



Universidad del Desarrollo
Facultad de Ingeniería

PRONÓSTICO DEL PRECIO PROMEDIO PARA EL POLIESTIRENO DE ALTO IMPACTO

FERNANDO ANDRÉS PAVEZ STUARDO

PROFESOR(ES) GUÍA: MAURICIO VARAS VALDÉS, PhD
HÉCTOR VALDÉS GONZÁLEZ, PhD

PROYECTO DE GRADO PRESENTADO A LA FACULTAD DE INGENIERÍA DE LA
UNIVERSIDAD DEL DESARROLLO PARA OPTAR AL GRADO ACADÉMICO DE MAGÍSTER EN
INGENIERÍA INDUSTRIAL Y DE SISTEMAS

SANTIAGO – CHILE
2021



Universidad del Desarrollo
Facultad de Ingeniería

PRONÓSTICO DEL PRECIO PROMEDIO PARA EL POLIESTIRENO DE ALTO IMPACTO

POR: FERNANDO ANDRÉS PAVEZ STUARDO

Proyecto de Grado presentado a la Comisión integrada por los profesores:

PROFESORES GUIA: Mauricio Varas Valdés, PhD y Héctor Valdés González,
PhD

PROFESOR INTEGRANTE 1: Cristian Candia

PROFESOR INTEGRANTE 2: José Luis Salazar Navarrete

Para completar las exigencias del Grado de Magíster en Ingeniería Industrial y de
Sistemas.

Noviembre, 2021

Santiago, Chile

DECLARACIÓN DE ORIGINALIDAD

Por medio de la presente, declaro que el trabajo titulado: **PRONÓSTICO DEL PRECIO PROMEDIO PARA EL POLIESTIRENO DE ALTO IMPACTO**, que presento a la Universidad del Desarrollo de Chile, es de mi autoría (o co-autoría) y no ha sido publicado previamente, ni está siendo considerado para publicación bajo otra filiación. En igual sentido, declaro que el trabajo de tesis y su contenido, son originales y que todos los datos y referencias a trabajos ya publicados con anterioridad han sido debidamente identificados, referenciados o citados en el documento, y que estas citas han sido incluidas en las referencias bibliográficas. Afirmo, asimismo, que los materiales presentados no se encuentran protegidos por derechos de autor; y en caso de que así lo estuvieran, me hago responsable de cualquier litigio o reclamo relacionado con la violación de derechos de propiedad intelectual, exonerando de toda responsabilidad a la Universidad del Desarrollo de Chile.

Finalmente, me comprometo a no someter este trabajo (o parte de este), a consideración en ninguna revista o congreso para publicación sin contar con la aprobación y haber pasado el debido proceso de revisión en Universidad del Desarrollo. En caso de que un artículo sea aprobado para su publicación, autorizo a la Universidad del Desarrollo a incluir dicho artículo en sus revistas, y a reproducirlo, editarlo, distribuirlo, exhibirlo y comunicarlo en el país y en el extranjero, por medios impresos, electrónicos, Internet o cualquier otro medio, para propósitos científicos y sin fines de lucro.



FERNANDO ANDRÉS PAVEZ STUARDO

Dedicatoria

*Dedico este trabajo a mis padres que
con su ayuda y comprensión en mis primeros años de estudio
me han impulsado a seguir esforzándome y creciendo profesionalmente,
y a Marcela con quien quiero seguir recorriendo el mundo.*

AGRADECIMIENTOS

Este trabajo es fruto de incontables horas de estudio, muchas de análisis, unas cuantas más de escritura y me siento honrado de poder agradecer, primeramente, a Dios por la fuerza que me ha dado en los momentos de debilidad y agobio, en segundo lugar, a nuestros profesores, en especial a Gustavo Contesse, Héctor Valdés y Mauricio Varas que sin su entrega no hubiese sido posible terminar este proyecto. Y a todos los compañeros que con sus intervenciones han complementado el proceso de aprendizaje.

En segunda instancia agradezco a Coexpan que ha entregado la información que sustenta la base de este estudio, a su gerente general y a todos los trabajadores que de una u otra forma hacemos que esta empresa siga creciendo.

A Raúl, Cristian, Paola y Jessica, mis compañeros más cercanos con quienes he pasado muchas horas resolviendo problemas, trabajando juntos para hacer que Coexpan sea un lugar mucho más agradable para nuestros colaboradores. Agradezco su comprensión y compromiso para conmigo y con todo el equipo que hay detrás.

A mi hermana y mis queridas sobrinas que me regalan sus sonrisas cuando más lo necesito. A mis padres que con su apoyo desde pequeño me han impulsado a salir adelante.

Finalmente, agradezco el apoyo de mi esposa, sin él, éste y otros proyectos no hubiesen visto nunca la luz.

PRONÓSTICO DEL PRECIO PROMEDIO PARA EL POLIESTIRENO DE ALTO IMPACTO

Fernando Andrés Pavez Stuardo

Bajo la supervisión de los Profesores Mauricio Varas Valdés, PhD y Héctor Valdés González, PhD, en la Universidad del Desarrollo de Chile

Resumen

Chile al no ser un productor de resinas plásticas, depende totalmente de las importaciones. Las grandes empresas pueden importar directamente del productor y las empresas más pequeñas sólo tienen la disponibilidad de compra a través de un trader con precios sobre la media del mercado internacional.

Este trabajo presenta una herramienta que permite organizar, presentar y describir un conjunto de datos cuya finalidad sea el pronóstico de precios del poliestireno de alto impacto. El objetivo de esta investigación es proponer un modelo estadístico que oriente al usuario en la compra del poliestireno de alto impacto y estimar el precio de compra para el siguiente ciclo. Para lograrlo se propone una aproximación cuantitativa, basada en la identificación de variables que intervienen en la estructura de costo de la materia prima y la elección de un modelo estadístico capaz de predecir el precio de ésta. El conjunto de modelos propuestos son ARIMA, Holt Winters, Regresión lineal y Redes neuronales. Los datos muestran que al comparar las predicciones efectuadas con el modelo y los precios históricos se obtiene una confiabilidad del 95%. Esto permite mejorar la toma de decisiones en el proceso de compra, disminuyendo de este modo el costo de la materia prima y con ello la optimización de los recursos financieros de la empresa.

Palabras claves: Métodos cuantitativos; Pronósticos; Series de tiempo; Regresión lineal múltiple; Estadística descriptiva.

HIGHLIGHTS

PRONÓSTICO DEL PRECIO PROMEDIO PARA EL POLIESTIRENO DE ALTO IMPACTO

FERNANDO ANDRÉS PAVEZ STUARDO

- Proponer un modelo explicativo de precios para la compra de poliestireno.
- Identificar las variables del proceso su dependencia e influencia en el precio.
- Minimiza el error de la predicción con datos de entrenamiento y validación.
- Encontrar un modelo que se ajuste al 95% confianza a la serie de tiempo.
- Un modelo econométrico puede proporcionar precios estimados para el poliestireno.

ÍNDICE GENERAL

Tabla de contenido

1	INTRODUCCIÓN	10
1.1	LA ECONOMETRÍA Y SU APLICACIÓN EN LOS PRONÓSTICOS DE PRECIOS	11
1.2	BREVE DISCUSIÓN DE LA LITERATURA	11
1.3	CONTRIBUCIÓN DEL TRABAJO	18
1.4	OBJETIVO GENERAL.....	19
1.4.1	<i>Objetivos específicos</i>	19
1.5	PROPUESTA METODOLÓGICA.....	19
1.6	ORGANIZACIÓN Y PRESENTACIÓN DE ESTE TRABAJO	21
2	INFORMACIÓN Y RESULTADOS	22
2.1	PROCEDIMIENTO DE RECOGIDA Y ANÁLISIS DE DATOS	22
2.2	CARGA DE PAQUETES DE R.....	24
2.3	TRATAMIENTO DE DATOS	24
2.4	ANÁLISIS E INTERPRETACIÓN DE LOS DATOS	25
2.5	ANÁLISIS BIVARIADO.....	26
2.6	ANÁLISIS MULTIVARIADO.....	28
2.7	REGRESIÓN LINEAL MÚLTIPLE.....	28
2.8	VALIDACIÓN DE RESULTADOS.....	29
2.9	VALIDACIÓN DE RESULTADOS MODELO CON VARIABLES DUMMY.....	33
2.10	ITERACIONES DEL BANCO DE MODELOS CONMUTADO MODELO ARIMA.....	35
2.11	DISCUSIÓN DE RESULTADOS	38
3	ARTÍCULO	41
4	CONCLUSIONES GENERALES.....	60
4.1	PROPUESTA PARA TRABAJOS FUTUROS	62
5	REFERENCIAS BIBLIOGRAFICAS	63
6	ANEXO: REPORTE DE PLAGIO.....	67

ÍNDICE DE FIGURAS Y TABLAS

FIGURA 1: COMPOSICIÓN DEL POLIESTIRENO DE ALTO IMPACTO	12
FIGURA 2: MODELO CONMUTADO	20
FIGURA 3: SERIES DE TIEMPO DE CADA VARIABLE	26
FIGURA 4: ANÁLISIS BIVARIADO DE LAS VARIABLES EN ESTUDIO (BD2)	27
FIGURA 5: PRIMERA VERSIÓN DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE.....	29
FIGURA 6: SEGUNDA VERSIÓN DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE	30
FIGURA 7: GRÁFICO DE RESIDUOS	30
FIGURA 8: MODELO REGRESIÓN LÍNEA MÚLTIPLE CON VARIABLES DUMMY.....	33
FIGURA 9: GRÁFICO COMPARATIVO PREDICCIÓN V/S VALOR REAL	35
FIGURA 10: AUTOCORRELACIÓN Y PRUEBA BOX-PIERCE DE LAS VARIABLES	37
TABLA 1: ANÁLISIS EXPLORATORIO DE VARIABLES, RSTUDIO	25
TABLA 2: PRUEBA CORRELACIÓN DE PEARSON	27
TABLA 3: EVENTOS TEMPORALES QUE EXPLICAN DATOS ATÍPICOS.....	31
TABLA 4: VALIDACIÓN DE MODELO CON VARIABLES FICTICIAS	34
TABLA 5 RESULTADO DE LA PRUEBA AUMENTADA DICKEY-FULLER.....	37

1 INTRODUCCIÓN

Históricamente las empresas han intentado lograr ventajas competitivas que les concedan mejorar su posición en el mercado, en estos últimos años el avance tecnológico ha permitido que el foco de atención sea la información con la cual generar conocimiento. En esta búsqueda la predicción de precios ha sido ampliamente estudiada por la industria en general.

En la actualidad existe una gran variedad de estudios enfocados al pronóstico de variables macroeconómicas, y son muchos los modelos que se utilizan para ello, en las técnicas actuales es común ver desarrollos en Data Science, Deep Learning y Machine Learning para extraer conocimiento de los datos en sus diferentes formas, sean estructurados o no estructurados, empleando técnicas dentro el contexto de las matemáticas, estadística e informática obteniendo con ello rapidez y exactitud en el análisis.

La tecnología computacional de hoy en día permite reducir los tiempos de cálculo haciendo más factible la incorporación de modelos propios de la minería de datos en la predicción de variables estocásticas en tiempo real, pudiendo absorber la variabilidad producida por los acontecimientos en la economía mundial y el impacto que tiene en la subida o bajada del precio de un bien.

Para este trabajo se propone un banco de modelos estadísticos conmutados que a través de iteraciones sucesivas sea capaz de seleccionar el más apto para la predicción de precios con un nivel de confianza del 95%.

En Chile no existe una planta petroquímica destinada a la fabricación de poliestireno de alto impacto lo que hace aún más complejo el desarrollo del mercado del plástico. Todo el material es importado principalmente de Argentina, Brasil, Colombia, México y China. Este mercado es altamente complejo en cuanto a la fijación y predicción de precios, la alta variabilidad del petróleo y sus derivados producen variaciones en los precios internacionales y locales difíciles de prever

En la industria del plástico, el valor de la materia prima representa gran parte del costo de la operación, por lo que disponer de información con algunos meses de anticipación puede mejorar la toma de decisiones sobre compras futuras y/o la elaboración de contratos en base a variables que intervienen en el proceso de fabricación de la materia prima. Las empresas

productoras de lámina plástica en Chile están en constante estado de alerta ante las subidas y bajadas del precio del petróleo dificultando aún más la gestión de compras.

El modelo propuesto es capaz de proyectar un valor futuro para el precio del poliestireno de alto impacto a partir de información propia y de libre acceso en la industria del plástico, tomando en cuenta sucesos aleatorios en las observaciones e inferencias de las variables del entorno, con esto es factible pronosticar el precio y decidir la conveniencia de una compra spot o en su defecto adquirir contratos teniendo como base de cálculo alguna de las variables del mercado.

1.1 La econometría y su aplicación en los pronósticos de precios

Entendida la problemática, es posible efectuar el siguiente cuestionamiento: ¿cómo el uso de los modelos econométricos aplicados en las series de tiempo de los precios del petróleo y sus derivados puede generar modelos explicativos capaces de pronosticar los precios futuros del poliestireno de alto impacto en Chile?

En efecto, disponer de la información histórica de las variables observadas permite al analista estimar valores futuros. Para el caso de estudio se propone la utilización de un banco de modelos conmutados cuyo algoritmo sea capaz de seleccionar de entre todos al más preciso, teniendo como base la historia de la serie temporal.

1.2 Breve discusión de la literatura

En la gestión de compra de cualquier materia prima, el precio siempre ha sido una variable importante, sino la más importante en la toma de decisión, más aun en el mercado del plástico donde el valor de la materia prima representa gran parte del costo del producto final. Predecir adecuadamente dicho precio, puede efectuarse con la ayuda de tecnología computacional, que permite que los tiempos de cálculo se reduzcan haciendo más factible la incorporación de modelos propios de la minería de datos.

En este sentido, la atención se centra en encontrar un modelo adecuado con el cual se pueda estimar, con una confianza del 95%, el precio de compra del poliestireno, tomando en cuenta sucesos aleatorios de observaciones, logrando modelar patrones y extraer inferencias acerca de la población de estudio (Coy Mondragón, 2021). El resultado de este análisis permitirá al

usuario disponer de información con un horizonte de algunos meses de anticipación, mejorando con ello la toma de decisión y logrando una mejor gestión de compras.

Poliestireno de alto impacto

Es un polímero termoplástico que se obtiene de la polimerización del monómero estireno más un 14% de polibutadieno (John Scheirs, 2003).

El estireno es un hidrocarburo aromático proveniente de un anillo de benceno con un sustituyente etileno. El polibutadieno es un elastómero que se obtiene de la polimerización del Butadieno. El Butadieno es un alqueno que se produce de la destilación del petróleo.

La representación de las distintas interacciones en las materias primas que conforman al poliestireno de alto impacto se puede ver en la Figura 1 siendo el petróleo el origen de la cadena productiva, pasando por los distintos componentes y sus transformaciones hasta llegar al producto en estudio.

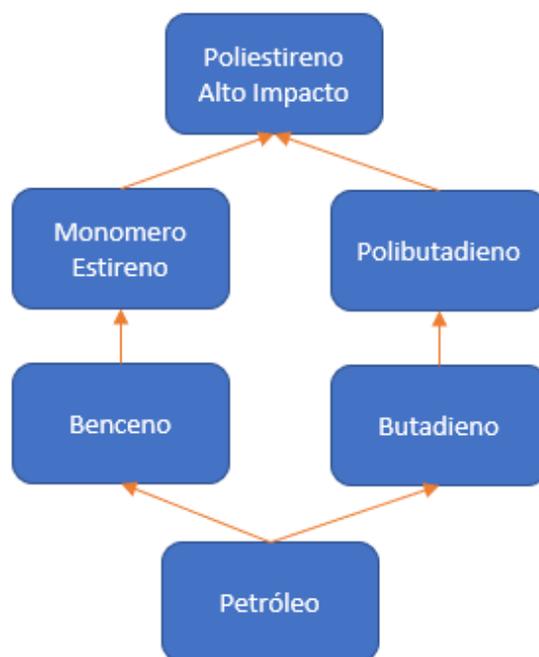


Figura 1: Composición del poliestireno de alto impacto
(Fuente: Elaboración propia)

Alternativas en el ecodiseño para el reemplazo del poliestireno

Desarrollos del ecodiseño en la industria del embalaje han estado impulsando la búsqueda de nuevas alternativas de resinas más amigables con el ambiente. Una de las resinas que se encuentran disponibles en el mercado es el PLA (ácido poliláctico), que es un polímero biodegradable cuya base se obtiene a partir de almidón de maíz, yuca o caña de azúcar. En la actualidad, el PLA ha despertado gran interés por ser un posible sustituto a las resinas termoplásticas en la industria del packaging. Entre sus propiedades mecánicas se encuentran: baja resistencia al impacto, dureza, rigidez, elasticidad son similares a las del tereftalato de polietileno (PET), presenta buena resistencia a la torsión, alto módulo de flexión y alta transparencia hacen que la película plástica sea comparable a una película de celofán.

Por sus características el PLA puede ser un buen competidor de los plásticos de origen petroquímicos, y además puede ser biodegradable al 100% en medio acuoso, temperatura ambiente y un rango de ph entre 5 y 8 (Mexpolimeros, 2020). Su alto valor en comparación con las resinas de origen petroquímico aún lo hacen poco atractivo para la industria láctea. Para estos formatos, el poliestireno de alto impacto (PS) sigue siendo el más requerido.

Mercado chileno del plástico para el packaging de alimentos

El plástico, a pesar de ser un material contaminante para el medio ambiente, aún es usado en aplicaciones industriales. En la industria de alimentos, la principal aplicación es la de contenedor primario. Son muchas las características que hacen al plástico el embalaje perfecto, entre las que se incluyen: su bajo costo, la resistencia al impacto, y la capacidad de contener alimentos sin alterar sus propiedades.

Hasta hace un par de años era impensado alargar la vida de frutas, verduras, carnes, quesos y otros, en la actualidad con los nuevos aditivos es posible crear barreras de oxígeno, mantener atmósferas inertes, aditivos que absorben humedad, otros que filtran el paso de la luz UV, etc., todo esto enfocado en prolongar la vida de productos alimenticios (Industries, 2019).

La guerra ambiental contra el uso del plástico

Desde el año 2018, Chile ha mostrado preocupación medioambiental en cuanto al uso del plástico en productos de un solo uso. Ejemplo de ello es la Ley 21.100 sobre la prohibición de

bolsas plásticas para el comercio (Ley 21.100), y luego, con la incorporación de la Ley 20.920 que establece el marco para la gestión de residuos, la responsabilidad extendida del productor y fomento al reciclaje (LEY 20920). Estas claras señales han hecho que el mercado del packaging esté dando un vuelco en el uso de ciertas resinas, tal es el caso del poliestireno que está dejando de ser el material preferido para las nuevas soluciones de empaque por no contar con un mercado de reciclaje claro y definido.

La industria del plástico en Chile aún está en período de ajustes, no existe claridad sobre el uso de resinas recicladas salvo en el PET que tiene una economía circular. A pesar de todas estas controversias Chile sigue siendo un consumidor bastante pequeño en comparación con otros mercados latinoamericanos y las decisiones que se tomen al interior del país acerca de la resina predominante no afectarán el precio de mercado de la resina en estudio.

El precio del petróleo y su efecto en el precio de la resina

Existe correlación entre el petróleo y las distintas resinas plásticas derivadas de él, las fluctuaciones hacen que los precios del mercado sean impactados. Estudios de Duncan Seddon & Associates (Seddon, 2013) confirman esta tesis obteniendo un coeficiente de correlación (R^2) de 0.81 en el período comprendido entre 2000 y 2013.

Los últimos datos de American Chemistry Council muestran un mayor impulso en junio 2021 y se pronostica un segundo semestre con una tendencia creciente que impacta directamente en el precio de las materias primas y petróleo. La escasez del suministro en el mercado europeo de acrilonitrilo-butadieno-estireno sin duda provoca una desestabilización en los precios del cono sur debido a las importaciones y la creciente demanda de este tipo de productos.

El valor de los petroquímicos es altamente influenciado por el precio del petróleo. La caída o aumento de los precios dependerá del nivel de stock en los mercados. Su regulación depende en gran medida de las variables exógenas, ejemplo de esto se encuentra en la tormenta que afectó al norte de América en marzo 2021 cuya consecuencia fue la escasez de monómero estireno y en el segundo trimestre la escasez del polibutadieno, materias primas vitales para la fabricación del poliestireno de alto impacto.

Repositorios de precios

El poder de la información y el valor que ésta tiene en el mercado ha sido el impulsor para que algunas empresas estén dispuestas a invertir y otras a pagar por información concentrada en un solo lugar. PLATTS (Platts, 2021), ICIS (ICIS, 2021) e INVESTING (Investing, 2021) son referentes en cuanto a calidad de información y en algunos mercados sirven como referencia para la elaboración de contratos, monitoreando precios de productos básicos, consiguiendo así transparencia en la transacción. Se publican precios spots (precio contado para entrega inmediata) de diversas materias primas, curvas forward que pueden ser consultadas sobre una base coherente y fidedigna de información. Es importante destacar que los precios que sustentan la base de estudio se obtienen de estos repositorios.

Chile y su dependencia en los mercados

Chile tiene una economía abierta con una alta dependencia del comercio internacional. Sus principales exportaciones son materias primas sin mucha elaboración, como, por ejemplo: cobre, filetes de pescado, pulpa de celulosa, vino y frutas. Las principales importaciones, por otra parte, son: petróleo, gas de petróleo, aparatos eléctricos y vehículos a motor. La exportación de bienes y servicios bajó en 3,5% en 2020 y se prevé que suba en 0,5% en 2021, mientras que las importaciones bajaron en 11,7% en 2020 y se prevé que crezcan en 10,4%. Estas bajas se explican principalmente por la crisis económica mundial producto de la pandemia por COVID 19 (Entreprises, 2021).

En el caso del plástico, las importaciones chilenas provienen principalmente de China, Estados Unidos, México, Brasil, Colombia y Argentina. Las empresas se ven forzadas a adquirir estos productos tanto en el mercado local o a través de importaciones directas. Los precios a los que se acceden suelen estar sujetos a bandas establecidas por contrato, compras spot, o una mezcla de ellos.

Mecanismos para la estabilización de precios sobre bandas referenciales

Un mercado convulsionado con muchas alzas y pocas bajas en los precios ha hecho que las empresas cambien su estrategia y opten por contratos forward. Este método puede beneficiar o perjudicar la operación. Sin embargo, en mercados tan volátiles con el de las resinas plásticas es conveniente contar con una fórmula de precios que asegure el

abastecimiento y permita tanto al vendedor como al comprador mantener estabilidad con miras hacia el futuro en una relación de largo plazo.

La base para estos contratos se encuentra en las materias primas con la que se produce el poliestireno, que suele ser el polibutadieno o monómero de estireno, más una variable que puede incluir el flete u otro que convenga a ambas partes, teniendo al mercado chino como referencia.

Pronósticos de precios: Teoría y definiciones

Pronosticar es el arte y ciencia de predecir acontecimientos futuros (Cohen, 2017) Generalmente los pronósticos son utilizados para apoyar las decisiones gerenciales. Se basan en la historia de los acontecimientos pasados, para traer al tiempo presente un valor futuro de comportamiento bajo una probabilidad de ocurrencia. El pronóstico también tiene una temporalidad que puede ser afectada por sucesos que ocurren en otras variables ajenas al proceso.

Modelos del conmutador

ARIMA

Es un modelo autorregresivo integrado de promedio móvil (acrónimo del inglés autoregressive integrated moving average). ARIMA utiliza variaciones y regresión de datos estadísticos, cuyo objetivo es encontrar un valor estimado del futuro. Es un modelo dinámico de series de tiempo, las estimaciones son explicadas en base a la información del pasado y no por variables independientes como en una regresión lineal (Douglas Montgomery, 2008).

Holt Winters

Es un modelo basado en suavizamiento exponencial. Este método puede adaptarse fácilmente a los cambios y tendencias, así como también a patrones estacionales.

Regresión lineal

Es un modelo asociativo con enfoque cuantitativo que permite pronosticar una variable dependiente con una o más variables independientes a través de una ecuación lineal. Puede ser usada para estimar precios de productos o servicios, estimar demandas entre otras (Betancourt, 2021).

La regresión lineal múltiple considera linealidad en la relación, normalidad y homogeneidad de la muestra. La función (1) que describe a la regresión es.

$$y' = a + b_1x_1 + b_2x_2 + b_ix_i \quad (1)$$

Donde y es la variable endógena o dependiente, x las variables exógenas o independientes, b los coeficientes estimados del modelo y a es la intersección o constante.

Redes neuronales artificiales - RNA

Consiste en un conjunto de unidades (neuronas artificiales) interconectadas entre sí, las que comúnmente se distribuyen en capas. Un sistema típico cuenta con tres capas. En la primera se encuentran las neuronas de entrada, las cuales envían información a través del campo de neuronas, la sinapsis ocurre entre la segunda y tercera capa donde se envía a la salida como información procesada. Los sistemas más complejos tienen más capas. Estos sistemas aprenden y se forman así mismos en lugar de ser programados y sobresalen en áreas donde la detección de soluciones o características es difícil de expresar con la programación convencional, dentro de las desventajas de los RNA (Jones, 2019) es necesario ajustar hiperparámetros como el número de capas ocultas, funciones de activación, número de iteraciones, optimizador y número de neuronas por capa y todo ello ocurre en una caja negra sin mayor intervención del analista.

La importancia de un modelo cuantitativo y su aplicación a gran escala

El objetivo de este trabajo es concentrar la búsqueda de un modelo estadístico dentro de un banco de modelos conmutados con la finalidad de conocer, comprender y simular múltiples alternativas de precios para el poliestireno que ingresa a Chile. La lógica de selección está basada en la interpretación del error y en el ajuste que logre la predicción sobre la información de mercado imperante en el momento.

La era digital y la cantidad abrumadora de información que minuto a minuto se está generando en todos los sectores de la sociedad permiten hoy tomar decisiones más acertadas que hace 15 años atrás. Una desviación puede rápidamente ser corregida evitando así problemas en abastecimiento o sobre precio en compras de materias primas. El término información en tiempo real (Emilio Ontiveros, 2017) trae consigo nuevos desafíos en cuanto al análisis de este nuevo conocimiento y será función de este modelo conmutado ayudar a precisar la información y seleccionar el modelo que mejor se adapte a las circunstancias del ambiente tan convulsionado. Mucho de estos sucesos que día a día se van gestando en el mercado podrán ir corrigiendo el análisis pudiendo en cada iteración encontrar el mejor valor en cuanto al precio del poliestireno.

En la literatura revisada sobre pronóstico de precios para resinas de poliestireno no se ha encontrado un trabajo similar donde a través de modelos econométricos se pueda proyectar un precio futuro de este producto, sin embargo, Carlos Fernández de la Universidad Nacional de Colombia (Fernández Múnera, 2020) desarrolló un trabajo que aborda el análisis de precios para el polipropileno en ambientes coberturistas enfocado en la disminución del riesgo en la decisión. Si bien es cierto, los materiales y precios que se tratan en este estudio son distintos a los de esta tesis, los modelos estadísticos usados pueden ser extrapolados para este trabajo.

Finalmente, tras la revisión de las principales contribuciones que aportan o han aportado a la línea de trabajo de este proyecto, es posible indicar que una oportunidad de desarrollo se encuentra en el hecho que no existe, para el caso de Coexpan Chile S.A en la industria del plástico (empresa especialistas en la fabricación de láminas de plástico rígido y productos termoformados para brindar soluciones para la industria del embalaje), información con suficiente certeza, respecto de una herramienta que permita pronosticar el precio promedio para el poliestireno de alto impacto. Lo que autoriza la siguiente como contribución para este proyecto de grado.

1.3 Contribución del trabajo

Habiendo recorrido las bases teóricas fundamentales para este estudio, es importante mencionar que la principal motivación para realizarlo ha sido la falta de un modelo que permita explicar los precios del poliestireno de alto impacto en Chile. Ante esto, se propone un sistema conmutado de modelos capaz de iterar y encontrar uno que permita estimar el precio de compra del poliestireno en base a las realizaciones de ciertas variables estocásticas. En este sentido, la presente tesis contribuye al apoyo de la toma de decisiones en este contexto, al permitir a un analista obtener un valor estimado del precio del poliestireno con un 95% de confianza.

De acuerdo con lo mencionado anteriormente, este trabajo considera los siguientes como objetivo general y objetivos específicos para este trabajo de tesis.

1.4 Objetivo general

Proponer un modelo estadístico que oriente al usuario en la compra del poliestireno de alto impacto, para la estimación de un precio objetivo de compra basado en datos históricos y con ello optimizar el aprovisionamiento futuro de los materiales.

1.4.1 Objetivos específicos

- Estudiar la relación entre la variable dependiente e independientes.
- Identificar patrones válidos, novedosos y potencialmente útiles en las variables de entrada al sistema de modelos conmutados.
- Verificar estadísticamente que el modelo seleccionado cumple con las hipótesis para aceptarlo.

1.5 Propuesta metodológica

La metodología utilizada en este estudio comprende las siguientes cinco etapas.

1. **Análisis del problema:** en esta primera etapa se analizan las dificultades actuales para conseguir un precio estimado de compra que incorpore a las variables que afectan el precio final del bien.
2. **Revisión bibliográfica:** en esta etapa se revisa la literatura, haciendo hincapié en casos de éxito y fracaso en cuanto al uso de modelos estadístico para la predicción de precios futuros.
3. **Extracción, transformación y carga de datos KDD (Knowledge Discovery in Databases):** esta etapa tiene la finalidad de seleccionar desde un conjunto de datos, la información necesaria para ser usada en los modelos.
 - i. Análisis cuantitativo.
 - ii. Tratamiento de datos.
 - iii. Análisis descriptivo.
 - iv. Análisis bivariado.
 - v. Cálculo del modelo.

4- **Selección de modelos estadísticos:** en el proceso de iteración el banco de modelos conmutados es capaz de seleccionar al que sea capaz de entregar la mejor estimación (ver Figura 2).

En esta etapa el analista dispone de cuatro modelos (ARIMA, Holt Winters, Regresión lineal múltiple y Redes neuronales). La previa selección de modelos econométricos permite al conmutador partir desde una base predefinida, sin embargo, no existe restricción para adherir nuevos modelos y probar su funcionalidad en el caso que los existentes no sean capaces de predecir con cierta precisión los precios para el poliestireno de alto impacto. Los modelos que sirven de base para el conmutador concentran un conjunto de algoritmos capaces de extraer información valiosa de los metadatos pudiendo detectar patrones de comportamiento, subidas o bajadas de alguna variable de interés para crear un nuevo repositorio de información con la cual a través de estadística deducir relaciones entre variables y generar predicciones (Perez, 2015).

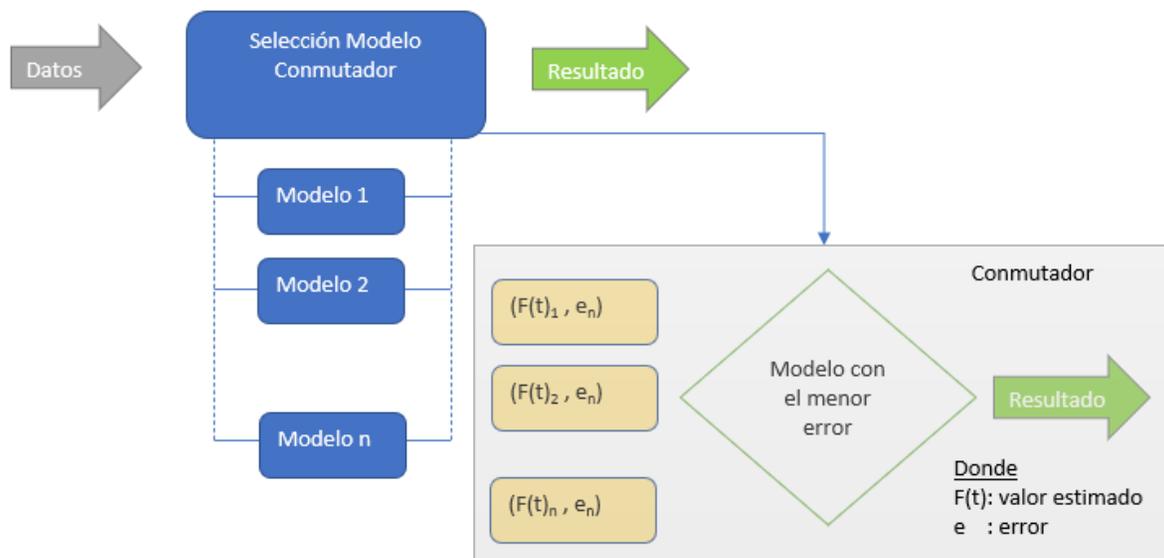


Figura 2: Modelo conmutado
Fuente: Elaboración propia)

5- **Validación:** esta etapa involucra evaluar la información estadística de los errores, analizar el p-valor, y comparar de forma gráfica el comportamiento del modelo con los datos históricos. Para todos los efectos, los análisis de significancia se realizan con un nivel de confianza del 95%, y el rechazo de la hipótesis nula (H_0) es con un p valor inferior al 5%.

1.6 Organización y presentación de este trabajo

Este proyecto de grado posee cuatro capítulos principales y se organiza como sigue:

Capítulo 1: Se exponen los principales conceptos contextualizados, desde la literatura se construye una discusión que otorga pertinencia y centra la atención en el foco de la investigación, pudiendo de esta manera visualizar una contribución clara a la luz de los antecedentes, además se propone un objetivo el cual a través de un marco metodológico permite su implementación.

Capítulo 2: Está asociado a la captación de información, propuesta de modelos y, por último, la explicación de los resultados.

Capítulo 3: El proyecto de grado, se presenta en formato resumido en un artículo académico que se estructura de la siguiente manera:

1. Título
2. Resumen
3. Introducción
4. Metodología
5. Resultados
 - a. Discusión de resultados
6. Conclusiones
7. Referencias

Capítulo 4: Finalmente, se exponen las conclusiones generales derivadas de este trabajo, y una dirección para la investigación futura, la cual considera aquellas preguntas no contestadas durante el desarrollo de este trabajo, se presentan en este capítulo.

Referencias generales

2 INFORMACIÓN Y RESULTADOS

Para abordar este trabajo de investigación se ha optado por una aproximación cuantitativa, que permite considerar la siguiente estructura para la presentación de la información y sus análisis:

2.1 Procedimiento de recogida y análisis de datos

Este trabajo emplea la información disponible en repositorios de datos para la industria del plástico, específicamente en el precio del poliestireno de alto impacto y en los compuestos que lo componen. Por tal motivo a partir del año 2020 en forma mensual se ha ido recogiendo información de precios de las materias primas vinculadas en la fabricación de esta resina para su posterior análisis. En particular se operó a través de la gestión de bases de datos, modelos y simulaciones que permitieran el análisis de esta situación.

El método utilizado en este estudio es de carácter cuantitativo que se apoya en la metodología KDD (Knowledge Discovery in Databases) cuya finalidad es el descubrimiento automático de datos a través de reglas y patrones para que el usuario pueda realizar el análisis.

Fechas en que se recogieron los datos:

La información para este proyecto de título fue extraída desde el repositorio ICIS entre el 01 de agosto de 2020 y 15 de julio de 2021, los datos anteriores a esta fecha fueron entregados por el encargado de compras de la empresa Coexpan S.A., esta información es enviada mensualmente desde la casa matriz ubicada en España.

Coherencia con lo planificado:

En una etapa inicial se esperaba contar con una mayor cantidad de información con la cual poder construir un modelo más preciso y robusto. A lo largo del camino el objetivo debió modificarse y ser acotado a la construcción de un banco de modelos conmutados con la finalidad de seleccionar el más adecuado dentro de varias alternativas. El precio del petróleo sufre variaciones diarias lo que a su vez gatilla alzas o bajas de los productos que dependen de él, pero no con la misma velocidad en términos comparativos, por ello disponer de esta información con una frecuencia

menor al precio promedio mensual aportaría mayor riqueza al análisis y, tal vez, una mayor precisión.

Fortalezas y debilidades del proyecto:

Fortalezas:

- Proyecto de rápida implementación
- Proceso extrapolable a otras situaciones
- Proceso escalable y con posibilidad de un mayor manejo de datos
- Permite dar respuesta a la pregunta de investigación

Las debilidades:

- Para generalizar resultados, la muestra debe ser mayor
- Considerar obtener información de otros repositorios
- Analizar la utilización de otros modelos econométricos apoyados en información cualitativa del entorno

Población y muestras

Además de lo planteado en el marco metodológico, en la sección de recogida de información con la que se efectuará el estudio, se hace notar que la muestra contiene sólo promedios mensuales los que reúnen los acontecimientos diarios en cuanto a los precios de las materias primas, tanto del petróleo como de sus derivados, especialmente el monómero estireno y el benceno ambos compuesto para la elaboración del poliestireno de alto impacto.

Instrumento.

Como se indicó anteriormente, en la recogida de información fue necesario la utilización de una ETL (Extract, Transform and Load) cuyo proceso permite a las empresas extraer datos desde múltiples fuentes, reordenarlos, limpiarlos y depurarlos para finalmente cargarlos en una nueva base de datos de donde en segunda instancia son analizados por intermedio del proceso KDD, para luego ejecutar

el modelamiento y posteriormente realizar la estimación en base a los modelos predefinidos.

2.2 Carga de paquetes de R

Como se ha indicado anteriormente, el proceso ETL ha permitido disponer de información depurada y ordenada en una nueva base de datos para posteriormente comenzar con los cálculos estadísticos en Rstudio, para ello se ha tenido que utilizar los siguientes paquetes de las librerías para sustentar todo el análisis.

```
library(rstudioapi)      # Área de trabajo
library(readxl)          # Abrir archivos xlsx
library(tidyverse)       # Tratamiento de datos
library(psych)           # Análisis descriptivo
library(skimr)           # Análisis exploratorio
library(car)             # Supuestos estadísticos
library(lmtest)          # Supuestos estadísticos
library(tseries)         # Supuestos estadísticos
library(sandwich)        # Errores robustos
library(forecast)        # Series de tiempo
library(corrplot)        # Gráfico correlación
```

2.3 Tratamiento de datos

Debido a que las oscilaciones entre los valores de las series en diferentes momentos de tiempo, $(x_t - x_{t-h})$ son proporcionales al nivel inicial. Para evitar esta proporcionalidad, y conseguir que la tasa de crecimiento en cada observación sea independiente de la magnitud de dicha observación, se utiliza la transformación logarítmica dando como resultado bd2.

```
bd2 <- read_excel("Data Precio RStudio24092021.xlsx")[,-1] %>%
mutate_all(funs(log(./lag(.,1)))) %>% slice(-1)
```

#Generar variables de Series de Tiempo

```
Styrene.ts<-ts(bd2$Styrene, start=c(2016,1),frequency = 12)
Benzene.ts<-ts(bd2$Benzene, start=c(2016,1),frequency = 12)
Brent.ts<-ts(bd2$Brent, start=c(2016,1),frequency = 12)
PS.ts<-ts(bd2$PS_Spot, start=c(2016,1),frequency = 12)
```

2.4 Análisis e interpretación de los datos

El banco de modelos conmutados está compuesto por cuatro modelos. El criterio de selección se basa tanto en el error como también en el ajuste de la predicción que éstos tengan con los datos disponibles.

Antes del análisis exploratorio se crean dos bases de datos. La primera contiene información es escala natural bd1 (directamente del repositorio sin modificación), mientras que la segunda se aplica una variación mensual bd2.

En la Tabla 1 se presenta el análisis exploratorio realizado a ambas bases de datos. En particular, se muestra información sobre mínimos, máximos, media, mediana, 1° cuartil y 3° cuartil, de cada una de las variables. Esta metodología proporciona una síntesis de la información contenida en la muestra con el objetivo de mejorar la precisión, sencillez, y ordenamiento de los datos, como también extraer las características más representativas de ésta, junto a ello en la Figura 3 se presenta en forma gráfica la serie temporal de las variables con el objetivo de visualizar las alzas y bajas de precios en el tiempo.

Tabla 1: Análisis exploratorio de variables, RStudio

Resumen (bd1: escala natural)

	Styrene	Benzene	Brent	Ps_Spot
Min.	451,1	285,3	19,8	828
1st Qu.	893,7	615,6	46,8	1070
Median	1011,9	736,4	55,1	1256
Mean	1043,0	727,8	55,16	1243
3rd Qu.	1243,2	849,7	63,7	1426
Max.	1881,3	1230,4	78,9	1562

Resumen (bd2: variación mensual)

	Styrene	Benzene	Brent	Ps_Spot
Min.	-0,365240	-0,661114	-0,616390	-0,154577
1st Qu.	-0,051592	-0,023493	-0,025970	0,023981
Median	0,000309	0,006196	0,030310	0,007669
Mean	0,007763	0,012009	0,011210	0,006888
3rd Qu.	0,070337	0,061253	0,675200	0,037764
Max.	0,402808	0,297185	0,467480	0,266161

(Fuente: Elaboración propia)

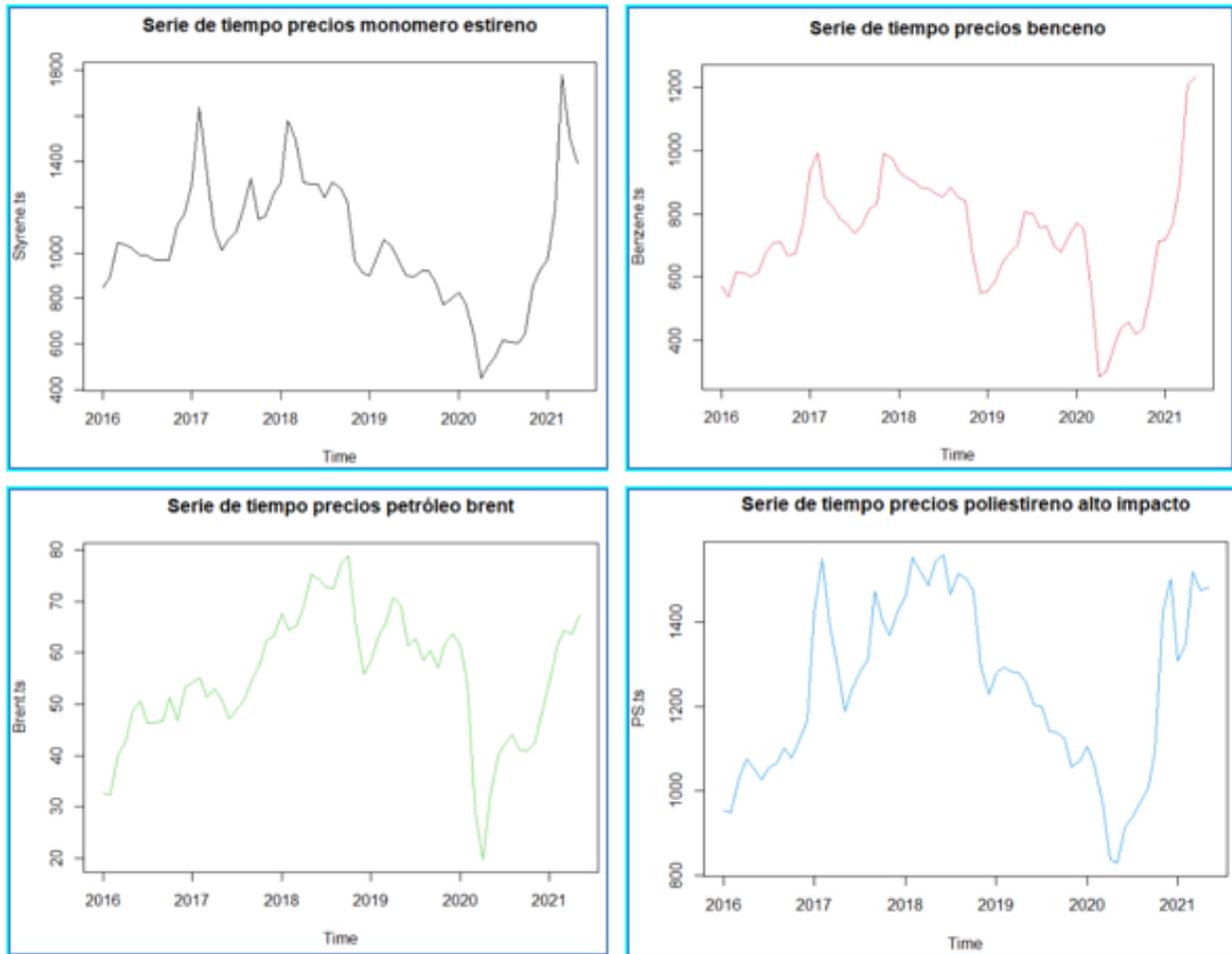


Figura 3: Series de tiempo de cada variable
(Fuente: Elaboración propia)

2.5 Análisis bivariado

Este análisis permite determinar en qué medida es posible predecir el valor de una variable dependiente en el caso de conocer el valor de otra independiente. En la Figura 4, se muestra bajo la diagonal los diagramas de dispersión bivariados, en la diagonal los histogramas y sobre ella la correlación de Pearson que muestra una relación simple sin hacer afirmaciones sobre causa o efecto.

Para complementar el análisis gráfico se emplea la prueba de Pearson. El objetivo es medir la relación estadística entre dos variables continuas, específicamente, la relación del poliestireno con el resto variables independientes.

Considerando un nivel de confianza del 95% se presenta la siguiente hipótesis de contraste.

H₀: correlación = 0 v/s **H_a**: correlación ≠ 0.

Los valores de la prueba de hipótesis se muestran en la Tabla 2, siendo para todos valores de p menores a 0.05. Por lo tanto, existe evidencia suficiente para rechazar la hipótesis nula. De esta forma, es posible inferir que existe correlación significativa entre las variables.

Tabla 2: Prueba correlación de Pearson

	Estadísticos	Benzene	Brent	Styrene
Ps_Spot	test statistic	6,1288	3,0167	9,3586
	df	62	62	62
	p-value	6,734E-08	0,0037030	1,813E-13
	cor	0,6142274	0,3577636	0,76519

(Fuente: Elaboración propia)

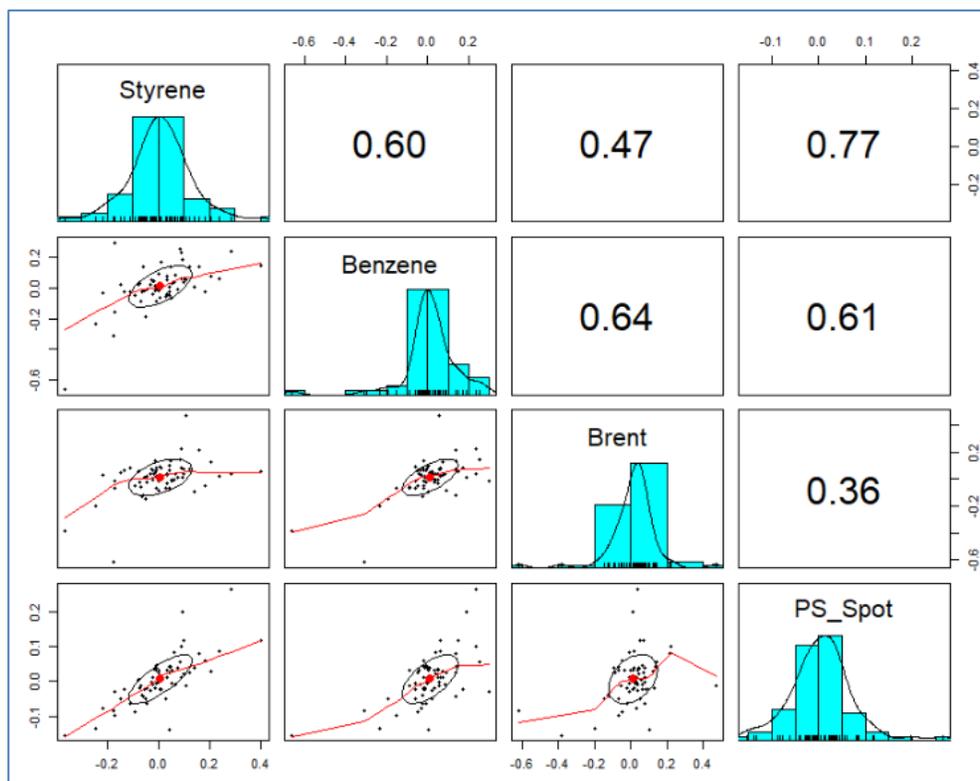


Figura 4: Análisis bivariado de las variables en estudio (bd2)
(Fuente: Elaboración propia)

2.6 Análisis multivariado

Permite comprender a través de métodos estadísticos la información relevante de diferentes variables, entrega una visión general realista y precisa de lo que está sucediendo a nivel de la información, resulta una potente y confiable prueba de significación siempre y cuando se tenga claro que se espera en su utilización.

A la luz de la información contenida en las series de tiempo, los análisis previos realizados en ellas y apoyado en el uso del banco de modelos conmutados se ha obtenido como resultado que la regresión línea múltiple es el modelo estadístico adecuado para ser empleado en el pronóstico de precios para el poliestireno de alto impacto. En la sección 2.10 es posible apreciar los cálculos que apoyan esta selección desestimando el uso de los otros modelos (ARIMA, Holt Winters y Redes neuronales).

2.7 Regresión lineal múltiple

Este modelo permite explicar la relación existente entre una variable dependiente y otras variables independientes. Para el análisis se considera la base de datos con variaciones mensuales bd2, con el fin de evitar un problema de regresión espuria considerando variables independientes no estacionarias.

Previo a la elaboración del modelo se evalúa la existencia de correlación entre variables. Para ello, se utilizó el coeficiente de correlación de Pearson expuesto en la Tabla 2 y un gráfico de análisis bivariado, que se muestra en la Figura 4.

En la Figura 5 se ilustra el modelo de regresión lineal múltiple para todas las variables disponibles. En una primera aproximación se identifica que la variable de precios de “petróleo Brent” no resultó ser estadísticamente significativa. Por lo tanto, se eliminó y se calibró nuevamente el modelo. Los resultados de esta nueva iteración se muestran en la Figura 6.

En esta nueva iteración no se aprecia que existan nuevas variables con poca significancia por lo que es posible inferir que el nuevo modelo explica a la variable dependiente lo siendo válido para ser usado en el pronóstico de precios del poliestireno de alto impacto, a su vez el

valor de R2 ajustado es 61,17% lo que demuestra que la variabilidad actual del precio es explicado por el modelo de regresión múltiple con dos variables independientes válidas que no generan una regresión espuria.

Reemplazando los valores del modelo en la ecuación (1) se tiene.

$$y^{\wedge}'=0,002631+0,126285*(Benzene)+0,353044*(Styrene) \quad (2)$$

LM (formula = PS_Spot ~ Brent + Benzene + Styrene, data = bd2)				
Residuals				
Min	1Q	Median	3Q	Max
-0,14943	-0,01988	-0,00187	0,02004	0,12946
Coefficients	Estimate	Std.Error	t value	pr(> t)
(Intercept)	0,002903	0,005435	0,0534	0,59519
Brent	-0,079569	0,053074	-1,4990	0,13906
Benzene	0,170078	0,057888	2,9380	0,00468 **
Styrene	0,365057	0,055805	6,5420	1,51E-08 ***
Códigos	*** (0)	** (0,001)	* (0,01)	. (0,01)
Residual standart error: 0,04328 on 61 degrees of freedom				
Multiple R-squared: 0,6376 Adjusted R-squared: 0,6195				
F-statistic: 35,2 on 3 and 60 DF,				
p-value : 2,996E-13				

Figura 5: Primera versión del modelo de regresión lineal múltiple
(Fuente: Elaboración propia)

2.8 Validación de resultados

Análisis gráfico

En cuanto al análisis de residuos, la Figura 7 muestra la existencia de valores atípicos. Específicamente, las observaciones 12, 58 y 60 las deben ser sometidas a un análisis puntual para determinar si son errores de digitación o corresponden a algún fenómeno temporal. El período de estas observaciones corresponde a enero 2017, noviembre 2020 y enero 2021, respectivamente.

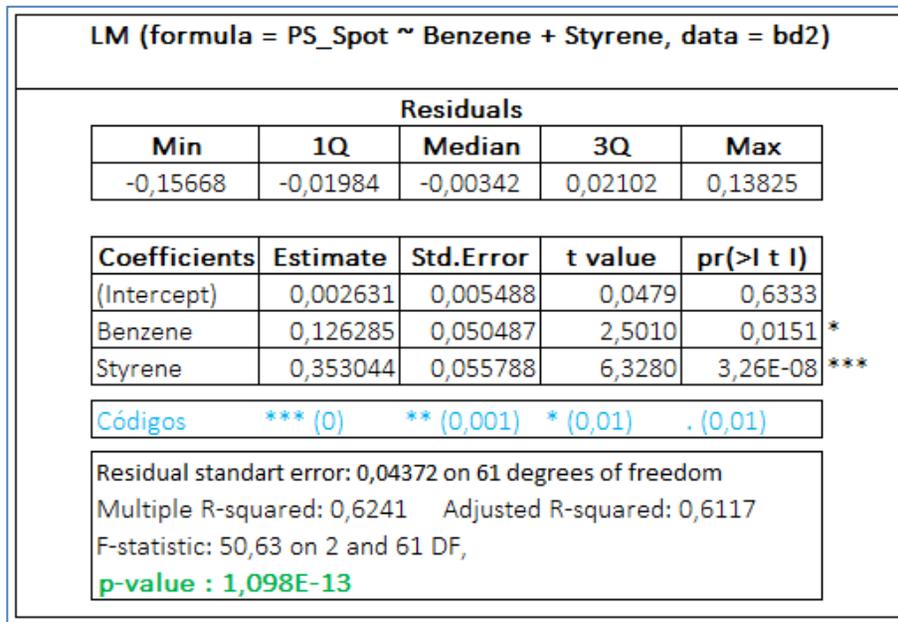


Figura 6: Segunda versión del modelo de regresión lineal múltiple (Fuente: Elaboración propia)

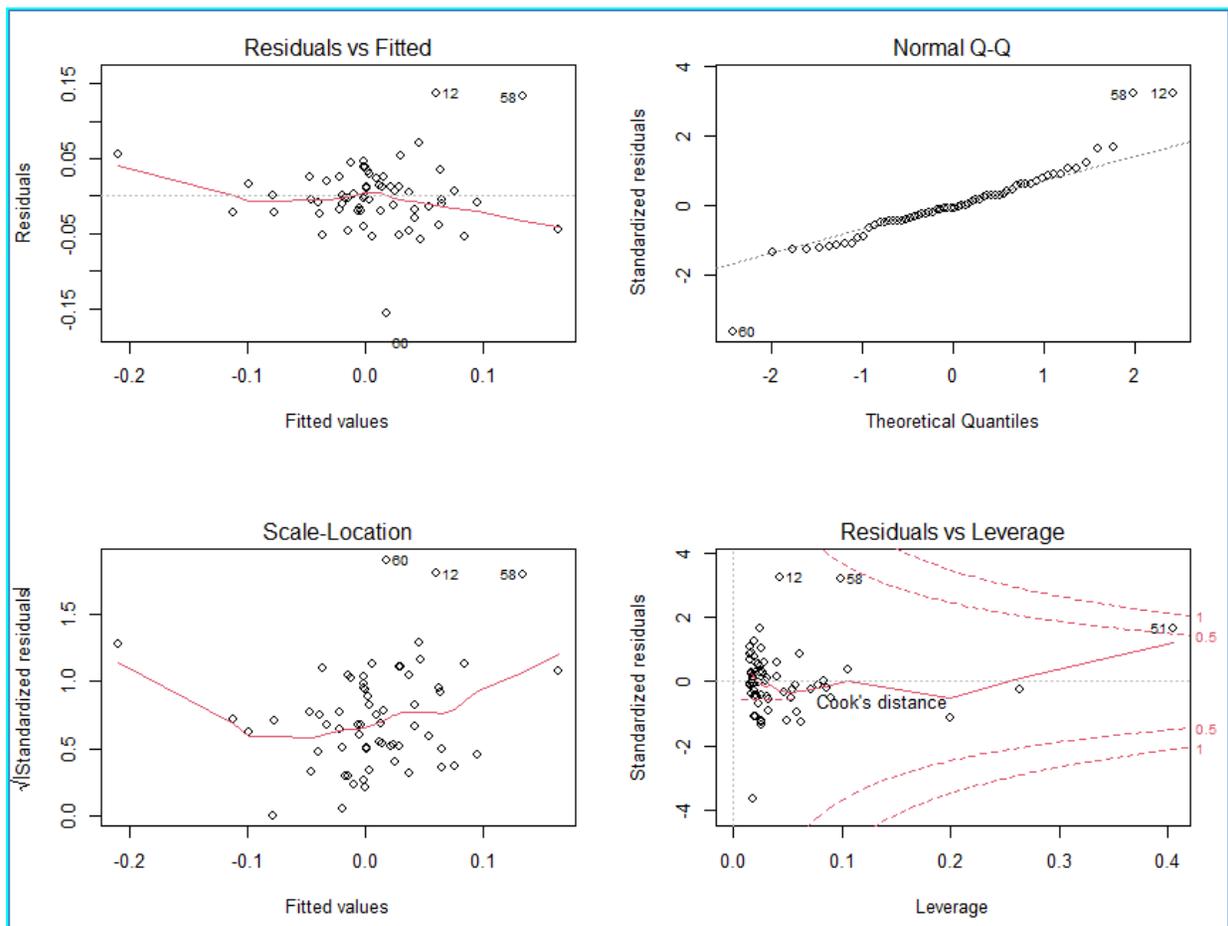


Figura 7: Gráfico de residuos

(Fuente: Elaboración propia)

Informes de distintos medios dan a conocer los eventos que ocasionan los datos atípicos señalados en el párrafo anterior, en la Tabla 3 se resumen los acontecimientos.

Tabla 3: Eventos temporales que explican datos atípicos.

Enero 2017. El aumento sostenido en el precio de las materias primas, principalmente el precio del petróleo Brent que en 12 meses aumentó un 77% y 79%, lo que desencadena un alza en sus derivados (Datosmacro, 2017).
Octubre 2020. La peor tormenta de nieve que batió récords en Estados Unidos, la ola de frío cruzó de norte a sur (Ortiz M. , 2020).
Diciembre 2020. Ascend informa de una parada no planificada en Florida. El alza de precios en las resinas podría acelerarse debido a un fallo de energía en la instalación de Pensacola ocasionado por las tormentas de Laura y Delta (Europe, 2020).

(Fuente: Elaboración propia)

Normalidad

Se utilizó la prueba de Jarque-Bera para comprobar si los residuos tienen la asimetría y la curtosis de una distribución normal.

La hipótesis para contrastar la normalidad de los residuos es la siguiente:

H₀: Normalidad v/s **H_a**: No normalidad.

Los resultados de la prueba son:

X-squared = 35,504; df = 2; p-value = 1,952e-08. Entonces, es posible inferir que existe evidencia suficiente para rechazar H₀ por tanto no se cumple este supuesto.

Homocedasticidad

Esta propiedad implica que la varianza estimada de los residuos depende de los valores de las variables independientes. Para el cálculo de la homocedasticidad, se utilizó la prueba Studentized Breusch-Pagan.

La hipótesis para contrastar la Homocedasticidad es la siguiente:

H₀: Homocedasticidad v/s **H_a**: Heterocedasticidad.

Los resultados de la prueba son BP = 3,7494; df = 2; p-value = 0,1534. Dado que el p-value es menor al 5%, existe evidencia para concluir que se cumple la homocedasticidad.

Autocorrelación

Para este análisis se utiliza la prueba Durbin-Watson. El objetivo es detectar la presencia de autocorrelación en los residuos del modelo múltiple.

Las hipótesis para contrastar la Autocorrelación es las siguientes:

H₀: Independencia v/s **H_a**: Autocorrelación.

Los resultados de la prueba son DW = 1,7908; p-value = 0,1852. Por lo tanto, existe evidencia estadística para no rechazar H₀, lo que permite concluir que los términos del error son independientes positivamente.

Multicolinealidad

Para evaluar el grado de dependencia de las variables independientes, se realizó el cálculo del indicador de Factor de Inflación de la Varianza (VIF), el cual mide el efecto combinado que tienen las dependencias entre los regresores sobre la varianza. Los valores de la prueba son Benzene = 1,553213; Styrene = 1,553213.

En el análisis bivariado de la Figura 4, se deduce que existe correlación entre las variables explicativas. Esto es común en las variables de una regresión múltiple. Por lo tanto, y considerando los valores observados en la prueba, existe evidencia suficiente para concluir que no existe multicolinealidad.

Con el objetivo de lograr un mejor ajuste del modelo, se decide trabajar con variables ficticias con el fin de medir el factor cualitativo, y así contrastar si el efecto de éste es relevante en el modelo de regresión. Se definen tres variables ficticias representadas por d1, d2, d3, para cada uno de los períodos señalados anteriormente ver Tabla .3.

Este último modelo se expone en la Figura 8. Nótese que existe un mejor ajuste en comparación a los dos anteriores, obteniendo un p-value de 2,2E-16 y un valor de R2 ajustado de 81,9%. Por lo tanto, existe evidencia suficiente para concluir que el modelo es válido para ser usado en el pronóstico de precios del poliestireno y que la incorporación de variables ficticias resultó ser significativa.

2.9 Validación de resultados modelo con variables dummy

Este acápite tiene por objetivo validar las pruebas del nuevo modelo con variables ficticias, para ello se ejecutarán las mismas pruebas que en el modelo anterior. Análisis completo en Tabla 4.

LM (formula = PS_Spot ~ Brent + Benzene + Styrene + + d1 + d2 + d3, data = bd2)				
Residuals				
Min	1Q	Median	3Q	Max
-0,05477	-0,01887	0,00000	0,01649	0,07836
Coefficients	Estimate	Std.Error	t value	pr(> t)
(Intercept)	0,001002	0,003826	0,262	0,7944
Benzene	0,087524	0,034891	2,509	0,0149 *
Styrene	0,326430	0,038972	8,376	1,44E-11 ***
d1	0,149733	0,030556	4,900	8,05E-06 ***
d2	0,151352	0,031499	4,805	1,13E-05 ***
d3	-0,153795	0,030143	-5,102	3,88E-06 ***
Códigos	*** (0)	** (0,001)	* (0,01)	. (0,01)
Residual standart error: 0,02986 on 58 degrees of freedom				
Multiple R-squared: 0,8333, Adjusted R-squared: 0,819				
F-statistic: 58 on 5 and 58 DF,				
p-value : 2,2E-16				

Figura 8: Modelo regresión línea múltiple con variables dummy
(Fuente: Elaboración propia)

Tabla 4: Validación de modelo con variables ficticias

Normalidad		Homocedasticidad	
<i>(prueba de Jarque Bera)</i>		<i>(prueba Studentized Breusch-Pagan)</i>	
H ₀ : Normalidad	H _a : No normalidad	H ₀ : Homocedasticidad	H _a : Heteroscedasticidad
X-squared	0,38933	BP	2,7078
df	2	df	5
p-value	0,8231	p-value	0,7449
<p>A un nivel de significancia del 5% la hipótesis nula H₀ de normalidad no se rechaza, puesto que el valor de p-value > 0,05 por tanto se cumple este supuesto</p>		<p>A un valor de significancia usual del 5% la hipótesis nula de homocedasticidad no se rechaza, puesto que el valor de p-value > 0,05, por tanto, se cumple este supuesto.</p>	
Autocorrelación		Multilinealidad	
<i>(prueba Durbin-Watson)</i>		<i>(prueba factor de inflación de la varianza VIF)</i>	
H ₀ : Independencia	H _a : Autocorrelación	<p>El objetivo es no encontrar multicolinealidad entre las variables independientes. En los resultados del cálculo del indicador todos los VIF son menores a 5, por lo tanto, no se detecta multicolinealidad entre las variables independientes (máximo valor obtenido fue de 1,591009)</p>	
DW	2,1196		
p-value	0,6664		
<p>A un valor de significancia usual del 5% la hipótesis nula de independencia no se rechaza, puesto que el valor de p-value > 0.05 por lo tanto los términos del error son independientes.</p>			

(Fuente: Elaboración propia)

Predicciones

En la Figura 9 se presenta un gráfico de predicción de precios para el poliestireno basado en el modelo de regresión lineal múltiple con variables ficticias (d1, d2, d3). Al comparar el valor real con la predicción, se aprecia que existe un ajuste significativo, lo que se corrobora con el R² ajustado de los modelos.

La incorporación de variables ficticias de tipo temporal recoge información que no puede ser visualizada directamente sobre las observaciones, complementando el análisis previo elaborado en la Figura 7.

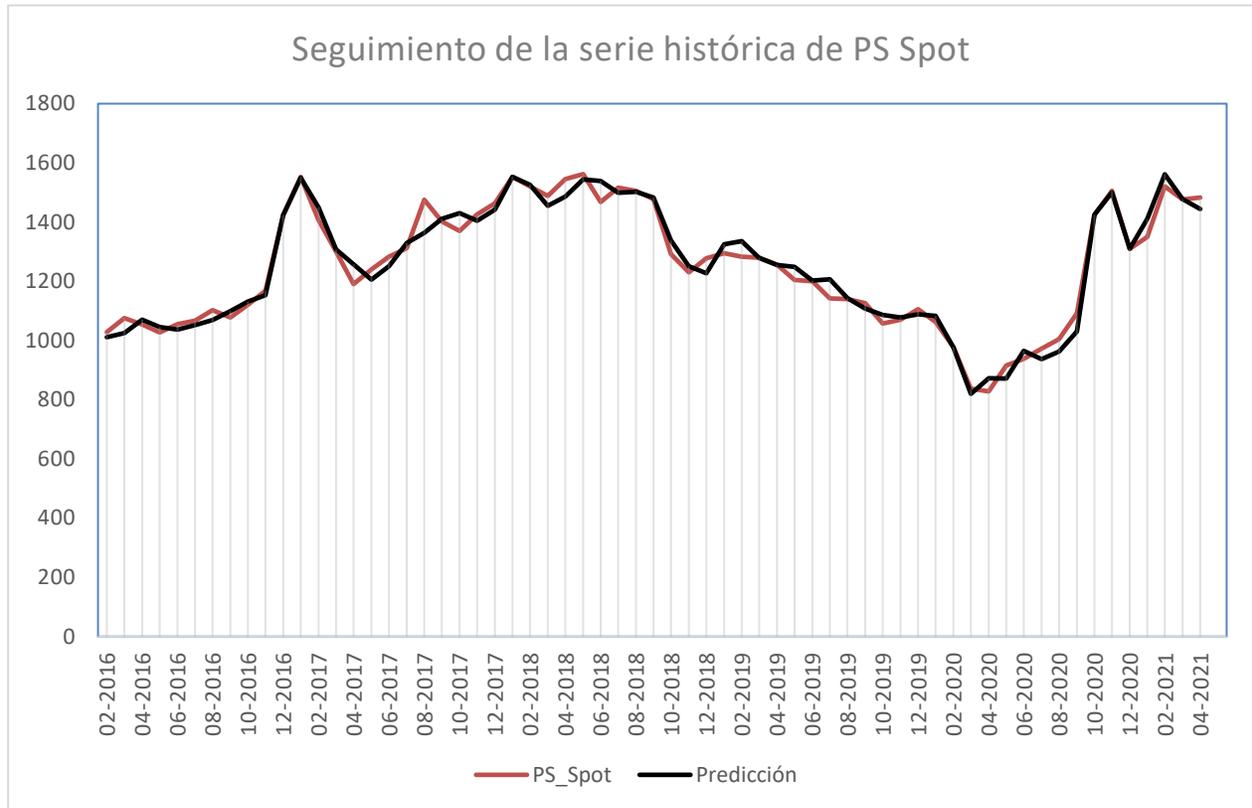


Figura 9: Gráfico comparativo predicción v/s valor real (Fuente: Elaboración propia)

2.10 Iteraciones del banco de modelos conmutado Modelo ARIMA

Prueba Dickey-Fuller

Busca determinar la existencia o no de raíces unitarias en una serie de tiempo, en la Tabla 5 se presentan los valores de la prueba con el objetivo de tener evidencia para contrastar la siguiente hipótesis.

H₀: La serie no es estacionaria v/s **H_a**: La serie es estacionara.

Como resultado de la prueba se puede inferir que existe evidencia suficiente en los valores de p-value para aceptar la hipótesis nula evidenciando con ello, la existencia de raíces unitarias por lo tanto las variables no son estacionarias.

Para apoyar la conclusión obtenida con Dickey-Fuller, se realiza una prueba gráfica de autocorrelación (es una característica que consiste en que, elementos cercanos en el espacio o en el tiempo se parecen más entre sí que con respecto a elementos más lejanos, solamente por el hecho de estar cerca) y una prueba numérica con Box-Pierce bajo la siguiente hipótesis.

H₀: Los residuos no poseen autocorrelación entre si v/s **H_a**: Los residuos poseen autocorrelación entre si.

Al analizar la información de la Figura 10 se concluye que existe evidencia suficiente para rechazar H₀. Los datos se distribuyen de forma independiente de modo que cualquier correlación observada en los datos es resultado de la aleatoriedad del proceso.

En la comparación de ambas pruebas se puede inferir que las series no son estacionarias, están guiadas por la tendencia, la cual se aprecia en el estadístico de la Figura 10 que va creciendo conforme avanza el tiempo.

Sobre el análisis de los datos se puede concluir que los rezagos de la serie (PS.ts) no tienen un impacto sobre el valor actual de ésta, el tiempo deja de ser significativo no existiendo una dependencia con el período anterior, por lo tanto, el modelo ARIMA no es recomendado para la utilización de pronóstico de precios para el poliestireno de alto impacto.

Modelo Holt Winters

Este modelo utiliza una técnica de suavizado exponencial, basado en estimaciones recursivas a partir de los datos históricos, este método necesita que en la serie de tiempo exista tendencia y estacionalidad (ver Tabla 5), la cual no se encuentra presente en la serie.

Al no existir relación temporal en los datos no tiene sentido la aplicación del modelo ya que los resultados esperados pierden significancia y no representan la realidad del suceso en estudio.

2.11 Discusión de resultados

A lo largo de esta investigación queda de manifiesto que las primeras aproximaciones de solución al problema de pronósticos no han dado los resultados esperados. La elección a priori de modelos sin el conocimiento acabado de la información ha ocasionado entrar en un bucle de iteraciones forzando resultados que no fueron significativos. Tal como lo dicen Rodríguez y Gonzalez en su libro Principios de Econometría (David E. Rodríguez Guevara, 2019) y Ocerin en su libro Modelos econométricos y series temporales (Ocerin, 1998) los primeros pasos es determinar la naturaleza de los datos para luego definir los modelos factibles y tomar una decisión en base al ajuste de éstos.

El aprendizaje supervisado y no supervisado son las dos grandes áreas que cubren estos modelos. En primer lugar, no es posible afirmar que uno sea mejor que otro sólo conociendo la teoría que los sustenta o de lo atractivo que pueden llegar a ser en cuanto a su alcance, la decisión se basa en formular la pregunta adecuada y luego encontrar la respuesta mediante la revisión, el reconocimiento de los datos incorrectos, y la interpretación coherente de los resultados. Las herramientas analíticas permiten descubrir pautas y tendencias nuevas, encontrar ideas a partir de datos aparentemente inconexos y descubrir relaciones interesantes desde el punto de vista estadístico. Una predicción con escasa información tiene riesgos de quedar sesgada o en su defecto sobre ajustada. Un proceso analítico de calidad puede ayudar al analista en la toma de decisión y concluir si el modelo estadístico responde con una confianza del 95% a la pregunta de origen. La escasez de información en tiempo real puede causar sesgo al observador apartándolo del análisis de estocasticidad de las variables y los estímulos externo de la economía, perdiendo significancia y posibles regresiones espurias; sin embargo, no impide que éste pueda proponer un modelo que permita predecir el futuro de una variable dependiente si la información disponible tuvo un preprocesamiento de calidad. Estos hallazgos son coherentes con los resultados indicados por (Alliance, 2015) quien sostiene que la calidad de datos y número de ellos es fundamental para entrenar modelos econométricos.

En la búsqueda de un modelo aceptable, el uso de la minería de datos juega un rol preponderante al unir las diversas técnicas en el análisis de datos, desde la recolección y

depuración de la información, pasar por un aprendizaje supervisado de los modelos hasta lograr un aprendizaje automático y finalmente una predicción significativa.

Durante la ejecución de este trabajo se han evidenciado algunas limitantes que han impedido alcanzar el máximo potencial de los modelos escogidos, dentro de ellos destacan:

- El alcance y restricción de cada técnica dependen directamente de la estructura de los datos (Alliance, 2015).
- Se debe considerar el modelamiento matemático que sustenta a cada modelo ya que existe uno adecuado para cada problema (David E. Rodríguez Guevara, 2019).
- Para este trabajo el tamaño de la muestra comprende 65 períodos confiables de información, que resultan escasos al momento de modelar una red neuronal, solo para el entrenamiento se necesita al menos un 70% de la muestra siendo ésta bastante reducida para alcanzar el potencial de la herramienta con una alta probabilidad de obtener un modelo sobre ajustado.
- El recorrido de la muestra es mensual para una mejor adaptación debería ser al menos diaria para determinar si la correlación observada depende del período anterior o si realmente responde a sucesos aleatorios. En su libro digital (Alliance, 2015) describe la importancia de las observaciones como píxeles dentro de una pantalla, al estar aislados éstos entregan una pequeña cantidad de información, pero cuando se los combina con una cantidad suficiente de píxeles en el orden correcto, el analista puede localizar significados nuevos y, a veces inesperados; por lo tanto, restringir la data solo a sucesos mensuales no es la mejor decisión que puede tomar el dueño del dato.

Es importante discutir elementos respecto a las variables dummy relacionados con los outliers encontrados en la serie. En primer lugar, estos sucesos corresponden a fenómenos naturales y económicos (oferta/demanda) que impactaron directamente al precio del petróleo y a sus derivados, en algunos casos la escasez de éstos impulsó la tendencia alcista de los precios en la industria del plástico. También, es importante destacar que las variables económicas no proceden de un modelo experimental, siendo los agentes económicos los que hacen que las variables se influyan mutuamente tal como lo dice Morales (Morales, 2010) la alta correlación hace difícil precisar a partir de la información muestral los coeficientes

asociados a dichas variables, y como consecuencia existe una precisión reducida en los modelos. Por lo tanto, para un mejor ajuste en el modelo de regresión lineal múltiple se decidió incorporar aspectos cualitativos dotándolo de mayor significancia y separando la correlación de estos fenómenos para aumentar la capacidad de predicción. Además, se realizaron pruebas de predicción para ambos modelos en cuyos resultados el R2 ajustado pasó de 61,17% a 81,9%. Sobre la base de los dos modelos propuesto, no es posible concluir que sólo con la adición de variables ficticias un modelo sea más preciso que otro que no las tiene, la evidencia demuestra que la alta correlación entre variables hace difícil discriminar la relevancia para explicar la variable dependiente y por consiguiente una imprecisión en los resultados.

Finalmente, la actualización constante de la historia de compras permite que la simulación sea más precisa obteniendo más y mejor información del nuevo escenario lo que permite al analista ajustar el modelo y tomar una decisión, donde las estimaciones por si solas son incapaces de hacerlo basado sólo en los registros históricos.

3 ARTÍCULO

El presente apartado, recoge la investigación contextualizada motivo de este proyecto de grado, y es presentada en formato de artículo académico. Se trata de un artículo conciso, escrito en el formato típico de revistas especializadas o de conferencias, de acuerdo con reglas específicas definidas por la dirección del programa.

El artículo, ha sido cuidadosamente redactado con el fin de que se haga fácilmente entendible y logre expresar de un modo claro y sintético lo que se pretende comunicar, considerando las citas y referencias respectivas de los estudios que lo fundamentan. El trabajo realizado, se sintetiza entonces como artículo, para facilitar al trabajo de quienes puedan estar interesados en consultar la obra original.

Este trabajo, considera y discute, a través de un proyecto aplicado, desarrollado en un contexto de realidad profesional, la integración de herramientas y conocimientos que se han adquirido en las líneas de desarrollo del programa. Lo que se consolida en una investigación profesional contextualizada a la realidad profesional que se expone, la que se relacionada con líneas y ámbitos específicos abordados en el plan de estudios del programa, permitiendo integrar, de manera adecuada, los conocimientos teóricos y metodológicos desarrollados en él.

PRONÓSTICO DEL PRECIO PROMEDIO PARA EL POLIESTIRENO DE ALTO IMPACTO

Fernando Andrés Pavez Stuardo

*Graduado del programa de Magister en Ingeniería Industrial y de Sistemas, Facultad de Ingeniería, Universidad de Desarrollo,
ferandres45@gmail.com*

Resumen:

Chile al no ser un productor de resinas plásticas, depende totalmente de las importaciones. Las grandes empresas pueden importar directamente del productor y las empresas más pequeñas sólo tienen la disponibilidad de compra a través de un trader con precios sobre la media del mercado internacional.

Este trabajo presenta una herramienta que permite organizar, presentar y describir un conjunto de datos cuya finalidad sea el pronóstico de precios del poliestireno de alto impacto. El objetivo de esta investigación es proponer un modelo estadístico que oriente al usuario en la compra del poliestireno de alto impacto y estimar el precio de compra para el siguiente ciclo. Para lograrlo se propone una aproximación cuantitativa, basada en la identificación de variables que intervienen en la estructura de costo de la materia prima y la elección de un modelo estadístico capaz de predecir el precio de ésta. El conjunto de modelos propuestos son ARIMA, Holt Winters, Regresión lineal y Redes neuronales. Los datos muestran que al comparar las predicciones efectuadas con el modelo y los precios históricos se obtiene una confiabilidad del 95%. Esto permite mejorar la toma de decisiones en el proceso de compra, disminuyendo de este modo el costo de la materia prima y con ello la optimización de los recursos financieros de la empresa.

Palabras claves: Métodos cuantitativos; Pronósticos; Series de tiempo; Regresión lineal múltiple; Estadística descriptiva.

1. Introducción

En la gestión de compra de cualquier materia prima, el precio siempre ha sido una variable importante, sino la más importante en la toma de decisión, más aun en el mercado del plástico donde el valor de la materia prima representa gran parte del costo del producto final. Predecir adecuadamente dicho precio, puede efectuarse con la ayuda de tecnología computacional, que permite que los tiempos de cálculo se reduzcan haciendo más factible la incorporación de modelos propios de la minería de datos.

En este sentido, la atención se centra en encontrar un modelo adecuado con el cual se pueda estimar, con una confianza del 95%, el precio de compra del poliestireno, tomando en cuenta sucesos aleatorios de observaciones, logrando modelar patrones y extraer inferencias acerca de la población de estudio (Coy Mondragón, 2021). El resultado de este análisis permitirá al usuario disponer de información con un horizonte de algunos meses de anticipación, mejorando

con ello la toma de decisión y logrando una mejor gestión de compras.

Poliestireno de alto impacto

Es un polímero termoplástico que se obtiene de la polimerización del monómero estireno más un 14% de polibutadieno (John Scheirs, 2003).

El estireno es un hidrocarburo aromático proveniente de un anillo de benceno con un sustituyente etileno. El polibutadieno es un elastómero que se obtiene de la polimerización del Butadieno. El Butadieno es un alqueno que se produce de la destilación del petróleo.

La representación de las distintas interacciones en las materias primas que conforman al poliestireno de alto impacto se puede ver en la Figura 1 siendo el petróleo el origen de la cadena productiva, pasando por los distintos componentes y sus transformaciones hasta llegar al producto en estudio.

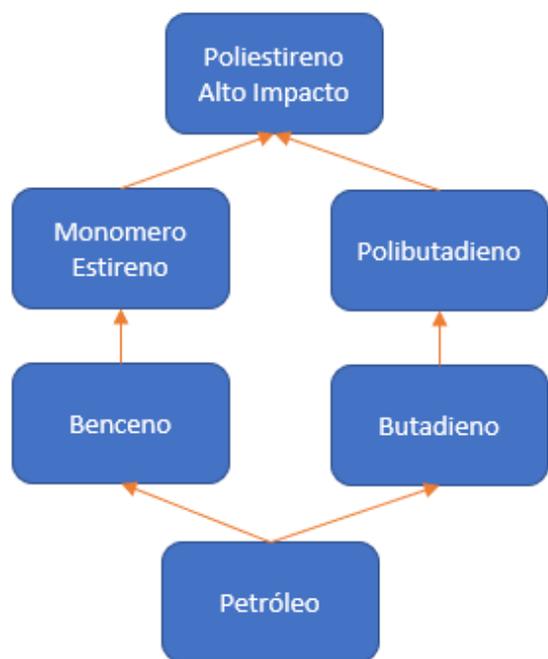


Figura 1: Composición del poliestireno de alto impacto (Fuente: Elaboración propia)

Alternativas en el ecodiseño para el reemplazo del poliestireno

Desarrollos del ecodiseño en la industria del embalaje han estado impulsando la búsqueda de nuevas alternativas de resinas más amigables con el ambiente. Una de las resinas que se encuentran disponibles en el mercado es el PLA (ácido poliláctico), el cual es un polímero biodegradable cuya base se obtiene a partir de almidón de maíz, yuca o caña de azúcar. En la actualidad, el PLA ha despertado gran interés por ser un posible sustituto a las resinas termoplásticas en la industria del packaging. Entre sus propiedades mecánicas se encuentran: baja resistencia al impacto, dureza, rigidez, elasticidad son similares a las del tereftalato de polietileno (PET), presenta buena resistencia a la torsión, alto módulo de flexión y alta transparencia hacen que la película plástica sea comparable a una película de celofán.

Por sus características el PLA puede ser un buen competidor de los plásticos de origen petroquímicos, y además puede ser biodegradable al 100% en medio acuoso, temperatura ambiente y un rango de pH entre 5 y 8 (Mexpolimeros, 2020). Su alto valor en comparación con las resinas de origen petroquímico aún lo hacen

poco atractivo para la industria láctea. Para estos formatos, el poliestireno de alto impacto (PS) sigue siendo el más requerido.

Mercado chileno del plástico para el packaging de alimentos

El plástico, a pesar de ser un material contaminante para el medio ambiente, aún es usado en aplicaciones industriales. En la industria de alimentos, la principal aplicación es la de contenedor primario. Son muchas las características que hacen al plástico el embalaje perfecto, entre las que se incluyen: su bajo costo, la resistencia al impacto, y la capacidad de contener alimentos sin alterar sus propiedades.

Hasta hace un par de años era impensado alargar la vida de frutas, verduras, carnes, quesos y otros, en la actualidad con los nuevos aditivos es posible crear barreras de oxígeno, mantener atmósferas inertes, aditivos que absorben humedad, otros que filtran el paso de la luz UV, etc., todo esto enfocado en prolongar la vida de productos alimenticios (Industries, 2019).

La guerra ambiental contra el uso del plástico

Desde el año 2018, Chile ha mostrado preocupación medioambiental en cuanto al uso del plástico en productos de un solo uso. Ejemplo de ello es la Ley 21.100 sobre la prohibición de bolsas plásticas para el comercio (Ley 21.100), y luego, con la incorporación de la Ley 20.920 que establece el marco para la gestión de residuos, la responsabilidad extendida del productor y fomento al reciclaje (LEY 20920). Estas claras señales han hecho que el mercado del packaging esté dando un vuelco en el uso de ciertas resinas, tal es el caso del poliestireno que está dejando de ser el material preferido para las nuevas soluciones de empaque por no contar con un mercado de reciclaje claro y definido.

La industria del plástico en Chile aún está en período de ajustes, no existe claridad sobre el uso de resinas recicladas salvo en el PET que tiene una economía circular. A pesar de todas estas controversias Chile sigue siendo un consumidor bastante pequeño en comparación con otros mercados latinoamericanos y las decisiones que se tomen al interior del país acerca de la resina predominante no afectarán el precio de mercado de la resina en estudio.

El precio del petróleo y su efecto en el precio de la resina

Existe correlación entre el petróleo y las distintas resinas plásticas derivadas de él, las fluctuaciones hacen que los precios del mercado sean impactados.

Estudios de Duncan Seddon & Associates (Seddon, 2013) confirman esta tesis obteniendo un coeficiente de correlación (R2) de 0.81 en el período comprendido entre 2000 y 2013.

Los últimos datos de American Chemistry Council muestran un mayor impulso en junio 2021 y se pronostica un segundo semestre con una tendencia creciente que impacta directamente en el precio de las materias primas y petróleo. La escasez del suministro en el mercado europeo de acrilonitrilo-butadieno-estireno sin duda provoca una desestabilización en los precios del cono sur debido a las importaciones y la creciente demanda de este tipo de productos.

El valor de los petroquímicos es altamente influenciado por el precio del petróleo. La caída o aumento de los precios dependerá del nivel de stock en los mercados. Su regulación depende en gran medida de las variables exógenas, ejemplo de esto se encuentra en la tormenta que afectó al norte de América en marzo 2021 cuya consecuencia fue la escasez de monómero estireno y en el segundo trimestre la escasez del polibutadieno, materias primas vitales para la fabricación del poliestireno de alto impacto.

Repositorios de precios

El poder de la información y el valor que ésta tiene en el mercado ha sido el impulsor para que algunas empresas estén dispuestas a invertir y otras a pagar por información concentrada en un solo lugar. PLATTS (Platts, 2021), ICIS (ICIS, 2021) e INVESTING (Investing, 2021) son referentes en cuanto a calidad de información y en algunos mercados sirven como referencia para la elaboración de contratos, monitoreando precios de productos básicos, consiguiendo así transparencia en la transacción. Se publican precios spots (precio contado para entrega inmediata) de diversas materias primas, curvas forward que pueden ser consultadas sobre una base coherente y fidedigna de información. Es importante destacar que los precios que sustentan la base de estudio se obtienen de estos repositorios.

Chile y su dependencia en los mercados

Chile tiene una economía abierta con una alta dependencia del comercio internacional. Sus principales exportaciones son materias primas sin mucha elaboración, como, por ejemplo: cobre, filetes de pescado, pulpa de celulosa, vino y frutas. Las principales importaciones, por otra parte, son: petróleo, gas de petróleo, aparatos eléctricos y

vehículos a motor. La exportación de bienes y servicios bajó en 3,5% en 2020 y se prevé que suba en 0,5% en 2021, mientras que las importaciones bajaron en 11,7% en 2020 y se prevé que crezcan en 10,4%. Estas bajas se explican principalmente por la crisis económica mundial producto de la pandemia por COVID 19 (Empresas, 2021).

En el caso del plástico, las importaciones chilenas provienen principalmente de China, Estados Unidos, México, Brasil, Colombia y Argentina. Las empresas se ven forzadas a adquirir estos productos tanto en el mercado local o a través de importaciones directas. Los precios a los que se acceden suelen estar sujetos a bandas establecidas por contrato, compras spot, o una mezcla de ellos.

Mecanismos para la estabilización de precios sobre bandas referenciales

Un mercado convulsionado con muchas alzas y pocas bajas en los precios ha hecho que las empresas cambien su estrategia y opten por contratos forward. Este método puede beneficiar o perjudicar la operación. Sin embargo, en mercados tan volátiles con el de las resinas plásticas es conveniente contar con una fórmula de precios que asegure el abastecimiento y permita tanto al vendedor como al comprador mantener estabilidad con miras hacia el futuro en una relación de largo plazo.

La base para estos contratos se encuentra en las materias primas con la que se produce el poliestireno, que suele ser el polibutadieno o monómero de estireno, más una variable que puede incluir el flete u otro que convenga a ambas partes, teniendo al mercado chino como referencia.

Pronósticos de precios: Teoría y definiciones

Pronosticar es el arte y ciencia de predecir acontecimientos futuros (Cohen, 2017). Generalmente los pronósticos son utilizados para apoyar las decisiones gerenciales. Se basan en la historia de los acontecimientos pasados, para traer al tiempo presente un valor futuro de comportamiento bajo una probabilidad de ocurrencia. El pronóstico también tiene una temporalidad que puede ser afectada por sucesos que ocurren en otras variables ajenas al proceso.

Modelos del conmutador

ARIMA

Es un modelo autorregresivo integrado de promedio móvil (acrónimo del inglés autoregressive integrated

moving average). ARIMA utiliza variaciones y regresión de datos estadísticos, cuyo objetivo es encontrar un valor estimado del futuro. Es un modelo dinámico de series de tiempo, las estimaciones son explicadas en base a la información del pasado y no por variables independientes como en una regresión lineal (Douglas Montgomery, 2008).

Holt Winters

Es un modelo basado en suavizamiento exponencial. Este método puede adaptarse fácilmente a los cambios y tendencias, así como también a patrones estacionales.

Regresión lineal

Es un modelo asociativo con enfoque cuantitativo que permite pronosticar una variable dependiente con una o más variables independientes a través de una ecuación lineal. Puede ser usada para estimar precios de productos o servicios, estimar demandas entre otras (Betancourt, 2021).

La regresión lineal múltiple considera linealidad en la relación, normalidad y homogeneidad de la muestra. La función (1) que describe a la regresión es.

$$y' = a + b_1x_1 + b_2x_2 + b_ix_i \quad (1)$$

Donde y es la variable endógena o dependiente, x las variables exógenas o independientes, b los coeficientes estimados del modelo y a es la intersección o constante.

Redes neuronales artificiales - RNA

Consiste en un conjunto de unidades (neuronas artificiales) interconectadas entre sí, las que comúnmente se distribuyen en capas. Un sistema típico cuenta con tres capas. En la primera se encuentran las neuronas de entrada, las cuales envían información a través del campo de neuronas, la sinapsis ocurre entre la segunda y tercera capa donde se envía a la salida como información procesada. Los sistemas más complejos tienen más capas. Estos sistemas aprenden y se forman así mismos en lugar de ser programados y sobresalen en áreas donde la detección de soluciones o características es difícil de expresar con la programación convencional, dentro de las desventajas de los RNA (Jones, 2019) es necesario ajustar hiperparámetros como el número de capas ocultas, funciones de activación, número de iteraciones, optimizador y número de neuronas por capa y todo ello ocurre en una caja negra sin mayor intervención del analista.

La importancia de un modelo cuantitativo y su aplicación a gran escala

El objetivo de este trabajo es concentrar la búsqueda de un modelo estadístico dentro de un banco de modelos conmutados con la finalidad de conocer, comprender y simular múltiples alternativas de precios para el poliestireno que ingresa a Chile. La lógica de selección está basada en la interpretación del error y en el ajuste que logre la predicción sobre la información de mercado imperante en el momento.

La era digital y la cantidad abrumadora de información que minuto a minuto se está generando en todos los sectores de la sociedad permiten hoy tomar decisiones más acertadas que hace 15 años atrás. Una desviación puede rápidamente ser corregida evitando así problemas en abastecimiento o sobre precio en compras de materias primas. El término información en tiempo real (Emilio Ontiveros, 2017) trae consigo nuevos desafíos en cuanto al análisis de este nuevo conocimiento y será función de este modelo conmutado ayudar a precisar la información y seleccionar el modelo que mejor se adapte a las circunstancias del ambiente tan convulsionado. Mucho de estos sucesos que día a día se van gestando en el mercado podrán ir corrigiendo el análisis pudiendo en cada iteración encontrar el mejor valor en cuanto al precio del poliestireno.

En la literatura revisada sobre pronóstico de precios para resinas de poliestireno no se ha encontrado un trabajo similar donde a través de modelos econométricos se pueda proyectar un precio futuro de este producto, sin embargo, Carlos Fernández de la Universidad Nacional de Colombia (Fernández Múnera, 2020) desarrolló un trabajo que aborda el análisis de precios para el polipropileno en ambientes coberturistas enfocado en la disminución del riesgo en la decisión. Si bien es cierto, los materiales y precios que se tratan en este estudio son distintos a los de esta tesis, los modelos estadísticos usados pueden ser extrapolados para este trabajo.

Entendida la problemática, y la revisión de la literatura presentada, es posible efectuar el siguiente cuestionamiento: ¿cómo el uso de los modelos econométricos aplicados en las series de tiempo de los precios del petróleo y sus derivados puede generar modelos predictivos capaces de pronosticar los precios futuros del poliestireno de alto impacto en Chile?

En efecto, disponer de la información histórica de las variables observadas permite al analista estimar

valores futuros, para el caso de estudio se propone la utilización de un banco de modelos conmutados cuyo algoritmo sea capaz de seleccionar de entre todos al más preciso, teniendo como base la historia de la serie temporal.

El foco de este trabajo está en predecir el valor de una resina que está compuesta por tres principales elementos derivados del petróleo (IG Group, 2019), cuyo precio fluctúa en base a la oferta y demanda existiendo períodos de estabilidad, así como otros altamente volátiles. En la búsqueda de precisión se debe tener en cuenta no perder la simpleza y no generar un sobre ajuste del modelo.

Habiendo recorrido las bases teóricas fundamentales para este estudio, es importante mencionar que la principal motivación para realizarlo ha sido la falta de un modelo que permita predecir los precios del poliestireno de alto impacto en Chile. Ante esto, se propone un sistema conmutado de modelos capaz de iterar y encontrar uno que permita estimar el precio de compra del poliestireno en base a las realizaciones de ciertas variables estocásticas. En este sentido, la presente tesis contribuye al apoyo de la toma de decisiones en este contexto, al permitir a un analista obtener un valor estimado del precio del poliestireno con un 95% de confianza.

Entendido esto, el objetivo general de este proyecto es proponer un modelo estadístico que oriente al usuario en la compra del poliestireno de alto impacto, para la estimación de un precio objetivo de compra basado en

datos históricos y con ello optimizar el aprovisionamiento futuro de materiales.

2. Metodología

La metodología utilizada en este estudio comprende las siguientes cinco etapas.

1. **Análisis del problema:** en esta primera etapa se analizan las dificultades actuales para conseguir un precio estimado de compra que incorpore a las variables que afectan el precio final del bien.
2. **Revisión bibliográfica:** en esta etapa se revisa la literatura, haciendo hincapié en casos de éxito y fracaso en cuanto al uso de modelos estadístico para la predicción de precios futuros.
3. **Extracción, transformación y carga de datos KDD (Knowledge Discovery in Databases):** esta etapa tiene la finalidad de seleccionar desde un conjunto de datos, la información necesaria para ser usada en los modelos.
 - I. Análisis cuantitativo.
 - II. Tratamiento de datos.
 - III. Análisis descriptivo.
 - IV. Análisis bivariado.
 - V. Cálculo del modelo.
4. **Selección de modelos estadísticos:** en el proceso de iteración el banco de modelos conmutados es capaz de seleccionar al que sea capaz de entregar la mejor estimación (ver Figura 2).

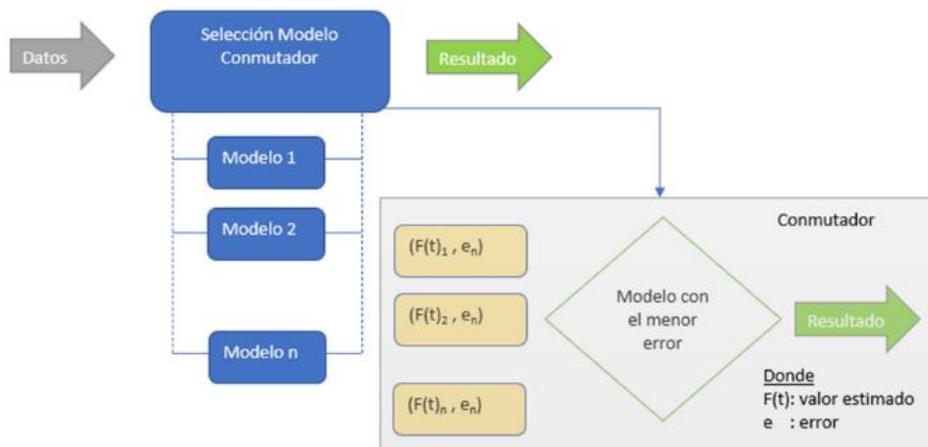


Figura 2: Modelo conmutado
Fuente: Elaboración propia)

En esta etapa el analista dispone de cuatro modelos (ARIMA, Holt Winters, Regresión lineal múltiple y Redes neuronales). La previa selección de modelos econométricos permite al conmutador partir desde una base predefinida, sin embargo, no existe restricción para adherir nuevos modelos y probar su funcionalidad en el caso que los existentes no sean capaces de predecir con cierta precisión los precios para el poliestireno de alto impacto. Los modelos que sirven de base para el conmutador concentran un conjunto de algoritmos capaces de extraer información valiosa de los metadatos pudiendo detectar patrones de comportamiento, subidas o bajadas de alguna variable de interés para crear un nuevo repositorio de información con la cual a través de estadística deducir relaciones entre variables y generar predicciones (Perez, 2015)

- 5- Validación: esta etapa involucra evaluar la información estadística de los errores, analizar el p-valor, y comparar de forma gráfica el comportamiento del modelo con los datos históricos. Para todos los efectos, los análisis de significancia se realizan con un nivel de confianza del 95%, y el rechazo de la hipótesis nula (H_0) es con un p valor inferior al 5%.

3. Resultados

El banco de modelos conmutados está compuesto por cuatro modelos El criterio de selección se basa tanto en el error como también en el ajuste de la predicción que éstos tengan con los datos disponibles.

3.1 Tratamiento de datos y análisis exploratorio

Antes del análisis exploratorio se crean dos bases de datos. La primera contiene información en escala natural bd1 (directamente del repositorio sin modificación), mientras que en la segunda se aplica una variación mensual debido a que las oscilaciones entre los valores de las series en diferentes momentos de tiempo, $(x_t - x_{t-h})$ son proporcionales al nivel inicial. Para evitar esta proporcionalidad, y conseguir que la tasa de crecimiento en cada observación sea independiente de la magnitud de dicha observación, se utiliza la transformación logarítmica dando como resultado bd2. (Prog. RStudio `bd2<- read_excel("Data Precio.xlsx")[-1] %>% mutate_all(funs(log(./lag(.,1)))) %>% slice(-1)`)

En la Tabla 1 se presenta el análisis exploratorio realizado a ambas bases de datos. En particular, se muestra información sobre mínimos, máximos, media, mediana, 1° cuartil y 3° cuartil, de cada una de las variables. Esta metodología proporciona una síntesis de la información contenida en la muestra con el objetivo de mejorar la precisión, sencillez, y ordenamiento de los datos, como también extraer las características más representativas de ésta, junto a ello en la Figura 3 se presenta en forma gráfica la serie temporal de las variables con el objetivo de visualizar las alzas y bajas de precios en el tiempo.

3.2 Análisis bivariado

Este análisis permite determinar en qué medida es posible predecir el valor de una variable dependiente en el caso de conocer el valor de otra independiente. En la Figura 4, se muestra bajo la diagonal los diagramas de dispersión bivariados, en la diagonal los histogramas y sobre ella la correlación de Pearson que muestra una relación simple sin hacer afirmaciones sobre causa o efecto.

Para complementar el análisis gráfico se emplea la prueba de Pearson. El objetivo es medir la relación estadística entre dos variables continuas, específicamente, la relación del poliestireno con el resto variables independientes.

Considerando un nivel de confianza del 95% se presenta la siguiente hipótesis de contraste.

H_0 : correlación = 0 v/s H_a : correlación \neq 0.

Los valores de la prueba de hipótesis se muestran en la Tabla 2, siendo para todos valores de p menores a 0.05. Por lo tanto, existe evidencia suficiente para rechazar la hipótesis nula. De esta forma, es posible inferir que existe correlación significativa entre las variables.

Tabla 1: Análisis exploratorio de variables, RStudio

Resumen (bd1: escala natural)

	Styrene	Benzene	Brent	Ps_Spot
Min.	451,1	285,3	19,8	828
1st Qu.	893,7	615,6	46,8	1070
Median	1011,9	736,4	55,1	1256
Mean	1043,0	727,8	55,16	1243
3rd Qu.	1243,2	849,7	63,7	1426
Max.	1881,3	1230,4	78,9	1562

Resumen (bd2: variación mensual)

	Styrene	Benzene	Brent	Ps_Spot
Min.	-0,365240	-0,661114	-0,616390	-0,154577
1st Qu.	-0,051592	-0,023493	-0,025970	0,023981
Median	0,000309	0,006196	0,030310	0,007669
Mean	0,007763	0,012009	0,011210	0,006888
3rd Qu.	0,070337	0,061253	0,675200	0,037764
Max.	0,402808	0,297185	0,467480	0,266161

(Fuente: Elaboración propia)

Tabla 2: Prueba correlación de Pearson

	Estadísticos	Benzene	Brent	Styrene
Ps_Spot	test statistic	6,1288	3,0167	9,3586
	df	62	62	62
	p-value	6,734E-08	0,0037030	1,813E-13
	cor	0,6142274	0,3577636	0,76519

(Fuente: Elaboración propia)

3.3 Análisis multivariado

Permite comprender a través de métodos estadísticos la información relevante de diferentes variables, entrega una visión general realista y precisa de lo que está sucediendo a nivel de la información, resulta una potente y confiable prueba de significación siempre y cuando se tenga claro que se espera en su utilización.

A la luz de la información contenida en las series de tiempo, los análisis previos realizados en ellas y apoyado en el uso del banco de modelos conmutados se ha obtenido como resultado que la regresión línea múltiple es el modelo estadístico adecuado para ser empleado en el pronóstico de precios para el poliestireno de alto impacto. En la sección 3.7 es posible apreciar los cálculos que apoyan esta selección desestimando el uso de los otros modelos (ARIMA, Holt Winters y Redes neuronales).

3.4 Regresión lineal múltiple

Este modelo permite hacer pronósticos basado en la relación existente entre una variable dependiente y otras variables independientes. Para el análisis se considera la base de datos con variaciones mensuales bd2, con el fin de evitar un

problema de regresión espuria considerando variables independientes no estacionarias.

Previo a la elaboración del modelo se evalúa la existencia de correlación entre variables. Para ello, se utilizó el coeficiente de correlación de Pearson expuesto en la Tabla 2 y un gráfico de análisis bivariado, que se muestra en la Figura 4.

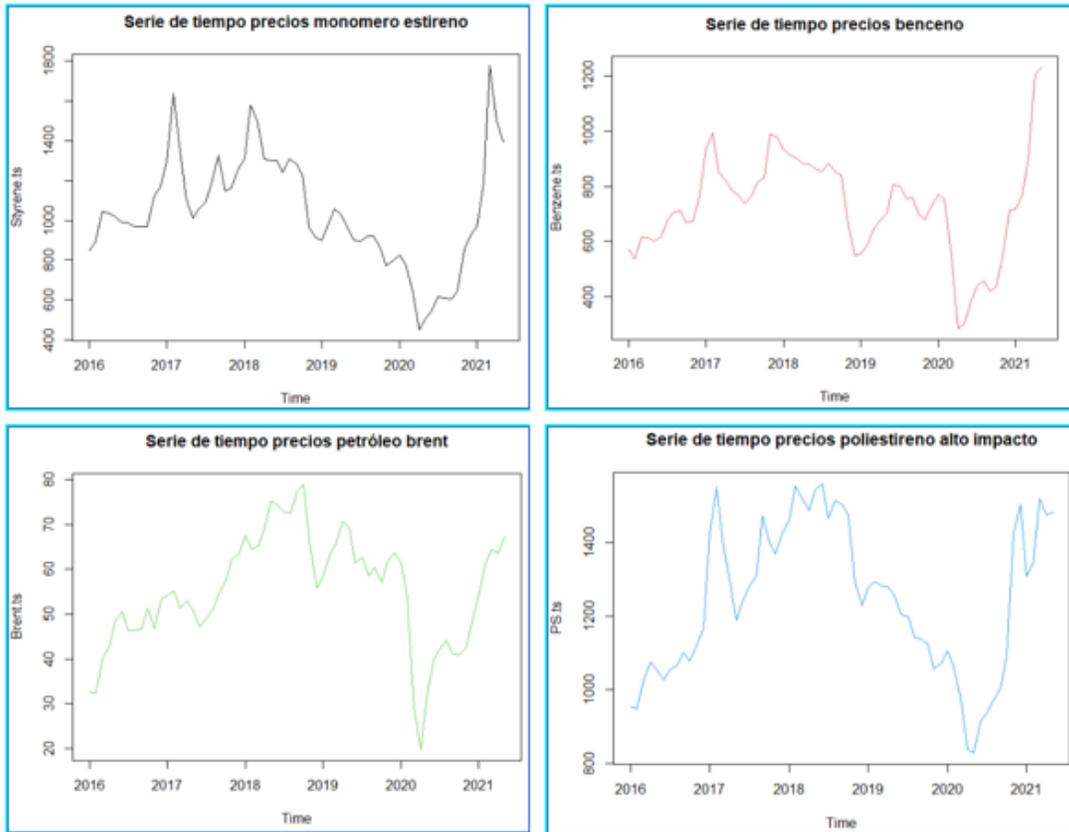


Figura 3: Series de tiempo de cada variable
(Fuente: Elaboración propia)

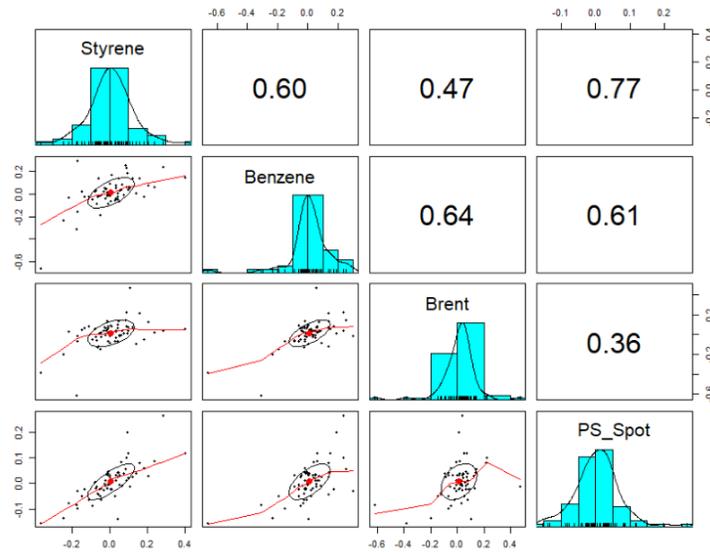


Figura 4: Análisis bivariado de las variables en estudio (bd2)
(Fuente: Elaboración propia)

En la Figura 5 se ilustra el modelo de regresión lineal múltiple para todas las variables disponibles. En una primera aproximación se identifica que la variable de precios de “petróleo Brent” no resultó ser estadísticamente significativa. Por lo tanto, se eliminó y se calibró nuevamente el modelo. Los resultados de esta nueva iteración se muestran en la Figura 6.

En esta nueva iteración no se aprecia que existan nuevas variables con poca significancia por lo que es posible inferir que el nuevo modelo explica a la variable dependiente lo siendo válido para ser usado en el

pronóstico de precios del poliestireno de alto impacto, a su vez el valor de R2 ajustado es 61,17% lo que demuestra que la variabilidad actual del precio es explicado por el modelo de regresión múltiple con dos variables independientes válidas que no generan una regresión espuria.

Reemplazando los valores del modelo en la ecuación (1) se tiene.

$$y^{\wedge}=0,002631+0,126285*(Benzene)+0,353044*(Styrene) \tag{2}$$

Residuals				
Min	1Q	Median	3Q	Max
-0,14943	-0,01988	-0,00187	0,02004	0,12946

Coefficients	Estimate	Std.Error	t value	pr(> t)
(Intercept)	0,002903	0,005435	0,0534	0,59519
Brent	-0,079569	0,053074	-1,4990	0,13906
Benzene	0,170078	0,057888	2,9380	0,00468 **
Styrene	0,365057	0,055805	6,5420	1,51E-08 ***

Códigos	*** (0)	** (0,001)	* (0,01)	. (0,01)
---------	---------	------------	----------	----------

Residual standart error: 0,04328 on 61 degrees of freedom
Multiple R-squared: 0,6376 Adjusted R-squared: 0,6195
F-statistic: 35,2 on 3 and 60 DF,
p-value : 2,996E-13

Figura 5: Primera versión del modelo de regresión lineal múltiple (Fuente: Elaboración propia)

3.5 Validación de resultados

Análisis gráfico

En cuanto al análisis de residuos, la Figura 7 muestra la existencia de valores atípicos. Específicamente, las observaciones 12, 58 y 60 las deben ser sometidas a un análisis puntual para determinar si son errores de digitación o corresponden a algún fenómeno temporal. El período de estas observaciones corresponde a enero 2017, noviembre 2020 y enero 2021, respectivamente.

Informes de distintos medios dan a conocer los eventos que ocasionan los datos atípicos señalados en el párrafo anterior, en la Tabla 3 se resumen los acontecimientos.

LM (formula = PS_Spot ~ Benzene + Styrene, data = bd2)

Residuals				
Min	1Q	Median	3Q	Max
-0,15668	-0,01984	-0,00342	0,02102	0,13825

Coefficients	Estimate	Std.Error	t value	pr(> t)
(Intercept)	0,002631	0,005488	0,0479	0,6333
Benzene	0,126285	0,050487	2,5010	0,0151 *
Styrene	0,353044	0,055788	6,3280	3,26E-08 ***

Códigos *** (0) ** (0,001) * (0,01) . (0,01)

Residual standart error: 0,04372 on 61 degrees of freedom
Multiple R-squared: 0,6241 Adjusted R-squared: 0,6117
F-statistic: 50,63 on 2 and 61 DF,
p-value : 1,098E-13

Figura 6: Segunda versión del modelo de regresión lineal múltiple
(Fuente: Elaboración propia)

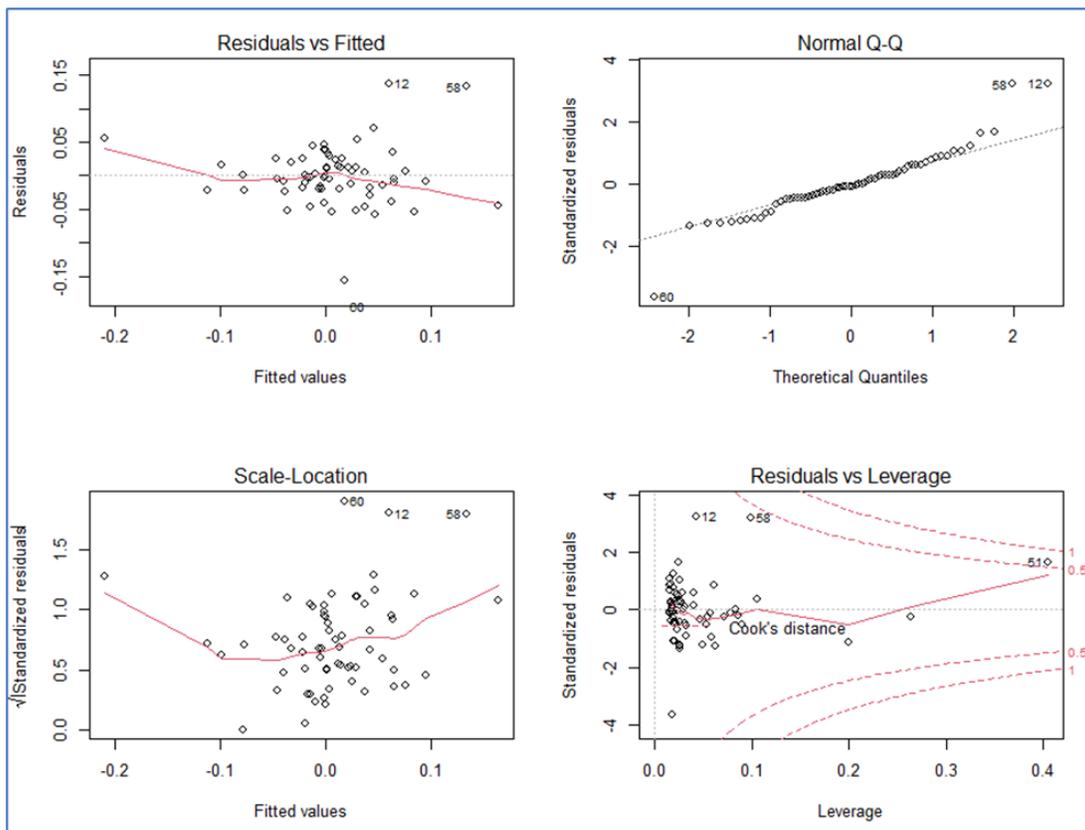


Figura 7: Gráfico de residuos
(Fuente: Elaboración propia)

Tabla 3: Eventos temporales que explican datos atípicos.

<p><i>Enero 2017.</i> <i>El aumento sostenido en el precio de las materias primas, principalmente el precio del petróleo Brent que en 12 meses aumentó un 77% y 79%, lo que desencadena un alza en sus derivados (Datosmacro, 2017).</i></p>
<p><i>Octubre 2020.</i> <i>La peor tormenta de nieve que batió récords en Estados Unidos, la ola de frío cruzó de norte a sur (Ortiz M. , 2020).</i></p>
<p><i>Diciembre 2020.</i> <i>Ascend informa de una parada no planificada en Florida. El alza de precios en las resinas podría acelerarse debido a un fallo de energía en la instalación de Pensacola ocasionado por las tormentas de Laura y Delta (Europe, 2020).</i></p>

(Fuente: Elaboración propia)

Normalidad

Se utilizó la prueba de Jarque-Bera para comprobar si los residuos tienen la asimetría y la curtosis de una distribución normal.

La hipótesis para contrastar la normalidad de los residuos es la siguiente:

H₀: Normalidad v/s H_a: No normalidad.

Los resultados de la prueba son:

X-squared = 35,504; df = 2; p-value = 1,952e-08. Entonces, es posible inferir que existe evidencia suficiente para rechazar H₀ por tanto no se cumple este supuesto.

Homocedasticidad

Esta propiedad implica que la varianza estimada de los residuos depende de los valores de las variables independientes. Para el cálculo de la homocedasticidad, se utilizó la prueba Studentized Breusch-Pagan.

La hipótesis para contrastar la Homocedasticidad es la siguiente:

H₀: Homocedasticidad v/s H_a: Heterocedasticidad.

Los resultados de la prueba son BP = 3,7494; df = 2; p-value = 0,1534. Dado que el p-value es menor al 5%, existe evidencia para concluir que se cumple la homocedasticidad.

Autocorrelación

Para este análisis se utiliza la prueba Durbin-Watson. El objetivo es detectar la presencia de autocorrelación en los residuos del modelo múltiple.

Las hipótesis para contrastar la Autocorrelación es las siguientes:

H₀: Independencia v/s H_a: Autocorrelación.

Los resultados de la prueba son DW = 1,7908; p-value = 0,1852. Por lo tanto, existe evidencia estadística para no rechazar H₀, lo que permite concluir que los términos del error son independientes positivamente.

Multicolinealidad

Para evaluar el grado de dependencia de las variables independientes, se realizó el cálculo del indicador de Factor de Inflación de la Varianza (VIF), el cual mide el efecto combinado que tienen las dependencias entre los regresores sobre la varianza. Los valores de la prueba son Benzene = 1,553213; Styrene = 1,553213.

En el análisis bivariado de la Figura 4, se deduce que existe correlación entre las variables explicativas. Esto es común en las variables de una regresión múltiple. Por lo tanto, y considerando los valores observados en la prueba, existe evidencia suficiente para concluir que no existe multicolinealidad.

Con el objetivo de lograr un mejor ajuste del modelo, se decide trabajar con variables ficticias con el fin de medir el factor cualitativo, y así contrastar si el efecto de éste es relevante en el modelo de regresión. Se definen tres variables ficticias representadas por d1, d2, d3, para cada uno de los períodos señalados anteriormente ver Tabla .3.

Este último modelo se expone en la Figura 8. Nótese que existe un mejor ajuste en comparación a los dos anteriores, obteniendo un p-value de 2,2E-16 y un valor de R² ajustado de 81,9%. Por lo tanto, existe evidencia

suficiente para concluir que el modelo es válido para ser usado en el pronóstico de precios del poliestireno y que la incorporación de variables ficticias resultó ser significativa.

3.6 Validación de resultados modelo con variables dummy

Este acápite tiene por objetivo validar las pruebas del nuevo modelo con variables ficticias, para ello se ejecutarán las mismas pruebas que en el modelo anterior. Análisis completo en Tabla 4.

Una variable dummy considera sustitutos numéricos a factores cualitativos en un modelo de regresión y así contrastar si el efecto de éste es relevante en el modelo.

LM (formula = PS_Spot ~ Brent + Benzene + Styrene + d1 + d2 + d3, data = bd2)				
Residuals				
Min	1Q	Median	3Q	Max
-0,05477	-0,01887	0,00000	0,01649	0,07836
Coefficients	Estimate	Std.Error	t value	pr(> t)
(Intercept)	0,001002	0,003826	0,262	0,7944
Benzene	0,087524	0,034891	2,509	0,0149 *
Styrene	0,326430	0,038972	8,376	1,44E-11 ***
d1	0,149733	0,030556	4,900	8,05E-06 ***
d2	0,151352	0,031499	4,805	1,13E-05 ***
d3	-0,153795	0,030143	-5,102	3,88E-06 ***
Códigos *** (0) ** (0,001) * (0,01) . (0,01)				
Residual standart error: 0,02986 on 58 degrees of freedom Multiple R-squared: 0,8333, Adjusted R-squared: 0,819 F-statistic: 58 on 5 and 58 DF, p-value : 2,2E-16				

Figura 8: Modelo regresión línea múltiple con variables dummy (Fuente: Elaboración propia)

Tabla 4: Validación de modelo con variables ficticias

Normalidad (prueba de Jarque Bera)		Homocedasticidad (prueba Studentized Breusch-Pagan)	
H ₀ : Normalidad	H _a : No normalidad	H ₀ : Homocedasticidad	H _a : Heteroscedasticidad
X-squared	0,38933	BP	2,7078
df	2	df	5
p-value	0,8231	p-value	0,7449
A un nivel de significancia del 5% la hipótesis nula H ₀ de normalidad no se rechaza, puesto que el valor de p-value > 0,05 por tanto se cumple este supuesto		A un valor de significancia usual del 5% la hipótesis nula de homocedasticidad no se rechaza, puesto que el valor de p-value > 0,05, por tanto, se cumple este supuesto.	

Autocorrelación	
(prueba Durbin-Watson)	
H₀: Independencia	H_a: Autocorrelación
DW	2,1196
p-value	0,6664
<p>A un valor de significancia usual del 5% la hipótesis nula de independencia no se rechaza, puesto que el valor de p-value > 0.05 por lo tanto los términos del error son independientes.</p>	

Multicolinealidad
(prueba factor de inflación de la varianza VIF)
<p>El objetivo es no encontrar multicolinealidad entre las variables independientes.</p> <p>En los resultados del cálculo del indicador todos los VIF son menores a 5, por lo tanto, no se detecta multicolinealidad entre las variables independientes (máximo valor obtenido fue de 1,591009)</p>

(Fuente: Elaboración propia)

Predicciones

En la Figura 9 se presenta un gráfico de predicción de precios para el poliestireno basado en el modelo de regresión lineal múltiple con variables ficticias (d1, d2, d3). Al comparar el valor real con la predicción, se aprecia que existe un ajuste significativo, lo que se corrobora con el R2 ajustado de los modelos.

La incorporación de variables ficticias de tipo temporal recoge información que no puede ser visualizada directamente sobre las observaciones, complementando el análisis previo elaborado en la Figura 7.

3.7 Iteraciones del banco de modelos conmutado Modelo ARIMA

Prueba Dickey-Fuller

Busca determinar la existencia o no de raíces unitarias en una serie de tiempo, en la Tabla 5 se presentan los valores de la prueba con el objetivo de tener evidencia para contrastar la siguiente hipótesis.

H₀: La serie no es estacionaria v/s H_a: La serie es estacionaria.

Como resultado de la prueba se puede inferir que existe evidencia suficiente en los valores de p-value para aceptar la hipótesis nula evidenciando con ello, la existencia de raíces unitarias por lo tanto las variables no son estacionarias.

Para apoyar la conclusión obtenida con Dickey-Fuller, se realiza una prueba gráfica de autocorrelación (es una

característica que consiste en que, elementos cercanos en el espacio o en el tiempo se parecen más entre sí que con respecto a elementos más lejanos, solamente por el hecho de estar cerca) y una prueba numérica con Box-Pierce bajo la siguiente hipótesis.

H₀: Los residuos no poseen autocorrelación entre si

H_a: Los residuos poseen autocorrelación entre si

Al analizar la información de la Figura 10 se concluye que existe evidencia suficiente para rechazar H₀. Los datos se distribuyen de forma independiente de modo que cualquier correlación observada en los datos es resultado de la aleatoriedad del proceso.

En la comparación de ambas pruebas se puede inferir que las series no son estacionarias, están guiadas por la tendencia, la cual se aprecia en el estadístico de la Figura 10 que va creciendo conforme avanza el tiempo.

Sobre el análisis de los datos se puede concluir que los rezagos de la serie (PS.ts) no tienen un impacto sobre el valor actual de ésta, el tiempo deja de ser significativo no existiendo una dependencia con el período anterior, por lo tanto, el modelo ARIMA no es recomendado para la utilización de pronóstico de precios para el poliestireno de alto impacto.

Modelo Holt Winters

Este modelo utiliza una técnica de suavizado exponencial, basado en estimaciones recursivas a

partir de los datos históricos, este método necesita que en la serie de tiempo exista tendencia y estacionalidad (ver Tabla 5), la cual no se encuentra presente en la serie.

Al no existir relación temporal en los datos no tiene sentido la aplicación del modelo ya que los resultados esperados pierden significancia y no representan la realidad del suceso en estudio.

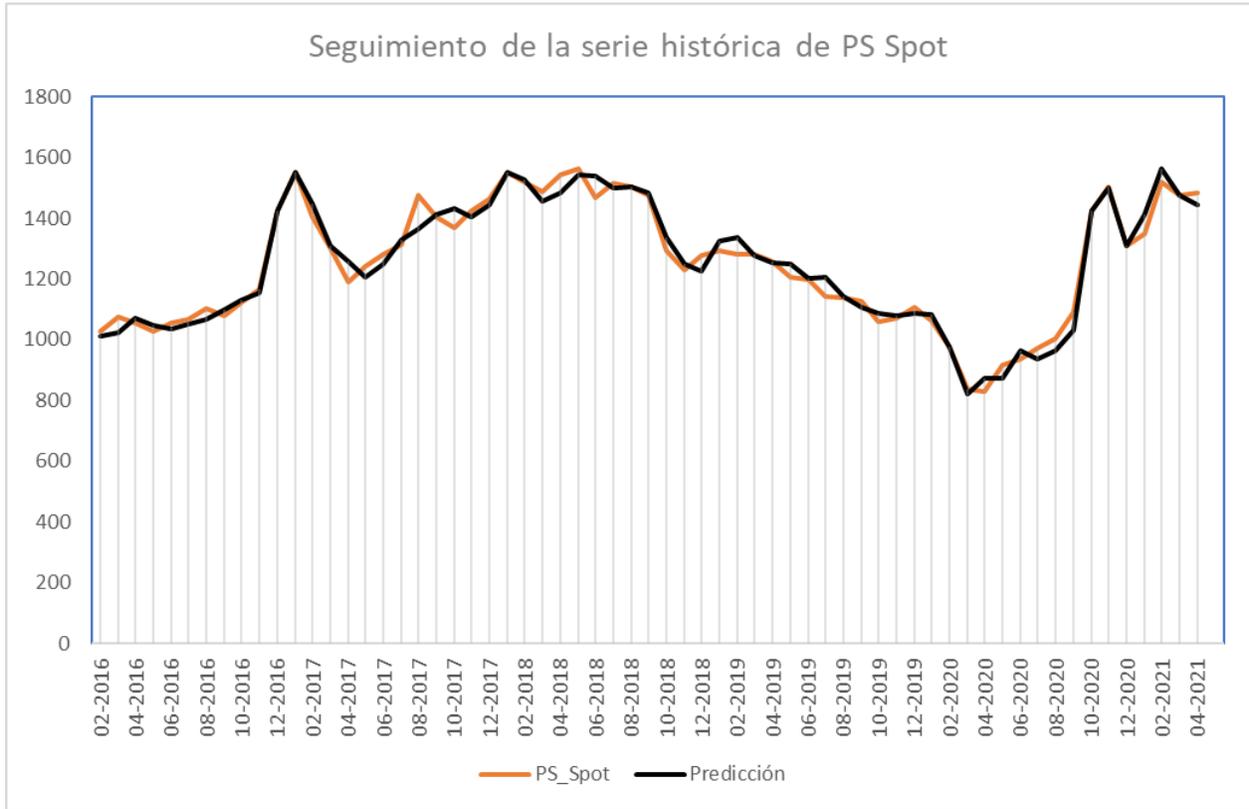


Figura 9: Gráfico comparativo predicción v/s valor real (Fuente: Elaboración propia)

Tabla 5 Resultado de la prueba aumentada Dickey-Fuller

Data	Styrene	Benzene	Brent	PS_Spot
Dickey-Fuller	-1,6507	-1,8647	-1,9757	-1,7353
Lag order	3	3	3	3
p-value	0,7166	0,6298	0,5849	0,6823
Alternative hypothesis	No Stationary			

(Fuente: Elaboración propia)

disponible tuvo un preprocesamiento de calidad. Estos hallazgos son coherentes con los resultados indicados por (Alliance, 2015) quien sostiene que la calidad de datos y número de ellos es fundamental para entrenar modelos econométricos.

En la búsqueda de un modelo aceptable, el uso de la minería de datos juega un rol preponderante al unir las diversas técnicas en el análisis de datos, desde la recolección y depuración de la información, pasar por un aprendizaje supervisado de los modelos hasta lograr un aprendizaje automático y finalmente una predicción significativa.

Durante la ejecución de este trabajo se han evidenciado algunas limitantes que han impedido alcanzar el máximo potencial de los modelos escogidos, dentro de ellos destacan.

- El alcance y restricción de cada técnica dependen directamente de la estructura de los datos (Alliance, 2015).
- Se debe considerar el modelamiento matemático que sustenta a cada modelo ya que existe uno adecuado para cada problema (David E. Rodríguez Guevara, 2019).
- Para este trabajo el tamaño de la muestra comprende 65 períodos confiables de información, que resultan escasos al momento de modelar una red neuronal, solo para el entrenamiento se necesita al menos un 70% de la muestra siendo ésta bastante reducida para alcanzar el potencial de la herramienta con una alta probabilidad de obtener un modelo sobre ajustado.
- El recorrido de la muestra es mensual para una mejor adaptación debería ser al menos diaria para determinar si la correlación observada depende del período anterior o si realmente responde a sucesos aleatorios. En su libro digital (Alliance, 2015) describe la importancia de las observaciones como píxeles dentro de una pantalla, al estar aislados éstos entregan una pequeña cantidad de información, pero cuando se los combina con una cantidad suficiente de píxeles en el orden correcto, el analista puede localizar significados nuevos y, a veces inesperados; por lo tanto, restringir la data solo a sucesos mensuales no es la mejor decisión que puede tomar el dueño del dato.

Es importante discutir elementos respecto a las variables dummy relacionados con los outliers encontrados en la serie. En primer lugar, estos sucesos corresponden a fenómenos naturales y económicos

(oferta/demanda) que impactaron directamente al precio del petróleo y a sus derivados, en algunos casos la escasez de éstos impulsó la tendencia alcista de los precios en la industria del plástico. También, es importante destacar que las variables económicas no proceden de un modelo experimental, siendo los agentes económicos los que hacen que las variables se influyan mutuamente tal como lo dice Morales (Morales, 2010) la alta correlación hace difícil precisar a partir de la información muestral los coeficientes asociados a dichas variables, y como consecuencia existe una precisión reducida en los modelos. Por lo tanto, para un mejor ajuste en el modelo de regresión lineal múltiple se decidió incorporar aspectos cualitativos dotándolo de mayor significancia y separando la correlación de estos fenómenos para aumentar la capacidad de predicción. Además, se realizaron pruebas de predicción para ambos modelos en cuyos resultados el R² ajustado pasó de 61,17% a 81,9%. Sobre la base de los dos modelos propuestos, no es posible concluir que sólo con la adición de variables ficticias un modelo sea más preciso que otro que no las tiene, la evidencia demuestra que la alta correlación entre variables hace difícil discriminar la relevancia para explicar la variable dependiente y por consiguiente una imprecisión en los resultados.

Finalmente, la actualización constante de la historia de compras permite que la simulación sea más precisa obteniendo más y mejor información del nuevo escenario lo que permite al analista ajustar el modelo y tomar una decisión, donde las estimaciones por sí solas son incapaces de hacerlo basado sólo en los registros históricos.

4. Conclusiones

Este trabajo establece que el uso de modelos econométricos aplicados en las series de tiempo de los precios del petróleo y sus derivados puede generar modelos predictivos capaces de pronosticar los precios futuros del poliestireno de alto impacto. En la búsqueda de estos modelos el uso de las herramientas analíticas sobre las series de tiempo, la iteración continua entre distintos modelos y la validación de todos los elementos y valores de los parámetros dan como resultado un modelo sin sesgo y con estimaciones más precisas. Por lo tanto, el modelo propuesto es capaz de explicar con un 95% de confianza un precio estimado de compra para el poliestireno de alto impacto.

En efecto, la evidencia muestra que los mercados del plástico están altamente influenciados por los precios

de los commodities, cada uno de estos compuestos mantiene una relación de dependencia entre ellas y el entorno que termina afectando al precio final del producto, Es necesario reconocer que el mercado atraviesa un importante momento de transformación, nuevas regulaciones al uso de empaques, el rechazo a los plásticos de un solo uso y la aparición cada vez más fuerte de los discursos de sostenibilidad que desafían a las compañías a encontrar alternativas distintas. Esta transformación social, sin duda, cambiará la lógica de la formación de precios, generando nuevos incentivos hacia la preferencia en el uso de plásticos reciclados, como resultado de esto la nueva información se transforma en variables de entrada para modelos capaces de interpretar y entregar un valor preciso al analista.

Dicho lo anterior, este trabajo contribuye a la comprensión de cómo un modelo seleccionado, desde un banco de modelos, permite al analista obtener un valor estimado del precio del poliestireno de alto impacto con un 95% de confianza.

En la literatura, siempre estuvo presente la idea sobre la calidad y cantidad de la información, y la importancia del data owner (experto encargado de garantizar la correcta gestión de los datos) que con cuya expertiz es capaz de administrar y suministrar a los usuarios la información específica que se requiera para un estudio en particular. Sobre esta base la brecha detectada tiene relación con la cantidad de información dispuesta para realización de los ensayos en este proyecto.

La tecnología avanza a pasos agigantados y la tendencia actual es la automatización de procesos recurrentes, en estadística predictiva el machine learning (aprendizaje automático) está ganando cada vez más adeptos ya que este tipo de modelos son capaces de aprender sin supervisión constante de un analista lo que difiere en gran medida con lo que se propone en este trabajo, donde todo el proceso ocurre en forma manual.

Referencias

- Alexandru Tatomir1, C. M. (2018). *Conceptual model development using a generic Features, Events, and Processes (FEP) database for assessing the potential impact of hydraulic fracturing on groundwater aquifers*. ADGEO.
- Alliance, B. T. (2015). *¿Por qué son tan importante los datos? Washington: BSA The Software Alliance*.
- AMBIENTE, M. D. (1 de Junio de 2016). *LEY 20920 Establece marco para la gestión de residuos, la responsabilidad extendida del productor y fomento al reciclaje*. Chile.
- AMBIENTE, M. D. (03 de 08 de 2018). *Ley 21100 Prohíbe la entrega de bolsas plásticas de comercio en todo el territorio nacional*. Chile. Obtenido de <http://bcn.cl/2fck6>
- Antonio Pulido San Roman, J. P. (2001). *Modelo Econométricos*. Madrid: Pirámide.
- Betancourt, D. (26 de Febrero de 2021). *Ingenio Empresa*. Obtenido de www.ingenioempresa.com/regresion-lineal
- Cohen, R. (2017). *Gestión de Recursos FI UBA, Pronósticos*.
- Coy Mondragón, G. E.-B. (2021). *Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales*. *Revista Mutis*, 11(1).
- Datosmacro. (31 de Enero de 2021). *Datosmacro*. Obtenido de [www.datosmacro.com: https://datosmacro.expansion.com/materias-primas/brent?dr=2017-01](https://datosmacro.expansion.com/materias-primas/brent?dr=2017-01)
- David E. Rodríguez Guevara, G. J. (2019). *Principios de Econometría*. Medellín, Colombia: Instituto Tecnológico Metropolitano.
- Douglas C. Montgomery, C. L. (2008). *Introduction to Time Series Analysis and Forecasting*. Canadá: Wiley Interscience.
- Emilio Ontiveros, V. L. (2017). *Economía de los Datos, Riqueza 4.0*. Madrid, España: Editorial Ariel S. A.
- Empresas, E. (25 de Agosto de 2021). *santandertrade.com*. Obtenido de <https://santandertrade.com/es/portal/analizar-mercados/chile/cifras-comercio-exterior>
- Europe GmbH, I. M. (07 de Marzo de 2021). *IG Group*. Obtenido de <https://www.ig.com/es/estrategias-de-trading/los-7-factores-que-afectan-el-precio-del-petroleo-190307>
- Europe, P. I. (03 de Septiembre de 2020). *ategi*. Obtenido de [www.ategi.com: https://www.ategi.com/2020/12/17/precios-al-alza-en-polimeros/](https://www.ategi.com/2020/12/17/precios-al-alza-en-polimeros/)
- Fernández Múnera, C. D. (28 de Mayo de 2021). *Herramienta de datos para el análisis de precios de polipropileno en ambientes coberturistas*. Medellín.

- ICIS. (06 de Agosto de 2021). ICIS Independent Commodity Intelligence Services. Obtenido de <https://www.icis.com/explore/>
- ICIS. (06 de Agosto de 2021). Independent commodity intelligence services. Obtenido de <https://www.icis.com/explore/resources/news/2021/07/23/10524881/topic-page-coronavirus-oil-price-direction-impact-on-chemicals>
- Industries, K. (17 de Mayo de 2021). knauf Industries. Obtenido de <https://knauf-industries.es/ventajas-de-usar-envases-de-plastico-para-alimentos/>
- Investing. (06 de Agosto de 2021). Investing. Obtenido de <https://es.investing.com/>
- John Scheirs, D. P. (2003). *Modern Styrenic Polymers*. John Wiley & Sons.
- Jones, H. (2019). *Las Redes Neuronales: Una guía esencial para principiantes de las redes neuronales artificiales y su papel en el aprendizaje automático y la inteligencia artificial*. Bravex Publications.
- Maldonado, S. (2019). *Modelos analíticos y minería de datos*. Santiago: e-book.
- Mercurio, E. (28 de Agosto de 2021). www.emol.cl. Obtenido de <https://www.emol.com/noticias/Economia/2020/08/28/996390/pib-tendencial-15-Presupuesto.html>
- Mexpolimeros. (4 de Marzo de 2021). Mexpolimeros. Obtenido de <https://www.mexpolimeros.com/pla.html>
- Morales, A. (2010). *Análisis de Regresión*. Madrid: @copyright Alfonso Novales.
- Ocerin, J. C. (1998). *Modelos econométricos y series temporales*. Barcelona, España: Reverté S.A.
- Ortiz, C. J. (2018). *Técnicas de IA aplicadas al pronóstico de la demanda en distribuidores*. Eae.
- Ortiz, M. (23 de Septiembre de 2021). *La opinión*. Obtenido de www.laopinion.com: <https://laopinion.com/2020/10/23/tormenta-de-nieve-en-octubre-batira-records-en-estados-unidos/>
- Perez, M. (2015). *Minería de Datos a Través de Ejemplos*. Mexico: Alfaomega.
- Platts, S. G. (06 de Agosto de 2021). S&P Global. Obtenido de <https://www.spglobal.com/en/>
- Ramírez, F. O. (2007). *Introducción a las Series de Tiempo*. Medellín: Universidad de Medellín.
- Roa, C. (Junio 2020). *Proyección de precio de largo plazo del cobre: un enfoque metodológico crítico*. Santiago: *Journal of Mining Engineering and Research*. 2 (1): 36-46. June, 2020.
- Seddon, D. (05 de Marzo de 2021). Duncan Seddon & Associates Pty. Ltd. Obtenido de <https://www.duncanseddon.com/c8-the-impact-of-oil-price-on-styrene-price/>
- Siegel, E. (2013). *Analítica Predictiva. Predecir el Futuro Utilizando Big Data*. Anaya Multimedia.
- Thomas T. Nagle, K. R. (2002). *Estrategia y Tácticas de Precios: Una Guía Para Tomar Decisiones Rentables*. Alhambra.
- Timarán-Pereira, S. R.-A.-Z.-T. (2016). *El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas*. Bogotá: Ediciones Universidad Cooperativa de Colombia.
- Timarán-Pereira, S. R.-A.-Z.-T.-P. (2016). *Descubrimiento de patrones de desempeño académico*. Bogotá: Ediciones Universidad Cooperativa de Colombia. Obtenido de <https://ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230-1?inline=1>
- Wichern, J. E.-D. (2009). *Pronósticos en los negocios* (9ª Edición). Pearson.

4 CONCLUSIONES GENERALES

Este trabajo establece que el uso de modelos econométricos aplicados en las series de tiempo de los precios del petróleo y sus derivados puede generar modelos predictivos capaces de pronosticar los precios futuros del poliestireno de alto impacto. En la búsqueda de estos modelos el uso de las herramientas analíticas sobre las series de tiempo, la iteración continua entre distintos modelos y la validación de todos los elementos y valores de los parámetros dan como resultado un modelo sin sesgo y con estimaciones más precisas. Por lo tanto, el modelo propuesto es capaz de explicar con un 95% de confianza un precio estimado de compra para el poliestireno de alto impacto.

En efecto, la evidencia muestra que los mercados del plástico están altamente influenciados por los precios de los commodities, cada uno de estos compuestos mantiene una relación de dependencia entre ellas y el entorno que termina afectando al precio final del producto, Es necesario reconocer que el mercado atraviesa un importante momento de transformación, nuevas regulaciones al uso de empaques, el rechazo a los plásticos de un solo uso y la aparición cada vez más fuerte de los discursos de sostenibilidad que desafían a las compañías a encontrar alternativas distintas. Esta transformación social, sin duda, cambiará la lógica de la formación de precios, generando nuevos incentivos hacia la preferencia en el uso de plásticos reciclados, como resultado de esto la nueva información se transforma en variables de entrada para modelos capaces de interpretar y entregar un valor preciso al analista.

También es importante considerar los siguientes aspectos que apoyaron la toma de decisión en la selección del modelo estadístico adecuado para obtener una mejor precisión en la predicción:

- Dado que todas las variables de este estudio poseen un origen común fue necesario estudiar la relación entre ellas y verificar la no existencia de una

dependencia entre las variables independientes y la variable dependiente, ya que de lo contrario no es posible determinar con exactitud la predicción del modelo y solo sería factible explicar su comportamiento global.

- Se identificó patrones válidos, novedosos y potencialmente útiles en las variables de entrada al sistema de modelos conmutados para la generación de variables ficticias que puedan explicar de mejor forma los valores fuera de rango y no tratarlos como anomalías de la investigación.
- Se verificó estadísticamente en base a los criterios de validación establecidos en el diseño del modelo conmutado, para ello se realizaron las pruebas de normalidad, homocedasticidad, autocorrelación y multicolinealidad para verificar la validez de la predicción.

Dicho lo anterior, este trabajo contribuye a la comprensión de cómo un modelo seleccionado, desde un banco de modelos, permite al analista obtener un valor estimado del precio del poliestireno de alto impacto con un 95% de confianza.

En la literatura, siempre estuvo presente la idea sobre la calidad y cantidad de la información, y la importancia del data owner (experto encargado de garantizar la correcta gestión de los datos) que con cuya expertiz es capaz de administrar y suministrar a los usuarios la información específica que se requiera para un estudio en particular. Sobre esta base la brecha detectada tiene relación con la cantidad de información dispuesta para realización de los ensayos en este proyecto.

La tecnología avanza a pasos agigantados y la tendencia actual es la automatización de procesos recurrentes, en estadística predictiva el machine learning (aprendizaje automático) está ganando cada vez más adeptos ya que este tipo de modelos son capaces de aprender sin supervisión constante de un analista lo que difiere en gran medida con lo que se propone en este trabajo, donde todo el proceso ocurre en forma manual.

4.1 Propuesta para trabajos futuros

Como continuación de este trabajo de tesis, hay varias líneas de desarrollo que quedan pendientes, y en las que es posible continuar trabajando; algunas de ellas, están más directamente relacionadas con este trabajo de tesis y son el resultado de preguntas que han ido surgiendo durante el proceso de investigación, otras son más tangenciales a la investigación. A continuación, se proponen trabajos futuros que pueden investigarse como conclusión de esta investigación.

Para un trabajo futuro sería importante revisar de cerca la Ley 20.920 que establece el marco para la gestión de residuos, la responsabilidad extendida del productor y fomento al reciclaje, ya que esta puede hacer cambiar la industria del packaging en Chile reemplazando el poliestireno de alto impacto por resinas mas amigables con el medio ambiente.

Ampliar la búsqueda de información, no sólo en los repositorios, sino, que también sobre los acontecimientos que estén provocando alzas y bajas en los precios que no puedan ser explicados por el aumento o baja en la demanda.

Implementar herramientas que permitan capturar información en tiempo real, cuyo objetivo es aumentar el tamaño muestral.

En base a lo anterior ampliar el banco de modelos conmutados con el objetivo de obtener nuevas propuestas de modelos y lograr una mejor precisión en el uso de otras técnicas cuantitativas.

Con el apoyo en el análisis bivariado de los datos es posible incluir los siguientes modelos para un futuro trabajo (S)ARIMA(X) y regresiones lineales simples para cada una de las categorías que el analista pueda encontrar al momento de analizar los datos (Ramesh, 2021).

5 REFERENCIAS BIBLIOGRAFICAS

- Alexandru Tatomir¹, C. M. (2018). Conceptual model development using a generic Features, Events, and Processes (FEP) database for assessing the potential impact of hydraulic fracturing on groundwater aquifers. ADGEO.
- Alliance, B. T. (2015). ¿Por qué son tan importante los datos? Washington: BSA The Software Alliance.
- AMBIENTE, M. D. (1 de Junio de 2016). LEY 20920 Establece marco para la gestión de residuos, la responsabilidad extendida del productor y fomento al reciclaje. Chile.
- AMBIENTE, M. D. (03 de 08 de 2018). Ley 21100 Prohíbe la entrega de bolsas plásticas de comercio en todo el territorio nacional. Chile. Obtenido de <http://bcn.cl/2fck6>
- Antonio Pulido San Roman, J. P. (2001). Modelo Económicos. Madrid: Pirámide.
- Betancourt, D. (26 de Febrero de 2021). Ingenio Empresa. Obtenido de www.ingenioempresa.com/regresion-lineal.
- Cohen, R. (2017). Gestión de Recursos FI UBA, Pronósticos.
- Coy Mondragón, G. E.-B. (2021). Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales. Revista Mutis, 11(1).
- Datosmacro. (31 de Enero de 2021). Datosmacro. Obtenido de www.datosmacro.com: <https://datosmacro.expansion.com/materias-primas/brent?dr=2017-01>
- David E. Rodríguez Guevara, G. J. (2019). Principios de Econometría. Medellín, Colombia: Instituto Tecnológico Metropolitano.
- Douglas C. Montgomery, C. L. (2008). Introduction to Time Series Analysis and Forecasting. Canadá: Wiley Interscience.
- Emilio Ontiveros, V. L. (2017). Economía de los Datos, Riqueza 4.0. Madrid, España: Editorial Ariel S. A.

Entreprises, E. (25 de Agosto de 2021). santandertrade.com. Obtenido de <https://santandertrade.com/es/portal/analizar-mercados/chile/cifras-comercio-exterior>

Europe GmbH, I. M. (07 de Marzo de 2021). IG Group. Obtenido de <https://www.ig.com/es/estrategias-de-trading/los-7-factores-que-afectan-el-precio-del-petroleo-190307>

Europe, P. I. (03 de Septiembre de 2020). ategi. Obtenido de www.ategi.com: <https://www.ategi.com/2020/12/17/precios-al-alza-en-polimeros/>

Fernández Múnera, C. D. (28 de Mayo de 2021). Herramienta de datos para el análisis de precios de polipropileno en ambientes coberturistas. Medellín.

ICIS. (06 de Agosto de 2021). ICIS Independent Commodity Intelligence Services. Obtenido de <https://www.icis.com/explore/>

ICIS. (06 de Agosto de 2021). Independent commodity intelligence services. Obtenido de <https://www.icis.com/explore/resources/news/2021/07/23/10524881/topic-page-coronavirus-oil-price-direction-impact-on-chemicals>

Industries, K. (17 de Mayo de 2021). knauf Industries. Obtenido de <https://knauf-industries.es/ventajas-de-usar-envases-de-plastico-para-alimentos/>

Investing. (06 de Agosto de 2021). Investing. Obtenido de <https://es.investing.com/>

John Scheirs, D. P. (2003). Modern Styrenic Polymers. John Wiley & Sons.

Jones, H. (2019). Las Redes Neuronales: Una guía esencial para principiantes de las redes neuronales artificiales y su papel en el aprendizaje automático y la inteligencia artificial. Bravex Publications.

Maldonado, S. (2019). Modelos analíticos y minería de datos. Santiago: e-book.

Mercurio, E. (28 de Agosto de 2021). www.emol.cl. Obtenido de <https://www.emol.com/noticias/Economia/2020/08/28/996390/pib-tendencial-15-Presupuesto.html>

Mexpolimeros. (4 de Marzo de 2021). Mexpolimeros. Obtenido de <https://www.mexpolimeros.com/pla.html>

- Morales, A. (2010). Análisis de Regresión. Madrid: @copyright Alfonso Novales.
- Ocerin, J. C. (1998). Modelos econométricos y series temporales. Barcelona, España: Reverté S.A.
- Ortiz, C. J. (2018). Técnicas de IA aplicadas al pronóstico de la demanda en distribuidores. Eae.
- Ortiz, M. (23 de Septiembre de 2021). La opinión. Obtenido de www.laopinion.com: <https://laopinion.com/2020/10/23/tormenta-de-nieve-en-octubre-batira-records-en-estados-unidos/>
- Perez, M. (2015). Minería de Datos a Través de Ejemplos. Mexico: Alfaomega.
- Platts, S. G. (06 de Agosto de 2021). S&P Global. Obtenido de <https://www.spglobal.com/en/>
- Ramesh, P. S. (13 de Diciembre de 2021). Towards Data Science. Obtenido de <https://towardsdatascience.com/keep-it-simple-keep-it-linear-a-linear-regression-model-for-time-series-5dbc83d89fc3>
- Ramírez, F. O. (2007). Introducción a las Series de Tiempo. Medellín: Universidad de Medellín.
- Roa, C. (Junio 2020). Proyección de precio de largo plazo del cobre: un enfoque metodológico crítico. Santiago: Journal of Mining Engineering and Research. 2 (1): 36-46. June, 2020.
- Seddon, D. (05 de Marzo de 2021). Duncan Seddon & Associates Pty. Ltd. Obtenido de <https://www.duncanseddon.com/c8-the-impact-of-oil-price-on-styrene-price/>
- Siegel, E. (2013). Analítica Predictiva. Predecir el Futuro Utilizando Big Data. Anaya Multimedia.
- Thomas T. Nagle, K. R. (2002). Estrategia y Tácticas de Precios: Una Guía Para Tomar Decisiones Rentables. Alhambra.
- Timarán-Pereira, S. R.-A.-Z.-T. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño

académico con árboles de decisión en las competencias genéricas. Bogotá: Ediciones Universidad Cooperativa de Colombia.

Timarán-Pereira, S. R.-A.-Z.-T.-P. (2016). Descubrimiento de patrones de desempeño académico. Bogotá: Ediciones Universidad Cooperativa de Colombia. Obtenido de

<https://ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230-1?inline=1>

Wichern, J. E.-D. (2009). Pronósticos en los negocios (9ª Edición). Pearson.

6 ANEXO: REPORTE DE PLAGIO

El reporte de posibilidad de plagio de este trabajo, con otros trabajos publicados entrega un porcentaje de similitud de: 1%



Indicador de plagio

Índice de plagio: 1%

Date: 19 November 2021

Información documento	Índice de plagio
Documento Artículo Fernando Pavez V010 HVC - Correcciones MV Final sin bibliografía.docx	1% Texto en la web
Tamaño 478,4 KB	
Nº palabras 3.895	
Subido el 19/11/2021	
Fuentes encontradas	
Fuente	Índice de Plagio
PLA Acido Poliláctico - Polímeros termoplásticos, elastómeros y aditivos https://www.mexpolimeros.com/pla.html	0%
Cifras del comercio exterior en Chile - Santandertrade.com https://santandertrade.com/es/portal/analizar-mercados/chile/cifras-comercio-exterior	0%
Red neuronal artificial - Wikipedia, la enciclopedia libre https://es.wikipedia.org/wiki/Red_neuronal_artificial	0%
Prueba de Dickey-Fuller - Wikipedia, la enciclopedia libre https://es.wikipedia.org/wiki/Prueba_de_Dickey-Fuller	0%