizaron las diferentes variables para saber si todas se ajustan entre ellas frente al índice COLCAP y logran llegar al objetivo de determinar si los datos recolectados pueden medir el impacto y la importancia que tienen para la predicción de la variable dependiente, evaluando así la viabilidad y confiabilidad del modelo propuesto.

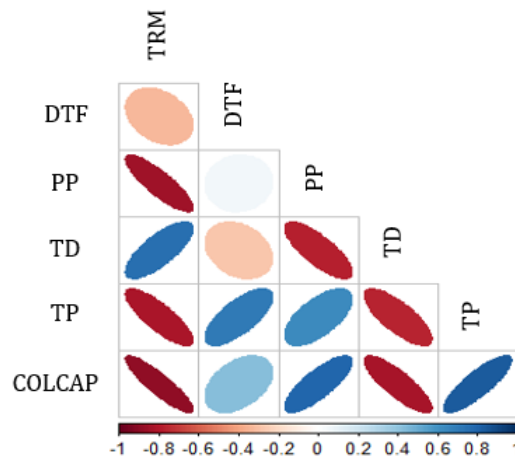


Fig. 14. Diagrama de correlación entre indicadores de entrada y el índice COLCAP

Haciendo uso del análisis de regresión múltiple que se puede visualizar de forma completa y detallada en el ANEXO 5, en el cual se plantea un modelo donde se desea analizar si el índice COLCAP (Y) está siendo explicado por la tasa representativa del mercado (TRM), el depósito a término fijo (DTF), el precio del petróleo, la tasa de desempleo en Estados Unidos y la tasa prime. A partir de esto se pasa a observar cual es la mejor relación de las transformaciones de las variables independientes con respecto a el índice COLCAP, luego se eliminan las variables que generan sesgo en el modelo, lo que muestra que las variables que aportan significativamente corresponden a la tasa representativa del mercado (X1), el depósito a término fijo (X2), el precio del petróleo (X3), la tasa de desempleo en Estados Unidos (X4), la tasa prime (X5) y el estado de confinamiento (X6); por otro lado, en la Fig. 15 se obtuvo el coeficiente de determinación R^2 ajustado, definido como el porcentaje de variación de la variable dependiente que está siendo explicado por las variables independientes del modelo, en este caso el $R^2 = 0.8971$, lo cual indica que el 89.71% de la variación en el índice COLCAP está estrechamente relacionada con las variables presentadas.

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.947546911
Coefficiente de determinación R ²	0.897845148
R ² ajustado	0.897140632
Error típico	65.5399328
Observaciones	731

Fig. 15. Coeficiente de determinación (R² ajustado)

De acuerdo con lo anterior, y la secuencia lógica del modelo de regresión múltiple, se plantean tres supuestos de ANOVA con un nivel de confianza del 95%, en donde también se tendrá en cuenta el gráfico 4 en 1 mostrado en la siguiente figura:

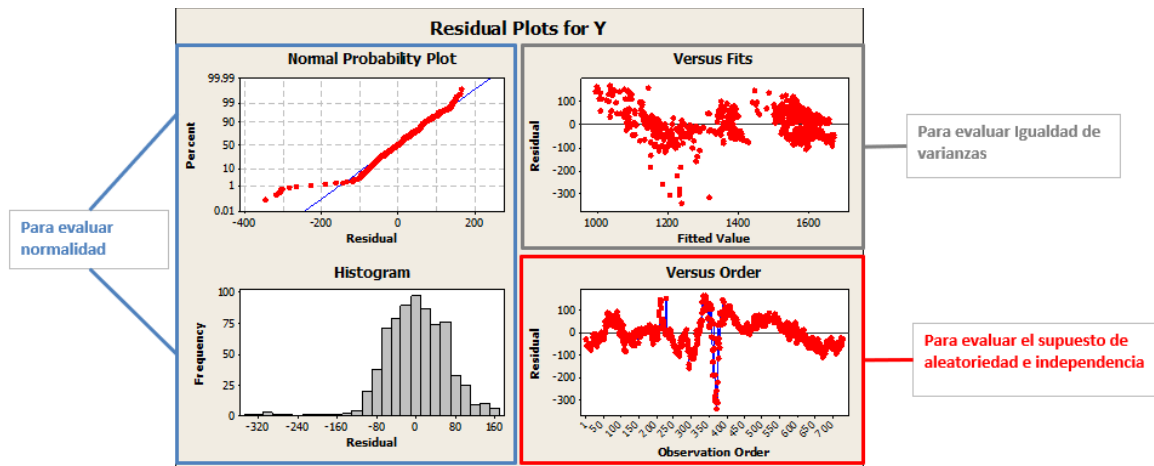


Fig. 16. Gráfico 4 en 1 para evaluación de supuestos

Planteamiento de hipótesis de prueba de bondad de ajuste a la distribución normal

H₀: Los residuos se ajustan a un comportamiento normal

H_a: Los residuos no se ajustan a un comportamiento normal

Regla de decisión: Rechazo H₀ si p-valor < α (para un nivel de confianza del 95%)

Decisión: Como el P-valor = 0.005 < α = 0.05 **Rechazo H₀**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos no se ajustan a un comportamiento normal. Lo que quiere decir, que hubo factores de ruido que afectaron la información tomada para el experimento y que dañaron el comportamiento normal en las variables.

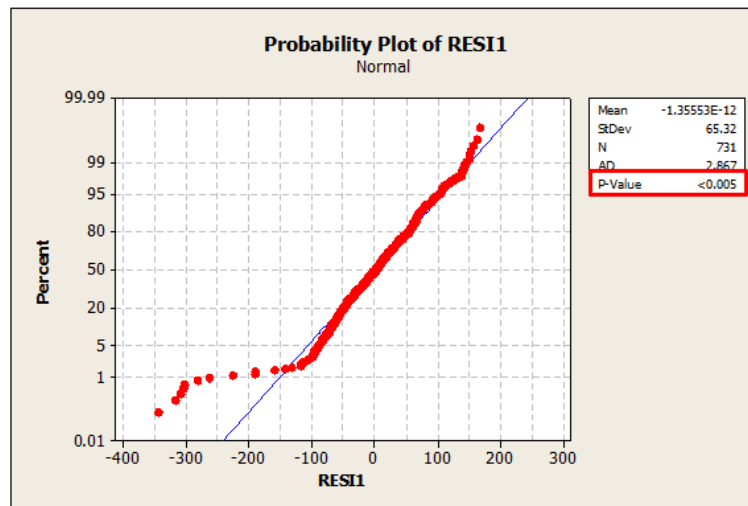


Fig. 17. Prueba de normalidad para los residuos

Planteamiento de hipótesis de prueba de igualdad de varianzas "homocedasticidad"

Ho: Los residuos se tienen varianzas estadísticamente iguales

Ha: Los residuos no tienen varianzas estadísticamente iguales

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.002116 < \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos no tienen varianzas estadísticamente iguales. Esto significa que las parejas de datos no oscilan aleatoriamente por encima y por debajo de cero con una variación estadísticamente igual, lo que permite decir que los datos no son comparables entre ellos y hay variables que tienen variación significativamente diferente.

ANÁLISIS DE VARIANZA					
		Grados de libertad de cuadrado de los cuadrados		F	P-Valor
Regresión	1	30635949.7	30635949.7	9.58337837	0.002116024
Residuos	363	1160431040	3196779.73		
Total	364	1191066990			

Fig. 18. Prueba de igualdad de varianzas

Planteamiento de hipótesis de prueba de aleatoriedad e independencia

Ho: Los residuos son aleatorios e independientes

Ha: Los residuos no son aleatorios ni independientes

Regla de decisión:

Si $D.W. < dl$ Rechazo Ho

Si $dl < D.W. < du$ Cumple con reserva

Si $D.W. > du$ No rechazo Ho

Decisión: Como $D.W. = 0.206295 < dl = 1.718$ **Rechazo Ho**

Conclusión: A partir del estadístico Durbin Watson calculado en Minitab con un nivel de confianza del 95%, se puede afirmar que los residuos no son aleatorios e independientes, lo que quiere decir que si existe autocorrelación y sesgos entre las variables analizadas.

Datos	
n =	731
k =	5
dl =	1.718
du =	1.820

Fig. 19. Datos para encontrar el valor dl y du en tablas

Dato arrojado en Minitab
Durbin-Watson statistic = 0.206295

Fig. 20. Prueba de aleatoriedad e independencia para los residuos (Minitab)

Al desarrollar los supuestos y determinar si los datos se ajustan o no a una distribución normal, si tienen o no varianzas iguales, y si son aleatorios e independientes; se parte del hecho que estos deben cumplirse para así afirmar que se tiene un ajuste válido y confiable en el modelo de regresión. Finalmente se rechaza cada uno de los supuestos mencionados con anterioridad y se concluye

que, a partir de los resultados obtenidos, el modelo planteado para realizar la predicción del índice COLCAP no es confiable, robusto y aplicable para este caso.

Por lo cual, el uso de la herramienta de regresión múltiple no logra cumplir su objetivo principal, de ajustar el modelo de forma curvilínea entre las variables independientes, convirtiéndose en una opción inadecuada para realizar las predicciones del índice COLCAP en el proyecto. Por lo tanto, se debe precisar la información recolectada y buscar otro tipo de alternativas para poder llevar a cabo un estudio robusto, confiable y verídico; lo que conlleva al surgimiento de una oportunidad de trabajo para el diseño y desarrollo de un modelo de impacto que logre medir y predecir el índice COLCAP relacionando de manera óptima cada una de las variables planteadas.

Por otra parte, se decide hacer la evaluación de otro modelo de regresión múltiple en donde la variable “confinamiento” se divida en tres estados (pre-confinamiento, confinamiento y post-confinamiento), el cual se podrán analizar de forma más profunda en el ANEXO 8. También se realiza la evaluación de cada una de las variables de forma independiente en modelos de regresión simple para conocer su comportamiento y verificar que está pasando con cada una de ellas. A partir de esto, se encuentra que todas las variables se ajustan a un modelo POLINÓMICO² debido a que su coeficiente de determinación es el más alto y además se identifica que ninguna de las variables cumple con los tres supuestos. Los resultados de estos modelos se encuentran de forma más detallada en el ANEXO 9.

En la TABLA VII se muestran los resultados de los supuestos para cada modelo de regresión planteado.

TABLA VII.
RESULTADOS DE LOS SUPUESTOS DE ANOVA EN MODELOS DE REGRESIÓN

MODELO	Prueba de normalidad Ho: Las medidas se ajustan a un comportamiento normal vs. Ha: Las medidas no se ajustan a un comportamiento normal		Prueba de igualdad de varianzas "homocedasticidad" Ho: Los residuos se tienen varianzas estadísticamente iguales vs. Ha: Los residuos no tienen varianzas estadísticamente iguales		Prueba de aleatoriedad e independencia Ho: Los residuos son aleatorios e independientes vs. Ha: Los residuos no son aleatorios ni independientes			
	P - valor	Decisión: Rechazo Ho si p-valor < $\alpha=0.05$	P - valor	Decisión: Rechazo Ho si p-valor < $\alpha=0.05$	Durbin Watson	dl	du	Decisión: Si $D.W. < dl$ Rechazo Ho Si $dl < D.W. < du$ Cumple con reserva Si $D.W. > du$ No rechazo Ho
Regresión múltiple con 2 estados (con confinamiento y sin confinamiento)	0.005	Rechazo Ho	0.002116	Rechazo Ho	0.206295	1.718	1.820	Rechazo Ho
Regresión múltiple con 3 estados (pre-confinamiento, confinamiento y post-confinamiento)	0.005	Rechazo Ho	0.002539	Rechazo Ho	0.207284	1.707	1.831	Rechazo Ho
Regresión simple para la tasa representativa del mercado (TRM)	0.005	Rechazo Ho	0.027	Rechazo Ho	0.130427	1.748	1.789	Rechazo Ho
Regresión simple para el depósito a término fijo (DTF)	0.005	Rechazo Ho	0.000	Rechazo Ho	0.039270	1.748	1.789	Rechazo Ho

Regresión simple para el precio del petróleo	0.005	Rechazo Ho	0.000	Rechazo Ho	0.096230	1.748	1.789	Rechazo Ho
Regresión simple para la tasa de desempleo en USA	0.005	Rechazo Ho	0.000	Rechazo Ho	0.105517	1.748	1.789	Rechazo Ho
Regresión simple para la tasa prime	0.005	Rechazo Ho	0.000	Rechazo Ho	0.138886	1.748	1.789	Rechazo Ho

Teniendo en cuenta lo anterior, es una oportunidad el diseño de un modelo que cumpla con cada uno de los parámetros deseados y a su vez logre su objetivo de manera que sea útil y aplicable, ofreciendo finalmente una mejor manera de asumir y entender situaciones y fenómenos como la pandemia del COVID-19 desde el ámbito económico y estadístico.

B. Revisión de literatura

Los modelos de evaluación de impacto económico permiten hacer una simplificación de las relaciones que existen entre las variables que explican gran parte del funcionamiento de la economía, por lo cual, son herramientas esenciales que ayudan a estudiar el comportamiento de situaciones o fenómenos que se viven en la actualidad, los cuales pueden ser complejos de analizar. Por medio de este tipo de modelos, se puede tener un enfoque exclusivo en variables de mayor relevancia, lo que proporciona una simplificación del análisis que se desea realizar, permitiendo una mejor aplicación en el estudio de casos reales.

Considerando el problema del cálculo del índice COLCAP en el mercado, se tiene que es uno de los indicadores bursátiles de mayor importancia para referenciar el mercado nacional, por lo cual, es fundamental realizar un análisis adecuado de su comportamiento, teniendo en cuenta cualquier situación que pueda presentarse en el país y que a su vez impacte de forma positiva o negativa en la economía. Al hacer una revisión en la literatura, se encuentran algunos modelos que permiten evaluar o calcular el comportamiento del índice COLCAP, pero ninguno hace referencia al impacto del virus SARS-CoV-2 ni al confinamiento que se dio a partir de él.

*TABLA VIII.
REVISIÓN DE LITERATURA*

Autores	Problema tratado	Método de solución propuesto	Aporte
Susan Juliette León Crisanchó, Alfredo Trespalacios Carrasquilla [22]	Factores macroeconómicos que influyen en la volatilidad del índice COLCAP	Se analizaron diferentes estudios que involucraron variables económicas en la predicción de series financieras especialmente en índices bursátiles para tres medidas de volatilidad: los rendimientos al cuadrado, el VIMA y la desviación estándar; con las cuales se estimaron 6 modelos econométricos, identificando que sí existen factores macroeconómicos nacionales e internacionales que influyen sobre la volatilidad del índice y subsecuentemente la incertidumbre del mercado.	En el proyecto se crearon modelos de regresión de las variables a través del método de mínimos cuadrados para cada una de las medidas de volatilidad relacionadas a las variables escogidas que influyen en el índice COLCAP.
Alonso Triana Rodríguez, Luis Alejandro Lozano [23]	Análisis de la relación que tiene el índice COLCAP con los fundamentales de la macroeconomía que representan la actividad nacional, y como estos tienen consecuencia en el índice bursátil.	Se analizaron diferentes estudios que involucran variables económicas que influyen sobre índices bursátiles para extraer las más representativas a nivel nacional, mediante un modelo econométrico se identifica que si existen factores macroeconómicos nacionales que influyen sobre la volatilidad del índice.	En el proyecto se evidencia una fuerte relación del índice COLCAP con el petróleo WTI, de esta manera se evidencia la gran importancia de esta variable en el modelo, es así, que se comprende en cierta medida las correlaciones positivas entre estos. El índice de precios al consumidor (IPC) es la variable descartada por no presentar normalidad ni poder acogerse al modelo trabajado.

Ángela Ramos, Camilo Escamilla, Sebastián Peña, Valentina Olmos [24]	Identificar qué factores mueven el mercado de valores colombiano, específicamente las acciones que integran el índice COLCAP, permitiendo la discusión teórica y la evaluación del comportamiento de sus acciones.	Se presenta un modelo econométrico donde, se evalúan los efectos de cada una de las variables como resultado de una parte metodológica y de contexto. A través de la prueba Dickey Fuller (1979), se determina la naturaleza de los datos trabajados en la investigación, para establecer en la misma el modelo que más se ajuste, ya sea un modelo VAR (modelos autorregresivos vectoriales) o un modelo VEC (modelo de corrección de error vectorial). También se utilizan modelos de regresión lineal múltiple por el método de mínimos cuadrados ordinarios.	Finalmente, teniendo en cuenta dos variables externas (Tasa Desempleo EE.UU y Tasa Prime), es evidente cómo el índice COLCAP no se encuentra afectado en mayor proporción por estas variables externas, sin embargo, se puede evidenciar como las variables internas (DTF, Precio del petróleo Brent y TRM) son variables fundamentales para determinar el comportamiento del índice COLCAP, es decir, la repercusión de las variables internas para este índice son de mayor influencia que las variables externas.
Roobilson Rojas Rey [25]	Determinar las variables macroeconómicas que influyen en el valor del índice de las acciones más transadas en Colombia.	Para determinar el grado de relación lineal entre las variables se calcula el índice de correlación de Pearson. Se realiza regresión lineal con las variables macroeconómicas, seleccionando las más significativas. El modelo APT resulta ser una opción complementaria para la valoración de los activos financieros de Colombia, debido a la actividad de las empresas que cotizan en la BVC y al tipo de economía del país, las variables macroeconómicas permiten prever su valor y comportamiento.	Las variables independientes TRM e Inflación siempre explican representativamente al índice independientemente de las demás variables que acompañen la regresión. No es viable incorporar todas las variables macroeconómicas para explicar el valor del índice, debido a que solo algunas revelan el comportamiento de la variable dependiente. Variables como la tasa de intervención, el precio del petróleo WTI y la producción industrial, la tasa de desempleo y la producción de petróleo también explican el comportamiento del índice, sin embargo, su probabilidad varía dependiendo las demás independientes incluidas. Por otra parte, existe una relación entre las variables macroeconómicas y el COLCAP, algunas más influyentes y representativas.

Al momento de realizar la revisión de literatura a temas relacionados con el problema identificado, se encuentra que existe muy poca información que aborde el mismo objeto de estudio o que tengan un proceso de investigación similar, y los pocos que han sido encontrados manejan metodologías y herramientas semejantes para resolver el problema. Esto se debe a que, a nivel local, los temas relacionados a la pandemia del COVID-19 han sido enfocados en términos de salud pública y no en modelos de medición de impacto para la economía.

C. Exploración de ideas y selección de alternativa

Es fundamental realizar una exploración de ideas para contar con varias alternativas, para esto se realizó una lluvia de ideas basada en la revisión de la literatura de la sección anterior, se utilizó el método BrainWriting que es muy sencillo, cada integrante planteo unas propuestas de solución y en una reunión se hizo el debate BrainWriting para seleccionar y dejar las mejores alternativas. A continuación, se presenta las diferentes ideas seleccionadas:

TABLA IX.
ANÁLISIS DE LAS ALTERNATIVAS

Alternativas	Análisis
Modelo de Regresión Múltiple	La regresión múltiple trata de ajustar modelos lineales o curvilíneos entre una variable dependiente y más de una variable independiente [28]. Este modelo se puede ajustar a nuestro proyecto ya que sabemos que todas las variables se correlacionan, se debe verificar si el modelo realmente explica el índice COLCAP para que se pueda desarrollar por este modelo.
Modelo Económico Diff-Diff	Método estadístico que permite estimar y evaluar el efecto causal de una variable de tratamiento sobre una variable resultado, es una de las técnicas más utilizadas para analizar el efecto o impacto de un cambio sobre un sistema [26]. Es un modelo económico matemáticamente muy básico, puede que se ajuste ya que trabaja variables cualitativas, pero es posible que la construcción del modelo necesite una educación más robusta.
Modelo de Serie de Tiempos (ARIMA)	ARIMA (Autoregressive Integrated Moving Average), destinados a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un

	papel fundamental, un tipo de modelo que produce pronósticos con base en una síntesis de los patrones históricos en los datos [27]. Modelo que se ajusta muy bien al proyecto ya que puede evaluar la variable confinamiento a través del tiempo.
Modelo de Media Móvil Simple	Una media móvil nos muestra el valor medio del precio de un activo en un número de sesiones determinado [29]. La fórmula de media móvil simple toma el promedio de datos de un período de tiempo y "mueve" el período en la serie de datos, un punto de datos cada vez [30] y la fórmula de media móvil exponencial es una media móvil de datos que da más peso a los datos más recientes del período y menos peso a los datos más antiguos del período [31]. Son modelos matemáticamente sencillos de trabajar y se les puede involucrar una variable muy importante que es el tiempo de confinamiento, haciendo que estos dos modelos se adapten bien al proyecto.
Modelo de Media Móvil Exponencial	

1) Método AHP

Con el fin de encontrar la mejor alternativa que se ajuste en el desarrollo del proyecto se aplicará el método de Análisis de Jerarquía Analítica (AHP) que se podrá visualizar detalladamente en el ANEXO 6, donde se utilizarán como alternativas los modelos mencionados en la TABLA IX. Por consenso, se determinaron los criterios más relevantes para evaluar cada modelo en base a las metas del proyecto.

- **Facilidad:** El modelo debe estar a la altura de los conocimientos de los estudiantes para poder aplicarlo.
- **Seguridad:** El modelo debe ser confiable en sus predicciones.
- **Adaptabilidad:** Que tan adaptable es el modelo en diferentes escenarios con circunstancias similares.
- **Precisión:** El modelo debe arrojar información precisa, de lo contrario no servirá para el desarrollo del proyecto.

A continuación, se presenta en la TABLA X la evaluación de los cuatro criterios mencionadas para determinar su prioridad por el método visual, después se observará la TABLA XI donde se presentan los resultados de los porcentajes relativos correspondientes a cada criterio, para argumentar cuál es el de mayor importancia al momento de evaluar las alternativas.

TABLA X.
MÉTODO VISUAL PARA LOS CRITERIOS

1/5	1/4	1/3	1/2	1	2	3	4	5
Precisión		Seguridad	Adaptabilidad	Facilidad				

TABLA XI.
PORCENTAJE RELATIVO PARA CADA CRITERIO

Facilidad	8%
Seguridad	23%
Adaptabilidad	14%
Precisión	54%

En la TABLA XII se muestra la escala de puntuación que se usará para evaluar cada alternativa en base a los criterios planteados.

TABLA XII.
ESCALA DE PUNTUACIÓN PARA LOS CRITERIOS DE EVALUACIÓN

Criterio	Calificación				
	1	2	3	4	5
Facilidad de uso	Muy Complejo	Complejo	Medianamente complejo	Ligeramente complejo	Poco complejo
Seguridad	Muy inseguro	Inseguro	Medianamente seguro	Seguro	Muy seguro
Adaptabilidad	Muy inadaptable	Inadaptable	Medianamente adaptable	Adaptable	Muy adaptable
Precisión de los resultados	Muy impreciso	Impreciso	Medianamente preciso	Preciso	Muy preciso

Por último, se hizo la evaluación para encontrar la mejor alternativa siguiendo el mismo procedimiento, pero esta vez teniendo en cuenta la escala de puntuación. Según los resultados de la TABLA XIII se puede concluir que por el método de Análisis de Jerarquía Analítica (AHP) la mejor alternativa para realizar el proyecto es el modelo de series de tiempo (ARIMA).

TABLA XIII.
RESULTADOS DE LA METODOLOGÍA AHP

Alternativas	Gran total	Mejor Alternativa
Modelo de regresión múltiple	10%	Modelo de serie de tiempo (ARIMA)
Modelo económico Diff-Diff	8%	
Modelo de series de tiempo (ARIMA)	41%	
Modelo media móvil simple	23%	
Modelo media móvil exponencial	18%	

2) *Comparación de Modelos por medio del error cuadrático medio (MSE)*

Se realizaron los modelos de media móvil simple, media móvil exponencial, regresión múltiple y económico Diff-Diff; con el fin de comparar los pronósticos y errores cuadráticos obtenidos de cada uno, es posible ver más detalladamente los análisis y las gráficas de cada uno en el ANEXO 10. El objetivo es evaluar cuál modelo puede realizar un buen pronóstico del índice COLCAP para el proyecto, por lo que se realizó una tabla con los datos de los errores cuadráticos medios obtenidos. A continuación, se presenta la TABLA XIV con los valores:

TABLA XIV.
RESULTADOS DE LA COMPARACIÓN DE MODELOS POR MEDIO DEL MSE

Modelo	Error Cuadrático Medio (MSE)
Media Móvil Simple	\$1.007
Media Móvil Exponencial	\$ 216,74
Diff-Diff	\$37.806
Regresión múltiple con 2 niveles	\$4.257,37
Regresión múltiple con 3 niveles	\$4.214,20

Al hacer la comparación de los errores cuadráticos medios analizados en la TABLA XIV, se concluye que, dentro de los modelos evaluados el de media móvil exponencial es quien presenta el menor error, sin embargo, es necesario continuar con la realización del Modelo de Series de Tiempo ARIMA, el cual se ejecutará a lo largo del semestre para su respectivo análisis de resultados y evaluación de su error.

D. Objetivos

Objetivo general:

Diseñar un modelo estadístico que considere como variable de entrada el confinamiento debido al virus SARS-CoV-2, que permita predecir el comportamiento del índice COLCAP para medir el impacto económico en la Bolsa de Valores de Colombia (BVC).

Objetivos específicos:

- Analizar la influencia de las variables de entrada en el índice COLCAP durante el periodo marzo 2019-marzo 2021.

- Mostrar la infografía de la influencia de las variables de entrada en el modelo haciendo uso de un tablero de mando (dashboard) en el periodo marzo 2019-marzo 2021.
- Construir un modelo adaptable a situaciones extremas como la pandemia COVID-19, donde la variable confinamiento sea evaluada para realizar la predicción del índice COLCAP y sea ajustable a este modelo.

E. Plan de trabajo (PdT)

TABLA XV.
PLAN DE TRABAJO

Objetivo General					
Diseñar un modelo estadístico que considere como variable de entrada el confinamiento debido al virus SARS-CoV-2, que permita predecir el comportamiento del índice COLCAP para medir el impacto en la Bolsa de Valores de Colombia (BVC).					
Objetivo Específico	Actividad	Área IISE	Herramientas de Ingeniería Industrial	Entregable (alcance)	Fecha entrega
Analizar la influencia de las variables de entrada en el índice COLCAP durante el periodo marzo 2019-marzo 2021.	Evaluación del comportamiento de los datos de entrada a través de estadística descriptiva.	5. Quality & Reliability Engineering	Teoría de probabilidades	Anexo 4 – Análisis de estadística descriptiva.	29/04/2021
	Análisis de los índices de las variables de entrada a través del modelo de regresión múltiple.	5. Quality & Reliability Engineering	Regresión y diseño de experimentos	Anexo 5- Regresión múltiple para los indicadores de las variables de entrada.	07/05/2021
Mostrar la infografía de la influencia de las variables de entrada en el modelo haciendo uso de un tablero de mando (dashboard) en el periodo marzo 2019-marzo 2021.	Investigación del método.	11. Information engineering.	Seminario de investigación	N.A.	16/08/2021
	Elaboración de un dashboard para facilitar el interactuar del usuario con el modelo.	11. Information engineering.	N.A.	Tabla dashboard.	25/08/2021
Construir un modelo adaptable a situaciones extremas como la pandemia COVID-19, donde la variable confinamiento sea evaluada para realizar la predicción del índice COLCAP y sea ajustable a este modelo.	Definición del modelo que ha sido seleccionado para realizar la predicción del índice (modelo base).	11. Information engineering.	Estadística aplicada	N.A.	09/09/2021
	Construcción del modelo con las variables de entrada definidas en el proyecto.	12.1 Product Design & Development	Estadística aplicada - Regresión y diseño de experimentos	Modelo final.	25/10/2021
	Diseño de una encuesta de satisfacción dirigida a los clientes.	5. Quality & Reliability Engineering	N.A.	Resultados de la encuesta.	12/11/2021
	Evaluación del error relativo porcentual de los resultados arrojados por el modelo.	5. Quality & Reliability Engineering	Estadística aplicada	Cálculo de los resultados al comparar los datos reales y estimados.	05/11/2021
	Conclusiones del modelo a partir de los resultados entregados.	11. Information engineering.	Seminario de investigación	Conclusiones y recomendaciones.	16/11/2021

El soporte del plan de trabajo por medio del cronograma detallado se encuentra en el ANEXO 7.

V.DISEÑAR

A. Desarrollo del diseño de la solución

De acuerdo con la fase anterior, se analizaron 5 alternativas para el diseño del modelo estadístico del impacto económico que genera el COVID-19 al valor del índice COLCAP. Los modelos se evaluaron a través del error cuadrático medio, en el cual se concluyó que el modelo que obtuvo el menor error fue el de la Media Móvil Exponencial. Se procede con la sexta y última alternativa, el modelo de serie de tiempo ARIMA, para el correspondiente análisis, evaluación de su error y comparación con las demás alternativas.

Las series de tiempo ARIMA son modelos paramétricos que tratan de obtener la representación de la serie en términos de la interrelación temporal de sus elementos. Este tipo de modelos que caracterizan las series como sumas o diferencias, ponderadas

o no, de variables aleatorias o de las series resultantes [33]. Éste es un modelo dinámico de series temporales, es decir, los valores futuros vienen explicados por los datos del pasado y no por variables independientes.

La sigla ARIMA significa Promedios Móviles Autorregresivos Integrados, en donde:

- AR hace referencia al componente de autorregresión
- MA hace referencia al componente de promedios móviles
- I hace referencia a la diferenciación.

La notación clásica es ARIMA (p, d, q) donde:

p: Valor que controla el número de componentes autorregresivas (AR) del modelo.

d: Valor que controla el número de integraciones (I) del modelo.

q: Valor que controla el número de componentes media móvil (MA) del modelo.

La fórmula general de ARIMA es la siguiente:

$$Y_t = \underbrace{(Y_t^d - Y_t)}_I + 0 + \underbrace{\sum_{i=1}^p \phi_i Y_{t-i}^d}_{AR} + \underbrace{\sum_{i=1}^q \theta_i \epsilon_{t-i}}_{MA} + \epsilon_t$$

Para la identificación del modelo primero es necesario identificar a simple vista qué modelos se pueden ajustar a la función de autorregulación, procediendo a calcular una medida que defina cuál de los observados sería la mejor opción. En este caso el patrón de correlación es el que se debe comparar con determinados patrones para asociar el modelo ARIMA correspondiente.

Los modelos ARIMA son orientados a estudiar procesos estacionarios, los cuales tienen una función de probabilidad que no varía en el tiempo (la media y la varianza no cambian).

Una vez obtenido el modelo, se evalúa a través de los parámetros p, d, q. Para entrar en la verificación del modelo se evalúan los p-valores y se grafican los residuos para ver si realmente no muestran patrones. Si el modelo se aproxima satisfactoriamente a la serie observada, los residuos deben tender a comportarse como ruido blanco, lo que se comprobaría mediante las funciones de autocorrelación de los residuos (ACF, ACFP) [34].

PASOS EN R DESARROLLADOS PARA PREDECIR EL ÍNDICE COLCAP:

Para plantear y desarrollar el modelo en el lenguaje de programación R-Studio se debe tener en cuenta la información obtenida anteriormente, en donde la base de datos generada nos proporciona el valor del índice COLCAP como variable a tratar.

En este modelo fueron requeridas librerías que permitieron que las funciones se ejecutarán correctamente a través del software usado:

```
library(tseries)
library(timsac)
library(ggplot2)
library(changepoint)
```

Como se mencionó anteriormente, se requiere de la base de datos obtenida, la cual proporciona la información necesaria con las fechas y precios del índice para el modelo estadístico, al tener la información de entrada en el software se generó el gráfico correspondiente, en el cual se visualiza el precio del índice COLCAP, con el fin de corroborar que la Fig. 21 y la información concuerde con los datos a trabajar.

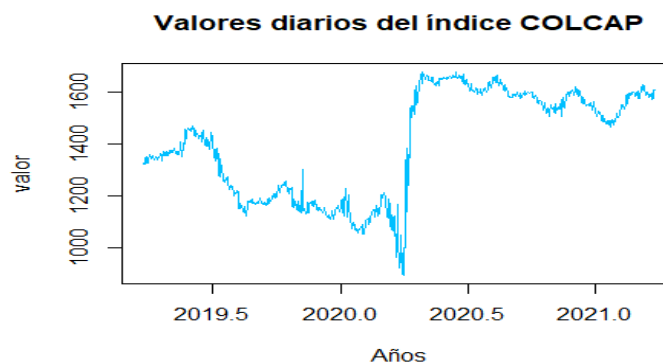


Fig. 21. Gráfico de serie de tiempo del índice COLCAP en R-Studio

Al tener la información, y antes de realizar cualquier EDA (análisis exploratorio de datos), se deben comprender los tres componentes de una serie temporal:

- *Tendencia*: Un aumento o disminución a largo plazo de los datos se denomina tendencia. No tiene por qué ser necesariamente lineal. Es el patrón subyacente en los datos a lo largo del tiempo.
- *Estacional o Periódico*: Cuando una serie está influenciada por factores estacionales, es decir, un trimestre del año, mes o días de una semana, la estacionalidad existe en la serie y siempre es de un período fijo y conocido. Por ejemplo: un aumento repentino en las ventas durante el verano.
- *Cíclico*: Cuando los datos muestran subidas y caídas que no son del período fijo.

En R-Studio se pueden obtener los componentes de la serie temporal como se muestra en la Fig. 22, por lo tanto, se procedió a verificar que la serie sea estacionaria, si no es el caso, entonces no se podrá pronosticar más adelante y se debe proceder de manera diferente.

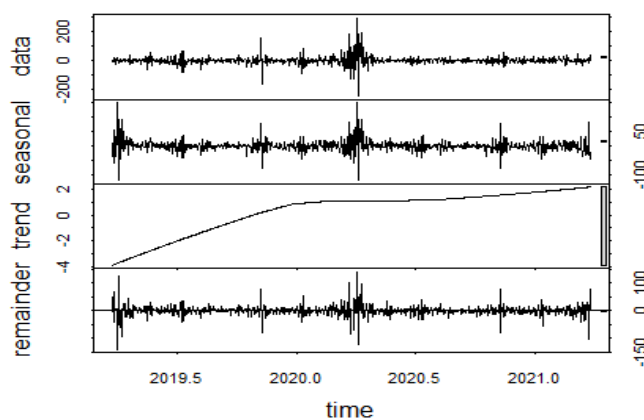


Fig. 22. Componentes de la serie temporal del índice COLCAP

A través de la prueba de hipótesis se planteó que:

Ho: La serie de tiempo no es estacionaria

Ha: La serie de tiempo es estacionaria

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de significancia del 0.05)

Decisión: Como el $P\text{-valor} = 0.3717 > \alpha = 0.05$ **No rechazo Ho**

Conclusión: Con un nivel de significancia del 0.05, se puede afirmar que en este caso la serie de tiempo no es estacionaria.

Teniendo en cuenta lo anterior, se debe buscar la estacionariedad, para facilitar el análisis.

Para conseguir estacionariedad se debe:

- Diferencia los datos – se calculan las diferencias entre observaciones consecutivas.
- Aplica el logaritmo o raíz cuadrada a los datos de la serie para estabilizar la varianza no constante.
- Si los datos contienen una tendencia, ajusta algún tipo de curva a los datos y luego modela los residuales de ese ajuste.
- Prueba de raíz unitaria: esta prueba se usa para descubrir la primera diferencia o regresión que se debe usar en los datos de tendencias para hacerla estacionaria.

Para proceder en el modelo de manera correcta se realiza una diferenciación entre los valores del precio, para lograr la estacionariedad requerida. Las veces que se debe diferenciar son definidas por R-Studio, el cual retorna, solo una diferenciación, y así se obtiene la Fig. 23.

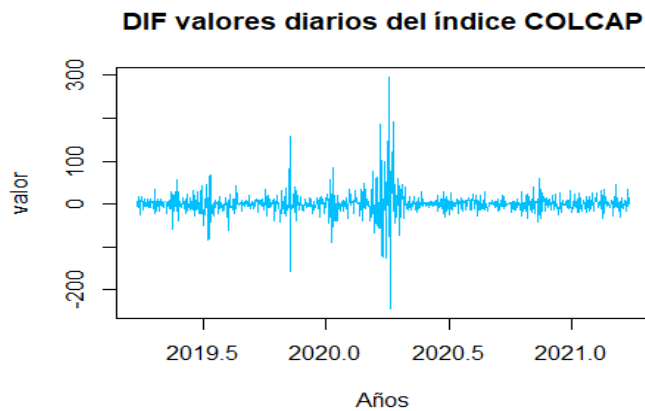


Fig. 23. Gráfico de diferencia en diferencia del valor del índice COLCAP en R-Studio

Análisis Exploratorio de Datos (EDA)

Se inicia haciendo un ‘lag plot’ (gráfica de rezagos) que se usa para evaluar cuando los valores de un conjunto de datos o una serie de tiempo es aleatoria. Si los datos son aleatorios, el lag plot no mostrará un comportamiento identificable, y si los datos no son aleatorios, el lag plot, mostrará un comportamiento identificable [42]. Estos también, ayudan a identificar outliers.

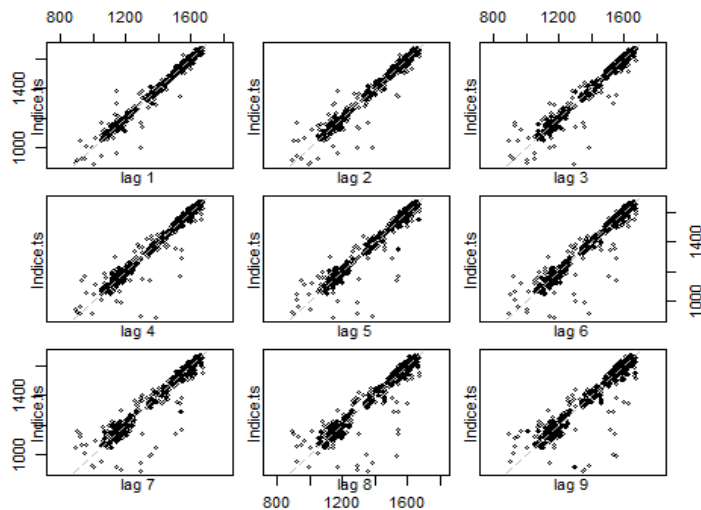


Fig. 24. Gráfico de rezagos en la serie de valores diarios del índice COLCAP.

Ahora se hace uso de un **análisis de autocorrelación** para examinar la dependencia en serie, este se utiliza para estimar qué valor en el pasado tiene una correlación con el valor actual.

Como se observa en la Fig. 25 la autocorrelación continúa disminuyendo a medida que aumenta el retraso, lo que confirma que no existe una asociación lineal entre las observaciones separadas por retrasos mayores.

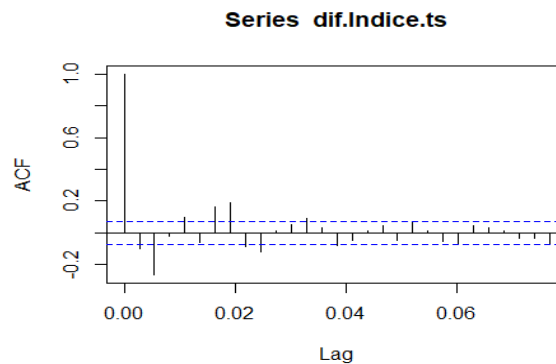


Fig. 25. Función de autocorrelación simple de una serie temporal

Así mismo se aplica la función de autocorrelación parcial en la Fig. 26.

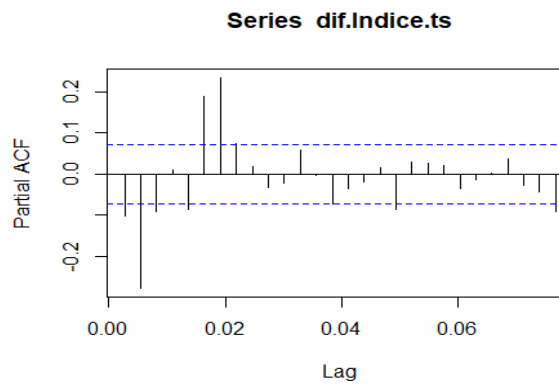


Fig. 26. Función de autocorrelación parcial

Ajuste del modelo

Una vez que se tienen los datos listos y se han satisfecho todos los supuestos del modelo para determinar el orden que se ajustará a los datos, se deben identificar los parámetros: p , d y q , que son enteros positivos que hacen referencia al orden de las partes medias autorregresivas, integradas y móviles del modelo, respectivamente como se mencionó con anterioridad [35].

Bajo el esquema Box-Jenking, para identificar qué valores de p y q serán apropiados, se ejecutan las funciones `acf()` y `pacf()`

La forma de `acf()` para definir valores de p y q es mirando los gráficos y repasando la tabla para determinar qué tipo de modelo seleccionar y cuáles serán los valores de p , d y q . En la literatura se conoce que este tipo de modelos carecen de una buena identificabilidad.

El proceso de ajuste es un proceso recursivo y se necesita ejecutar la función `arima()` con diferentes valores (p , d , q) para encontrar el modelo más optimizado y eficiente. Sin embargo, acá vamos a hacer uso del algoritmo en R-Studio.

Elección del mejor modelo

R-Studio utiliza la estimación de máxima verosimilitud (MLE) para ajustar el modelo ARIMA. Intenta maximizar la probabilidad logarítmica para valores dados de p , d y q al encontrar estimaciones de parámetros para maximizar la probabilidad de obtener los datos que hemos observado. A través del Criterio de Información de Akaike (AIC) se seleccionan los que tienen el criterio más bajo de ese valor. Prueba el criterio de información bayesiano (BIC) de Schwarz e investiga los modelos con los

valores BIC más bajos. Al estimar los parámetros del modelo utilizando la estimación de máxima verosimilitud, es posible aumentar la probabilidad agregando parámetros adicionales, lo que puede resultar en un ajuste excesivo [35].

Función `auto.arima()`: el paquete de pronóstico proporciona dos funciones: `ets()` y `auto.arima()` para la selección automática de modelos exponenciales y modelos ARIMA. Esta función en R-Studio utiliza una combinación de pruebas de raíz unitaria, minimización del AIC y MLE para obtener un modelo ARIMA.

La prueba KPSS se usa para determinar el número de diferencias (d) En el algoritmo *Hyndman-Khandakar* para el modelado ARIMA automático.

Los p , d y q luego se eligen minimizando el AICc. El algoritmo utiliza una búsqueda por pasos para recorrer el espacio del modelo, para seleccionar el mejor modelo con el AICc más pequeño.

Si $d = 0$, entonces se incluye la constante c ; si $d \geq 1$ entonces la constante c se pone a cero. Las variaciones en el modelo actual se consideran variando p y / o q del modelo actual en ± 1 e incluyendo / excluyendo c del modelo actual. El mejor modelo considerado hasta ahora (ya sea el modelo actual o una de estas variaciones) se convierte en el nuevo modelo actual.

En este caso el modelo ajustado que R-Studio retorna es un ARIMA (2,1,4).

B. Validación del diseño propuesto

Para el pronóstico, los modelos ARIMA proporcionan, además de una predicción puntual, la distribución de probabilidad completa para los futuros valores de la serie. Considerando como predicción óptima la que tiene un error cuadrático medio mínimo [34]. Dado a esto, se presenta a continuación la Tabla XVI con los errores cuadráticos medios de las cinco alternativas anteriormente mostradas añadiendo la del modelo de series de tiempo ARIMA para evaluar todas las alternativas:

TABLA XVI.
RESULTADOS DE LA COMPARACIÓN DE TODAS LAS ALTERNATIVAS DE MODELOS POR MEDIO DEL MSE

Modelo	Error Cuadrático Medio (MSE)
Media Móvil Simple	\$1.007
Media Móvil Exponencial	\$ 216,74
Diff-Diff	\$37.806
Regresión múltiple con 2 niveles	\$4.257,37
Regresión múltiple con 3 niveles	\$4.214,20
Serie de tiempo ARIMA	\$807,79

Después de evaluar y analizar todas las alternativas de modelo que se adecuan para el desarrollo del proyecto, se obtuvo el resultado del error cuadrático medio (MSE) de cada uno, donde se puede evidenciar que los modelos que mejor se ajustan al proyecto son el de series de tiempo ARIMA y el Media Móvil Exponencial.

Sin embargo, el modelo de Media Móvil Exponencial es un cálculo del precio medio o ponderado durante un período de tiempo específico que da más peso a los datos de precios más recientes [31], por lo tanto, para evaluar el impacto económico en la Bolsa de Valores de Colombia, queremos conocer de forma clara cuál es el comportamiento autorregresivo que se presenta, debido a que este modelo ajusta los datos día a día, no es conveniente teniendo en cuenta que no considera el pasado. Para evaluar este tipo de impacto y poder detectar de manera adecuada los cambios que se presentan, como se mencionó anteriormente es importante tener en cuenta el componente autorregresivo que permite conocer el pasado de una serie de datos y también el

promedio móvil, que es la parte que me indica la aleatoriedad e incertidumbre, por lo cual podemos identificar que el modelo que mejor se ajusta con el objetivo del proyecto es el de series de tiempo ARIMA.

A partir de lo anterior, se plantean supuestos para evaluar el modelo ARIMA. Para que un modelo sea considerado adecuado, es necesario que sus residuos estén normalmente distribuidos, es decir que tengan un ajuste normal e independientemente distribuido con media 0 y varianza mínima, por lo anterior se realiza la verificación:

Media cero de los residuos:

Ho: La media de los residuos es cero

Ha: La media de los residuos no es cero

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de significancia de 0.05)

Decisión: Como el $P\text{-valor} = 0.6855 > \alpha = 0.05$ **No rechazo Ho**

Conclusión: Con un nivel de significancia de 0.05, se puede afirmar que se cumple con el supuesto y en este caso la media de los residuos es igual a cero.

Independencia de los residuos:

Ho: Los residuos son independientes

Ha: Los residuos no son independientes

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de significancia de 0.05)

Decisión: Como el $P\text{-valor} = 0.9902458 > \alpha = 0.05$ **No rechazo Ho**

Conclusión: Con un nivel de significancia de 0.05, se puede afirmar que los residuos que son independientes, por consecuencia no existe dependencia entre los residuos del modelo presentado.

Distribución normal:

A partir de la Fig. 27 se puede observar que los datos presentados no están sobre la línea de tendencia, pero como los gráficos son subjetivos, se debe validar el supuesto con una prueba de hipótesis.

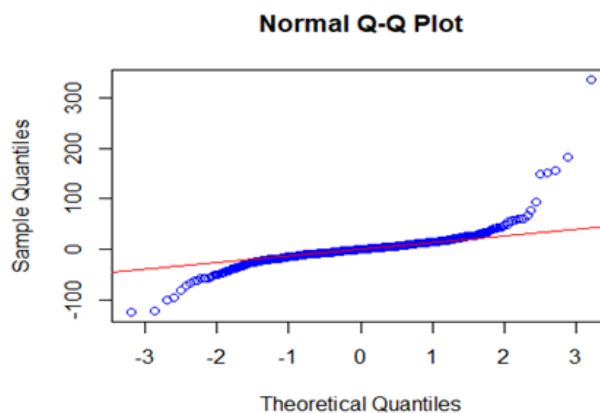


Fig. 27. Prueba para validar el supuesto de distribución normal

Al realizar la prueba *Shapiro–Wilk*, se plantea que:

Ho: Los residuos presentan una distribución normal

Ha: Los residuos presenta una distribución no normal

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de significancia de 0.05)

Decisión: Como el P-valor = $2.2e-16 < \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de significancia de 0.05, se puede afirmar que los datos no se ajustan a un comportamiento normal.

A través del uso del paquete "fitdistrplus" y de sus funciones en R-Studio, se realiza una prueba de bondad de ajuste y encontramos que se corrobora lo observado inicialmente en él, que la distribución de los errores en t-student, lo cual se puede visualizar de mejor forma en el ANEXO 11 (Código en R-Studio).

Es así como se concluyen los diferentes supuestos donde se debe tener en cuenta que, el ajuste de distribución normal que se buscaba no está determinado totalmente por la prueba realizada, note que al tener errores distribuidos normalmente es equivalente a tener observaciones distribuidas normalmente para cualquier modelo de serie temporal lineal, así que no es necesario asumir la normalidad de los errores.

A menudo se utiliza la máxima verosimilitud para estimar los parámetros del modelo, donde se asume normalidad y, aun así, ésta da buenos resultados incluso con datos no normales.

Según el artículo *Maximum Likelihood Estimates of Non-Gaussian ARMA Models* estudia modelos ARMA con errores distribuidos por Student-t y Laplace.: "Consideramos un algoritmo de máxima verosimilitud aproximada para estimar los parámetros de procesos autorregresivos de media móvil posiblemente no causales y no invertibles conducidos por ruido independiente idénticamente distribuido no gaussiano. La estimación de máxima verosimilitud aproximada normalizada tiene un máximo global que es consistente y eficiente. Las estimaciones y su matriz de covarianza asintótica asociada se calculan con una subrutina implementada en FORTRAN 77" [36].

Después de obtener el modelo, se analizan los datos, por medio de la función CPT.MEAN del paquete CHANGE.POINT los cuales calculan el posicionamiento óptimo y (potencialmente) el número de puntos de cambio para los datos utilizando el método especificado por el usuario [37]. A partir del modelo planteado se identificó que en el dato 372 se refleja un cambio en la media.

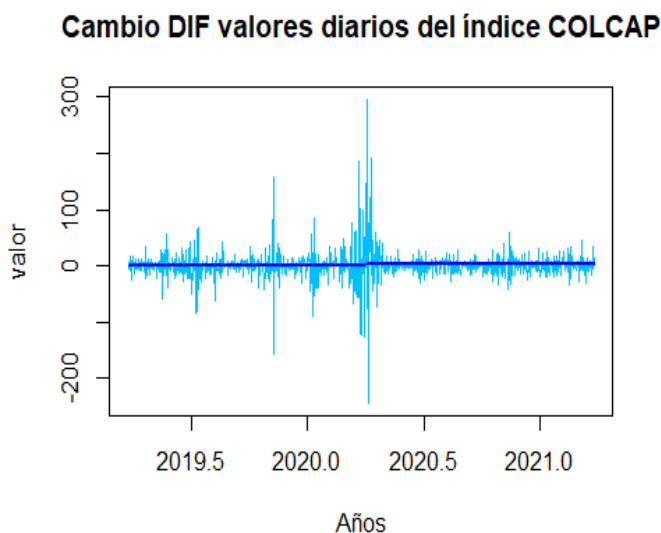


Fig. 28. Gráfica de puntos de cambio en R-Studio

Con esta función, detectamos técnicamente que existe un pequeño cambio en la serie, lo cual nos confirma el impacto, existen otros métodos (bajo otros supuestos) para detectar estos puntos de cambio, para más información ver artículo "Cambios estructurales en series de tiempo: una revisión del estado del arte" [43].

Con el fin de validar el resultado, se buscaron otras formas de evaluar la veracidad en los valores de diferencia del modelo, por ejemplo, a través de la detección de outliers, los cuales logran mostrar grandes variaciones donde se evidenció un efecto en el precio y la tendencia estudiada:

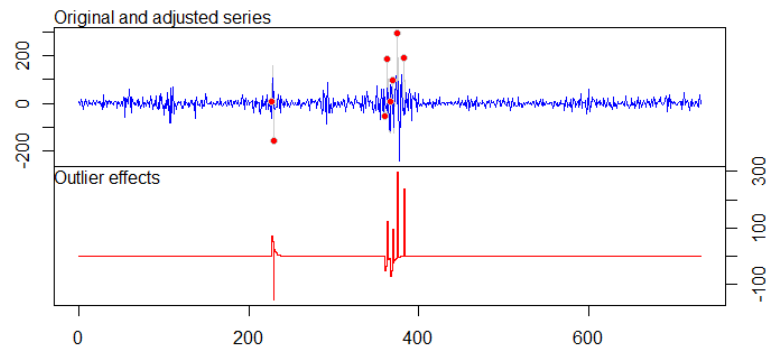


Fig. 29. Validación de los puntos de cambio

En la Fig. 29 se pueden identificar, casos que hacen referencia a valores atípicos que afectan a la serie en un momento del tiempo. Los outliers identifican a valores atípicos de causa desconocida, así que se detectaron según el efecto. En este caso tienen forma brusca y duración temporal, lo que los vuelve outliers aditivos. Y también se encuentran de forma gradual y duración temporal, lo que los vuelve un cambio temporal [38].

Outlier aditivo (AO): Es un suceso (efecto externo) que afecta a una serie en un sólo instante temporal. Gráficamente se representa como un pico aislado.

Outlier cambio temporal (TC): Es un suceso que tiene un impacto inicial y cuyo efecto decae exponencialmente de conformidad con un factor de amortiguación. Similar al outlier aditivo cuyos efectos desaparecen después de varios periodos, es decir, un pico que tarda varios periodos en desaparecer.

Para los outliers AO y TC, el efecto es transitorio [38]. De esta forma es posible validar como el modelo de series de tiempo ARIMA tiene la capacidad de detectar el impacto económico que tuvo la Bolsa de Valores de Colombia entre el periodo de marzo 2019 hasta marzo 2021.

Predecir utilizando un modelo ARIMA

Los pronósticos se muestran como una línea azul, con los intervalos de predicción del 80% como un área sombreada oscura, y los intervalos de predicción del 95% como un área sombreada clara. Este es el proceso general para analizar datos de series temporales y pronosticar valores de las series existentes utilizando ARIMA. Se procede a predecir los siguientes 12 días, desde el 26 de marzo del 2021 al 6 de abril del 2021, teniendo en cuenta que es el valor de la diferencia anteriores:

Predicción DIF valores diarios del índice COLCAP

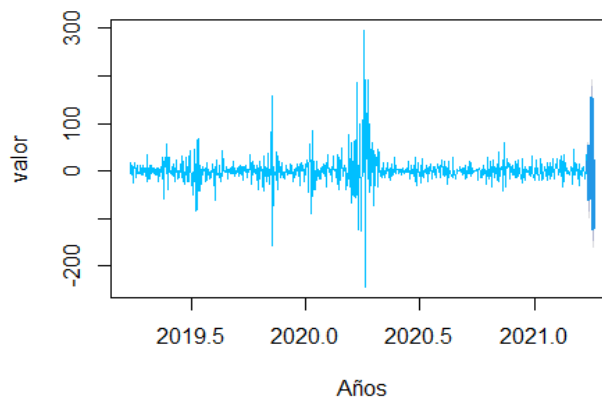


Fig. 30. Predicción del modelo para el índice COLCAP

Finalmente, para poder realizar una comparación entre el modelo generado y la base de datos real se validó a través de la predicción, teniendo en cuenta un gráfico generado a partir de la TABLA XVII, el cual muestra el ajuste y la similitud entre los valores, para así poder relacionar la aplicabilidad y validez del modelo.

TABLE XVII.
INTERVALOS DE CONFIANZA DEL ÍNDICE COLCAP EN R-STUDIO

Pronóstico índice COLCAP generado por R-Studio							Índice COLCAP real	
#	Fecha	Forecast	Lo 80	Hi 80	Lo 95	Hi 95	Valor	Diferencia
0	25/03/2021						1,320	
1	26/03/2021	1	- 25,489,538	27,365,473	- 3,947,941,437	4,135,535	1,314	- 6
2	27/03/2021	19	- 7,410,854	45,444,157	- 2,140,073,039	5,943,403	1,309	- 6
3	28/03/2021	55	28,555,035	81,410,046	1,456,515,816	9,539,992	1,303	- 6
4	29/03/2021	-61	- 87,213,057	- 34,358,045	- 10,120,293,351	- 2,036,817	1,321	18
5	30/03/2021	-21	- 47,271,591	5,583,421	- 6,126,146,727	1,957,330	1,327	6
6	31/03/2021	-1	- 27,679,330	25,175,682	- 4,166,920,660	3,916,556	1,317	- 11
7	1/04/2021	64	37,442,144	90,297,157	2,345,226,748	10,428,703	1,318	1
8	2/04/2021	3	- 23,444,404	29,410,609	- 3,743,428,071	4,340,049	1,316	- 2
9	3/04/2021	154	127,767,740	180,622,753	11,377,786,366	19,461,263	1,314	- 2
10	4/04/2021	40	13,958,962	66,813,975	-0.03091469	8,080,385	1,312	- 2
11	5/04/2021	-123	- 149,329,702	- 96,474,689	- 16,331,957,913	- 8,248,481	1,324	12
12	6/04/2021	24	- 2,667,634	50,187,379	- 1,665,751,129	6,417,726	1,324	0

En la Fig. 31 se observa como los datos pronosticados tienen fluctuaciones, sin embargo, estos se encuentran dentro de los intervalos de confianza tanto del 80% como del 95%.

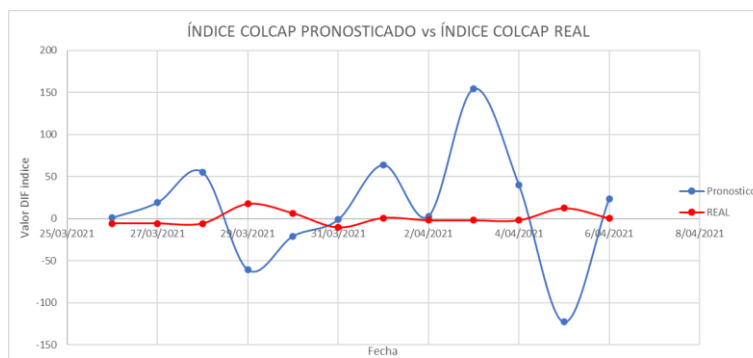


Fig. 31. Comparación entre el pronóstico y los datos reales

Por otra parte, en la Fig. 32 se presenta el pronóstico de la serie original, en donde la tendencia general de la serie se cumple y los valores del pronóstico coinciden o están muy cerca de los reales. Los datos que no coinciden están dentro del intervalo de confianza del 80% y 95%. Con esto se puede concluir que el modelo hace una buena predicción, al estar dentro de los intervalos de confianza y se considera que el modelo se ajusta, cumpliendo así con dos de los indicadores de desempeño (KPI's) que son confiabilidad del modelo, con la meta de obtener un error de máximo el 10% el cual se ha demostrado en toda esta etapa de validación teniendo en cuenta el error cuadrático medio (MSE), pronóstico, outliers y supuestos, también está la calidad de modelo con una meta de alcanzar un porcentaje de satisfacción del 70% el cual se validó por medio de una encuesta realizada a 14 personas del semillero de investigación S.E.D.E. y que se podrá visualizar en el ANEXO 12. De igual forma, con el fin de cumplir con el objetivo de mostrar una infografía sobre la influencia que tienen las variables de entrada en el índice COLCAP se realizó un dashboard, el cual puede ser visualizado por medio del siguiente enlace <https://tabsoft.co/3Ge42OH>.

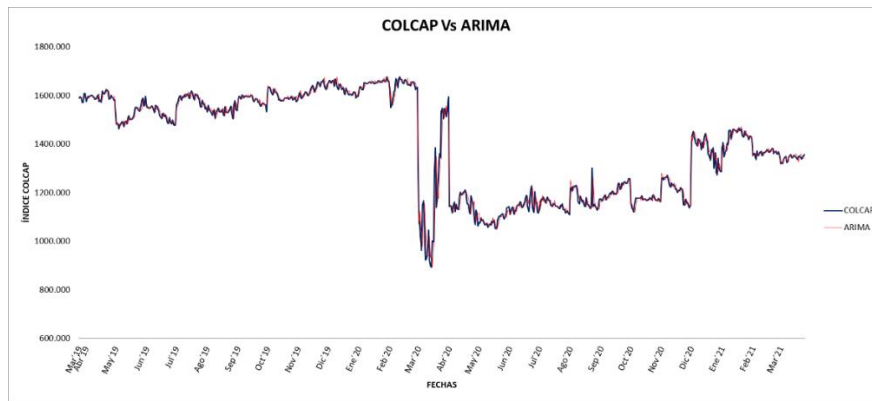


Fig. 32. Pronóstico de la serie de tiempo ARIMA

VI. VERIFICAR

A. Medición de los impactos y resultados

Para entrar a realizar la medición de impactos y resultados con el objetivo de conocer la aplicación del modelo en otro caso de estudio, se hizo uso de la información suministrada por el Indicador Mensual de Actividad Económica (IMAE), el cual ha sido diseñado por la Universidad Javeriana Cali y el Banco de la República que tiene como función ser una herramienta de apoyo al diagnóstico y la toma de decisiones de forma temprana dentro de la economía del departamento, permitiendo elaborar un pronóstico para la tasa de crecimiento del PIB regional, a través del monitoreo de las condiciones de la economía del Valle del Cauca [39]. Esto tiene como fin dar a conocer como el modelo propuesto cumple uno de los objetivos del proyecto donde se plantea que será adaptable a otros indicadores que pasen por situaciones extremas como la pandemia COVID-19.

Para la evaluación del indicador se tomó la base de datos suministrada en la página de la Javeriana Cali, donde dan una serie de datos mensuales los cuales se toman desde enero 2002 hasta junio 2021. Al tener la información de entrada en el software se generó el gráfico correspondiente para tener una mejor visualización del comportamiento del IMAE a través de los años.

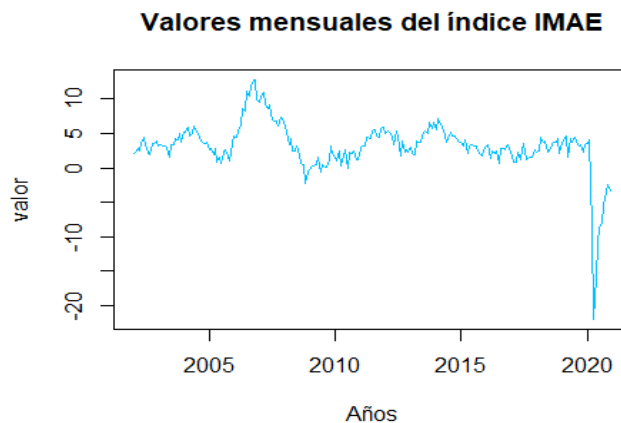


Fig. 33. Gráfico de serie de tiempo del IMAE VALLE en R-Studio

A partir de lo anterior, se plantean supuestos para evaluar el modelo ARIMA.

Media cero de los residuos:

H₀: La media de los residuos es cero

H_a: La media de los residuos no es cero

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de significancia de 0.05)

Decisión: Como el $P\text{-valor} = 0.6855 > \alpha = 0.05$ **No rechazo H_0**

Conclusión: Con un nivel de significancia de 0.05, se puede afirmar que se cumple con el supuesto y en este caso la media de los residuos es igual a cero.

Independencia de los residuos:

H_0 : Los residuos son independientes

H_a : Los residuos no son independientes

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de significancia de 0.05)

Decisión: Como el $P\text{-valor} = 0.9902458 > \alpha = 0.05$ **No rechazo H_0**

Conclusión: Con un nivel de significancia de 0.05, se puede afirmar que los residuos que son independientes, por consecuencia no existe dependencia entre los residuos del modelo presentado.

Distribución normal:

A partir de la Fig. 34 se puede observar que los datos presentados se encuentran en su mayoría sobre la línea de tendencia y solo se visualizan 3 datos atípicos, sin embargo, como los gráficos son subjetivos, se debe validar el supuesto con una prueba de hipótesis.

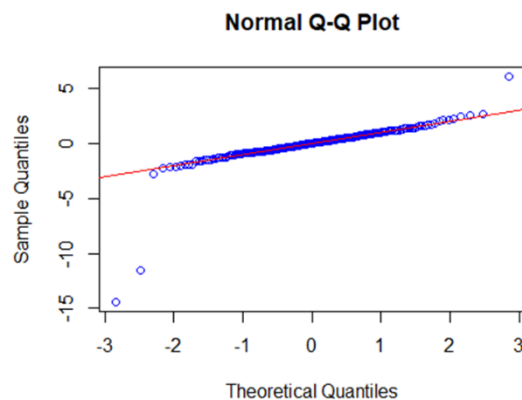


Fig. 34. Prueba para validar el supuesto de distribución normal del IMAE Valle

Al realizar la prueba *Shapiro–Wilk*, se plantea que:

H_0 : Los residuos presentan una distribución normal

H_a : Los residuos presenta una distribución no normal

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de significancia de 0.05)

Decisión: Como el $P\text{-valor} = 2.2e-16 < \alpha = 0.05$ **Rechazo H_0**

Conclusión: Con un nivel de significancia de 0.05, se puede afirmar que los datos no se ajustan a un comportamiento normal.

Al evaluar el supuesto de normalidad para el modelo ARIMA se debe recordar y tener en cuenta que el ajuste de distribución normal que se buscaba no está determinado totalmente por la prueba realizada, ya que al tener errores distribuidos normalmente

es equivalente a tener observaciones distribuidas normalmente para cualquier modelo de serie temporal lineal, por lo cual no es necesario asumir la normalidad de los errores.

Parte importante de la realización del modelo no es solo dar a conocer la eficiencia al predecir los datos, también es importante que este permita detectar los puntos de cambio en la serie, lo que permite conocer si se generó un impacto económico debido a situaciones como el COVID-19. Para esto, se analizan los datos por medio de los CPT.MEAN donde a partir del modelo se identificó que en el caso del indicador IMAE Valle el dato 219 refleja un cambio en la media y se observa en la Fig. 35.

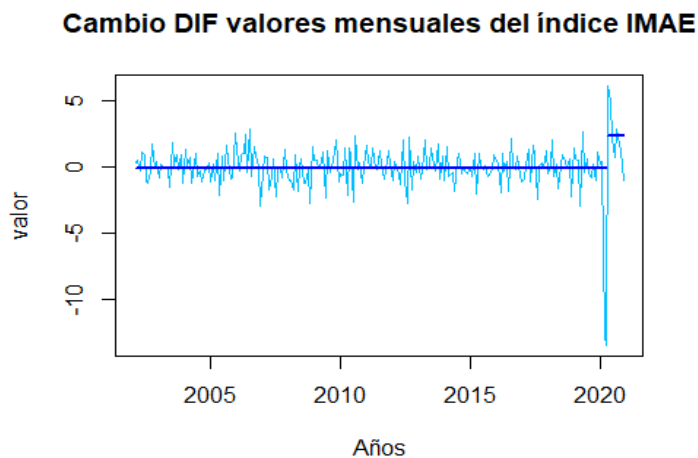


Fig. 35. Gráfica de puntos de cambio en R-Studio del IMAE Valle

Para validar el pronóstico entregado por R-Studio, y la veracidad de los valores de diferencia del modelo, se analizaron los puntos críticos de cambios (outliers) en la Fig.36.

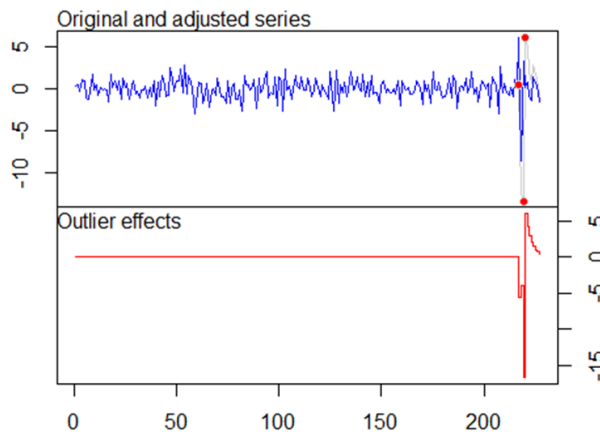


Fig. 36. Validación de los puntos de cambio del IMAE Valle

Aquí se pueden identificar casos que hacen referencia a valores atípicos que afectan a la serie en un momento del tiempo que pudieron ser generados por el COVID-19 y el paro nacional, validando como el modelo de series de tiempo ARIMA tiene la capacidad de detectar el impacto económico que tuvo el Valle del Cauca a causa de las situaciones mencionadas anteriormente.

Pronóstico del indicador IMAE Valle:

Teniendo en cuenta que el indicador que se evaluó es mensual se pasan a predecir 3 meses. Este pronóstico se muestra como una línea azul, con los intervalos de predicción del 80% como un área sombreada oscura, y los intervalos de predicción del 95% como un área sombreada clara.

Predicción DIF valores mensuales del IMAE

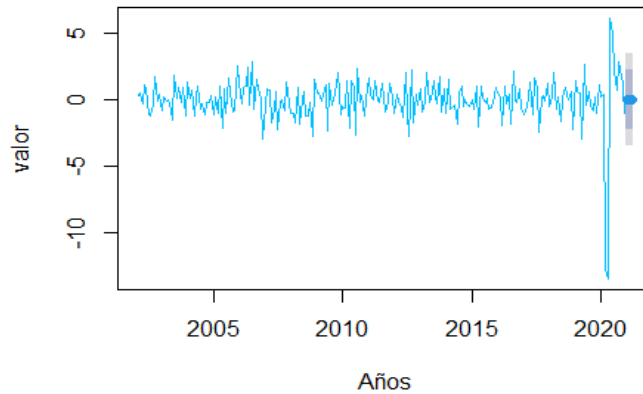


Fig. 37. Predicción del modelo para el IMAE Valle

Finalmente, por medio de la TABLA XVIII, se muestran los valores de los intervalos de confianza al 80% y al 95% que permite conocer como los datos pronosticados si se encuentran dentro de ellos, dando aplicabilidad y validez del modelo.

TABLA XVIII.
INTERVALOS DE CONFIANZA DEL IMAE VALLE EN R-STUDIO

Mes	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
ene-21	-0.02387376	-2.269253	2.221506	-3.457885	3.410138
feb-21	-0.02387376	-2.269253	2.221506	-3.457885	3.410138
mar-21	-0.02387376	-2.269253	2.221506	-3.457885	3.410138

Por otra parte, se realizó un breve análisis financiero por medio del cual se presenta información de los costos que puede generar la implementación del modelo al crear una interfaz que sea más amigable para los usuarios que deseen hacer uso del modelo si se efectúa, permitiendo suministrar flexibilidad a las necesidades de cada uno y ahorrar tiempo al momento de ingresar los datos y obtener resultados, ya que no deben entrar directamente al código en R-Studio y correrlo de forma manual.

Se debe tener en cuenta que el costo de una aplicación varía según la complejidad y las características, por lo cual entre más funciones se deseen, mayor será el costo. En este caso se analizó la implementación de un *shiny* que consiste en un paquete de R para la construcción de cuadros de mando *web* interactivos, que permite crear interfaces para algoritmos o acceder y manipular tablas de datos a través de controles de HTML: *sliders*, botones, etc. [40]

TABLA XIX.
COSTO DE IMPLEMENTAR O PROGRAMAR UN SHINY PARA R-STUDIO

Tipo de actividad	Costo	Unidades
Cursos en línea	13.99 - 84.99	Dólares
Contratación de profesionales	2.000.000 - 3.000.000	Pesos colombianos
Semilleros de investigación	Gratuito	N.A.

Como se observa en la TABLA XIX existen diferentes opciones que permiten hacer la creación de un *shiny*, como la contratación de un profesional, la realización en un semillero de investigación y el aprendizaje autónomo-colaborativo por medio de cursos.

B. Estandarización de la solución

En este apartado se presenta la estandarización del modelo que ha sido generado en el programa R-Studio. Para su realización se toma como herramienta la creación de un manual o guía de uso, que tiene como objetivo darle al lector u otra organización una base sólida para poner en práctica este modelo estadístico y para tener una mayor comprensión del proceso que se debe seguir al ingresar los datos al programa, el cual tiene por nombre *“Manual para el Desarrollo del Modelo Estadístico para Medir el Impacto Económico del COVID-19 a través de la Predicción del Índice COLCAP en R-Studio”* que podrá ser analizado de forma completa en el ANEXO 13.

En la Fig. 38 se puede observar un diagrama de flujo que permite visualizar la información planteada en el manual de forma más concreta.

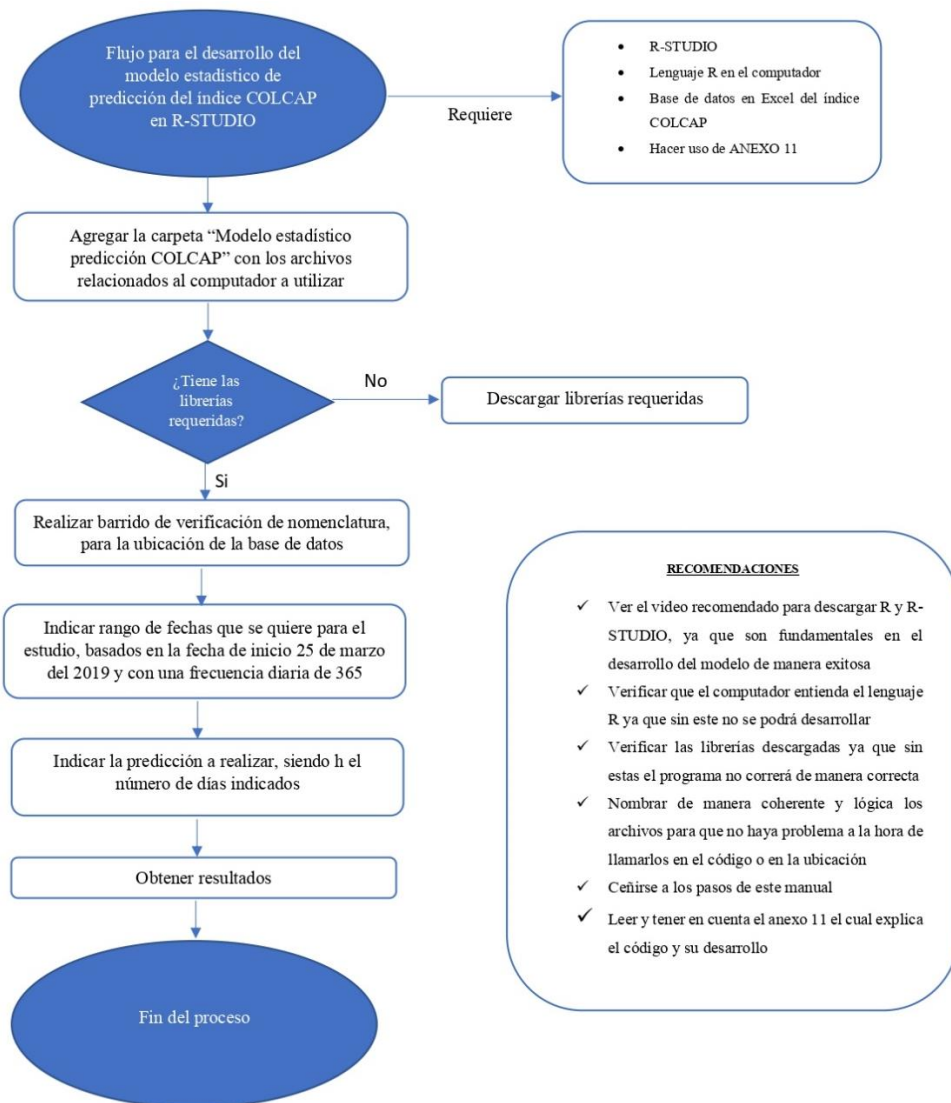


Fig. 38. Diagrama de flujo del proceso a seguir para correr el modelo en R-Studio

C. Conclusiones

- Se diseñó un modelo estadístico de series de tiempo ARIMA, que incluye como una de las variables de entrada el confinamiento ocasionado por la aparición del virus SARS-COV-2. El modelo logró predecir el comportamiento del índice COLCAP y se percibió un ajuste confiable del pronóstico.
- Se analizó la influencia de las variables de entrada haciendo uso de análisis de regresión múltiple, en el cual se plantea que el índice COLCAP está siendo explicado por la tasa representativa del mercado (TRM), el depósito a término fijo (DTF), el precio del petróleo, la tasa de desempleo en Estados Unidos, la tasa prime y la variable definida como confinamiento.
- Se evaluó el comportamiento de la base de datos con estadística descriptiva para las variables de entrada, para evaluar la correlación con la variable de salida, la distribución de cada una y la diferencia de la variable en las tres etapas de estudio (pre-confinamiento, confinamiento y post-confinamiento).
- Se mostró infografía sobre el modelo y sus variables de entrada en un dashboard o tablero de mando, lo que facilitó la forma de interactuar de los usuarios con el modelo permitiendo el cumplimiento de uno de los objetivos planteados en el proyecto.
- Se evaluó el error cuadrático medio de la variable respuesta obtenida del modelo y se obtuvo un buen ajuste del pronóstico. A pesar de que el modelo de series de tiempo ARIMA no tuviera el resultado más bajo, se estableció que para evaluar este tipo de impacto y poder detectar de manera adecuada los cambios que se presentan, es importante tener en cuenta el componente autorregresivo que permite conocer el pasado de una serie de datos y el promedio móvil que es la parte que me indica la aleatoriedad e incertidumbre, lo cual lo hace un modelo más adecuado al momento de implementarlo y cumplir los objetivos del proyecto.
- Se diseñó una encuesta dirigida a los clientes que contiene tres preguntas de satisfacción sobre temas tales como el cumplimiento de objetivos, el ajuste y la infografía presentada en el tablero de mando y el cual nos indicó una satisfacción por encima del 70%.
- Se construyó un modelo adaptable y ajustable que logró pronosticar el Indicador Mensual de Actividad Económica (IMAE) con respecto a la misma situación de confinamiento causada por el virus SARS-COV-2.

D. Recomendaciones

- Se recomienda para futuras aplicaciones del modelo toda la evaluación estadística requerida para que se pueda ajustar correctamente y se obtengan resultados confiables y precisos.
- Es clave que cualquier caso que se quiera aplicar al modelo cuente con una etapa de confinamiento o cualquier situación en la que la economía se detenga, como por ejemplo el paro nacional para así poder tener la evaluación de un periodo de tiempo previo y posterior al caso.
- Aunque se cuenta con un manual con el paso a paso para la implementación del modelo, se recomienda el desarrollo de alguna herramienta o aplicativo para que el uso del modelo sea mucho más fácil e intuitivo, ideal una plataforma en donde solo se ingresen datos para evitar cualquier tipo de confusión.
- Una buena consolidación y presentación de los resultados en la que cualquier persona que no necesariamente conozca de estadística pueda entender y analizar toda la información que se muestra.

VII.GLOSARIO

- **Organización para la Cooperación y el Desarrollo Económicos (OCDE):** Es una organización internacional que tiene como misión el diseño de mejoras políticas para una mejor vida [13].
- **Índice bursátil:** En la bolsa de valores, este índice representa un indicador de la evolución del mercado en función del comportamiento de las cotizaciones. Lo que permite la constitución de una base para identificar la percepción del mercado frente al comportamiento de las empresas y la economía [14].
- **Criterio de Información de Akaike (AIC):** Este criterio, definido por Akaike (1974) como “An Information Criterion”, se basa en la medida de información de Kullback-Leibler (1951), la cual permite interpretar la distancia entre dos distribuciones a partir de la log-verosimilitud de un modelo, como tal, el AIC proporciona un medio para la selección del modelo [39].
- **Criterio de información bayesiano (BIC):** El criterio BIC fue propuesto por Schwarz (1978) y ha sido uno de los métodos más populares usado para la selección de modelos. Este es un criterio de evaluación de modelos en términos de sus probabilidades posteriores [39].
- **IMAE:** Es el Indicador Mensual de Actividad Económica, que permite elaborar un pronóstico para la tasa de crecimiento del PIB regional, a través del monitoreo de las condiciones de la economía del Valle del Cauca (y de otras regiones) y dar cuenta de la situación en tiempo real, información útil para la toma de decisiones de los diferentes sectores de la economía y del gobierno regional [41].

VIII.REFERENCIAS

- [1] "¿Qué empresas cotizan en el COLCAP 2021?", Rankia, 2020. [En línea]. Disponible: <https://www.rankia.co/blog/analisis-colcap/1591934-que-empresas-cotizan-colcap-2021>. [Accedido: 03-mar-2021]
- [2] C. Tiempo, "¿Cómo va la recuperación de las acciones en la bolsa Colombia?", Portafolio.co, 2020. [En línea]. Disponible: <https://www.portafolio.co/economia/como-va-la-recuperacion-de-las-acciones-en-la-bolsa-colombia-547853>. [Accedido: 03-mar-2021]
- [3] "Decreto 1730 de 1991 - EVA - Función Pública", Funcionpublica.gov.co, 2021. [En línea]. Disponible: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=1514>. [Accedido: 07-mar-2021]
- [4] BVC, "Bolsa de Valores de Colombia, Un país todos los valores", Bvc.com.co, 2021. [En línea]. Disponible: <https://www.bvc.com.co/pps/tibco/portalbvc/Home/Mercados/enlinea/acciones#>. [Accedido: 05-mar-2021]
- [5] K. N. Sánchez, E. N. Cáceres & C. R. Rodríguez ¿Qué le aporta Colombia a la OCDE y viceversa?. [En línea]. Disponible en: <http://hdl.handle.net/20.500.12010/8235>. [Accedido: 14-mar-2021]
- [6] Reglamento de estudiantes, Consejo Directivo Pontificia Universidad Javeriana, Acuerdo N.º 567.
- [7] Congreso De La República, ley 510. 1999.
- [8] Congreso De La República, ley 45. 1990.
- [9] Congreso De La República, ley 715. 2001.
- [10] Congreso De La República, ley 819. 2003.
- [11] Congreso De La República, ley 35. 1993.
- [12] Congreso De La República, ley 1314. 2009.
- [13] OCDE, "Quiénes Somos", oecd.org. 2021 [En línea]. Disponible: <http://www.oecd.org/acerca/>. [Accedido: 19-mar-2021]
- [14] BVC, "Índices Bursátiles", Bvc.com.co. 2021. [En línea]. Disponible: <https://www.bvc.com.co/pps/tibco/portalbvc/Home/Mercados/descripciongeneral/indicesbursatiles?action=dummy>. [Accedido: 19-mar-2021]
- [15] E. S.A.S., "Ecopetrol y Bancolombia son las que más han perdido en capitalización bursátil", Diario La República, 2020. [En línea]. Disponible: <https://www.larepublica.co/finanzas/ecopetrol-y-bancolombia-son-las-que-mas-han-perdido-en-capitalizacion-bursatil-2976968>. [Accedido: 21- Apr- 2021].
- [16] Banco Mundial, "Evaluando Impactos: Midiendo el Impacto", Cega.berkeley.edu, 2009. [En línea]. Disponible: [http://cega.berkeley.edu/assets/cega_events/42/Measuring_Impacts_\(SP\).pdf](http://cega.berkeley.edu/assets/cega_events/42/Measuring_Impacts_(SP).pdf). [Accedido: 22- Apr- 2021].
- [17] "Tasa de Cambio Representativa del Mercado- TRM", Superintendencia Financiera de Colombia, 2021. [En línea]. Disponible: <https://www.superfinanciera.gov.co/inicio/informes-y-cifras/cifras/establecimientos-de-credito/informacion-periodica/diaria/tasa-de-cambio-representativa-del-mercado-trm-60819>. [Accedido: 25- May- 2021].
- [18] Banco de la República "Depósitos a término fijo - Enciclopedia | Banrepcultural", Enciclopedia.banrepcultural.org, 2021. [En línea]. Disponible: https://enciclopedia.banrepcultural.org/index.php/Dep%C3%B3sitos_a_t%C3%A9rmino_fijo. [Accedido: 25- May- 2021].
- [19] "Cómo se fija el precio del petróleo y cómo esto afecta a tu bolsillo - BBC News Mundo", BBC News Mundo, 2021. [En línea]. Disponible: <https://www.bbc.com/mundo/noticias-internacional-49745960>. [Accedido: 14- Apr- 2021].
- [20] "Tasa de desempleo (Unemployment rate) de los Estados Unidos", eFXto Forex, 2021. [En línea]. Disponible: [https://efxto.com/diccionario/tasa-de-desempleo-unemployment-rate-estados-unidos#:~:text=Tasa%20de%20desempleo%20\(Unemployment%20rate\)%20E2%80%93%20Estados%20Unidos,-%C2%BFQu%C3%A9%20es%20el&text=Es%20un%20indicador%20del%20mercado,que%20forman%20la%20fuerza%20laboral](https://efxto.com/diccionario/tasa-de-desempleo-unemployment-rate-estados-unidos#:~:text=Tasa%20de%20desempleo%20(Unemployment%20rate)%20E2%80%93%20Estados%20Unidos,-%C2%BFQu%C3%A9%20es%20el&text=Es%20un%20indicador%20del%20mercado,que%20forman%20la%20fuerza%20laboral). [Accedido: 21- Apr- 2021].
- [21] Banco de la República. "Prime", Banrep.gov.co, 2021. [En línea]. Disponible: <https://www.banrep.gov.co/es/glosario/prime>. [Accedido: 21 - Apr- 2021].
- [22] S. León Cristancho y A. Trespalcios Carrasquilla, "Factores macroeconómicos que influyen en la volatilidad del índice accionario COLCAP", Pregrado, Universidad EAFIT, 2015.
- [23] A. Triana Rodríguez y L. Lozano, "Incidencia de los fundamentales en el índice COLCAP", Pregrado, Universidad Cooperativa de Colombia, 2019.
- [24] Á. Ramos Garay, C. Escamilla Ospina, S. Peña Ochoa and V. Olmos Gómez, "Factores que afectan el precio de las principales acciones de la bolsa de valores de Colombia", Pregrado, Universidad de La Salle, 2018.
- [25] R. Rojas Rey, "Relación entre las variables macroeconómicas y las acciones del índice COLCAP", Maestría en Finanzas

Corporativas, CESA – Colegio de Estudios Superiores de Administración, 2018.

- [26] Millones Barragán, J., Nakazaki Servigón, J., Trujillo Calagua, G., & Chafloque Céspedes, R. “APLICACIÓN DEL MODELO DIFF IN DIFF PARA MEDIR EL IMPACTO ECONÓMICO DEL TERMINAL PORTUARIO REGIONAL DE LAMBAYEQUE SOBRE EL EMPLEO EN LA REGIÓN. REVISTA CIENCIA Y TECNOLOGÍA”, El Desarrollo- UJCM, pp. 06-12, diciembre 2020.
- [27] RUELAS, J., BALDOMERO, L., RÍOS, N., & GARCÍA-PUCH, B, “Investigación y desarrollo de las energías renovables en Sonora y Baja California”, Revista De Investigación Y Desarrollo, Vol2(5), pp. 1-10, 2016.
[En línea]. Disponible en
https://www.ecorfan.org/spain/researchjournals/Investigacion_y_Desarrollo/vol2num5/Revista_de_Investigacion_y_Desarrollo_V2_N5.pdf#page=28 [Accedido: 03 - Jun- 2021].
- [28] Montero Granados. R, “Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada”. Universidad de Granada. España,2016.
- [29] Sáenz, F. Medias móvil simple, exponencial y ponderada: fórmulas y ejemplos,2016. [En línea]. Disponible: <https://www.rankia.cl/blog/analisis-ipsa/2039072-medias-movil-simple-exponencial-ponderada-formulas-ejemplos> [Accedido: 03 - Jun- 2021].
- [30] MSDN. Media móvil Simple,2021. [En línea]. Disponible: [https://docs.microsoft.com/es-es/previous-versions/dd456706\(v=vs.140\)](https://docs.microsoft.com/es-es/previous-versions/dd456706(v=vs.140)) [Accedido: 03 - Jun- 2021].
- [31] MSDN. Media móvil exponencial,2021. [En línea]. Disponible: <https://msdn.microsoft.com/es-es/library/dd456643.aspx> [Accedido: 03 - Jun- 2021].
- [32] C. Tiempo, "Petróleo WTI en terreno negativo por primera vez en la historia", Portafolio.co, 2021. [En línea]. Disponible: <https://www.portafolio.co/internacional/noticias-hoy-precios-del-petroleo-20-de-abril-de-2020-crudo-hoy-540051>. [Accedido: 02- Aug- 2021].
- [33] P. González, Análisis de series temporales: Modelos ARIMA. 2009.
- [34] S. de la Fuente, Series temporales: Modelo ARIMA.
- [35] S. Chatterjee, "Time Series Analysis Using ARIMA Model In R", DataScience+, 2021. [En línea]. Disponible: <https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/>.
- [36] M. Lehr and K. Lii, "Maximum Likelihood Estimates of Non-Gaussian ARMA Models", *University of California Riverside, USA*, 1996.
- [37] R. Killick, K. Haynes, I. Eckley, P. Fearnhead y J. Lee, "Methods for Changepoint Detection", *Cran.r-project.org*, 2006. [En línea]. Disponible: <https://cran.r-project.org/web/packages/changepoint/changepoint.pdf>. [Accedido: 08- Oct- 2021].
- [38] J. López-de-Lacalle, "Detection of Outliers in Time Series", *Cran.r-project.org*, 2019. [En línea]. Disponible: <https://cran.r-project.org/web/packages/tsoutliers/tsoutliers.pdf>. [Accedido: 08- Oct- 2021].
- [39] L. Amaya Jiménez, "Criterio de Akaike para la selección de modelos con transformaciones", *Repository.usta.edu.co*, 2018. [En línea]. Disponible: <https://repository.usta.edu.co/handle/11634/12544>. [Accedido: 12- Nov- 2021].
- [40] C. Bellosta, "R para profesionales de los datos: una introducción", *Datanalytics.com*, 2018. [En línea]. Disponible: https://www.datanalytics.com/libro_r/shiny.html. [Accedido: 15- Nov- 2021].
- [41] Pontificia Universidad Javeriana Cali, "¿Qué es el IMAE? | Pontificia Universidad Javeriana, Cali", *Www2.javerianacali.edu.co*, 2021. [En línea]. Disponible: <https://www2.javerianacali.edu.co/facultades/ciencias-economicas-y-administrativas/imae/que-es-el-imae#gsc.tab=0>. [Accedido: 15- Nov- 2021].
- [42] M. O'Hara-Wild, R. Hyndman, E. Wang, D. Cook, T. Talagala and L. Chhay, "Feature Extraction and Statistics for Time Series", 2021.
- [43] P. Sánchez, Cambios estructurales en series de tiempo: una revisión del estado del arte. 2008.

IX.ANEXOS

TABLA XX.
TABLA DE ANEXOS

No. Anexo	Nombre	Desarrollo (propio o terceros)	Tipo de archivo (PDF, HTLM, Excel, Word...)
1	Anexo 1_PD1 2021 102	Propio	Excel
2	Anexo 2_PD1 2021 102	Propio	PDF
3	Anexo 3_PD1 2021 102	Propio	Excel
4	Anexo 4_PD1 2021 102	Propio	PDF
5	Anexo 5_PD1 2021 102	Propio	PDF
6	Anexo 6_PD1 2021 102	Propio	Excel
7	Anexo 7_PD1 2021 102	Propio	Project
8	Anexo 8_PD1 2021 102	Propio	PDF
9	Anexo 9_PD1 2021 102	Propio	PDF
10	Anexo 10_PD1 2021 102	Propio	PDF
11	Anexo 11_PD2 2021 102	Propio	PDF
12	Anexo 12_PD2 2021 102	Propio	PDF
13	Anexo 13_PD2 2021 102	Propio	PDF