



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

UNIDAD ACADÉMICA PROFESIONAL HUEHUETOCA

**PRONÓSTICO DEL PRECIO DE ACCIONES DEL
S&P/BMV SECTOR SELECTO DE BIENES
DE CONSUMO BÁSICO A TRAVÉS DE
BOSQUES ALEATORIOS CON RAZONES
FINANCIERAS**

TESIS

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN ACTUARÍA

PRESENTA:

LEONARDO DANIEL GARCÍA GONZÁLEZ

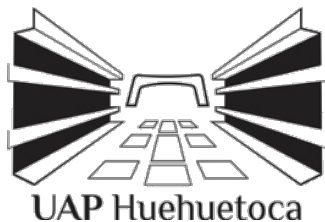
ASESOR:

MTR. EN ING. SAMUEL MARTÍNEZ BELLO

REVISORES:

DR. EN E. CHRISTIAN BUCIO PACHECO

DRA. EN E. MARÍA DEL PILAR LUGO MENDOZA



HUEHUETOCA, EDO. DE MÉXICO, MAYO DE 2025

Resumen

Este estudio implementa el método de Bosques Aleatorios para pronosticar los precios de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, utilizando razones financieras como variables predictoras. Se analiza la estructura y relevancia del índice en el mercado bursátil mexicano, destacando su impacto en la economía nacional. Además, se introduce el aprendizaje automático y los Bosques Aleatorios, explicando su funcionamiento y ventajas sobre otros modelos predictivos. La metodología se desarrolla en el software R con la paquetería randomForest, aplicándose a datos financieros de las empresas del índice.

Se realiza un análisis de correlación para seleccionar las variables más relevantes y estimar los modelos óptimos de Bosques Aleatorios. Para evaluar el rendimiento del modelo, se emplean diversas métricas y se aplica un intervalo de confianza al RMSE para enmarcar lo mejor posible el error. Finalmente, se presentan los pronósticos de cada una de las acciones que conforman el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico y se comentan los resultados obtenidos.

Palabras clave: Bosques Aleatorios, Variables Predictoras, Pronóstico de Precios.

Abstract

This study implements the Random Forest method to forecast stock prices in the S&P/BMV Consumer Staples Select Sector Index, using financial ratios as predictive variables. The research analyzes the structure and relevance of this index within the Mexican stock market, highlighting its impact on the national economy. Additionally, machine learning and Random Forests are introduced, explaining their functionality and advantages over other predictive models. The methodology is developed using R software with the randomForest package, applied to financial data from the index's constituent companies.

A correlation analysis is performed to select the most relevant variables and estimate optimal Random Forest models. Various metrics are used to assess model performance, and a confidence interval is applied to the RMSE to frame the error as accurately as possible. Finally, forecasts for each stock in the S&P/BMV Consumer Staples Select Sector Index are presented, and the results are discussed.

Keywords: Random Forests, Predictor Variables, Price Forecasting.

Índice general

Introducción	1
1 Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico	5
1.1 Visión general del índice	5
1.1.1 Estructura y componentes	6
1.2 Rendimiento y volatilidad	7
1.3 Relevancia del índice en el contexto de la Bolsa Mexicana de Valores	11
1.3.1 Características clave de la Bolsa Mexicana de Valores	12
1.3.2 El Índice S&P/BMV IPC: estructura y componentes	13
1.3.3 Comparativa entre el Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico y el S&P/BMV IPC	16
1.4 Impacto económico del Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico	19
1.5 Pronóstico del Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico	22
2 Bosques Aleatorios	25
2.1 Introducción al Machine Learning	26
2.1.1 Clasificación de los tipos de aprendizaje en Machine Learning	27
2.2 Árboles de decisión	28
2.2.1 Definición de un árbol de decisión para regresión	32
2.2.2 División de datos: Entrenamiento y prueba	33
2.2.3 Limitación de los árboles de decisión	34

2.3	Bosques aleatorios	36
2.3.1	Fundamentos de los bosques aleatorios	36
2.3.1.1	Bootstrap y Bagging	37
2.3.2	Definición de bosques aleatorios	38
2.3.3	Out of Bag Error (Error OOB)	39
2.3.4	Métricas de evaluación del modelo	40
2.3.5	Ajustes del modelo	42
3	Análisis de los resultados	45
3.1	Modelo de Bosques Aleatorios Propuesto	46
3.1.1	Formulación Matemática del Modelo	46
3.1.2	Base de Datos	48
3.1.3	Razones Financieras	49
3.1.4	Análisis de Correlación	52
3.2	Estimación de los bosques aleatorios	60
3.2.1	Resultados de los primeros modelos	61
3.2.2	Mejores modelos	67
3.2.3	Aplicación del intervalo de confianza al RMSE	80
3.3	Pronósticos	83
	Conclusiones	97
	Anexo	102
	Bibliografía	113

Índice de tablas

1.1	<i>Componentes del S&P/BMV del Sector Selecto de Bienes de Consumo Básico</i>	7
1.2	<i>Componentes del S&P/BMV IPC</i>	14
1.3	<i>Principales 10 componentes del S&P/BMV IPC</i>	15
1.4	<i>Comparativa de Rendimientos</i>	18
3.1	<i>Medidas financieras</i>	49
3.2	<i>Impacto de las Razones Financieras</i>	52
3.3	<i>Variables de entrada para cada acción</i>	59
3.4	<i>Resultados de los primeros modelos</i>	62
3.5	<i>Resultados del mejor modelo de bosques aleatorios para AC</i>	68
3.6	<i>Resultados del mejor modelo de bosques aleatorios para BIMBO A</i>	69
3.7	<i>Resultados del mejor modelo de bosques aleatorios para CHEDRAUI</i>	70
3.8	<i>Resultados del mejor modelo de bosques aleatorios para LA COMER UBC</i>	71
3.9	<i>Resultados del mejor modelo de bosques aleatorios para CUERVO</i>	72
3.10	<i>Resultados del mejor modelo de bosques aleatorios para FEMSA UBC</i>	73
3.11	<i>Resultados del mejor modelo de bosques aleatorios para GRUMA B</i>	74
3.12	<i>Resultados del mejor modelo de bosques aleatorios para HERDEZ</i>	75
3.13	<i>Resultados del mejor modelo de bosques aleatorios para KIMBER A</i>	76
3.14	<i>Resultados del mejor modelo de bosques aleatorios para KOFUBL</i>	77
3.15	<i>Resultados del mejor modelo de bosques aleatorios para SORIANA</i>	78
3.16	<i>Resultados del mejor modelo de bosques aleatorios para WALMEX</i>	79
3.17	<i>Intervalo de confianza al 95 % para el RMSE de cada acción</i>	81
3.18	<i>Pronóstico vs Valores Reales de AC</i>	84

3.19	<i>Pronóstico vs Valores Reales de BIMBO A</i>	85
3.20	<i>Pronóstico vs Valores Reales de CHEDRAUI</i>	86
3.21	<i>Pronóstico vs Valores Reales de LA COMER UBC</i>	87
3.22	<i>Pronóstico vs Valores Reales de CUERVO</i>	88
3.23	<i>Pronóstico vs Valores Reales de FEMSA UBC</i>	89
3.24	<i>Pronóstico vs Valores Reales de GRUMA B</i>	90
3.25	<i>Pronóstico vs Valores Reales de HERDEZ</i>	91
3.26	<i>Pronóstico vs Valores Reales de KIMBER A</i>	92
3.27	<i>Pronóstico vs Valores Reales de KOFUBL</i>	93
3.28	<i>Pronóstico vs Valores Reales de SORIANA</i>	94
3.29	<i>Pronóstico vs Valores Reales de WALMEX</i>	95

Índice de figuras

1.1	<i>Rendimientos diarios del índice en los últimos 10 años</i>	8
1.2	<i>Distribución de los rendimientos</i>	10
1.3	<i>Ponderación del S&P/BMV IPC por sector (2024)</i>	13
1.4	<i>Precios Históricos (datos expresados en base 100)</i>	17
2.1	<i>Ejemplo donde no se realiza mediante particiones recursivas</i>	29
2.2	<i>Ejemplo donde es realizada mediante particiones recursivas</i>	30
2.3	<i>Diagrama de árbol</i>	31
2.4	<i>Ejemplo de correlación de árboles de decisión</i>	35
3.1	<i>Análisis de correlación (1)</i>	53
3.2	<i>Análisis de correlación (2)</i>	54
3.3	<i>Análisis de correlación (3)</i>	55
3.4	<i>Análisis de correlación (4)</i>	56
3.5	<i>Análisis de correlación (5)</i>	57
3.6	<i>Disminución del error con base en el aumento de árboles en cada modelo (1)</i>	62
3.7	<i>Disminución del error con base en el aumento de árboles en cada modelo (2)</i>	63
3.8	<i>Disminución del error con base en el aumento de árboles en cada modelo (3)</i>	64
3.9	<i>Afinación del argumento mtry (1)</i>	65
3.10	<i>Afinación del argumento mtry (2)</i>	66
3.11	<i>Características del mejor modelo de AC</i>	68
3.12	<i>Características del mejor modelo de BIMBO A</i>	69
3.13	<i>Características del mejor modelo de CHEDRAUI</i>	70

3.14	<i>Características del mejor modelo de LA COMER UBC</i>	71
3.15	<i>Características del mejor modelo de CUERVO</i>	72
3.16	<i>Características del mejor modelo de FEMSA UBC</i>	73
3.17	<i>Características del mejor modelo de GRUMA B</i>	74
3.18	<i>Características del mejor modelo de HERDEZ</i>	75
3.19	<i>Características del mejor modelo de KIMBER A</i>	76
3.20	<i>Características del mejor modelo de KOFUBL</i>	77
3.21	<i>Características del mejor modelo de SORIANA</i>	78
3.22	<i>Características del mejor modelo de WALMEX</i>	79
3.23	<i>Dirección de los precios de AC</i>	84
3.24	<i>Dirección de los precios de BIMBO A</i>	85
3.25	<i>Dirección de los precios de CHEDRAUI</i>	86
3.26	<i>Dirección de los precios de LA COMER UBC</i>	87
3.27	<i>Dirección de los precios de CUERVO</i>	88
3.28	<i>Dirección de los precios de FEMSA UBC</i>	89
3.29	<i>Dirección de los precios de GRUMA B</i>	90
3.30	<i>Dirección de los precios de HERDEZ</i>	91
3.31	<i>Dirección de los precios de KIMBER A</i>	92
3.32	<i>Dirección de los precios de KOFUBL</i>	93
3.33	<i>Dirección de los precios de SORIANA</i>	94
3.34	<i>Dirección de los precios de WALMEX</i>	95

Introducción

El sector de bienes de consumo básico representa uno de los pilares más sólidos y resilientes de la economía mexicana. Este sector agrupa a las empresas más relevantes que producen y comercializan bienes esenciales para la población, como alimentos, bebidas, productos de higiene y otros artículos de primera necesidad. La demanda de estos productos se mantiene constante incluso en contextos económicos adversos, lo que convierte a este sector en un termómetro clave para medir el desempeño económico del país. En este sentido, el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico se erige como una herramienta indispensable para analizar el comportamiento de las principales empresas del sector, ofreciendo un panorama claro de su evolución y su impacto en la economía nacional.

El consumo privado, como componente principal del Producto Interno Bruto (PIB), representa aproximadamente el 65 % del PIB en México, según datos del Banco Mundial (2023). Este dato refleja la importancia del sector de bienes de consumo básico en la dinámica económica del país. Sin embargo, el sector no está exento de desafíos. Durante los últimos años, factores como la incertidumbre política, la desaceleración en la creación de empleos, episodios de inseguridad y la persistente inflación han afectado el consumo privado, generando una tendencia a la baja en varios trimestres, como lo señala INEGI (2023). Estos elementos han influido en el desempeño del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, el cual ha experimentado fluctuaciones significativas en los últimos años, incluyendo una caída pronunciada en los últimos meses de 2023 debido a factores macroeconómicos como

la inflación y el aumento de las tasas de interés.

Para los inversionistas, comprender el comportamiento de este índice es de vital importancia, ya que les permite identificar oportunidades de inversión en un sector que, por su naturaleza resiliente, ofrece cierta estabilidad incluso en escenarios económicos volátiles. Sin embargo, la volatilidad inherente a los mercados financieros y la complejidad de los factores que influyen en los precios de las acciones hacen que la predicción precisa de su comportamiento sea un desafío constante. Los métodos tradicionales de análisis financiero, como las regresiones lineales y los modelos econométricos, han demostrado ser insuficientes para capturar la complejidad y la volatilidad de los mercados financieros. Estos enfoques suelen asumir relaciones lineales entre las variables y no logran adaptarse adecuadamente a cambios abruptos en las condiciones del mercado. Además, son sensibles al ruido en los datos y carecen de la flexibilidad necesaria para modelar interacciones no lineales entre múltiples variables.

En contraste, las técnicas de aprendizaje automático, como los bosques aleatorios, han surgido como una alternativa robusta y eficaz para abordar estos desafíos. Los bosques aleatorios, una técnica basada en la combinación de múltiples árboles de decisión, han demostrado ser particularmente efectivos para manejar grandes volúmenes de datos, capturar relaciones no lineales y reducir el riesgo de sobreajuste. Esta metodología permite integrar diversas variables financieras, como razones financieras y datos históricos de precios, para generar pronósticos más precisos y confiables. Además, los bosques aleatorios son capaces de identificar patrones complejos en los datos que los métodos tradicionales no pueden detectar, lo que los convierte en una herramienta poderosa para el análisis financiero. Autores como Zhou et al. (2021) y Gu et al. (2020) han destacado la eficacia de los bosques aleatorios en la predicción de variables financieras, resaltando su capacidad para manejar datos no lineales y su robustez frente al ruido en los datos.

A pesar de su potencial, la aplicación de bosques aleatorios en el mercado financiero mexicano, específicamente en el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, ha sido poco explorada. Esta brecha en la investigación limita el aprovechamiento de herramientas

modernas de análisis que podrían aportar mayor certidumbre y transparencia al mercado financiero del país. Además, la falta de estudios aplicados a este sector específico restringe la comprensión de los factores clave que influyen en el desempeño de las empresas incluidas en el índice, lo que representa una oportunidad para los inversionistas que buscan optimizar sus estrategias de inversión.

La presente investigación tiene como objetivo principal implementar un modelo de bosques aleatorios para mejorar el pronóstico de los precios de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, utilizando razones financieras como variables predictoras. Este objetivo responde a la pregunta de investigación: ¿Cómo implementar un modelo de bosques aleatorios para mejorar el pronóstico de los precios de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, utilizando razones financieras como variables predictoras? Como hipótesis, se propone que el uso de un modelo de bosques aleatorios mejora significativamente la precisión de los pronósticos, permitiendo contrastar las estimaciones con los valores reales y evaluar su grado de exactitud.

Para los inversionistas, los resultados de esta investigación ofrecen una ventaja competitiva al proporcionar pronósticos más confiables y precisos sobre el comportamiento de las acciones del sector. Esto les permitirá tomar decisiones informadas, gestionar riesgos de manera más efectiva y capitalizar oportunidades en un mercado dinámico y en constante evolución. Además, el uso de técnicas avanzadas de aprendizaje automático puede ayudar a identificar tendencias y patrones que no son evidentes mediante métodos tradicionales, lo que se traduce en una mayor eficiencia en la asignación de recursos y en la optimización de carteras de inversión.

Para lograr este objetivo, la investigación se estructura en tres capítulos. El primer capítulo proporciona una visión general del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, detallando su estructura, componentes e importancia en la economía mexicana. Se analiza su rendimiento a lo largo del tiempo y su relación con otros índices bursátiles, como el S&P/BMV IPC, con el fin de evidenciar posibles oportunidades para los inversionistas.

Además, se destaca el papel del sector como un motor de crecimiento y estabilidad económica, especialmente en contextos de incertidumbre.

El segundo capítulo aborda la metodología empleada, con un enfoque en los conceptos de aprendizaje automático, árboles de decisión y bosques aleatorios. Se explica el funcionamiento de estas técnicas, sus ventajas y limitaciones, y se detalla el proceso paso a paso para la implementación del modelo en el contexto de esta investigación. Este capítulo sienta las bases teóricas y prácticas necesarias para comprender cómo se generarán los pronósticos que serán analizados en el siguiente apartado.

En el tercer capítulo se presenta el análisis de los resultados, donde se estima el modelo de bosques aleatorios utilizando la paquetería randomForest en el programa R. Se seleccionan las variables predictoras a partir de un análisis de correlación para evitar redundancias y garantizar la independencia entre ellas. Posteriormente, se estima el modelo para cada una de las acciones que conforman el índice, seleccionando los mejores modelos basados en su desempeño. A estos modelos se les aplica un intervalo de confianza al RMSE (Error Cuadrático Medio) para enmarcar con la mayor precisión posible el error del modelo y tenerlo presente en las conclusiones. Finalmente, se presentan los pronósticos generados para cada una de las acciones que integran el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

Por último, se presentan las conclusiones y se evalúa el desempeño del modelo, destacando su capacidad para predecir el comportamiento de los precios de las acciones y su contribución al análisis financiero en el mercado mexicano. Se analiza el impacto de las variables predictoras en los resultados y se contrastan las estimaciones con los valores reales para evaluar la precisión del modelo. Asimismo, los resultados de este estudio muestran una herramienta para los inversionistas para emplearla en sus estrategias de inversión, una gestión de riesgos de manera más efectiva y capitalizar las oportunidades que ofrece un sector tan dinámico y esencial como el de bienes de consumo básico. Finalmente, se discuten las limitaciones del estudio, así como las recomendaciones y posibles líneas de trabajo futuro derivadas de los hallazgos obtenidos.

Capítulo 1

Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico

En este capítulo se comienza con una visión general del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, detallando su estructura y los componentes que lo integran. Asimismo, se aborda la importancia del índice en el desempeño de la Bolsa Mexicana de Valores y su impacto directo en su principal indicador bursátil, el S&P/BMV IPC; todo esto con la finalidad de analizar el rendimiento a lo largo del tiempo entre dichos índices, con el objetivo de evidenciar posibles oportunidades para los inversionistas. Finalmente, se explica el impacto que tiene este índice en la economía mexicana, destacando su rol como un sector resiliente y necesario, fundamental para el crecimiento y estabilidad económica del país.

1.1. Visión general del índice

El índice S&P/BMV Sector Selecto de Bienes de Consumo Básico es un indicador del comportamiento específico de las empresas del sector selecto de bienes de consumo básico, que

se encuentran enlistadas en la Bolsa Mexicana de Valores (BMV), se lanzó por primera vez el 28 de Agosto de 2012 por S&P Dow Jones Indices teniendo como divisa de cálculo al peso mexicano (MXN). Su rebalanceo se hace de manera anual, específicamente en marzo, el índice sigue una metodología de ponderación de capitalización de mercado modificada, lo que significa que el peso asignado a cada empresa está en función de su valor de mercado, toda la información antes mencionada fue rescatada de S&P Dow Jones Indices (2024a).

1.1.1. Estructura y componentes

El objetivo principal del índice es reflejar el desempeño del mercado accionario de emisoras clave que producen y comercializan bienes esenciales para el consumo cotidiano. Es por ello, que las emisoras que se encuentran en el índice pasan por una selección rigurosa y específica, deben cumplir con ciertos criterios; ser líderes en el sector, que tengan una mayor capitalización bursátil, deben tener un historial consistente de operaciones en la BMV y mostrar un buen desempeño financiero. Por esta razón son pocas las empresas que llegan a este selecto índice, pues no muchas empresas cumplen con estos criterios y consistencia en el mercado.

En la Tabla 1.1 se muestran las 12 emisoras y su símbolo de cotización¹ en la Bolsa Mexicana de Valores. El orden en que se encuentran, es el que tiene mayor ponderación en el índice, siendo así que, Grupo Bimbo se convierte en la empresa más importante, con una ponderación de 12.7%, lo que refleja su relevancia en el sector, además su alta ponderación indica que las fluctuaciones en su desempeño financiero tienen un impacto significativo en el comportamiento general del índice. Por otro lado, la emisora con menor representación en el índice es Grupo Herdez con una ponderación $\leq 1.5\%$, en comparación con otras empresas del sector su participación es limitada. Pero a pesar de su menor peso, su inclusión subraya la importancia de la diversidad en la representación de empresas dentro del índice.

¹Un símbolo de cotización es un código único, generalmente una combinación de letras, que se utiliza para identificar a una empresa o activo en la bolsa de valores.

Tabla 1.1 Componentes del S&P/BMV del Sector Selecto de Bienes de Consumo Básico

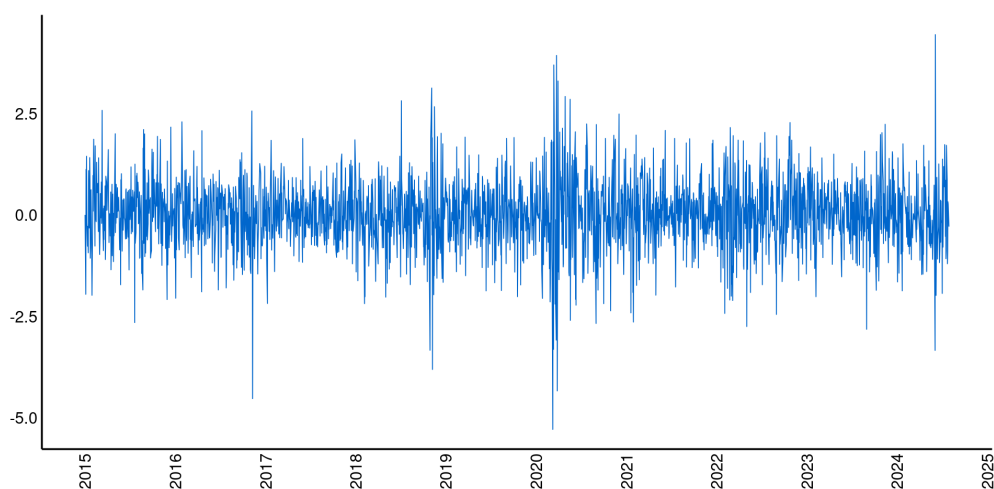
Componentes del ultimo Rebalanceo (2024)	
Empresa	Símbolo de cotización
1. Grupo Bimbo	BIMBOA
2. Fomento Económico Mexicano	FEMSAUB
3. Arca Continental	AC
4. Walmart de México	WALMEX
5. Gruma B	GRUMAB
6. Coca-Cola Femsa	KOFUBL
7. Kimberly Clark de México	KIMBERA
8. Grupo Comercial Chedraui	CHDRAUIB
9. La Comer	LACOMERUBC
10. Becele	CUERVO
11. Organización Soriana	SORIANAB
12. Grupo Herdez	HERDEZ

Fuente: Elaboración propia con datos de S&P Dow Jones Indices.

Es importante destacar que las 10 empresas principales constituyen el 97% de todo el índice, lo que refleja la concentración de valor en las compañías más representativas del sector. Este hecho pone de manifiesto la rigurosa selección de activos para integrar un indicador tan relevante como el S&P/BMV Sector Selecto de Bienes de Consumo Básico, asegurando que las empresas incluidas sean líderes en su ramo y esenciales para el análisis del desempeño económico del sector y del mercado financiero en general.

1.2. Rendimiento y volatilidad

Conocer el rendimiento del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico es fundamental debido a la naturaleza estable del sector que representa. A diferencia de otros sectores más expuestos a la volatilidad, este sector es menos susceptible a fluctuaciones extremas, dado que los bienes de consumo básico son esenciales para la población en cualquier circunstancia. A continuación, la Figura 1.1 muestra una evaluación de su rendimiento en estos últimos años, destacando su papel como indicador confiable dentro del mercado.

Figura 1.1 Rendimientos diarios del índice en los últimos 10 años

Fuente: Elaboración propia realizada con datos de S&P Dow Jones Índices.

El índice S&P/BMV Sector Selecto de Bienes de Consumo Básico muestra poca variación a lo largo del tiempo, con la mayoría de sus rendimientos situándose entre -2.5% y 2.5% , y una media de $\mu = 0.0218\%$. Esto evidencia un rango acotado de fluctuación que refleja la estabilidad intrínseca del sector. Esta estabilidad se atribuye a la baja elasticidad de la demanda de productos básicos: incluso en escenarios económicos adversos, los consumidores tienden a priorizar alimentos, artículos de higiene y otros productos esenciales. Además, el índice está compuesto por empresas grandes y consolidadas, caracterizadas por márgenes de ganancia estables y una menor exposición a riesgos sistémicos, lo que fortalece su capacidad de resistencia frente a la incertidumbre económica.²

Sin embargo, existen fluctuaciones en los rendimientos muy pronunciadas que corresponden a períodos específicos, como se describe a continuación:

- 10 de noviembre de 2016: Se registró una variación de -4.53% , principalmente debido al impacto de la victoria de Donald Trump en las elecciones presidenciales de Estados Unidos, llevadas a cabo el 8 de noviembre de 2016. Este evento generó incertidumbre

²El riesgo sistémico es el riesgo de que un evento o crisis afecte a todo el sistema financiero o económico de un país o región, no solo a una empresa o sector específico.

respecto a las políticas económicas, afectó el tipo de cambio y ocasionó fuga de capitales. A pesar de estas presiones, el sector selecto de bienes de consumo básico mostró mayor resiliencia en comparación con sectores como materiales y manufactura, que tuvieron retracciones de entre -5% y -7% , según Bekaert y Engstrom (2017). Además, autores como Cárdenas y Cruz (2017) analizaron cómo la elección de Trump afectó al mercado accionario mexicano y destacaron que, aunque hubo volatilidad generalizada, el sector de bienes de consumo básico mostró una respuesta más estable ante la incertidumbre política.

- 8 de noviembre de 2018: Se observó una caída de -3.81% en el índice del sector selecto de bienes de consumo básico de la BMV. Este descenso estuvo influenciado por la incertidumbre política y económica derivada de las elecciones intermedias en Estados Unidos, la guerra comercial entre EE.UU. y China, y preocupaciones sobre las políticas internas de México, factores que incrementaron la volatilidad de los mercados. Moreno-Brid y Paunovic (2018) analizaron cómo estas tensiones, en particular las comerciales y políticas internas, afectaron al mercado mexicano y destacaron que el sector de bienes de consumo básico mostró menor afectación en comparación con sectores como manufactura y materiales, más expuestos a los riesgos sistémicos.
- 9 de marzo de 2020: Esta fecha marcó el rendimiento *mín* del índice, con una caída de -5.29% , coincidiendo con el período más crítico de la pandemia de COVID-19. Durante este lapso, tanto la Bolsa Mexicana de Valores (BMV) como los mercados globales experimentaron caídas severas por el pánico financiero y las incertidumbres económicas. El S&P/BMV IPC, el principal índice de la Bolsa Mexicana de Valores, perdió aproximadamente un $24\% - 25\%$ desde mediados de febrero hasta finales de marzo. Los sectores más afectados fueron energía, con pérdidas de entre 30% y 35% , y materiales, con caídas de hasta 30% , de acuerdo con Rodríguez y Vargas (2020).

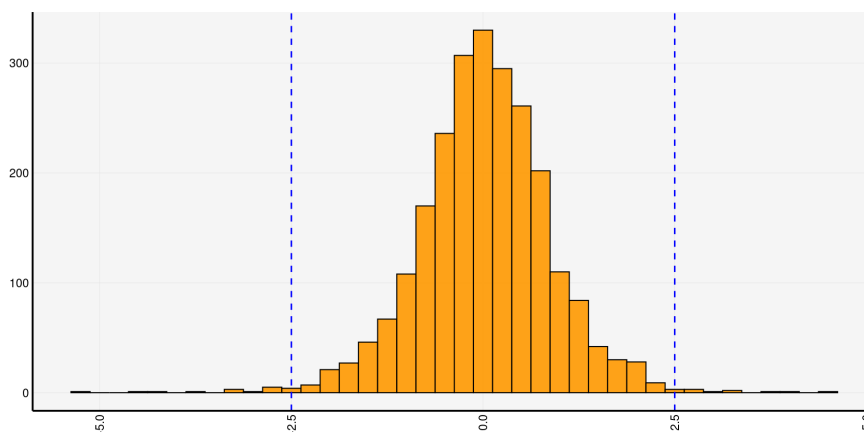
En contraste, el sector selecto de bienes de consumo básico mostró una mayor resistencia frente a la incertidumbre generada por la pandemia. Ramelli y Wagner (2020) analizaron la reacción del mercado al COVID-19 y concluyeron que sectores sensibles como energía

y transporte sufrieron fuertes caídas, mientras que sectores defensivos como bienes de consumo básico resistieron mejor, lo cual se reflejó en una disminución moderada del índice de -5.39% . Esto confirma la adaptabilidad de los sectores menos vulnerables durante crisis globales de gran impacto.

- 4 de junio de 2024: Se registró el máximo rendimiento diario del índice en los últimos 10 años, con un incremento del 4.44% . Este rendimiento fue impulsado por un rebote técnico tras la caída del "lunes negro" del 3 de junio de 2024, cuando el índice perdió un 3.34% debido a la reacción inicial del mercado frente a la victoria de Claudia Sheinbaum en las elecciones presidenciales. Según el medio digital Infobae (2024), el repunte del 4 de junio representó un ajuste en el mercado, a medida que los inversionistas procesaron las implicaciones del triunfo de Sheinbaum y estabilizaron sus posiciones. Este contexto político, junto con la naturaleza defensiva del sector selecto de bienes de consumo básico, favoreció su desempeño positivo en el índice.

Este patrón se puede explicar de mejor manera en la Figura 1.2 donde se aprecia más detalladamente la distribución de los rendimientos, con la finalidad de observar la frecuencia y magnitud de las variaciones en el S&P/BMV Sector Selecto de Bienes de Consumo Básico.

Figura 1.2 *Distribución de los rendimientos*



Fuente: Elaboración propia realizada con datos de S&P Dow Jones Índices.

La distribución de los rendimientos resalta características leptocúrticas, donde las colas capturan estas variaciones excepcionales y hay mayor concentración en la media. Por lo tanto, se observa que el 98.96 % de los rendimientos están acotados entre -2.5 % y 2.5 %, y solo el 1.04 % están fuera de ese rango generalmente asociados a eventos extraordinarios, además la distribución de los rendimientos confirma la concentración alrededor de la $\mu = 0.0218$ %, con una frecuencia de 348, lo que representa el 14.44 % de los rendimientos totales, por último se tiene una desviación estándar anual de 13.60 %.

En general, las fluctuaciones extremas observadas en el histograma así como la descripción de los eventos poco frecuentes, no representan el comportamiento típico del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, lo que confirma su estabilidad y baja volatilidad relativa.

1.3. Relevancia del índice en el contexto de la Bolsa Mexicana de Valores

La Bolsa Mexicana de Valores (BMV) es el principal mercado bursátil de México y un actor fundamental para el desarrollo económico y financiero del país. A través de sus índices representativos, proporciona una visión general del comportamiento del mercado, permitiendo a los inversionistas tomar decisiones informadas. Entre estos índices, el S&P/BMV IPC (Índice de Precios y Cotizaciones) destaca como el más representativo, ya que mide el desempeño de las empresas con mayor capitalización y liquidez en el mercado.

Sin embargo, la BMV también cuenta con índices sectoriales diseñados para reflejar el comportamiento de industrias específicas, como el S&P/BMV Sector Selecto de Bienes de Consumo Básico. Este índice agrupa a empresas cuya actividad principal está relacionada con productos esenciales para la vida cotidiana, lo que lo convierte en un termómetro del consumo básico en México. Su relevancia no solo radica en el comportamiento económico del

sector, sino también en su impacto directo en la estabilidad financiera y la percepción de los inversionistas en períodos de incertidumbre.

Para entender la importancia del S&P/BMV Sector Selecto de Bienes de Consumo Básico, es necesario explorar primero las características generales de la Bolsa Mexicana de Valores, después abordar el principal índice S&P/BMV IPC pues su estructura y componentes son la parte medular para entender la importancia del índice y porque lo convierte en un referente indiscutible para el análisis del mercado accionario mexicano.

1.3.1. Características clave de la Bolsa Mexicana de Valores

Actualmente la Bolsa Mexicana de Valores cuenta con 127³ empresas listadas en el mercado de capitales. Estas empresas abarcan diversos sectores, como bienes de consumo, industria, telecomunicaciones, servicios financieros y energía, reflejando la diversidad de la economía mexicana. Además, la BMV facilita la negociación de valores internacionales a través del Sistema Internacional de Cotizaciones (SIC), donde están listadas más de 3,000 emisoras extranjeras. Sin embargo, para el análisis del mercado local, el enfoque principal recae en las emisoras nacionales, debido a su representatividad en la economía mexicana.

Sin embargo, del total de las 127 empresas listadas en la BMV, no todas tienen un volumen de cotización significativo. Es por ello que se utiliza un índice que engloba a las 35 principales empresas, seleccionadas por ser las más grandes y líquidas del mercado. Este índice es el S&P/BMV IPC, que permite observar las tendencias generales y el desempeño de las emisoras más relevantes.

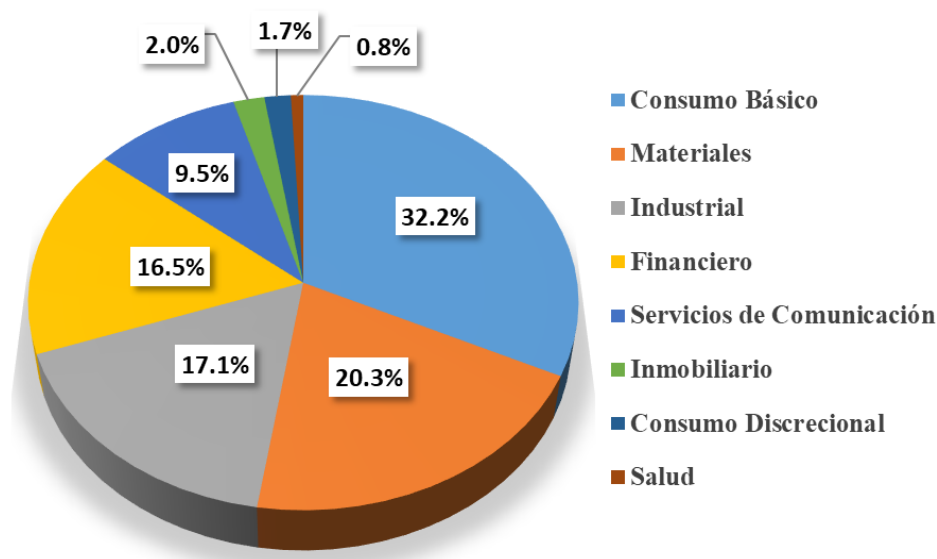
³Información sobre el número de empresas listadas en la Bolsa Mexicana de Valores, consultada de <https://www.bmv.com.mx/es/emisoras/informacion-de-emisoras> el 31 de marzo de 2025.

1.3.2. El Índice S&P/BMV IPC: estructura y componentes

EL S&P/BMV IPC busca medir el desempeño de las acciones de mayor tamaño y liquidez listadas en la Bolsa Mexicana de Valores. Su objetivo es proporcionar un índice amplio, representativo, pero al mismo tiempo fácilmente replicable, que abarque el mercado bursátil mexicano. Este índice tiene un cambio de muestra semi-anual en marzo y septiembre, esta compuesto de 35 empresas, y los componentes del índice son ponderados por capitalización de mercado modificada. Estos y más detalles del índice se encuentran en su Factsheet proporcionado por S&P Dow Jones Indices (2024b).

Se busca detallar y analizar la estructura del S&P/BMV IPC, así como observar cuales son las empresas que tienen mayor participación en el índice, e identificar cuales son los sectores a los que pertenecen, para hacer un contraste del peso en la construcción del índice.

Figura 1.3 Ponderación del S&P/BMV IPC por sector (2024)



Fuente: Elaboración propia realizada con datos del S&P Dow Jones Índices.

En primera instancia se presenta la Figura 1.3, que muestra la ponderación individual por sector en la que está distribuido el S&P/BMV IPC, donde se puede apreciar que la mayor participación en la composición del índice son las empresas del sector de consumo básico con

un 32.2 %, casi representando una tercera parte del índice total, en comparación con el sector de salud y atención médica que apenas llega a .8 %, representando menos del 1 % del total de las empresas que integran al S&P/BMV IPC. De esta manera se confirma la importancia de las empresas del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, destacándose como un sector clave en la construcción del principal índice de la Bolsa Mexicana de Valores.

A continuación se muestra la Tabla 1.2 donde se aprecian las 35 empresas que componen el S&P/BMV IPC, donde además se remarcan las empresas que pertenecen al S&P/BMV Sector Selecto de Bienes de Consumo Básico.

Tabla 1.2 *Componentes del S&P/BMV IPC*

Empresas enlistadas en el IPC, último Rebalanceo (2024)

1. Alfa A	19. Grupo Bimbo*
2. Alsea	20. Grupo Carso
3. América Móvil	21. Grupo Cementos de Chihuahua
4. Arca Continental*	22. Grupo Comercial Chedraui*
5. Banco del Bajío	23. Grupo Financiero Banorte
6. Becele*	24. Grupo Financiero Inbursa
7. Bolsa Mexicana de Valores	25. Grupo México
8. Cemex CPO	26. Grupo Televisa
9. Coca-Cola Femsa*	27. Industrias Peñoles
10. Corporación Inmobiliaria Vesta	28. Kimberly Clark de México*
11. El Puerto de Liverpool	29. La Comer*
12. Fomento Económico Mexicano*	30. Megacable Holdings
13. Genomma Lab Internacional	31. Orbia Advance Corporation
14. Gentera	32. Promotora y Operadora de Infraestructura
15. Gruma*	33. Quálitas Controladora
16. Grupo Aeroportuario del Centro Norte	34. Regional
17. Grupo Aeroportuario del Pacífico	35. Walmart de México*
18. Grupo Aeroportuario del Sureste	

*Empresas del índice S&P/BMV SSBCB

Fuente: Elaboración propia con datos del S&P Dow Jones Indices.

Es fácil ver la influencia de las empresas del S&P/BMV Sector Selecto de Bienes de Consumo Básico, en el principal índice de la Bolsa Mexicana de Valores, resaltando que de sus 35

empresas que lo conforman, diez pertenecen al índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, demostrando que sus empresas tienen un impacto directo en la estabilidad del S&P/BMV IPC, convirtiéndose en una parte medular en su composición y estructura.

Ahora bien, otro aspecto importante a resaltar es la ponderación individual para cada una de las 35 empresas que se muestran anteriormente, varía respecto a su capitalización de mercado, lo que significa que cada una recibe una ponderación diferente de acuerdo a sus características en el mercado. Es por eso que el S&P/BMV IPC tiene 10 principales componentes, en los cuales se concentra el 68.1% de la total de su composición. En la Tabla 1.3 se observa una lista más detallada de estas empresas, las cuales van numeradas con base en su peso de ponderación, de mayor a menor.

Tabla 1.3 Principales 10 componentes del S&P/BMV IPC

Componentes principales último Rebalanceo (2024)	
Empresa	Símbolo de cotización
1. Grupo Financiero Banorte	GFNORTEO
2. Grupo México	GMEXICOB
3. Fomento Económico Mexicano*	FEMSAUBD
4. América Móvil	AMXB
5. Walmart de México*	WALMEX
6. Cemex CPO	CEMEXCPO
7. Grupo Aeroportuario del Pacífico	GAPB
8. Grupo Aeroportuario del Sureste	ASURB
9. Arca Continental*	AC
10. Grupo Bimbo*	BIMBO A

*Empresas del índice SSBCB

Fuente: Elaboración propia con datos del S&P Dow Jones Indices.

Dentro del S&P/BMV IPC, el índice más representativo del mercado bursátil mexicano, Grupo Financiero Banorte se posiciona como la emisora con mayor peso, con una participación del 10.2%, siendo la única del sector financiero entre los principales componentes. Sin embargo, esta concentración no se replica en el S&P/BMV Sector Selecto de Bienes de Consumo Básico, donde destacan tres empresas relevantes dentro del índice, lo que subraya la impor-

tancia de este sector en la estructura del S&P/BMV IPC. La relevancia del sector de bienes de consumo básico se hace evidente al analizar la composición del principal índice bursátil del país, donde este sector representa el 32.2% de su ponderación total, consolidándose como un pilar fundamental en la Bolsa Mexicana de Valores.

En definitiva, el S&P/BMV Sector Selecto de Bienes de Consumo Básico juega un papel esencial en la estabilidad y desempeño de la Bolsa Mexicana de Valores. Su alta representatividad dentro del S&P/BMV IPC confirma que las empresas de este sector son clave en la estructura del índice y, por ende, en la economía del país. Más allá de reflejar el comportamiento de un sector específico, este índice permite evaluar la fortaleza y solidez del mercado en momentos de incertidumbre, convirtiéndose en una herramienta fundamental para inversionistas y analistas en la toma de decisiones estratégicas.

1.3.3. Comparativa entre el Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico y el S&P/BMV IPC

Como se ha mencionado previamente, el S&P/BMV IPC es el índice de referencia de la Bolsa Mexicana de Valores, compuesto por las 35 empresas con mayor tamaño y liquidez, mientras que el S&P/BMV Sector Selecto de Bienes de Consumo Básico agrupa a compañías cuyo negocio está basado en productos esenciales, lo que le otorga una relevancia particular dentro del panorama bursátil.

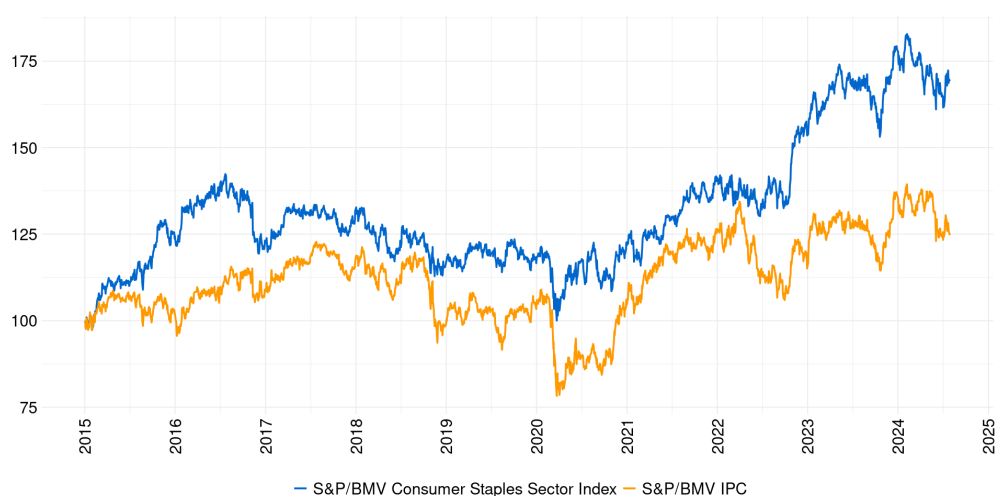
Resulta favorable para los inversionistas, hacer una comparativa entre estos dos índices, ya que su desempeño puede influir en la selección de activos según sus objetivos financieros y tolerancia al riesgo. Mientras que el S&P/BMV IPC ofrece una visión amplia del mercado y está expuesto a la volatilidad de múltiples sectores, el S&P/BMV Sector Selecto de Bienes de Consumo Básico tiende a mostrar un comportamiento más estable, dado que las empresas que lo integran operan en sectores menos sensibles a ciclos económicos adversos. En tiempos de incertidumbre o crisis, los inversionistas suelen buscar activos defensivos, y este índice puede

representar una opción atractiva debido a su menor exposición a fluctuaciones extremas del mercado.

Por lo tanto, se comienza examinando el comportamiento histórico de ambos índices, identificando sus diferencias en términos de rendimiento. Asimismo, se analiza cómo el S&P/BMV Sector Selecto de Bienes de Consumo Básico puede servir como una alternativa estratégica para aquellos inversionistas que buscan estabilidad en sus portafolios, minimizando riesgos sin sacrificar rendimientos sostenibles en el largo plazo.

A continuación se presenta la Figura 1.4, la cual muestra los precios históricos a lo largo de estos últimos años. En ella se observa una comparativa gráfica del crecimiento del precio de ambos índices, destacando que el S&P/BMV Sector Selecto de Bienes de Consumo Básico ha mostrado un crecimiento más notable en comparación con el S&P/BMV IPC.

Figura 1.4 *Precios Históricos (datos expresados en base 100)*



Fuente: Elaboración propia realizada con datos del S&P Dow Jones Índices.

Para ver una comparativa más técnica, la Tabla 1.4 muestra los rendimientos anualizados en distintos períodos, brindando una herramienta clave para evaluar la viabilidad de incluir este tipo de índices en portafolios de inversión. Esto permite identificar cómo un índice con menor riesgo y estabilidad relativa puede generar rendimientos consistentes, favoreciendo decisiones estratégicas en mercados volátiles.

Tabla 1.4 Comparativa de Rendimientos

Rend.	S&P/BMV SSBCB		S&P/BMV IPC	
	Total	Del Precio	Total	Del Precio
1 mes	1.03 %	0.82 %	-1.93 %	-2.09 %
3 meses	3.73 %	3.13 %	-4.67 %	-5.79 %
YTD	-2.18 %	-3.66 %	-7.53 %	-9.41 %
1 año	6.09 %	3.22 %	1.58 %	-1.95 %
3 años	10.68 %	7.67 %	2.68 %	-0.83 %
5 años	10.38 %	7.37 %	7.20 %	4.05 %
10 años	7.31 %	4.76 %	3.96 %	1.31 %

Nota: Los rendimientos están anualizados y la información esta actualizada hasta el 30 de Agosto de 2024.

Fuente: Elaboración propia con datos del S&P Dow Jones Indices.

El análisis de los rendimientos presentados entre el S&P/BMV Sector Selecto de Bienes de Consumo Básico (SSBCB) y el S&P/BMV IPC muestra diferencias significativas que reflejan las características propias de cada índice y las dinámicas de los sectores que los componen. Estas variaciones son evidentes en los horizontes de corto, mediano y largo plazo, y destacan al SSBCB como una opción más atractiva para inversionistas que priorizan estabilidad y rentabilidad constante.

En el corto plazo, el S&P/BMV Sector Selecto de Bienes de Consumo Básico muestra un desempeño positivo, con rendimientos de +1.03 % en 1 mes y +3.73 % en 3 meses. Por otro lado, el IPC experimenta pérdidas en el mismo periodo, con caídas de -1.93 % y -4.67 %, respectivamente. Este contraste se explica porque las empresas del SSBCB pertenecen al sector de bienes de consumo básico, el cual es menos susceptible a las fluctuaciones económicas debido a la naturaleza esencial de sus productos. En cambio, el IPC, al ser más diversificado, incluye sectores más expuestos a la volatilidad, como el energético y el tecnológico.

En el mediano plazo, la diferencia entre ambos índices sigue siendo notable. Mientras que el SSBCB acumula un rendimiento de -2.18 % en el año a la fecha (YTD), el IPC enfrenta

pérdidas más pronunciadas de -7.53% . Asimismo, a 1 año, el SSBCB destaca con un rendimiento positivo de $+6.09\%$, superando ampliamente al IPC, que apenas logra un $+1.58\%$. Este comportamiento evidencia la fortaleza del índice sectorial frente a condiciones adversas del mercado, ya que las empresas de bienes de consumo básico tienden a mantener ingresos estables incluso en entornos económicos desafiantes.

En el largo plazo, el SSBCB consolida su ventaja como una inversión sólida. A 3 años, genera un rendimiento total de $+10.68\%$, mientras que el IPC se queda en $+2.68\%$. Esta tendencia se mantiene a 5 y 10 años, donde el SSBCB sigue superando al IPC. Este desempeño sostenido resalta la capacidad del índice SSBCB para ofrecer retornos ajustados al riesgo más atractivos, gracias a la estabilidad inherente de su sector.

En conclusión, el S&P/BMV Sector Selecto de Bienes de Consumo Básico se posiciona como una alternativa superior para los inversionistas que buscan estabilidad y rentabilidad consistente, especialmente en un entorno económico incierto. Su enfoque en bienes de consumo básico lo convierte en una opción más resiliente y menos vulnerable a la volatilidad del mercado, en contraste con el S&P/BMV IPC, que refleja la exposición de sus sectores más cíclicos. Por ello, el SSBCB es ideal para aquellos que desean mitigar riesgos y obtener resultados sólidos a corto, mediano y largo plazo.

1.4. Impacto económico del Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico

El S&P/BMV Sector Selecto de Bienes de Consumo Básico desempeña un papel central en la economía mexicana, no solo como un indicador financiero, sino como un reflejo del consumo interno, que constituye la columna vertebral del Producto Interno Bruto (PIB) del país. De acuerdo con datos de CONCANACO (2024), el consumo privado representa aproximadamente el 71.9% del PIB nacional, subrayando la relevancia del sector de bienes de consumo básico.

Este índice, por tanto, no solo agrupa a las empresas más representativas del sector, sino que también actúa como un termómetro de la fortaleza económica y la estabilidad social.

El índice se distingue por incluir a empresas que producen y comercializan bienes esenciales, como alimentos, bebidas y productos de higiene, cuya demanda permanece constante incluso en tiempos de incertidumbre económica. Krugman (1997) subraya que los sectores relacionados con bienes de consumo básico tienen menor elasticidad de demanda, lo que los hace más resilientes ante las fluctuaciones del ciclo económico. En un contexto de alta volatilidad global, como el experimentado durante la pandemia de COVID-19, estos bienes han demostrado ser indispensables para mantener la estabilidad del mercado interno y evitar crisis de consumo.

Por su parte, el S&P/BMV Sector Selecto de Bienes de Consumo Básico también ha sido destacado por estudios académicos como un motor de crecimiento económico a través del empleo y la inversión. De acuerdo con datos de Research (2014), las empresas del sector de bienes de consumo básico representan una fuente importante de generación de empleo formal, especialmente en regiones con alta densidad poblacional, contribuyendo así a la reducción de desigualdades económicas. Además, la capacidad de estas empresas para atraer inversiones extranjeras directas reafirma su papel estratégico en la economía nacional.

En el ámbito bursátil, el S&P/BMV Sector Selecto de Bienes de Consumo Básico (SSBCB) destaca frente a su principal contraparte, el S&P/BMV IPC, al capturar de manera más efectiva las dinámicas del consumo interno y exhibir un desempeño financiero más sólido y estable. Un ejemplo claro de esto se observa en el análisis comparativo de rendimientos: en un período de un año, desde agosto de 2024, el SSBCB registra un rendimiento positivo del +6.09 %, mientras que el IPC presenta un avance más modesto del +1.58 %. Esta brecha se acentúa aún más en horizontes de tiempo más amplios; a tres años, el SSBCB acumula un rendimiento del +10.68 %, superando significativamente al IPC, que alcanza solo un +2.68 %. Sin embargo, más allá de estas cifras, el verdadero valor del SSBCB reside en su capacidad para brindar a los inversionistas una exposición directa a un sector estratégico y fundamen-

tal para la economía mexicana. Estas cifras, aunque relevantes, son solo un aspecto de la importancia del SSBCB, ya que su verdadero valor radica en su capacidad para ofrecer a los inversionistas una exposición directa a un sector crítico para la economía mexicana.

El papel de los inversionistas en este índice es crucial. Shiller (2000) argumenta que los mercados financieros bien estructurados pueden actuar como un puente entre los ahorros individuales y las necesidades de financiamiento de las empresas, fomentando así el crecimiento económico. Al invertir en el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, los inversionistas no solo buscan retornos atractivos, sino que también contribuyen al fortalecimiento de empresas clave para la estabilidad social y económica, impulsando la innovación, el desarrollo y el acceso a bienes esenciales.

Tras conocer el impacto del sector en la economía nacional, resulta indispensable analizar su estado actual, especialmente considerando los cambios estructurales, políticos y sociales que México ha enfrentado durante 2024. Según un artículo del periódico País (2024), el consumo privado ha mostrado signos de desaceleración desde 2023, con una tendencia a la baja en los meses de agosto y septiembre de 2024. Factores como la incertidumbre política, la disminución en la creación de empleos y episodios de inseguridad han afectado tanto a las empresas como a los puntos de venta. Además, la persistente inflación ha llevado a los consumidores a ser más cautelosos, reduciendo su gasto en bienes duraderos.

En este sentido, el S&P/BMV Sector Selecto de Bienes de Consumo Básico reafirma su importancia no solo como un indicador financiero, sino como un vehículo estratégico para entender y fortalecer el consumo interno. Su capacidad para resistir choques económicos y su papel central en la estabilidad del Producto Interno Bruto de México lo convierten en una herramienta indispensable tanto para inversionistas como para responsables de política económica, quienes deben reconocer su relevancia en la construcción de una economía más sólida y sostenible.

1.5. Pronóstico del Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico

Tras haber realizado un análisis previo sobre la importancia económica, las características del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico y un breve comparativo con el principal índice de la BMV, toca el turno de una de las partes medulares de este trabajo de investigación; abordar el objetivo de realizar pronósticos sobre las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

Los pronósticos en el mercado de acciones son una herramienta esencial para la toma de decisiones informadas para los inversionistas. La generación de pronósticos en el mercado de acciones es una práctica fundamental dentro de las finanzas y la economía, ya que permite anticipar la evolución futura de los precios de los activos financieros. En términos generales, un pronóstico es una estimación basada en modelos matemáticos, estadísticos y econométricos, que busca predecir el comportamiento de una variable en el tiempo a partir de datos históricos y factores explicativos relevantes (Hyndman & Athanasopoulos, 2018). En el contexto financiero, el pronóstico del precio de las acciones se apoya en el análisis de tendencias del mercado, indicadores macroeconómicos, razones financieras y otros factores determinantes del comportamiento bursátil.

La importancia de los pronósticos en el mercado de acciones radica en su capacidad para reducir la incertidumbre inherente a los mercados financieros y proporcionar información valiosa a inversionistas, analistas y reguladores. Como se menciona en la Hipótesis del Mercado Eficiente de Fama (1970), los precios de las acciones reflejan toda la información disponible, lo que implica que cualquier pronóstico debe considerar tanto datos históricos como expectativas futuras y factores externos que puedan influir en el mercado. Sin embargo, esta hipótesis ha sido debatida por diversos autores, como Shiller (2000), quien argumenta que los mercados pueden ser irracionales y que los precios pueden desviarse significativamente de su valor fundamental debido a factores psicológicos y conductuales. Esta dualidad entre

eficiencia e irracionalidad del mercado subraya la complejidad de predecir el comportamiento de los precios de las acciones y la necesidad de emplear metodologías robustas para abordar este desafío.

El análisis cuantitativo en la predicción de precios de acciones ha sido clave, trayendo como consecuencia modelos de regresión, series temporales, etc. Estos han sido de gran utilidad a lo largo del tiempo. Sin embargo, este análisis ha evolucionado significativamente con el desarrollo de modelos más sofisticados, como algoritmos de aprendizaje automático, entre ellos los bosques aleatorios y redes neuronales. Estos enfoques permiten capturar relaciones no lineales y patrones complejos en los datos financieros, mejorando la precisión de los pronósticos en comparación con los métodos tradicionales basados en técnicas econométricas clásicas (Bishop, 2006). Esta evolución ha sido fundamental para tener herramientas más sofisticadas para el pronóstico de precios, así como para muchas otras aplicaciones como la administración de portafolios de inversión y la asignación de activos, ya que los inversionistas utilizan estos modelos avanzados para identificar oportunidades de compra y venta, maximizando así su rentabilidad y reduciendo su exposición al riesgo.

En relación con eso, la metodología que se empleará en esta investigación se encuentra basada en bosques aleatorios, la cual será detallada en el siguiente capítulo. Allí se explicarán las características generales de esta técnica, así como el paso a paso para construir y validar el modelo, con el fin de generar pronósticos precisos y confiables para las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico. Este enfoque no solo busca contribuir a la comprensión de las dinámicas del sector, sino también ofrecer una herramienta valiosa para la toma de decisiones de inversión en un entorno financiero cada vez más complejo y competitivo.

Al emplear técnicas avanzadas como los bosques aleatorios, se pretende superar las limitaciones de los métodos tradicionales y proporcionar un análisis más robusto que permita a los inversionistas anticipar movimientos en los precios de las acciones, optimizando así sus estrategias de inversión y mejorando la gestión de sus recursos. De esta manera, el pronóstico del

precio de las acciones del índice se convierte en un elemento clave para fortalecer la gestión financiera y potenciar los resultados en el mercado bursátil.

Capítulo 2

Bosques Aleatorios

En el presente capítulo se abordará la metodología empleada en esta investigación, la cual se basa en el uso de bosques aleatorios para generar pronósticos sobre los precios de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico. Para contextualizar esta técnica, se comenzará con una introducción general al Machine Learning (aprendizaje automático), campo de la inteligencia artificial que permite a las computadoras aprender patrones a partir de datos. Dentro de este marco, se explicarán los principales tipos de aprendizaje, con especial énfasis en el aprendizaje supervisado, categoría a la que pertenecen los bosques aleatorios.

Posteriormente, se introducirán los árboles de decisión, que son la base fundamental para la construcción de los bosques aleatorios. Se explicarán sus principios, funcionamiento y limitaciones, lo cual permitirá entender por qué los bosques aleatorios surgen como una evolución natural para superar dichas limitaciones. Finalmente, se profundizará en la metodología de bosques aleatorios, detallando su estructura, ventajas y el proceso paso a paso para su implementación en el contexto de esta investigación. Este capítulo sentará las bases teóricas y prácticas necesarias para comprender cómo se generarán los pronósticos que serán analizados en los siguientes apartados.

2.1. Introducción al Machine Learning

Para entender mejor la metodología propuesta en esta investigación, resulta fundamental contextualizar al lector en el campo del Machine Learning (ML). El ML es una rama de la inteligencia artificial (IA) que se enfoca en el desarrollo de algoritmos capaces de aprender patrones a partir de datos, mejorar su desempeño con la experiencia y realizar predicciones o decisiones sin ser explícitamente programados para cada tarea. Según Mitchell (1997), “el Machine Learning es el estudio de algoritmos computacionales que mejoran automáticamente a través de la experiencia” (p. 1). Esta definición ha sido ampliada por autores más recientes, como Goodfellow et al. (2016), quienes destacan que el ML permite a los sistemas aprender representaciones de los datos, identificar patrones complejos y generalizar su conocimiento para aplicarlo en situaciones nuevas.

Los algoritmos de ML se basan en el análisis de grandes volúmenes de datos para identificar patrones y generar reglas de decisión. A diferencia de los programas tradicionales, donde las reglas son definidas explícitamente por el programador, en el ML los algoritmos aprenden estas reglas a partir de los datos de entrenamiento.

Como señala Bishop (2006),

“El Machine Learning se centra en el desarrollo de modelos que pueden aprender a realizar tareas específicas a partir de datos, en lugar de depender de reglas predefinidas” (p. 2).

Esta capacidad de aprendizaje automático es fundamental para aplicaciones que requieren razonamiento, adaptación y mejora continua, como la predicción de precios de acciones, el reconocimiento de imágenes o el procesamiento del lenguaje natural.

La diferencia entre los programas tradicionales y los algoritmos de ML radica en su enfoque para manejar los datos. En la programación tradicional, se establecen reglas fijas para procesar las entradas y obtener salidas específicas. En contraste, los algoritmos de ML utilizan

datos de entrenamiento para inferir reglas y ajustar sus parámetros, lo que les permite mejorar su desempeño con base en la experiencia. Como explica Domingos (2012), “el Machine Learning no se trata de programar computadoras para realizar tareas, sino de enseñarles a aprender de los datos” (p. 78). Esta capacidad de aprendizaje autónomo es lo que distingue al ML de otros enfoques computacionales.

2.1.1. Clasificación de los tipos de aprendizaje en Machine Learning

Los algoritmos de Machine Learning pueden clasificarse de diversas maneras según su enfoque o funcionalidad, como algoritmos basados en redes neuronales, algoritmos de regresión o métodos de agrupamiento. Sin embargo, su clasificación más fundamental se basa en el método de aprendizaje que emplean: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje profundo. A continuación, se describen estos tres tipos principales de aprendizaje.

- **Aprendizaje Supervisado:** El aprendizaje supervisado (Supervised Learning) es un enfoque en el que se entrena un modelo de Machine Learning utilizando un conjunto de datos etiquetados, es decir, datos que incluyen tanto las entradas (características) como las salidas deseadas (etiquetas). El objetivo es que el modelo aprenda a mapear las entradas a las salidas correctas, lo que le permitirá hacer predicciones precisas sobre nuevos datos no vistos. Como señala Hastie et al. (2009), “el aprendizaje supervisado se basa en la idea de que, a partir de ejemplos de entradas y salidas, el modelo puede generalizar para predecir resultados en datos futuros” (p. 2).
- **Aprendizaje No Supervisado:** El aprendizaje no supervisado (Unsupervised Learning) se utiliza para descubrir patrones o estructuras ocultas en datos no etiquetados. A diferencia del aprendizaje supervisado, este método no requiere datos con etiquetas predefinidas, lo que lo hace ideal para tareas de exploración y minería de datos. Según Géron (2019), “el aprendizaje no supervisado busca identificar similitudes y diferencias en los datos, agrupándolos o reduciendo su dimensionalidad para facilitar su interpreta-

ción” (p. 215). Los algoritmos de aprendizaje no supervisado se dividen principalmente en técnicas de agrupamiento¹ y análisis factorial².

- **Aprendizaje Profundo:** El aprendizaje profundo (Deep Learning) es una subárea del Machine Learning que se inspira en la estructura y funcionamiento del cerebro humano. Como explica Goodfellow et al. (2016), “el aprendizaje profundo permite a las máquinas aprender representaciones jerárquicas de los datos, lo que facilita la toma de decisiones en tareas complejas” (p. 5). Lo que se traduce en utilizar redes neuronales artificiales con múltiples capas para procesar datos y extraer patrones complejos, lo que lo hace especialmente útil para manejar datos no estructurados, como imágenes, texto o audio.

Los bosques aleatorios son una técnica de Machine Learning que se clasifica dentro del aprendizaje supervisado. Este método combina múltiples árboles de decisión para mejorar la precisión y robustez de las predicciones, reduciendo el riesgo de sobreajuste³ que suele afectar a los árboles individuales. Sin embargo, para comprender plenamente cómo funcionan los bosques aleatorios, es necesario abordar primero la incógnita de qué son los árboles de decisión y por qué surgieron como una solución inicial en el campo del aprendizaje automático.

2.2. Árboles de decisión

Los árboles de decisión son uno de los modelos más fundamentales y ampliamente utilizados en el campo del Machine Learning, particularmente dentro del aprendizaje supervisado. Su popularidad radica en su simplicidad, interpretabilidad y capacidad para manejar tanto problemas de clasificación como de regresión. Además, constituyen la base de modelos más

¹Agrupamiento: Organiza las observaciones en grupos basándose en similitudes o patrones.

²Análisis factorial: Busca reducir la dimensionalidad de los datos identificando los factores subyacentes que mejor los representan.

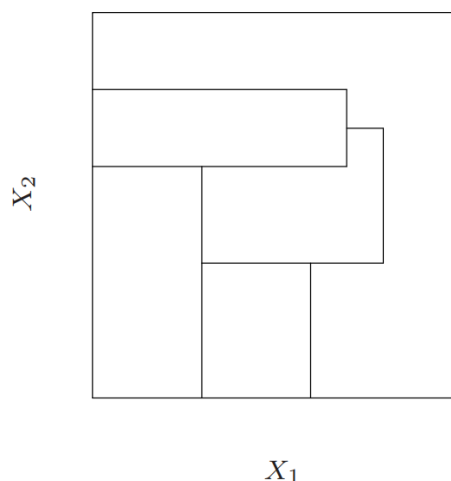
³El sobreajuste (*overfitting*, en inglés) ocurre en aprendizaje automático cuando un modelo aprende demasiado bien los detalles y el ruido del conjunto de entrenamiento, al punto de que pierde capacidad para generalizar a datos nuevos o no vistos.

sofisticados y perfeccionados, como son los Bosques Aleatorios, los cuales combinan múltiples árboles para mejorar la precisión y robustez de las predicciones.

El desarrollo de los árboles de decisión ha sido impulsado por diversos investigadores, pero uno de los algoritmos más influyentes es el Classification and Regression Trees (CART), introducido por Breiman et al. (1984). CART es un algoritmo que construye árboles de decisión dividiendo el conjunto de datos en subconjuntos de manera recursiva, seleccionando en cada paso la característica que mejor separa las clases (en clasificación) o minimiza la varianza (en regresión). Para ello, CART emplea criterios como la impureza de Gini en clasificación y la reducción de la varianza en regresión, estableciendo divisiones en los nodos de acuerdo con reglas basadas en valores umbral.

Para ilustrar el funcionamiento de los árboles de decisión, consideremos un problema de regresión con una variable de respuesta continua Y y entradas X_1 y X_2 , donde ambas entradas toman valores dentro del intervalo unitario. En la figura 2.1 se muestra una partición del espacio de características mediante líneas que son paralelas a los ejes de coordenadas. En cada elemento de la partición, podemos modelar Y con una constante diferente.

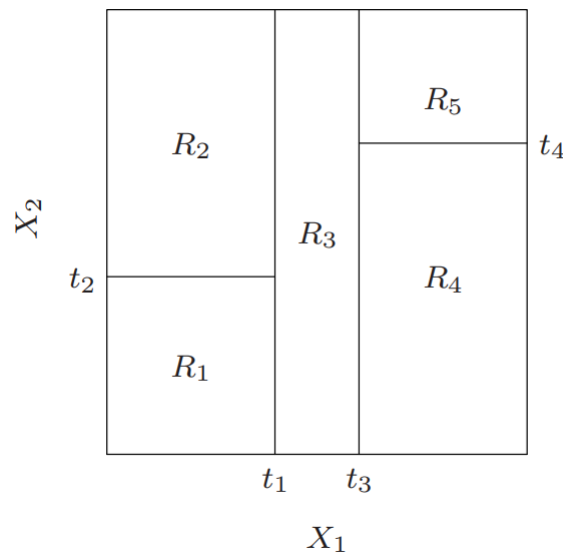
Figura 2.1 *Ejemplo donde no se realiza mediante particiones recursivas*



Fuente: Rescatada del libro “The Elements of Statistical Learning: Data Mining, Inference, and Prediction” (p. 306).

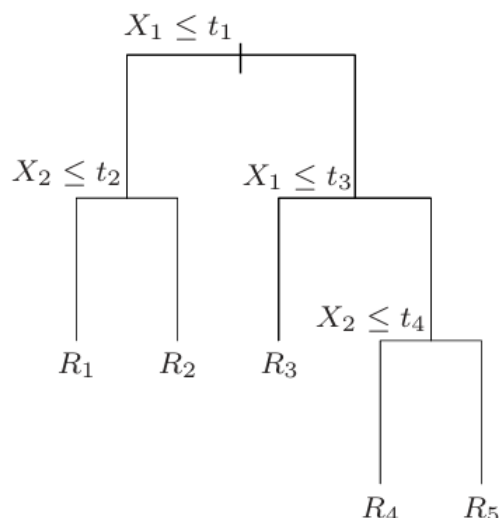
Sin embargo, hay un problema: aunque cada línea de partición tiene una descripción simple como $X_1 = c$, algunas de las regiones resultantes son complicadas de describir. Para simplificar las cosas, nos centramos en particiones binarias recursivas como las de la Figura 2.2. Primero dividimos el espacio en dos regiones y modelamos la respuesta mediante la media de Y en cada región. Elegimos la variable y el punto de división para lograr el mejor ajuste. Luego, una o ambas de estas regiones se dividen en dos regiones más, y este proceso se continúa hasta que se aplica alguna regla de parada.

Figura 2.2 Ejemplo donde es realizada mediante particiones recursivas



Fuente: Rescatada del libro “The Elements of Statistical Learning: Data Mining, Inference, and Prediction” (p. 306).

Una ventaja clave del árbol binario recursivo es su interpretabilidad. La partición del espacio de características se describe completamente mediante un solo árbol. Con más de dos entradas, las particiones como en la figura 2.2 son difíciles de dibujar, pero la representación del árbol binario funciona de la misma manera. Para explicar de mejor manera este proceso, es común representar las particiones binarias recursivas mediante un diagrama en forma de árbol.

Figura 2.3 Diagrama de árbol

Fuente: Rescatada del libro “The Elements of Statistical Learning: Data Mining, Inference, and Prediction” (p. 306).

En el ejemplo anterior de la figura 2.3, el resultado de este proceso es una partición en las cinco regiones R_j . Las observaciones que satisfacen la condición en cada unión se asignan a la rama izquierda, y las demás a la rama derecha. Los nodos terminales o hojas del árbol corresponden a las regiones R_1, R_2, \dots, R_5 . El modelo de regresión correspondiente predice Y con una constante c_j en la región R_j .

Por lo que, la estructura de un árbol de decisión consta de:

- **Nodo raíz:** donde se realiza la primera división de los datos.
- **Nodos internos:** donde se aplican reglas de partición en función de los predictores.
- **Hojas o nodos terminales:** donde se asigna la predicción final de acuerdo con la media de la variable respuesta en la región correspondiente.

2.2.1. Definición de un árbol de decisión para regresión

De acuerdo con el libro “*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*” de Hastie et al. (2009), los árboles de decisión para regresión se definen de la siguiente manera:

Definición 1. Dado un conjunto de datos $\{(X_i, Y_i)\}_{i=1}^n$, donde $X_i \in \mathbb{R}^p$ es un vector de variables predictoras y $Y_i \in \mathbb{R}$ es la variable respuesta, un árbol de decisión para regresión particiona el espacio de entrada \mathbb{R}^p en M regiones disjuntas $\{R_m\}_{m=1}^M$ de manera recursiva.

La predicción dentro de cada región R_m se realiza utilizando la media de los valores Y_i en esa región:

$$\hat{f}(X) = \sum_{m=1}^M c_m \mathbb{1}(X \in R_m) \quad (2.1)$$

donde

$$c_m = \frac{1}{|R_m|} \sum_{X_i \in R_m} Y_i$$

Aquí, $\mathbb{1}(X \in R_m)$ es la función indicadora que toma el valor de 1 si X pertenece a la región R_m y 0 en caso contrario.

Ahora bien, suponiendo que el criterio para minimizar el error es la suma de los residuos al cuadrado, es decir $\sum [y_i - f(x_i)]^2$, observamos que para dicha métrica el mínimo se alcanza en cada región R_m .

Para definir las regiones R_m , el árbol realiza divisiones sucesivas mediante umbrales en las variables predictoras. En cada nodo, se busca el predictor j y el punto de corte s que minimizan la suma de los errores cuadráticos dentro de las regiones:

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (2.2)$$

Donde:

$$R_1(j, s) = \{X \mid X_j \leq s\}, \quad R_2(j, s) = \{X \mid X_j > s\}.$$

El proceso continúa dividiendo cada nueva región hasta alcanzar un criterio de parada, como un número mínimo de observaciones por nodo o una reducción insuficiente en el error.

Una vez encontrada la mejor división, particionamos los datos en las dos regiones resultantes R_1 y R_2 , y repetimos el proceso de división en cada una de estas regiones. Este procedimiento se itera en todas las regiones resultantes.

La pregunta clave es determinar la profundidad o tamaño óptimo del árbol⁴. Un árbol excesivamente grande podría generar sobreajuste en los datos, mientras que un árbol demasiado pequeño podría no capturar las características significativas de la estructura subyacente.

2.2.2. División de datos: Entrenamiento y prueba

En el proceso de construcción de modelos de Machine Learning, es fundamental dividir el conjunto de datos disponibles en dos subconjuntos: entrenamiento y prueba. El conjunto de entrenamiento se utiliza para ajustar los parámetros del modelo, mientras que el conjunto de prueba sirve para evaluar su rendimiento en datos no vistos. Esta división es crucial para evitar el sobreajuste (overfitting), es decir, que el modelo memorice los datos de entrenamiento en lugar de generalizar patrones (James et al., 2013).

La proporción típica para esta división suele ser 70 %-30 % o 80 %-20 %, aunque la elección depende del tamaño del conjunto de datos y del problema específico. Según Hastie et al., 2009, una división adecuada garantiza que el modelo tenga suficientes datos para aprender sin sacrificar la capacidad de evaluar su rendimiento de manera robusta. Además, en casos donde los datos son limitados, técnicas como la validación cruzada son ampliamente recomendadas.

⁴El tamaño del árbol actúa como un parámetro de ajuste que determina la complejidad del modelo, y el tamaño óptimo del árbol debe ser seleccionado adaptativamente a partir de los datos.

En la validación cruzada, el conjunto de datos se divide en k subconjuntos (o "folds"), y el modelo se entrena y evalúa k veces, utilizando cada fold como conjunto de prueba una vez (Kohavi, 1995).

En el contexto de los árboles de decisión, la división de datos adquiere especial relevancia debido a su tendencia a sobreajustarse si no se controla su complejidad. Por ejemplo, un árbol demasiado profundo puede adaptarse perfectamente a los datos de entrenamiento, pero fallar al predecir nuevos datos. Esto se debe a que los árboles de decisión tienen una alta capacidad para capturar patrones específicos en los datos de entrenamiento, incluso si esos patrones son ruidosos o no son generalizables (Breiman et al., 1984). Por ello, es común utilizar técnicas como la poda para reducir la complejidad del árbol y mejorar su capacidad de generalización.

2.2.3. Limitación de los árboles de decisión

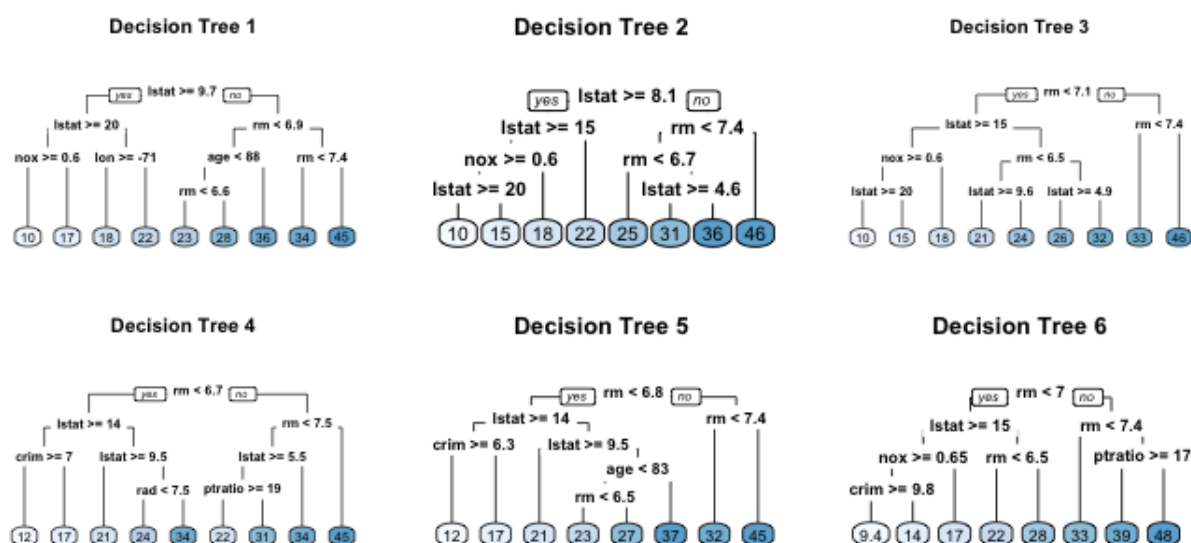
Los árboles de decisión, con su estructura intuitiva y capacidad para manejar tanto variables categóricas como numéricas, son una herramienta poderosa en el aprendizaje automático. Sin embargo, su simplicidad es también su talón de Aquiles. Un árbol de decisión básico realiza una única prueba lógica en cada nodo, lo que lo hace fácil de interpretar, pero también lo vuelve extremadamente sensible a pequeñas variaciones en los datos. Esta alta varianza significa que un ligero cambio en el conjunto de entrenamiento puede dar lugar a un árbol con divisiones completamente diferentes, comprometiendo su estabilidad y confiabilidad.

El problema radica en la naturaleza jerárquica de los árboles: un error en una división inicial se propaga hacia abajo, afectando todas las decisiones subsiguientes. Aunque técnicas como la poda o el uso de criterios de división más robustos pueden mitigar parcialmente este problema, la inestabilidad inherente persiste. Este es el precio que se paga por intentar modelar estructuras complejas con un enfoque tan sencillo y transparente.

Para comprender mejor la inestabilidad de los árboles de decisión, se presenta la figura 2.4

un ejemplo de Guide (2023), en la cual se crean seis árboles de decisión a partir de diferentes muestras de una base de datos. Vemos que la parte superior de los árboles tiene una estructura muy similar. Aunque existen 15 variables predictorias para dividir en los datos, todos los seis árboles tienen tanto las variables **lstat** como **rm** impulsando las primeras divisiones.

Figura 2.4 Ejemplo de correlación de árboles de decisión



Fuente: Imagen de UC Business Analytics R Programming Guide.

Esta característica se conoce como correlación de árboles y evita que el método reduzca óptimamente la varianza de los valores predictivos. Para reducir aún más la varianza, necesitamos minimizar la cantidad de correlación entre los árboles. Esto se puede lograr inyectando más aleatoriedad en el proceso de crecimiento del árbol.

Aunque los árboles de decisión son algoritmos muy potentes en el aprendizaje supervisado, pueden ser propensos a problemas como sesgos y sobreajuste. No obstante, cuando varios árboles de decisión se combinan en un conjunto, como en el algoritmo de Bosques Aleatorios, se obtienen predicciones más precisas, especialmente cuando los árboles individuales no están correlacionados entre sí.

2.3. Bosques aleatorios

Los bosques aleatorios (Random Forests) son una técnica de aprendizaje ensemble desarrollada por Leo Breiman en 2001. Este método combina múltiples árboles de decisión para mejorar la precisión y estabilidad del modelo. A diferencia de un árbol de decisión individual, que puede ser propenso al sobreajuste y a la alta varianza, los bosques aleatorios promedian las predicciones de varios árboles, lo que reduce la varianza y mejora la generalización. Para comprender esta técnica en su totalidad, es necesario analizar sus fundamentos, los mecanismos que la hacen efectiva y su aplicación práctica.

2.3.1. Fundamentos de los bosques aleatorios

Los bosques aleatorios se basan en dos mecanismos principales: el bagging (Bootstrap Aggregation) y la selección aleatoria de características. Estos mecanismos permiten reducir la varianza y mejorar la capacidad de generalización del modelo.

- **Bagging:** Esta técnica consiste en generar múltiples subconjuntos de los datos de entrenamiento mediante muestreo con reemplazo (Bootstrap). Cada subconjunto se utiliza para entrenar un árbol de decisión independiente. Al promediar las predicciones de estos árboles, se reduce la varianza y se mejora la estabilidad del modelo (James et al., 2013).
- **Selección aleatoria de características:** En cada división de un nodo, solo se considera un subconjunto aleatorio de las variables predictoras. Esto introduce diversidad en los árboles, evitando que todos se ajusten a los mismos patrones y reduciendo el riesgo de sobreajuste (Hastie et al., 2009).

Además, los bosques aleatorios utilizan la división de datos en conjuntos de entrenamiento y prueba, lo que permite evaluar el rendimiento del modelo de manera robusta. Gracias al

bagging, cada árbol se entrena con un subconjunto diferente de los datos, lo que maximiza el uso de la información disponible.

2.3.1.1. Bootstrap y Bagging

El Bootstrap es una técnica de remuestreo que permite cuantificar la incertidumbre de un estimador. En el contexto de los bosques aleatorios, se utiliza para generar múltiples muestras de entrenamiento a partir de los datos originales.

- **Definición de Bootstrap:** Sea Z nuestra muestra de datos original, de tamaño n . Una muestra Bootstrap Z_B se obtiene mediante extracciones con reemplazamiento sobre las observaciones de Z y tiene tamaño n .
- **Estimación del error estándar:** El error estándar de un estimador $\hat{\alpha}$ se calcula como:

$$SE_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1} \sum_{r=1}^B \left(\hat{\alpha}_r - \frac{1}{B} \sum_{k=1}^B \hat{\alpha}_k \right)^2} \quad (2.3)$$

- **Propiedades del Bootstrap:**
 1. La probabilidad de que una observación x_j no esté contenida en una muestra Bootstrap es $P(\{x_j \notin Y\}) = \left(1 - \frac{1}{n}\right)^n$.
 2. En una muestra Bootstrap, aproximadamente un tercio de las observaciones no están contenidas.

El Bagging (Bootstrap Aggregation) es una extensión del Bootstrap que tiene como objetivo reducir la varianza de un método de aprendizaje estadístico. En el caso de los bosques aleatorios, se aplica a los árboles de decisión de la siguiente manera:

- **Definición de Bagging:** Sea X nuestra muestra original. Se generan m muestras Bootstrap X_1, X_2, \dots, X_m , y se entrena un árbol de decisión en cada una. La predicción

final se obtiene promediando las predicciones de todos los árboles:

$$\hat{f}_{bag}(x) = \frac{1}{m} \sum_{k=1}^m \hat{f}_k(x) \quad (2.4)$$

- **Aplicación a árboles de decisión:** El bagging reduce la alta varianza de los árboles de decisión, mejorando la calidad de las predicciones.

2.3.2. Definición de bosques aleatorios

Los bosques aleatorios son una extensión del bagging aplicado a los árboles de decisión, diseñados para mejorar la precisión y reducir la varianza del modelo. La principal innovación de los bosques aleatorios radica en la selección aleatoria de las variables predictoras en cada división de un nodo, lo que reduce la correlación entre los árboles y mejora la generalización.

Definición 2. Sea $B = \{T_b\}_{b=1}^B$ un bosque aleatorio sobre X , nuestra muestra original. La predicción de regresión \hat{f}_{RF} de los bosques aleatorios es obtenida por:

$$\hat{f}_{RF}(x) = \frac{1}{B} \sum_{k=1}^B T_k(x) \quad (2.5)$$

Donde:

- $\hat{f}_{RF}(x)$: Es la predicción del modelo para un conjunto de características x .
- B : Es el número total de árboles en el bosque aleatorio.
- $T_k(x)$: Es la predicción generada por el k -ésimo árbol de decisión T_K cuando se evalúa con la entrada x .

Los bosques aleatorios mejoran la generalización del modelo introduciendo aleatoriedad en la construcción de los árboles. A continuación, se describen dos aspectos clave: la selección

aleatoria de predictores y la reducción de la varianza.

- **Selección aleatoria de predictores:** En cada división de un nodo, solo se considera un subconjunto aleatorio de las variables predictoras. Este subconjunto se elige de manera independiente para cada árbol, lo que introduce diversidad y evita que todos los árboles se ajusten a los mismos patrones. Para problemas de regresión, se recomienda utilizar $m \approx p/3$, donde p es el número total de predictores Breiman, 2001.
- **Reducción de la varianza:** Al promediar las predicciones de múltiples árboles, los bosques aleatorios reducen la varianza del modelo. La varianza de la predicción final se puede expresar como:

$$\text{Var}(\hat{f}_{RF}(x)) = \frac{1-\rho}{B}\sigma^2 + \rho\sigma^2 \quad (2.6)$$

donde ρ es la correlación entre los árboles, σ^2 es la varianza de un solo árbol, y B es el número de árboles. Al reducir ρ mediante la selección aleatoria de predictores, se minimiza la varianza del modelo. Además, al considerar una cantidad mayor de B , también se reduce significativamente la varianza, ya que el primer término de la ecuación tiende a cero a medida que B aumenta. Esto refuerza la importancia de utilizar un número suficiente de árboles para garantizar la estabilidad y precisión del modelo.

2.3.3. Out of Bag Error (Error OOB)

Una ventaja clave de los bosques aleatorios es la posibilidad de estimar el error de generalización sin necesidad de un conjunto de prueba separado. Esto se logra mediante el error Out of Bag (OOB), que utiliza las observaciones no incluidas en las muestras Bootstrap para evaluar el modelo.

- **Definición de observaciones OOB:** En cada muestra Bootstrap, aproximadamente un tercio de las observaciones no están incluidas. Estas observaciones se denominan “Out of Bag” (OOB) y se utilizan para evaluar el rendimiento del modelo.

- **Cálculo del error OOB:** Para problemas de regresión, el error OOB se calcula como el error cuadrático medio entre las observaciones OOB y sus predicciones:

$$Er_{OOB}(B) = \frac{1}{n} \sum_{i=1}^n \left(y_i - \hat{f}_{OOB}(x_i) \right)^2 \quad (2.7)$$

donde $\hat{f}_{OOB}(x_i)$ es la predicción del bosque aleatorio para la observación x_i , utilizando solo los árboles para los cuales x_i fue una observación OOB.

- **Importancia de las variables:** Los bosques aleatorios permiten cuantificar la importancia de las variables predictoras en problemas de regresión. La importancia se mide como la reducción promedio en el error cuadrático medio (MSE) cuando se utiliza una variable para dividir los nodos, promediada sobre todos los árboles.

Estas características los convierten en una herramienta poderosa y flexible para problemas de regresión, permitiendo obtener estimaciones más precisas y robustas.

2.3.4. Métricas de evaluación del modelo

Para evaluar de manera efectiva el modelo de bosques aleatorios, se emplean tres métricas clave: el Error Cuadrático Medio (MSE), la Raíz del Error Cuadrático Medio (RMSE) y el Coeficiente de Determinación (R^2). Cada una de estas métricas proporciona información valiosa sobre el desempeño del modelo y su capacidad para predecir con precisión los valores de la variable de respuesta.

- **El Error Cuadrático Medio (MSE)** es una métrica que calcula el promedio de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. Matemáticamente, se expresa como:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.8)$$

Esta métrica es especialmente útil porque penaliza los errores grandes de manera más significativa que los errores pequeños, lo que permite identificar modelos con predicciones muy alejadas de los valores reales. Sin embargo, al estar en unidades cuadradas, puede resultar difícil de interpretar directamente en el contexto de los datos.

- **La Raíz del Error Cuadrático Medio (RMSE)** es simplemente la raíz cuadrada del MSE. Esta transformación devuelve la métrica a las mismas unidades que la variable de respuesta, lo que facilita su interpretación. El RMSE proporciona una medida directa de la magnitud promedio del error en las predicciones del modelo. Un valor más bajo de RMSE indica un mejor ajuste del modelo a los datos, ya que refleja una menor discrepancia entre los valores observados y los predichos,

$$RMSE = \sqrt{MSE} \quad (2.9)$$

- **Coefficiente de Determinación (R^2)** mide la proporción de la varianza en la variable de respuesta que es explicada por el modelo. Se calcula como el complemento de la razón entre la suma de los cuadrados de los errores del modelo y la suma de los cuadrados de las desviaciones de los valores observados respecto a su media,

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2.10)$$

Un valor de R^2 cercano a 1 indica que el modelo explica casi toda la variabilidad de los datos, lo que sugiere un ajuste casi perfecto. Por el contrario, un valor cercano a 0 indica que el modelo no logra capturar adecuadamente la variabilidad de los datos, lo que sugiere un mal desempeño predictivo.

En conjunto, estas métricas proporcionan una visión completa del rendimiento del modelo en bosques aleatorios. El MSE y el RMSE ofrecen una medida cuantitativa del error de predicción, mientras que el R^2 brinda una perspectiva sobre la capacidad del modelo para explicar

la variabilidad de los datos. Estas métricas son fundamentales para comparar modelos, ajustar parámetros y garantizar que el modelo seleccionado sea el más adecuado para la tarea en cuestión.

2.3.5. Ajustes del modelo

Por último, se hablará acerca del proceso de ajuste del modelo de bosques aleatorios. El ajuste del modelo se hace con los datos de entrenamiento y se realizará utilizando la paquetería `randomForest` en el lenguaje de programación R. Esta paquetería, desarrollada por Breiman et al. (2018), permite implementar un enfoque de ensamblaje basado en la combinación de múltiples árboles de decisión. El ajuste del modelo es fundamental para optimizar su capacidad predictiva, ya que permite sintonizar los parámetros que controlan el comportamiento de los árboles individuales y el modelo en su conjunto, mejorando su capacidad para generalizar a nuevos datos y reduciendo el riesgo de sobreajuste.

La función `randomForest()` es el núcleo del paquete y permite ajustar un modelo de bosques aleatorios. Sus argumentos principales incluyen:

- **x**: Matriz de predictores de tamaño $n \times p$, donde n es el número de observaciones y p el número de variables predictoras.
- **y**: Vector de la variable respuesta. Si es de tipo `factor`, el modelo se ajusta para clasificación; si es numérico, para regresión. Si se omite, el algoritmo opera en modo no supervisado.
- **ntree**: Número de árboles en el bosque. Por defecto, se generan 500 árboles. Un valor más alto mejora la estabilidad del modelo, pero aumenta el tiempo de cómputo.
- **mtry**: Número de variables predictoras seleccionadas aleatoriamente en cada división. Por defecto, se usa \sqrt{p} para clasificación y $p/3$ para regresión. Este parámetro controla la diversidad de los árboles.

- **importance**: Si es **TRUE**, se calcula la importancia de las variables. Esto es útil para identificar predictores clave.
- **nodesize**: Tamaño mínimo de los nodos terminales. Por defecto, es 1 para clasificación y 5 para regresión. Valores más pequeños permiten árboles más complejos.
- **maxnodes**: Número máximo de nodos terminales en los árboles. Si no se especifica, no hay límite.
- **replace**: Indica si las muestras se toman con reemplazo. Por defecto, es **TRUE**, lo que permite el muestreo bootstrap.

La función `tuneRF()` automatiza la búsqueda del valor óptimo para el parámetro `mtry`. Sus argumentos principales son:

- **x**: Matriz de predictores.
- **y**: Vector de la variable respuesta.
- **mtryStart**: Valor inicial para `mtry`. Por defecto, se usa \sqrt{p} para clasificación y $p/3$ para regresión.
- **ntreeTry**: Número de árboles generados en cada iteración. Por defecto, es 50.
- **stepFactor**: Factor de incremento para `mtry` en cada iteración. Por ejemplo, un valor de 1.5 aumenta `mtry` en un 50% en cada paso.
- **improve**: Umbral de mejora relativa en el error OOB para detener la búsqueda. Por defecto, es 0.05.

La importancia de las variables se calcula utilizando la función `importance()` y se visualiza con `varImpPlot()`. Los argumentos clave incluyen:

- **type**: Tipo de importancia a calcular. Para clasificación, las opciones son 1 (disminución en la precisión) y 2 (disminución en el índice de Gini). Para regresión, solo se calcula la disminución en la precisión.
- **scale**: Si es TRUE, la importancia se escala para facilitar la comparación entre variables.

La salida de `importance()` es una matriz que incluye:

- **Mean Decrease Accuracy**: Mide cuánto disminuye la precisión del modelo al eliminar una variable.
- **Mean Decrease Gini**: Mide la contribución de una variable a la pureza de los nodos en los árboles.

Una vez ajustados todos los parámetros importantes de la función `randomForest()` y con el modelo entrenado usando los datos de entrenamiento, se procede a evaluar su desempeño generando las predicciones. Es fundamental asegurarse de que el modelo esté correctamente sintonizado para obtener predicciones precisas y robustas.

Capítulo 3

Análisis de los resultados

En este capítulo se realizará el análisis de los resultados del modelo de bosques aleatorios propuesto, el cual será estimado en el programa R utilizando la paquetería `randomForest`. Se utilizará como base de datos un conjunto de razones financieras obtenidas de los estados financieros de las empresas que componen el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, con el objetivo de determinar las variables predictoras del modelo. Estas variables serán sometidas a un análisis de correlación para evitar redundancias y garantizar la independencia entre ellas. Posteriormente, se estimará el modelo de bosques aleatorios para cada uno de los precios accionarios que componen el índice, seleccionando los mejores modelos basados en su desempeño. A estos modelos se les aplicará un intervalo de confianza al RMSE para enmarcar con la mayor precisión posible el error del modelo y tenerlo presente en las conclusiones. Finalmente, se presentarán los pronósticos generados para cada una de las acciones que integran el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

3.1. Modelo de Bosques Aleatorios Propuesto

En esta sección se presenta el modelo de bosques aleatorios utilizado para el pronóstico del precio de las acciones en el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico. El objetivo es estimar el precio de las acciones a partir de un conjunto de razones financieras que reflejan el desempeño y la estructura financiera de las empresas analizadas.

3.1.1. Formulación Matemática del Modelo

Dado un conjunto de datos de entrenamiento $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, donde:

- $\mathbf{x}_i \in \mathbb{R}^p$ representa el conjunto de razones financieras de la empresa en el periodo i ,
- $y_i \in \mathbb{R}$ es el precio de la acción correspondiente,

se construye un bosque aleatorio de regresión, el cual está compuesto por T árboles de decisión $h_t(X; \Theta_t)$. Cada árbol es entrenado con una muestra aleatoria de los datos mediante bagging y una selección aleatoria de variables en cada nodo. La predicción final del modelo se obtiene como el promedio de las predicciones de todos los árboles:

$$\hat{y}(X) = \frac{1}{T} \sum_{t=1}^T h_t(X; \Theta_t)$$

donde:

- $\hat{y}(X)$ es la estimación del precio de la acción.
- T es el número total de árboles en el bosque.
- $h_t(X; \Theta_t)$ es la predicción generada por el t -ésimo árbol, el cual se construye con un conjunto aleatorio de muestras y características Θ_t .

Para la estimación del pronóstico del precio de las acciones, el modelo utiliza un conjunto de razones financieras seleccionadas como variables predictoras X_i , definidas como:

$$X_i = \left\{ \begin{array}{l} (\text{Margen de utilidad})_1 \\ (\text{ROA})_2 \\ (\text{Razón de deuda})_3 \\ (\text{Razón corriente})_4 \\ (\text{Prueba ácida})_5 \\ (\text{Relación [Precio/Ganancias]})_6 \\ (\text{Relación [Precio/Valor Libros]})_7 \\ (\text{Crecimiento de ganancias})_8 \\ (\text{Rendimiento por dividendo})_9 \end{array} \right.$$

Cada una de estas variables proporciona información clave sobre la rentabilidad, liquidez y valuación de las empresas, permitiendo que el modelo capture patrones relevantes para la predicción del precio de las acciones.

Cabe mencionar que en la presente investigación se estimarán 12 modelos distintos, ya que el índice bajo estudio está compuesto por 12 activos financieros. Para cada empresa del índice, se construirá un modelo específico con la misma metodología descrita anteriormente, obteniendo así las siguientes formulaciones:

- Modelo 1: $\hat{y}_{AC}(X)$
- Modelo 2: $\hat{y}_{BIMBOA}(X)$
- Modelo 3: $\hat{y}_{CHEDRAUI}(X)$
- Modelo 4: $\hat{y}_{LACOMERUBC}(X)$
- Modelo 5: $\hat{y}_{CUERVO}(X)$

- Modelo 6: $\hat{y}_{FEMSAUBD}(X)$
- Modelo 7: $\hat{y}_{GRUMAB}(X)$
- Modelo 8: $\hat{y}_{HERDEZ}(X)$
- Modelo 9: $\hat{y}_{KIMBERA}(X)$
- Modelo 10: $\hat{y}_{KOFUBL}(X)$
- Modelo 11: $\hat{y}_{SORIANA}(X)$
- Modelo 12: $\hat{y}_{WALMEX}(X)$

Cada uno de estos modelos será evaluado de manera independiente, permitiendo pronosticar de manera correcta el precio de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

3.1.2. Base de Datos

En esta sección se describe la base de datos utilizada en la presente investigación, la cual está conformada por información financiera de las empresas que componen el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

El índice está compuesto por 12 empresas, cuyas empresas se presentan en la Tabla 1.1. Los datos fueron extraídos de la plataforma Economatica, asegurando que todas las empresas contarán con información continua y sin interrupciones, lo cual permite mejorar la precisión del modelo. Cabe destacar que los estados financieros de las empresas se reportan de manera trimestral, por lo que el período de estudio comprende desde el 31 de marzo de 2017 hasta el 31 de marzo de 2024, abarcando un total de 29 trimestres.

Cada acción del índice cuenta con una base de datos estructurada de manera homogénea, conteniendo 16 indicadores financieros y 29 observaciones trimestrales. Estos indicadores

fueron obtenidos a partir de los balances generales y estados de resultados de cada empresa. En la Tabla 3.1 se presentan los 16 elementos recopilados.

Tabla 3.1 *Medidas financieras*

Medidas financieras para cada empresa	
Precio de la Acción	Inventario Corriente
Utilidad Neta	Ganancia por Acción
Ventas Netas	Valor Contable por Acción
Activo Total	Dividendo Total
Deuda Total	Dividendo por Acción
Patrimonio Neto	Total de Activos en Circulación
Activo Corriente	Ganancias Actuales
Pasivo Corriente	Ganancias Anteriores

Fuente: Elaboración propia con datos extraídos de Economatica.

El objetivo de extraer estas 16 medidas financieras es calcular 9 razones financieras que servirán como variables predictoras en el modelo de bosques aleatorios para la estimación de precios de acciones. Esto se alinea con el propósito de la investigación, que consiste en realizar pronósticos para los precios de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, empleando como metodología el modelo estadístico de Bosques Aleatorios y utilizando razones financieras como variables predictoras. Con ello, se busca extraer conocimiento valioso que contribuya a la comprensión y análisis del comportamiento del sector.

3.1.3. Razones Financieras

El análisis de razones financieras es una herramienta fundamental en la evaluación del desempeño y la estabilidad de las empresas. Según Damodaran (2012), las razones financieras permiten a los inversionistas y analistas interpretar la información contable y traducirla en indicadores clave de rentabilidad, eficiencia, solvencia y crecimiento. Graham y Dodd (2009)

también enfatizan que estas métricas son esenciales para la valoración de empresas y la toma de decisiones de inversión, ya que proporcionan una visión clara de su situación financiera.

Para la presente investigación, se seleccionaron nueve razones financieras debido a su capacidad para ofrecer una visión integral de la salud financiera de una empresa y su potencial de crecimiento. Estas razones fueron calculadas a partir de los estados financieros trimestrales de las empresas analizadas, utilizando datos extraídos de la plataforma Economatica. Las fórmulas utilizadas para su cálculo provienen de Ross et al. (2019). A continuación, se presentan las razones seleccionadas junto con sus respectivas fórmulas:

- **Margen de Utilidad:**

$$\frac{\text{Utilidad Neta}}{\text{Ventas Netas}}$$

- **Retorno sobre Activos (ROA):**

$$\frac{\text{Utilidad Neta}}{\text{Activo Total}}$$

- **Razón Deuda:**

$$\frac{\text{Deuda Total}}{\text{Activo Total}}$$

- **Razón Corriente:**

$$\frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$$

- **Prueba Ácida:**

$$\frac{\text{Activo Corriente} - \text{Inventario Corriente}}{\text{Pasivo Corriente}}$$

- **Relación Precio-Ganancias (P/E):**

$$\frac{\text{Precio de la Accin}}{\text{Ganancia por Accin}}$$

- **Relación Precio-Valor en Libros (P/B):**

$$\frac{\text{Precio de la Accin}}{\text{Valor Contable por Accin}}$$

- **Crecimiento de Ganancias:**

$$\frac{\text{Ganancias Actuales} - \text{Ganancias Anteriores}}{\text{Ganancias Anteriores}}$$

- **Rendimiento por Dividendo:**

$$\frac{\text{Dividendo por Accin}}{\text{Precio de la Accin}}$$

La importancia de estas razones financieras radica en su capacidad para proporcionar información clave sobre la rentabilidad, el riesgo y la valuación de una empresa. De acuerdo con Brigham y Ehrhardt (2021), una combinación adecuada de estos indicadores es clave para la toma de decisiones estratégicas de inversión, ya que permiten identificar empresas con sólidos fundamentos financieros y un alto potencial de valorización en el mercado de acciones. La Tabla 3.2 muestra un resumen de la influencia de cada razón financiera en la valoración de una acción.

El análisis de estas razones financieras permite a los inversionistas obtener una visión holística de una empresa desde múltiples ángulos: rentabilidad, eficiencia, solvencia, valor y crecimiento. De acuerdo con Brigham y Ehrhardt (2021), una combinación adecuada de estos indicadores es clave para la toma de decisiones estratégicas de inversión, ya que permiten identificar empresas con sólidos fundamentos financieros y un alto potencial de valorización en el mercado de acciones.

Tabla 3.2 Impacto de las Razones Financieras

Razón Financiera	Importancia	Impacto
Margen de Utilidad	Evalúa la eficiencia en la generación de ganancias.	Un mayor margen indica mejor control de costos y mayor rentabilidad.
Retorno sobre Activos (ROA)	Mide la eficiencia en el uso de los activos.	Un ROA alto sugiere una gestión eficiente y mayor rentabilidad.
Razón Deuda	Indica el nivel de apalancamiento financiero.	Un alto endeudamiento puede aumentar el riesgo financiero.
Razón Corriente	Evalúa la liquidez de la empresa.	Un valor adecuado garantiza estabilidad en el corto plazo.
Prueba Ácida	Mide la capacidad de pago inmediato.	Un nivel adecuado indica solvencia sin depender de inventarios.
Relación Precio-Ganancias (P/E)	Evalúa la valoración de la empresa.	Una P/E baja puede indicar una oportunidad de inversión.
Relación Precio-Valor en Libros (P/B)	Compara el valor de mercado con el valor contable.	Una P/B baja puede indicar una acción infravalorada.
Crecimiento de Ganancias	Muestra la evolución de las utilidades.	Un crecimiento sostenido es indicativo de estabilidad y expansión.
Rendimiento por Dividendo	Mide la rentabilidad del dividendo.	Un alto rendimiento es atractivo para inversionistas que buscan ingresos pasivos.

Fuente: Elaboración propia con información de Graham y Dodd (2009).

3.1.4. Análisis de Correlación

Anteriormente se mencionó la importancia de las principales razones financieras que los inversionistas consideran para seleccionar acciones en el mercado. A continuación, se realiza un análisis de la correlación entre estas razones financieras, las cuales constituyen las variables que afectarán el pronóstico del precio de las acciones. Este análisis se lleva a cabo de manera individual para cada empresa, ya que el comportamiento de las variables puede variar significativamente entre diferentes acciones. Por esta razón, cada activo requiere un modelo de bosques aleatorios individual.

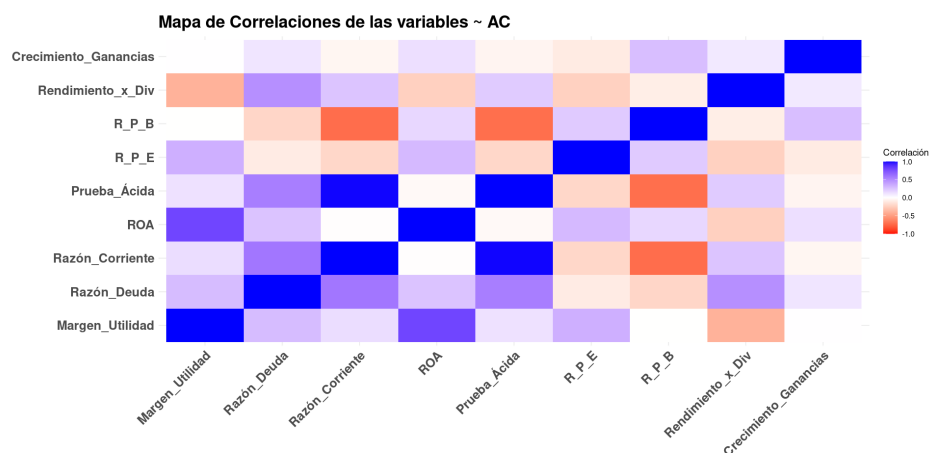
Si bien artículos como *Classification and Regression by Random Forest* (2002) sugieren que

un solo modelo de bosques aleatorios, correctamente aplicado y ajustado, puede generalizarse para pronosticar múltiples activos, otros autores como Kuhn y Johnson (2013) argumentan que, en la práctica, este enfoque no siempre es adecuado. Kuhn señala que, a pesar de compartir características similares, las variables no siempre se ajustan de la misma manera en diferentes contextos, lo que puede llevar a errores en los pronósticos. Por ello, es necesario desarrollar un modelo individual para cada acción.

Dicho lo anterior, la selección de variables se realizó mediante un análisis de correlación, ya que en los bosques aleatorios la correlación entre las variables de entrada puede tener un impacto significativo en el rendimiento del modelo. Este criterio se aplicó con el objetivo de evitar redundancia en la información. Aunque no existe un umbral estrictamente definido para la correlación en los Bosques Aleatorios, como regla general se recomienda evitar incluir variables con una correlación mayor a 0.8 o menor a -0.9 (James et al., 2013). Esto se debe a que, si las variables están altamente correlacionadas, los árboles pueden perder diversidad al utilizar información repetida.

A continuación, se presentan los mapas de calor individuales por acción, generados en R, en las Figuras 3.1, 3.2, 3.3, 3.4 y 3.5 los cuales proporcionan una perspectiva clara del grado de correlación entre las variables de entrada para cada acción.

Figura 3.1 *Análisis de correlación (1)*



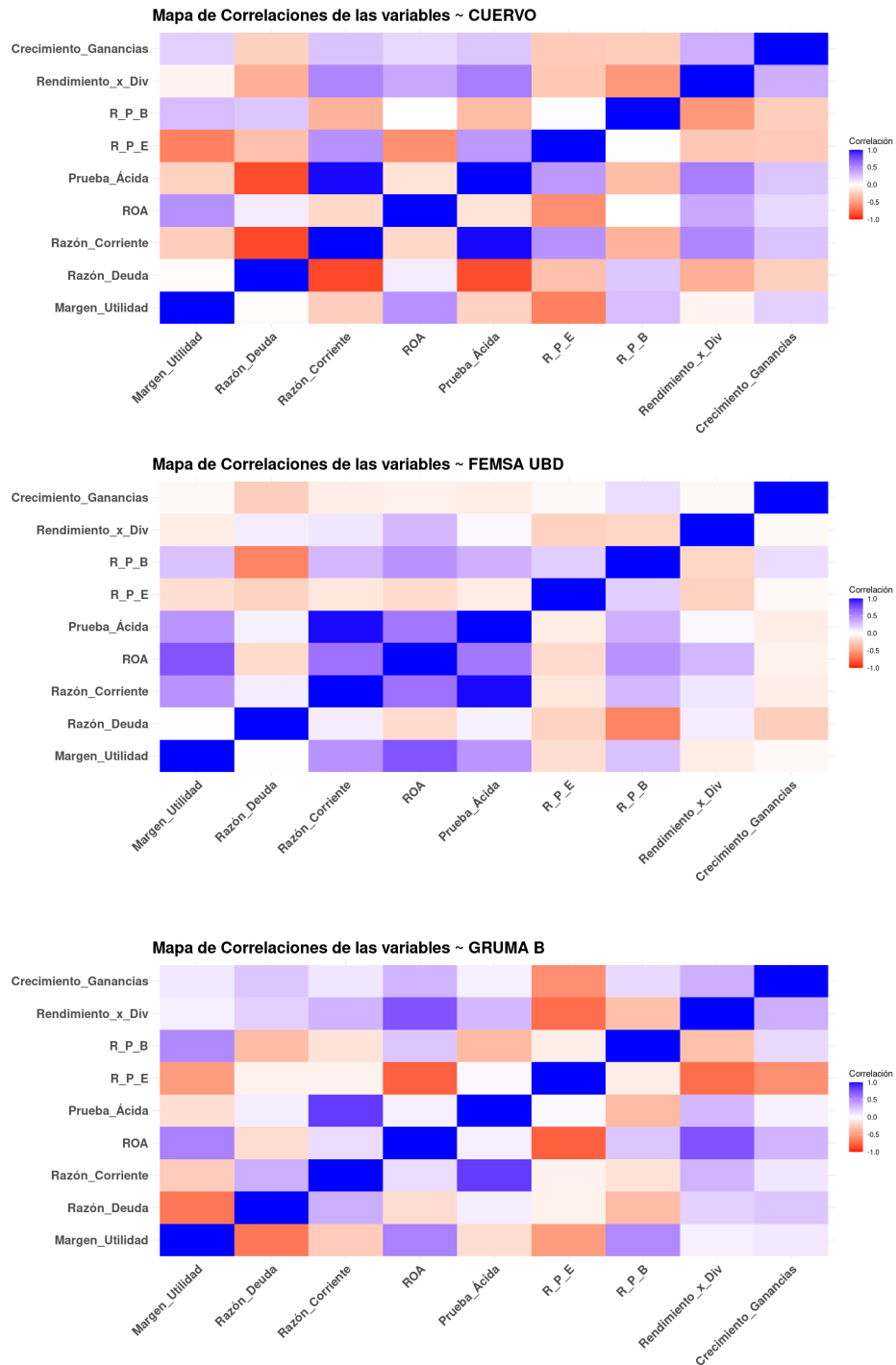
Fuente: Elaboración propia.

Figura 3.2 *Análisis de correlación (2)*



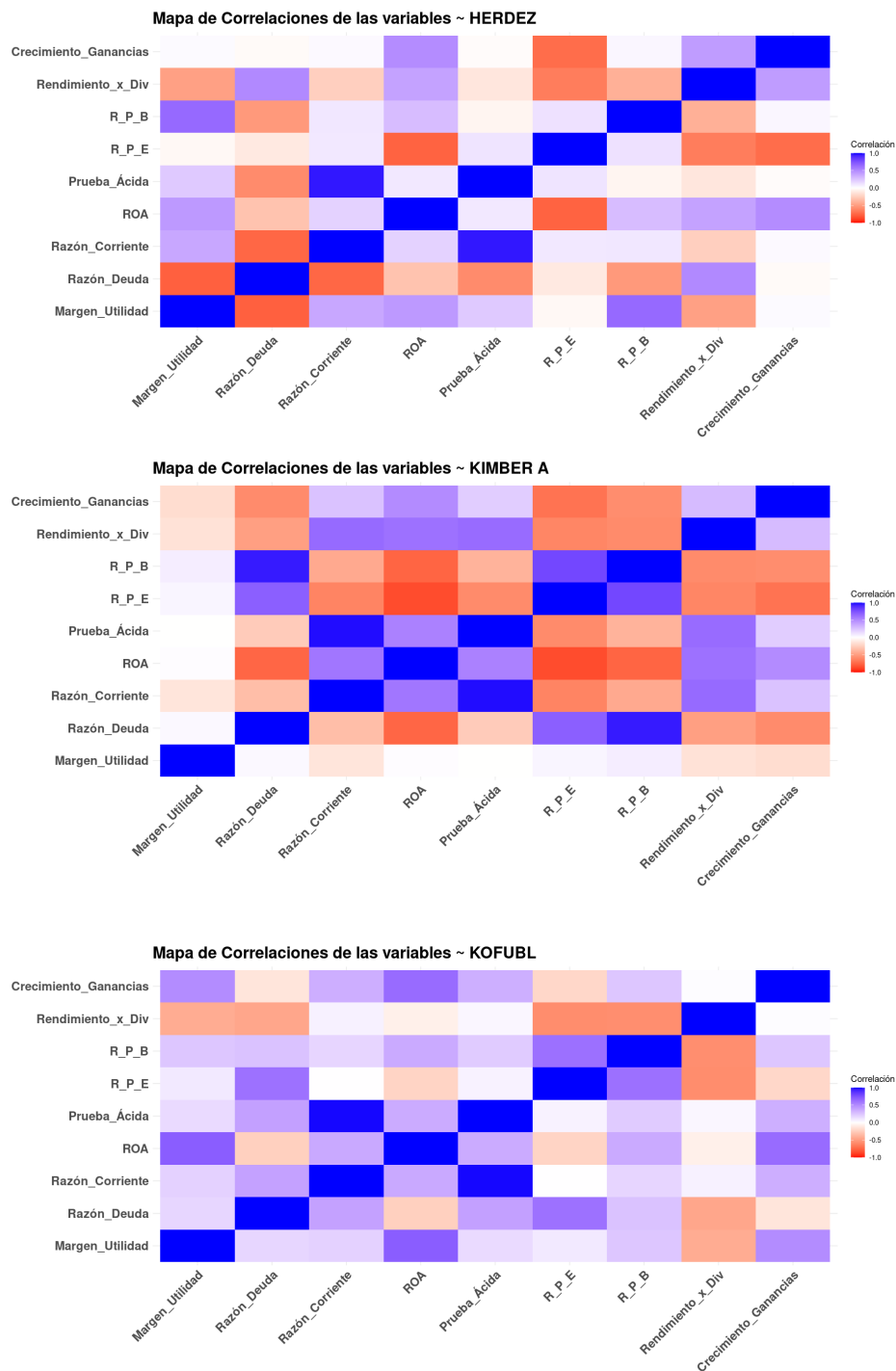
Fuente: Elaboración propia.

Figura 3.3 *Análisis de correlación (3)*



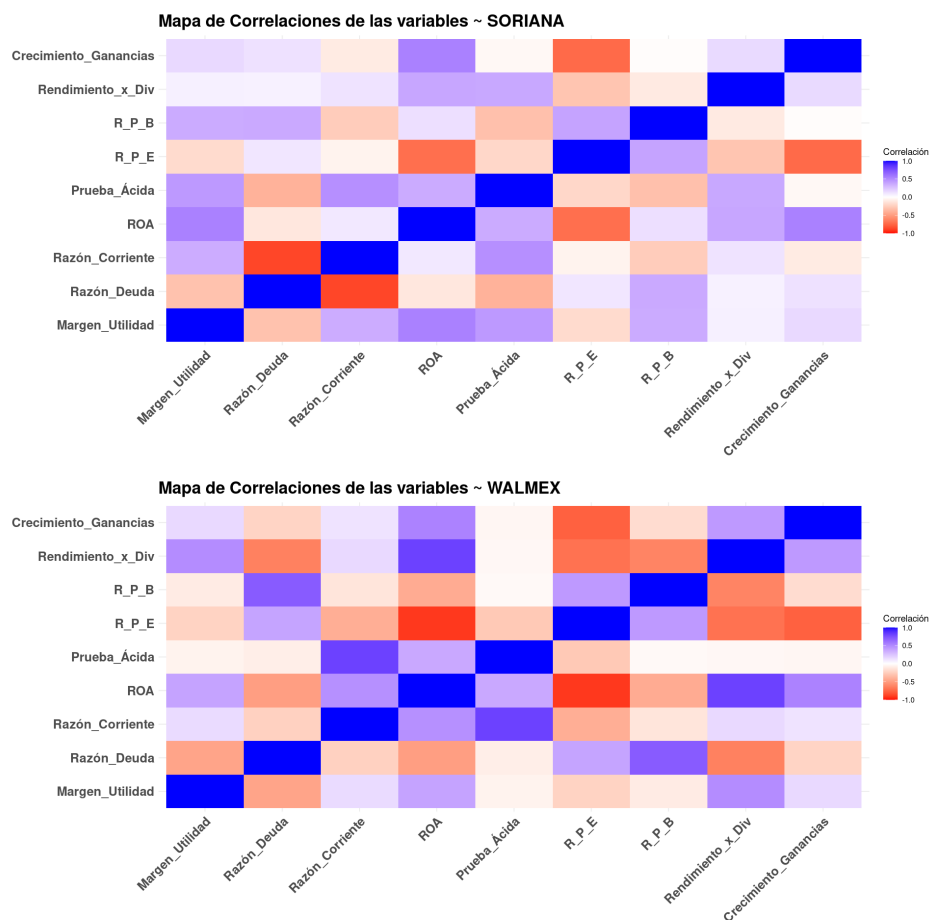
Fuente: Elaboración propia.

Figura 3.4 *Análisis de correlación (4)*



Fuente: Elaboración propia.

Figura 3.5 *Análisis de correlación (5)*



Fuente: Elaboración propia.

El análisis arrojó que para las acciones de GrumaB, Soriana, Walmex y Chredaui, todas las variables cumplen con el parámetro de correlación establecido, lo que indica que no presentan multicolinealidad significativa y pueden ser incluidas en el modelo sin mayor ajuste. Sin embargo, para otras acciones como AC, BIMBO A, LA COMER UBC, CUERVO, FEMSA UBD, HERDEZ, KIMBER A y KOFUBL, se identificaron pares de variables con correlaciones superiores a 0.8, lo que sugiere una posible redundancia en la información que podrían aportar al modelo.

- Prueba Ácida y Razón Corriente: Estas dos variables, que miden la liquidez a corto

plazo de una empresa, mostraron una correlación elevada ($\rho > 0.8$) en varias acciones. Dado que ambas evalúan la capacidad de la empresa para cubrir sus obligaciones a corto plazo, se decidió conservar la Razón Corriente por su enfoque más amplio en la salud financiera de la empresa, descartando la Prueba Ácida solo en los modelos donde esta correlación excede el umbral.

- Razón Deuda y Relación Precio-Valor en Libros R(P/B): En el caso específico de KIMBER A, se observó una correlación extremadamente alta ($\rho = 0.9520$) entre estas dos variables. La Razón Deuda evalúa el riesgo financiero y la capacidad de endeudamiento de la empresa, mientras que la Relación Precio-Valor en Libros R(P/B) es fundamental para determinar si una acción está sobrevalorada o infravalorada en relación con su valor contable. Dado que la Relación R(P/B) tiene un impacto más directo en la valoración de mercado de la acción, se decidió conservar esta variable y excluir la Razón Deuda únicamente en el modelo de KIMBER A.

La selección de variables se basó en dos criterios principales:

1. Relevancia para el precio de la acción: Se priorizaron aquellas variables que tienen un impacto más directo en la valoración de mercado, como la Relación Precio-Ganancias R(P/E) y la Relación Precio-Valor en Libros R(P/B), ya que estas proporcionan información clave sobre la percepción del mercado respecto a la empresa.
2. Enfoque en la salud financiera general: Variables como la Razón Corriente y el Retorno sobre Activos (ROA) fueron seleccionadas por su capacidad para ofrecer una visión integral de la solidez financiera de la empresa, lo cual es fundamental para predecir su desempeño futuro.

Una vez explicado, se presenta la Tabla 3.3 que resume las variables de entrada seleccionadas para cada acción, considerando las eliminaciones específicas donde se identificaron correlaciones elevadas:

Tabla 3.3 Variables de entrada para cada acción

AC	BIMBO A	CHEDRAUI
Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo Prueba Ácida
LA COMER UBC	CUERVO	FEMSA UBD
Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo
KOFUBL	HERDEZ	KIMBER A
Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo	Margen de Utilidad ROA Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo
GRUMA B	SORIANA	WALMEX
Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo Prueba Ácida	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo Prueba Ácida	Margen de Utilidad ROA Razón Deuda Razón Corriente R(P/E) R(P/B) Crecimiento de Ganancias Rendimiento por Dividendo Prueba Ácida

Fuente: Elaboración propia.

El proceso de selección de variables ha permitido identificar aquellas métricas financieras que, al no presentar multicolinealidad excesiva, pueden contribuir de manera significativa a la precisión del modelo de bosques aleatorios. La eliminación de variables redundantes, como la Prueba Ácida y la Razón Deuda, se realizó únicamente en los modelos donde estas presentaban correlaciones elevadas, asegurando que el modelo no se vea afectado por sesgos derivados de la correlación excesiva entre predictores. Con esta selección, se espera que el modelo pueda ofrecer predicciones robustas y confiables sobre el comportamiento del precio de las acciones.

3.2. Estimación de los bosques aleatorios

Para la construcción del modelo de pronóstico, en primera instancia, se dividen los datos en conjuntos de entrenamiento y prueba. La partición se realiza de manera aleatoria, asignando el 70 % de los datos al entrenamiento y el 30 % restante a la prueba. El conjunto de entrenamiento permite ajustar el modelo, mientras que el conjunto de prueba se utiliza para evaluar su capacidad predictiva, para de esta manera generar los pronósticos sobre los precios de las acciones.

Inicialmente, se aplica el modelo de bosques aleatorios utilizando su configuración por defecto. Esto permite obtener una primera aproximación del desempeño del modelo y sirve como referencia para las siguientes optimizaciones. Posteriormente, se lleva a cabo un ajuste más detallado, explorando y configurando cada uno de los parámetros del modelo, tales como el número de árboles, la profundidad máxima, la cantidad de variables consideradas en cada división y el criterio de división, entre otros. Este proceso de ajuste tiene como objetivo encontrar la mejor combinación de hiperparámetros que minimice el error de predicción y mejore la estabilidad del modelo.

Finalmente, para obtener una evaluación más precisa del error del modelo, se calcula la raíz

del error cuadrático medio (RMSE) y se le aplica un intervalo de confianza al 95 %. Este intervalo permite cuantificar la variabilidad en la estimación del error, proporcionando una mejor interpretación sobre la precisión del modelo en la predicción de los precios de las acciones del Índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

3.2.1. Resultados de los primeros modelos

El primer modelo de Bosques Aleatorios se aplicó a cada acción utilizando la configuración predeterminada de la función `randomForest`. En esta configuración, el número de árboles generados es de 500, mientras que el número de variables consideradas en cada división (`mtry`) se define como:

$$mtry = \frac{p}{3} \quad (3.1)$$

donde p representa el total de variables predictoras en el modelo. Dado que cada acción puede tener un conjunto distinto de variables predictoras, el valor de $mtry$ varía en cada modelo.

Esta primera implementación proporciona un punto de referencia para evaluar el desempeño del modelo sin ajustes adicionales. A partir de los resultados obtenidos, se analizará la calidad de ajuste del modelo y se identificarán posibles áreas de mejora. Posteriormente, se procederá a la optimización de los hiperparámetros con el objetivo de reducir el error de predicción y mejorar la precisión del modelo.

En la Tabla 3.4 se presentan los primeros resultados obtenidos para cada modelo aplicado a las diferentes acciones. Se incluyen el número total de árboles utilizados, el valor de $mtry$, el coeficiente de determinación R^2 como medida de ajuste del modelo, el error cuadrático medio (MSE) y la raíz del error cuadrático medio ($RMSE$). Estos indicadores permiten evaluar la calidad de las predicciones y sirven como base para la posterior optimización del modelo.

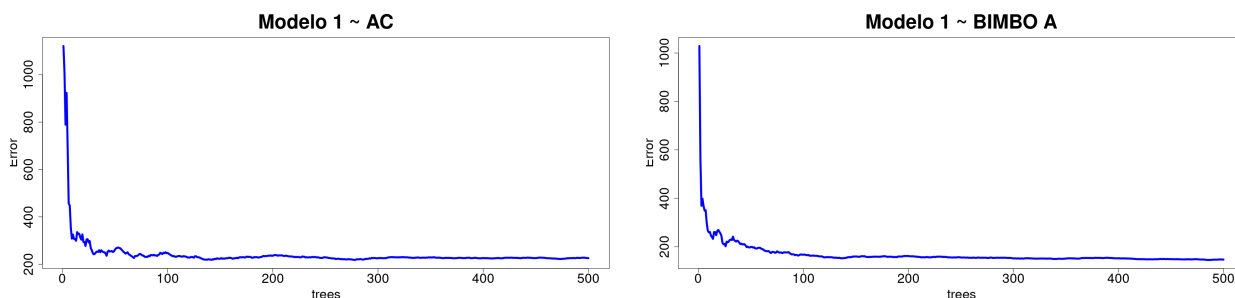
Tabla 3.4 Resultados de los primeros modelos

Acción	Núm. árboles	<i>mtry</i>	R^2	MSE	RMSE
AC	500	2	58.51 %	225.61	15.02
BIMBO A	500	2	54.30 %	146.79	12.11
CHEDRAUI	500	3	82.45 %	114.14	10.68
LA COMER UBC	500	2	63.97 %	24.70	4.96
CUERVO	500	2	85.38 %	11.02	3.31
FEMSA UBD	500	2	51.58 %	219.84	14.82
GRUMA B	500	3	31.78 %	785.85	28.03
HERDEZ	500	2	35.65 %	18.26	4.27
KIMBER A	500	2	14.46 %	10.54	3.24
KOFUBL	500	2	22.84 %	253.18	15.91
SORIANA	500	3	45.52 %	21.09	4.59
WALMEX	500	3	33.78 %	53.51	7.31

Fuente: Elaboración propia.

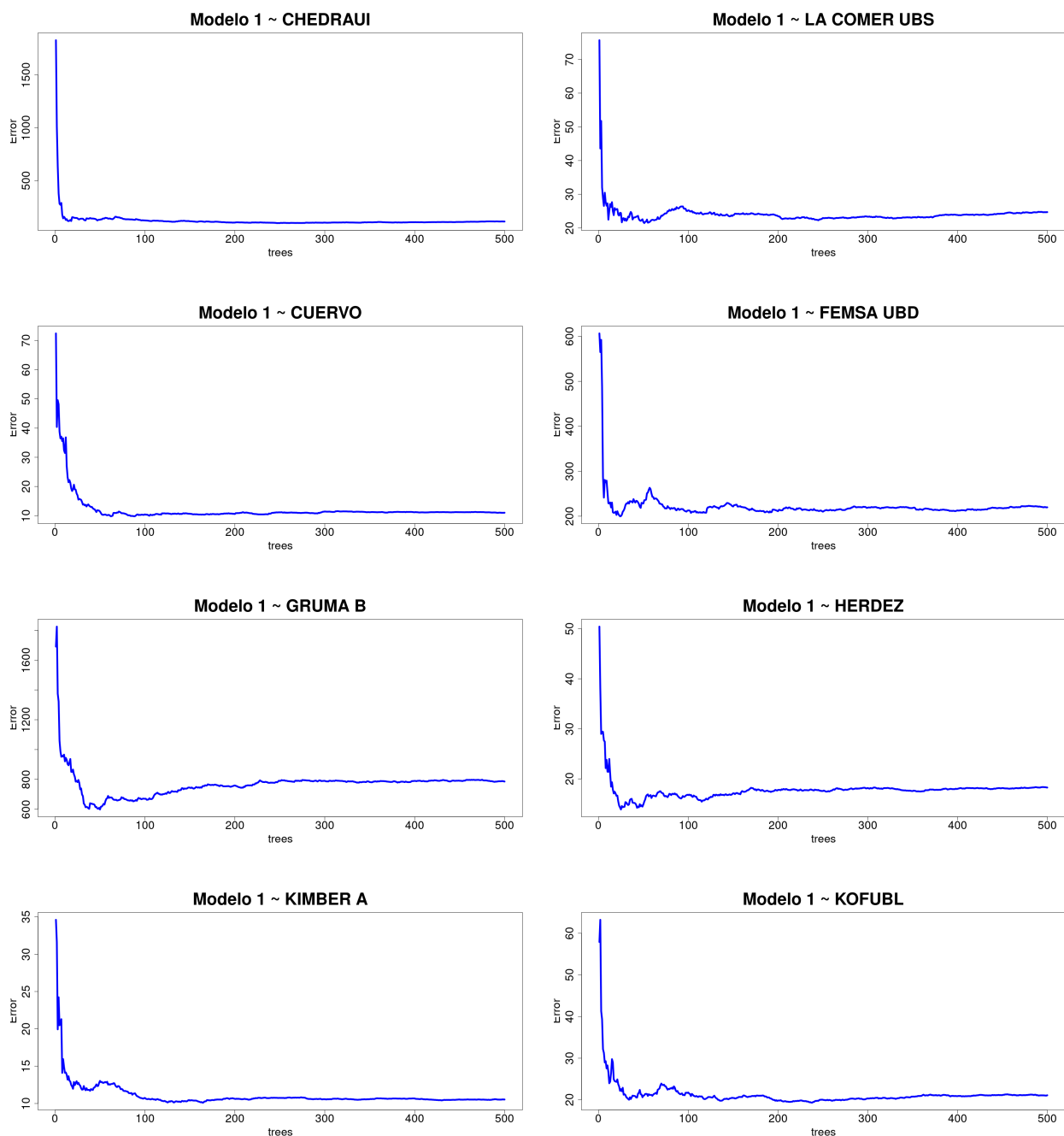
Como se puede observar en los resultados, el desempeño del modelo varía entre las distintas acciones. En primera instancia, se utiliza el número de árboles predeterminado de 500. En las Figuras 3.6, 3.7 y 3.8 se puede notar que, a medida que aumenta el número de árboles en los diferentes modelos, el error disminuye considerablemente hasta alcanzar un punto de estabilidad.

Figura 3.6 Disminución del error con base en el aumento de árboles en cada modelo (1)

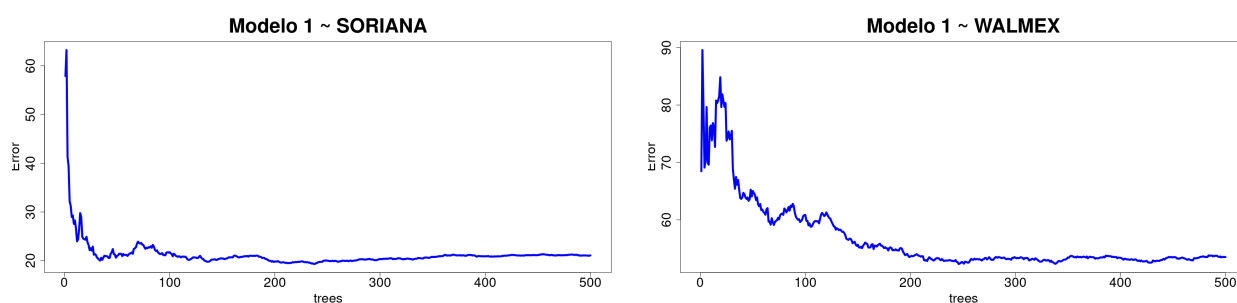


Fuente: Elaboración propia.

Figura 3.7 Disminución del error con base en el aumento de árboles en cada modelo (2)



Fuente: Elaboración propia.

Figura 3.8 Disminución del error con base en el aumento de árboles en cada modelo (3)

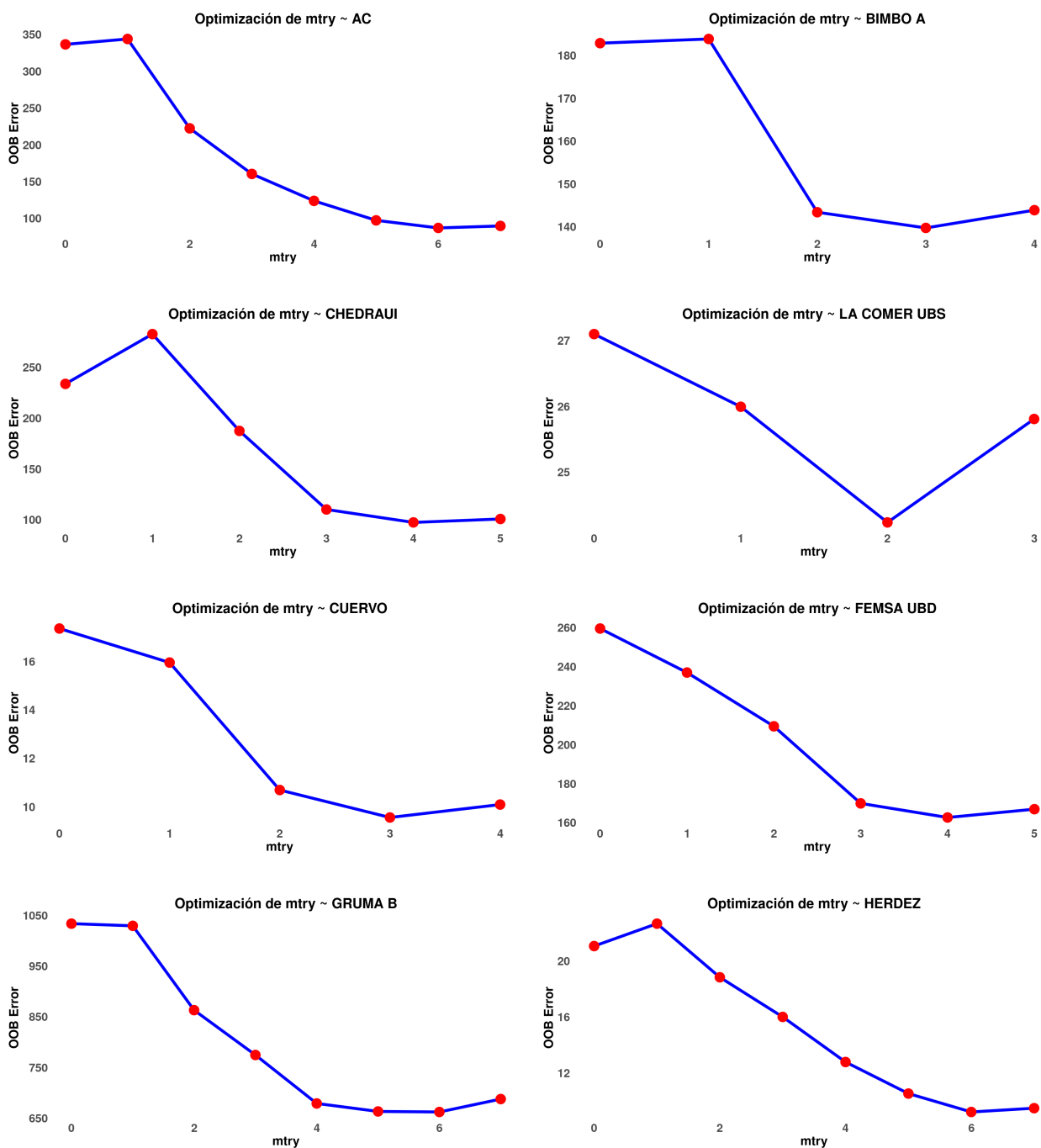
Fuente: Elaboración propia.

Es importante destacar que, independientemente de si se agregan o eliminan más árboles del valor predeterminado, el error se mantiene prácticamente constante una vez estabilizado. Esto indica que, más allá de cierto umbral, el incremento en la cantidad de árboles no mejora significativamente el desempeño del modelo, sino que solo aumenta el costo computacional sin aportar beneficios adicionales en la precisión de las predicciones.

Por otro lado, a diferencia del número de árboles, el hiperparámetro $mtry$ sí requiere ajustes específicos para cada modelo. De manera predeterminada, se toma como $mtry = p/3$ (ver 3.1), lo cual no siempre es la opción más óptima. Debido a que este parámetro define cuántas variables predictoras se consideran en cada división, su ajuste adecuado puede mejorar considerablemente la capacidad del modelo para captar patrones en los datos. Por esta razón, es fundamental optimizar $mtry$ para cada acción con el fin de obtener el mejor desempeño en términos de precisión y estabilidad del modelo.

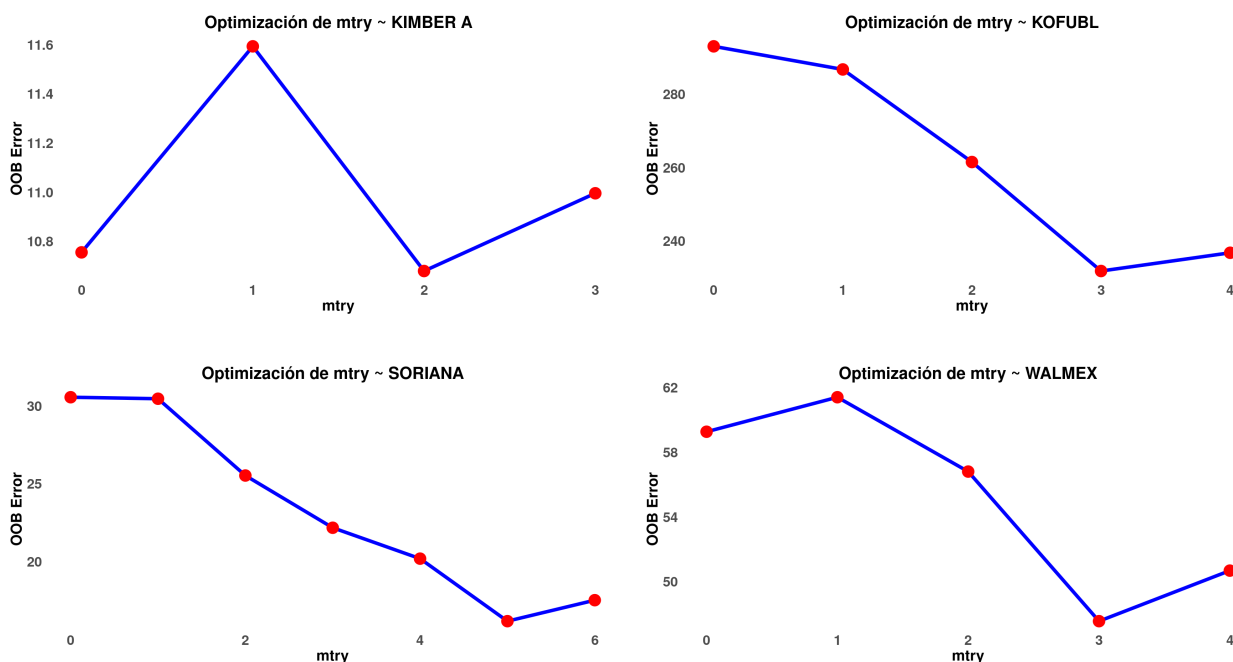
En las siguientes Figuras 3.9 y 3.10 se muestra el comportamiento del error a medida que se ajusta $mtry$. Para ello, se empleó la función `tuneRF`, que permite realizar una búsqueda iterativa del valor óptimo de este hiperparámetro en función de la reducción del error de predicción. A partir de un valor inicial, $mtry$ se ajustó progresivamente con base en un factor de modificación y un criterio de mejora mínima en el error, lo que permitió explorar distintas configuraciones y determinar cuál ofrecía el mejor desempeño para cada modelo.

Figura 3.9 Afinación del argumento mtry (1)



Fuente: Elaboración propia.

Figura 3.10 Afinación del argumento *mtry* (2)



Fuente: Elaboración propia.

A partir de este análisis, se identificó la necesidad de realizar un ajuste más preciso de los hiperparámetros del modelo. Las observaciones obtenidas en la optimización de *mtry* permitieron establecer un intervalo adecuado para generar distintas combinaciones de modelos en cada una de las acciones analizadas.

Para lograr esto, se empleó la técnica de búsqueda en rejilla (*Grid Search*)¹, la cual explora sistemáticamente diversas combinaciones de hiperparámetros con el objetivo de identificar la configuración que minimiza el error cuadrático medio (MSE) y proporciona el mejor ajuste del modelo. Este proceso garantiza que el modelo seleccionado posea un equilibrio entre precisión y estabilidad en la predicción.

¹De acuerdo con Bergstra y Bengio (2012), la búsqueda en rejilla es una técnica común en la optimización de hiperparámetros que consiste en evaluar combinaciones predefinidas de parámetros para seleccionar la mejor configuración según una métrica de desempeño.

Dado que el enfoque detallado de la búsqueda en rejilla y la configuración de los hiperparámetros se incluye en el Anexo 1, en la siguiente sección se presentan directamente los mejores modelos obtenidos para cada acción, listos para generar los pronósticos de las acciones.

3.2.2. Mejores modelos

Una vez realizado el ajuste de hiperparámetros y evaluadas las distintas configuraciones, en esta sección se presentan los mejores modelos obtenidos para cada una de las acciones analizadas. Estos modelos fueron seleccionados con base en su capacidad predictiva, asegurando que minimicen el error y maximicen el ajuste explicativo.

Para cada acción, se muestran diversas características relevantes del modelo, incluyendo las variables predictoras con mayor importancia en la estimación del precio, el coeficiente de determinación (R^2) que refleja el nivel de ajuste del modelo, y el histograma del error de predicción Out-of-Bag (OOB_RMSE). Además, se presentan comentarios generales sobre el desempeño de los modelos, permitiendo comprender mejor su capacidad de generalización y estabilidad.

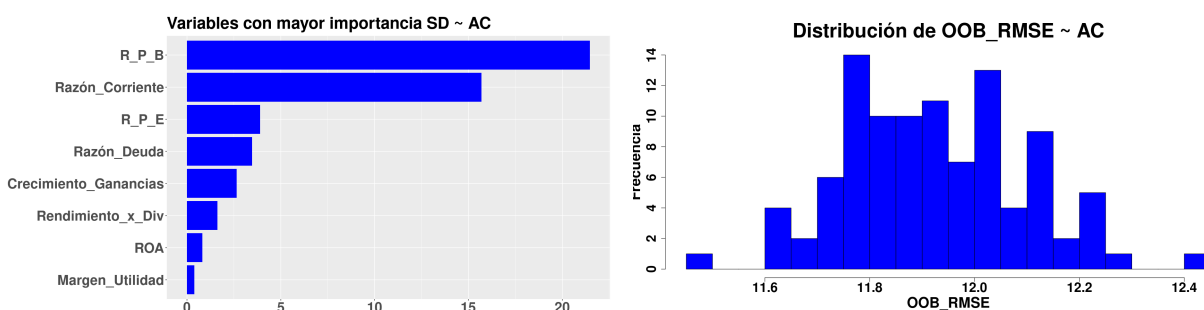
Tabla 3.5 Resultados del mejor modelo de bosques aleatorios para AC

R^2	MSE	RMSE
76.81 %	140.37	11.85

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para AC presenta un coeficiente de determinación R^2 del 76.81 %, lo que indica un alto nivel de varianza explicada en los datos. Las métricas de error, con un MSE de 140.37 y un RMSE de 11.85, sugieren que las predicciones del modelo son bastante precisas, con una desviación promedio de aproximadamente 11.85 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida de Razón Corriente, Relación del Precio con las Ganancias (R_P_E), y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.11 Características del mejor modelo de AC



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 11.5 y 12.4, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 58.51 %, un MSE de 225.61 y un RMSE de 15.02 , el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

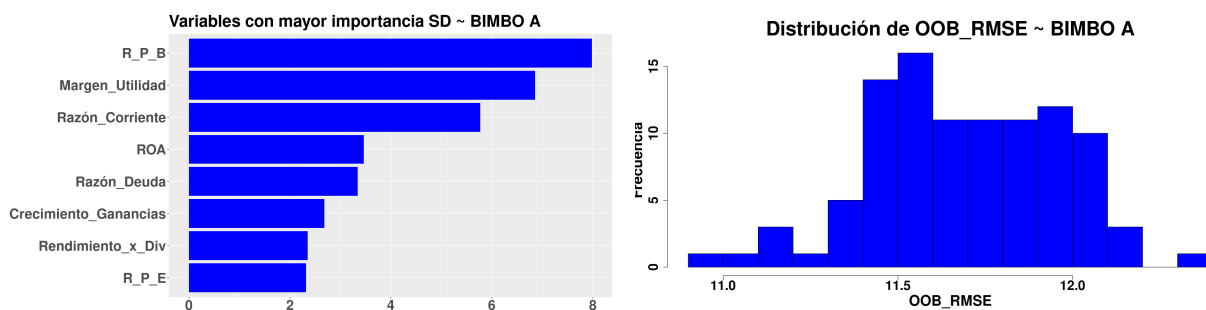
Tabla 3.6 Resultados del mejor modelo de bosques aleatorios para BIMBO A

R^2	MSE	RMSE
55.51 %	135.37	11.63

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para BIMBO A presenta un coeficiente de determinación R^2 del 55.51 %, lo que indica un nivel regular de varianza explicada en los datos. Las métricas de error, con un MSE de 135.37 y un RMSE de 11.63, sugieren que las predicciones del modelo pueden ser lo suficientemente precisas, con una desviación promedio de aproximadamente 11.63 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida de Margen de Utilidad, Razón Corriente, y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.12 Características del mejor modelo de BIMBO A



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 11.0 y 12.0, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 54.30 %, un MSE de 146.79 y un RMSE de 12.11, el mejor modelo muestra solo una pequeña mejora en todas las métricas. Por lo tanto, el ajuste de los hiperparámetros han contribuido muy poco a mejorar el rendimiento predictivo del modelo.

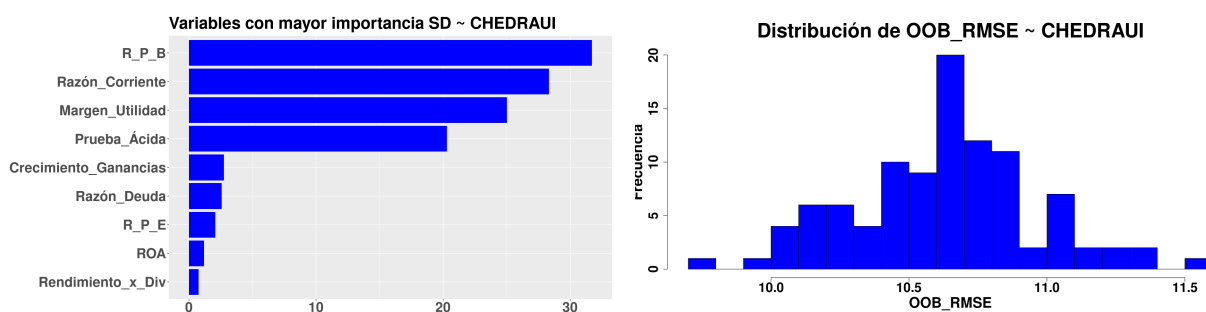
Tabla 3.7 Resultados del mejor modelo de bosques aleatorios para CHEDRAUI

R^2	MSE	RMSE
88.26 %	100.63	10.03

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para CHEDRAUI presenta un coeficiente de determinación R^2 del 88.26 %, lo que indica un alto nivel de varianza explicada en los datos. Las métricas de error, con un MSE de 100.63 y un RMSE de 10.03, sugieren que las predicciones del modelo son bastante precisas, con una desviación promedio de aproximadamente 10.03 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida de Razón Corriente, Margen Utilidad, y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.13 Características del mejor modelo de CHEDRAUI



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 10.0 y 11.5, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 82.45 %, un MSE de 114.14 y un RMSE de 10.68, el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

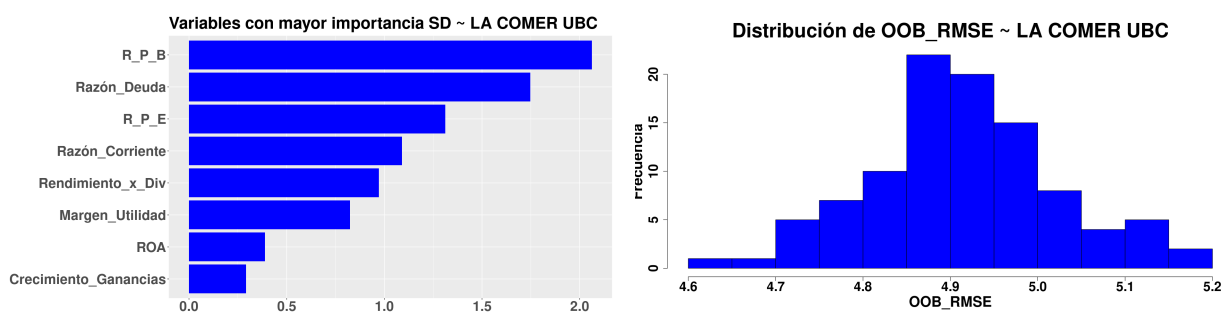
Tabla 3.8 Resultados del mejor modelo de bosques aleatorios para LA COMER UBC

R^2	MSE	RMSE
65.39%	23.73	4.87

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para LA COMER UBC presenta un coeficiente de determinación R^2 del 65.39 %, lo que indica un buen nivel de varianza explicada en los datos. Las métricas de error, con un MSE de 23.73 y un RMSE de 4.87, sugieren que las predicciones del modelo son considerablemente precisas, con una desviación promedio de aproximadamente 4.87 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida de Razón Corriente, Relación del Precio con las Ganancias (R_P_E), y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.14 Características del mejor modelo de LA COMER UBC



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 4.6 y 5.2, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 63.97 %, un MSE de 24.70 y un RMSE de 4.96, el mejor modelo muestra una ligera mejora en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han perfeccionado el modelo.

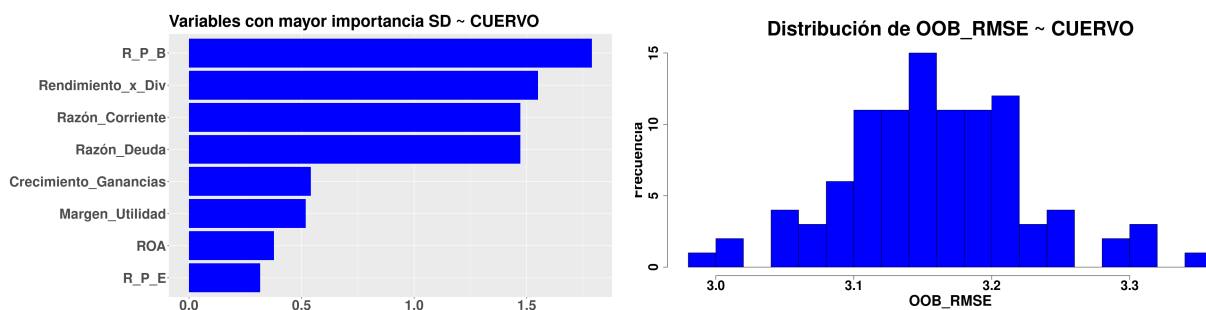
Tabla 3.9 Resultados del mejor modelo de bosques aleatorios para CUERVO

R^2	MSE	RMSE
84.87%	10.14	3.18

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para CUERVO presenta un coeficiente de determinación R^2 del 84.87%, lo que indica un alto nivel de varianza explicada en los datos. Las métricas de error, con un MSE de 10.14 y un RMSE de 3.18, sugieren que las predicciones del modelo son bastantes precisas, con una desviación promedio de aproximadamente 3.18 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), Rendimiento por Dividendo, Razón Corriente, y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.15 Características del mejor modelo de CUERVO



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 3.0 y 3.3, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 85.38%, un MSE de 11.02 y un RMSE de 3.31, el mejor modelo muestra una pequeña mejora en el RMSE Y MSE, pero en el R^2 se redujo un poco.

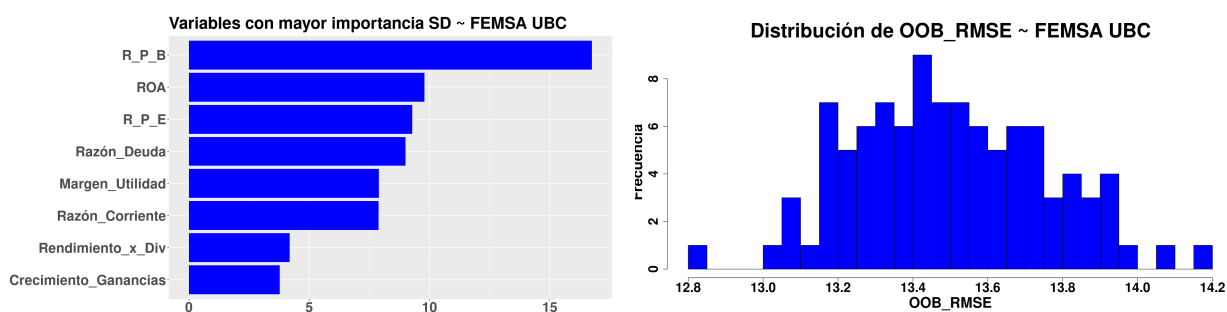
Tabla 3.10 Resultados del mejor modelo de bosques aleatorios para FEMSA UBC

R^2	MSE	RMSE
60.44 %	187.53	13.69

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para FEMSA UBC presenta un coeficiente de determinación R^2 del 60.44 %, lo que indica un nivel regular de varianza explicada en los datos. Las métricas de error, con un MSE de 187.53 y un RMSE de 13.69, sugieren que las predicciones del modelo son lo suficientemente precisas, con una desviación promedio de aproximadamente 13.69 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), ROA, Relación del Precio con las Ganancias (R_P_E), y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.16 Características del mejor modelo de FEMSA UBC



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 12.8 y 14.2, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 51.58 %, un MSE de 219.84 y un RMSE de 14.82, el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

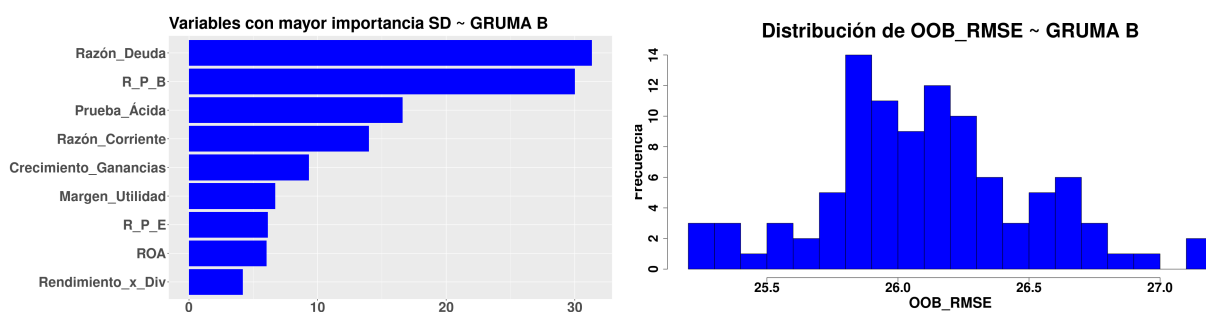
Tabla 3.11 Resultados del mejor modelo de bosques aleatorios para GRUMA B

R^2	MSE	RMSE
41.66 %	672.02	25.92

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para GRUMA B presenta un coeficiente de determinación R^2 del 41.66 %, lo que indica un nivel un poco bajo de varianza explicada en los datos. Las métricas de error, con un MSE de 672.02 y un RMSE de 25.92, sugieren que las predicciones del modelo son considerablemente precisas, con una desviación promedio de aproximadamente 25.92 unidades. Entre las variables más importantes destacan la Razón de Deuda, seguida de la Relación Precio con el Valor en Libros (R_P_B), Prueba Ácida, y otras que tienen un impacto significativo en las predicciones del modelo.

Figura 3.17 Características del mejor modelo de GRUMA B



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 25.5 y 27.0, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 31.78 %, un MSE de 785.85 y un RMSE de 28.03, el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

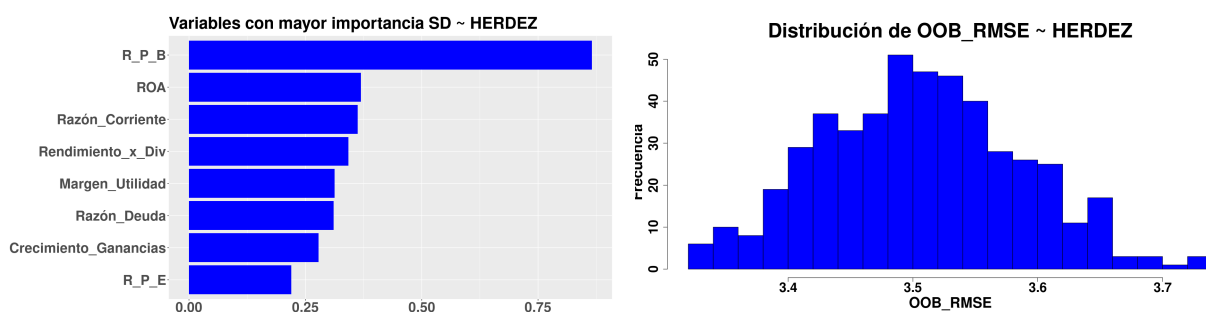
Tabla 3.12 Resultados del mejor modelo de bosques aleatorios para HERDEZ

R^2	MSE	RMSE
51.96 %	12.40	3.52

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para HERDEZ presenta un coeficiente de determinación R^2 del 51.96 %, lo que indica un nivel regular de varianza explicada en los datos. Las métricas de error, con un MSE de 12.40 y un RMSE de 3.52, sugieren que las predicciones del modelo pueden ser lo suficientemente precisas, con una desviación promedio de aproximadamente 3.52 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida de ROA, Razón Corriente, y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.18 Características del mejor modelo de HERDEZ



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 3.4 y 3.7, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 35.65 %, un MSE de 18.26 y un RMSE de 4.27, el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

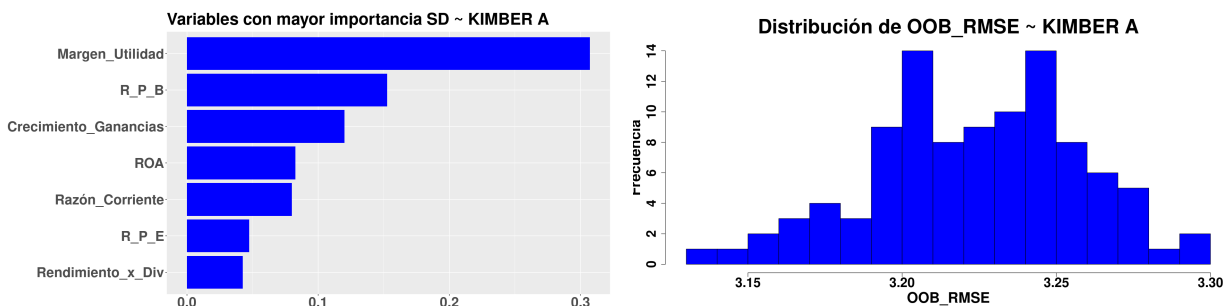
Tabla 3.13 Resultados del mejor modelo de bosques aleatorios para KIMBER A

R^2	MSE	RMSE
19.61 %	10.42	3.23

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para KIMBER A presenta un coeficiente de determinación R^2 de apenas el 19.61 %, lo que indica un nivel muy bajo de varianza explicada en los datos. Las métricas de error, con un MSE de 10.42 y un RMSE de 3.23, sugieren que las predicciones del modelo pueden llegar a no ser tan precisas; sin embargo, estas tienen una desviación promedio de aproximadamente 3.23 unidades. Entre las variables más importantes destacan el Margen de Utilidad, seguida de la Relación Precio con el Valor en Libros (R_P_B), Crecimiento de Ganancias y otras que tienen un impacto significativo en las predicciones del modelo.

Figura 3.19 Características del mejor modelo de KIMBER A



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 3.15 y 3.30, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 14.46 %, un MSE de 10.54 y un RMSE de 3.24, el mejor modelo muestra una pequeña mejora en todas las métricas, pero principalmente en el R^2 , aunque no de manera tan considerable.

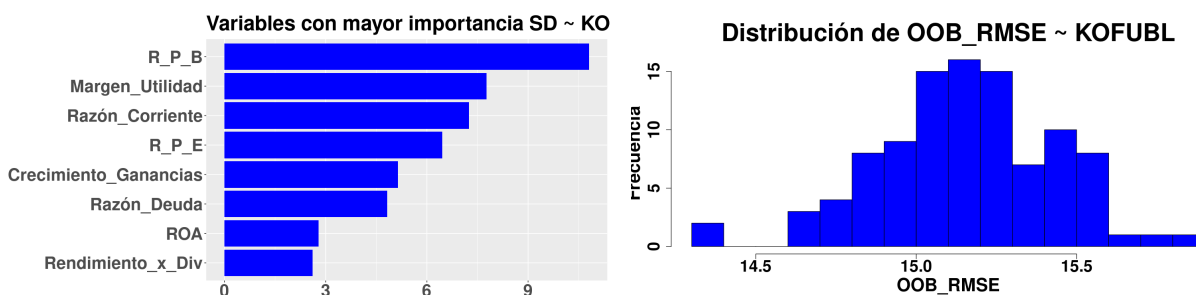
Tabla 3.14 Resultados del mejor modelo de bosques aleatorios para KOFUBL

R^2	MSE	RMSE
40.79 %	230.02	15.17

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para KOFUBL presenta un coeficiente de determinación R^2 del 40.79%, lo que indica un nivel un poco bajo de varianza explicada en los datos. Las métricas de error, con un MSE de 230.02 y un RMSE de 15.17, sugieren que las predicciones del modelo son considerablemente precisas, con una desviación promedio de aproximadamente 15.17 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida del Margen de Utilidad, Razón Corriente y otras que tienen un impacto significativo en las predicciones del modelo.

Figura 3.20 Características del mejor modelo de KOFUBL



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 14.5 y 15.5, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 22.84%, un MSE de 253.18 y un RMSE de 15.91, el mejor modelo muestra una pequeña mejora en todas las métricas, pero principalmente en el R^2 , que se refina de manera considerable.

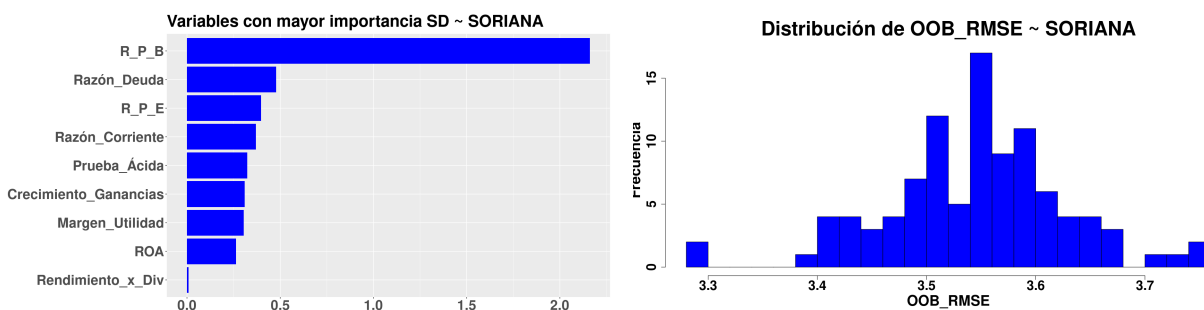
Tabla 3.15 Resultados del mejor modelo de bosques aleatorios para SORIANA

R^2	MSE	RMSE
76.86 %	14.26	3.78

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para SORIANA presenta un coeficiente de determinación R^2 del 76.86 %, lo que indica un alto nivel de varianza explicada en los datos. Las métricas de error, con un MSE de 14.26 y un RMSE de 3.78, sugieren que las predicciones del modelo son bastante precisas, con una desviación promedio de aproximadamente 3.78 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida de Razón de Deuda, Relación del Precio con las Ganancias (R_P_E), y otras, que tienen un impacto significativo en las predicciones del modelo.

Figura 3.21 Características del mejor modelo de SORIANA



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 3.3 y 3.7, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 45.52 %, un MSE de 21.09 y un RMSE de 4.59, el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

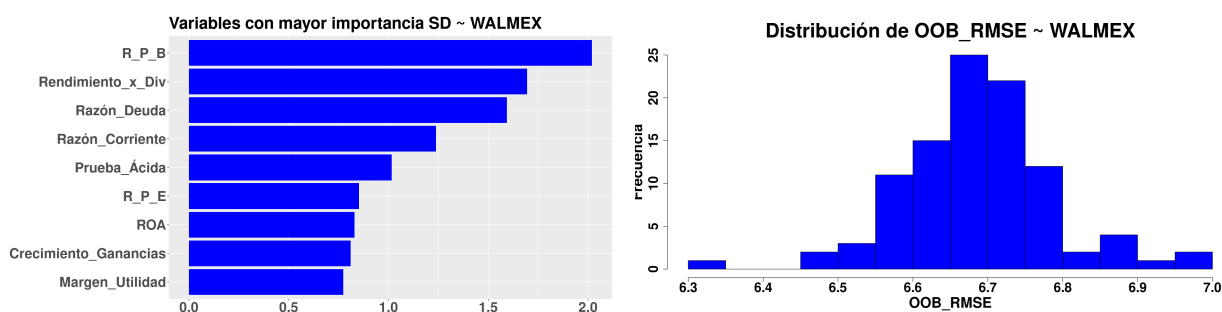
Tabla 3.16 Resultados del mejor modelo de bosques aleatorios para WALMEX

R^2	MSE	RMSE
50.34 %	44.80	6.69

Fuente: Elaboración propia.

El mejor modelo de bosques aleatorios para WALMEX presenta un coeficiente de determinación R^2 del 5034 %, lo que indica un nivel regular de varianza explicada en los datos. Las métricas de error, con un MSE de 44.80 y un RMSE de 6.69, sugieren que las predicciones del modelo pueden ser lo suficientemente precisas, con una desviación promedio de aproximadamente 6.69 unidades. Entre las variables más importantes destacan la Relación Precio con el Valor en Libros (R_P_B), seguida del Rendimiento por Dividendo, Razón de Deuda, Razón Corriente, Prueba Ácida, R_P_E, ROA, Crecimiento_Ganancias y Margen_Utilidad.

Figura 3.22 Características del mejor modelo de WALMEX



Fuente: Elaboración propia.

La distribución de OOB_RMSE, que oscila entre 6.3 y 7.0, indica una variación consistente del modelo. En comparación con el modelo inicial mostrado en la Tabla 3.4, que tenía un R^2 del 33.78 %, un MSE de 53.51 y un RMSE de 7.31, el mejor modelo muestra una mejora significativa en todas las métricas. Esto sugiere que el ajuste de los hiperparámetros y la selección de variables han contribuido a un mejor rendimiento predictivo.

3.2.3. Aplicación del intervalo de confianza al RMSE

Después de obtener los mejores modelos de bosques aleatorios para cada una de las acciones que conforman el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, así como sus respectivos valores de RMSE, es fundamental enmarcar mejor la incertidumbre asociada a estos errores. Para ello, se propone calcular un intervalo de confianza al 95 % para el RMSE, lo que permitirá obtener una estimación más precisa de su variabilidad.

Si asumimos que los residuos fuera de bolsa (OOB) siguen una distribución normal, podemos utilizar la distribución chi-cuadrado (χ^2) para calcular el intervalo de confianza del RMSE. Este enfoque se basa en la relación entre la varianza de los residuos OOB y la distribución χ^2 , lo que nos permite establecer límites de confianza que reflejen la incertidumbre en la estimación del error del modelo.

Fórmula del intervalo de confianza

El intervalo de confianza del $(1 - \alpha)$ % para la varianza (σ^2) de los residuos OOB se calcula como:

$$\left(\frac{(n-1) \cdot RMSE^2}{\chi_{1-\alpha/2, n-1}^2}, \frac{(n-1) \cdot RMSE^2}{\chi_{\alpha/2, n-1}^2} \right) \quad (3.2)$$

Donde:

- n es el número de observaciones.
- $RMSE$ es el valor del RMSE.
- $\chi_{\alpha/2, n-1}^2$ y $\chi_{1-\alpha/2, n-1}^2$ son los valores críticos de la distribución chi-cuadrado con $n - 1$ grados de libertad.

- α es el nivel de significancia.

Luego, tomamos la raíz cuadrada para obtener el intervalo de confianza para el RMSE:

$$\left(\sqrt{\frac{(n-1) \cdot RMSE^2}{\chi^2_{1-\alpha/2, n-1}}}, \sqrt{\frac{(n-1) \cdot RMSE^2}{\chi^2_{\alpha/2, n-1}}} \right) \tag{3.3}$$

La fórmula para el intervalo de confianza de la varianza se deriva de las propiedades de la distribución χ^2 (Casella & Berger, 2002).

A partir de la metodología descrita, se han calculado los intervalos de confianza al 95 % para el RMSE de cada una de las acciones analizadas. Estos intervalos se presentan a continuación en la Tabla 3.17, proporcionando una visión más clara de la incertidumbre en la estimación del error de los modelos de bosques aleatorios.

Tabla 3.17 Intervalo de confianza al 95 % para el RMSE de cada acción

Acción	RMSE	2.5 %	97.5 %
AC	11.85	9.21	16.62
BIMBO A	11.63	9.04	16.31
CHEDRAUI	10.03	7.80	14.07
LA COMER UBC	4.87	3.70	7.11
CUERVO	3.18	2.47	4.46
FEMSA UBD	13.69	10.64	19.20
GRUMA B	25.92	20.14	36.35
HERDEZ	3.52	2.74	4.94
KIMBER A	3.23	2.73	4.93
KOFUBL	15.17	11.80	21.28
SORIANA	3.78	2.94	5.30
WALMEX	6.69	5.20	9.38

Fuente: Elaboración propia.

En la tabla anterior, se presenta el intervalo de confianza al 95 % para el RMSE de cada acción. Este intervalo se calcula utilizando la distribución chi-cuadrado, que permite estimar

un rango en el que se espera que se encuentre el verdadero valor del RMSE con un 95 % de confianza. Sin embargo, para fines prácticos, solo nos interesa el límite superior del intervalo (97.5 %), ya que este valor expande el RMSE para definir una banda de error en la que los pronósticos pueden variar.

¿Por qué el 97.5 %?

- El límite superior del intervalo (97.5 %) representa el valor máximo que el RMSE podría tomar con un 95 % de confianza.
- Al utilizar este valor, estamos considerando el peor escenario posible en términos de error, lo que nos permite definir una banda de error más amplia y conservadora.
- Esta banda de error se expresa como:

$$RMSE_{97.5\%} < Error < RMSE_{97.5\%} \quad (3.4)$$

Donde:

- $RMSE_{97.5\%}$ es el límite superior del intervalo de confianza para el RMSE.
- El $Error$ es la diferencia entre el valor real y el pronóstico.

Ahora bien, una vez obtenidos estos intervalos, el siguiente paso es utilizar los modelos ajustados para realizar los pronósticos de las acciones que conforman el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico. Con ello, buscamos evaluar la capacidad predictiva de los modelos dentro de los rangos de incertidumbre establecidos, asegurando así una estimación más robusta de los precios futuros de las acciones.

3.3. Pronósticos

Tras la obtención de los intervalos de confianza para el RMSE, el siguiente paso es aplicar los modelos refinados de bosques aleatorios a los datos de prueba previamente separados al inicio del proceso. Con ello, generamos los pronósticos del precio de las acciones a un horizonte de un año, con periodicidad trimestral, respetando la frecuencia de nuestras observaciones.

Las fechas específicas de los pronósticos son:

- 30/06/2024
- 30/09/2024
- 31/12/2024
- 31/03/2025

Los precios históricos de las acciones fueron extraídos de Yahoo Finance, asegurando así la obtención de datos confiables y actualizados. Para evaluar la precisión de las predicciones, se compararán los valores pronosticados con los valores reales disponibles hasta la fecha. Además, se calculará la diferencia entre ambos para medir el error en cada punto de pronóstico.

Esta evaluación se complementará con gráficos individuales para cada acción, permitiendo una visualización clara del desempeño del modelo en la predicción de precios a futuro. La presentación de estos resultados no solo permitirá validar la efectividad del modelo, sino también identificar posibles ajustes para mejorar su capacidad predictiva en aplicaciones futuras.

Tabla 3.18 *Pronóstico vs Valores Reales de AC*

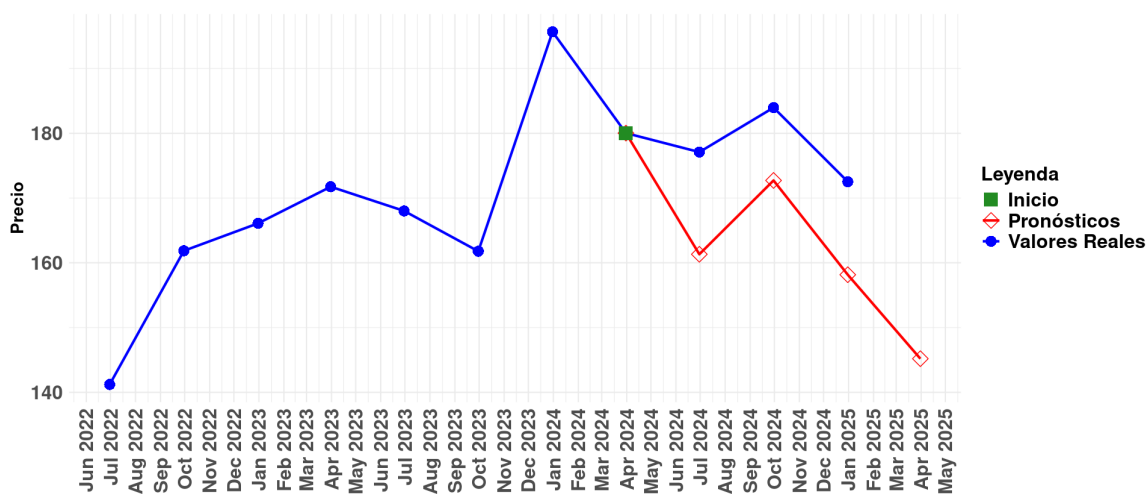
Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	161.32	177.1	15.78	16.62	76.81 %
30/09/2024	172.72	183.97	11.25		
31/12/2024	158.16	172.53	14.37		
31/03/2025	145.19	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 11.25 y 15.78, dentro del error permitido RMSE de 16.62. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.23, que presenta la dirección de los precios de AC para observar la tendencia que toma.

Figura 3.23 *Dirección de los precios de AC*

Gráfica de Pronósticos vs Valores Reales ~ AC



Fuente: Elaboración propia.

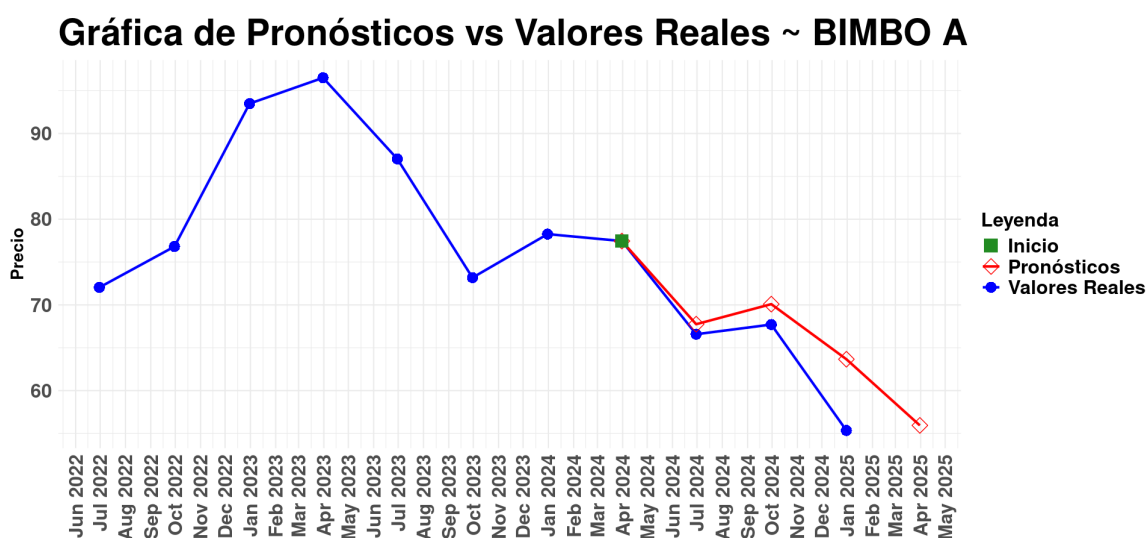
Tabla 3.19 *Pronóstico vs Valores Reales de BIMBO A*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	67.75	66.58	1.17	16.31	55.51 %
30/09/2024	70.09	67.71	2.38		
31/12/2024	63.66	55.35	8.31		
31/03/2025	55.94	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 1.17 y 8.31, dentro del error permitido RMSE de 16.31. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.24, que presenta la dirección de los precios de BIMBO A para observar la tendencia que toma.

Figura 3.24 *Dirección de los precios de BIMBO A*



Fuente: Elaboración propia.

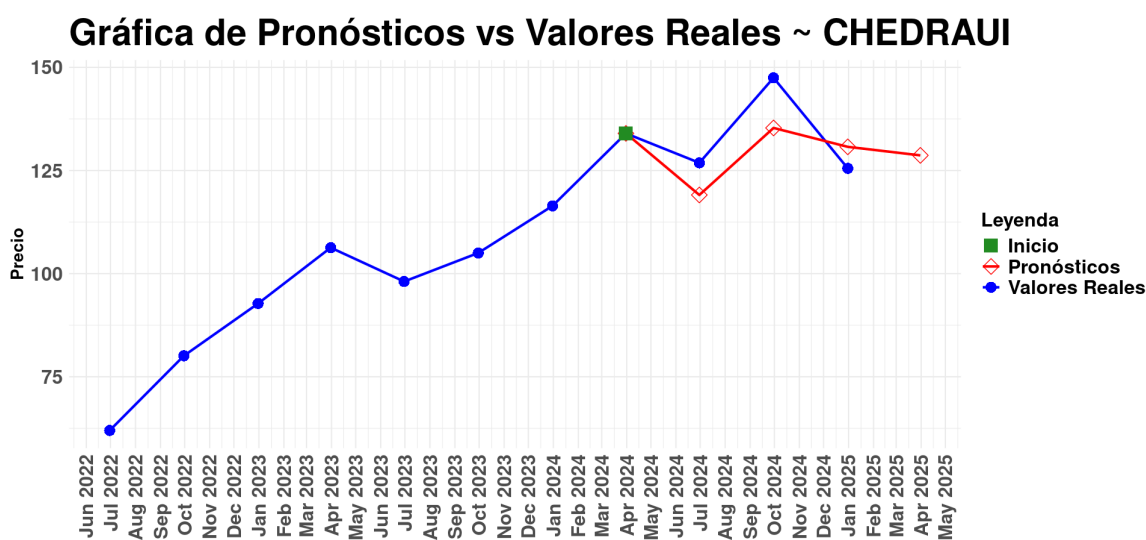
Tabla 3.20 *Pronóstico vs Valores Reales de CHEDRAUI*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	119.07	126.86	7.79		
30/09/2024	135.32	147.48	12.16	14.07	88.26 %
31/12/2024	130.74	125.52	5.22		
31/03/2025	128.67	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 5.22 y 12.16, dentro del error permitido RMSE de 14.07. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.25, que presenta la dirección de los precios de CHEDRAUI para observar la tendencia que toma.

Figura 3.25 *Dirección de los precios de CHEDRAUI*



Fuente: Elaboración propia.

Tabla 3.21 *Pronóstico vs Valores Reales de LA COMER UBC*

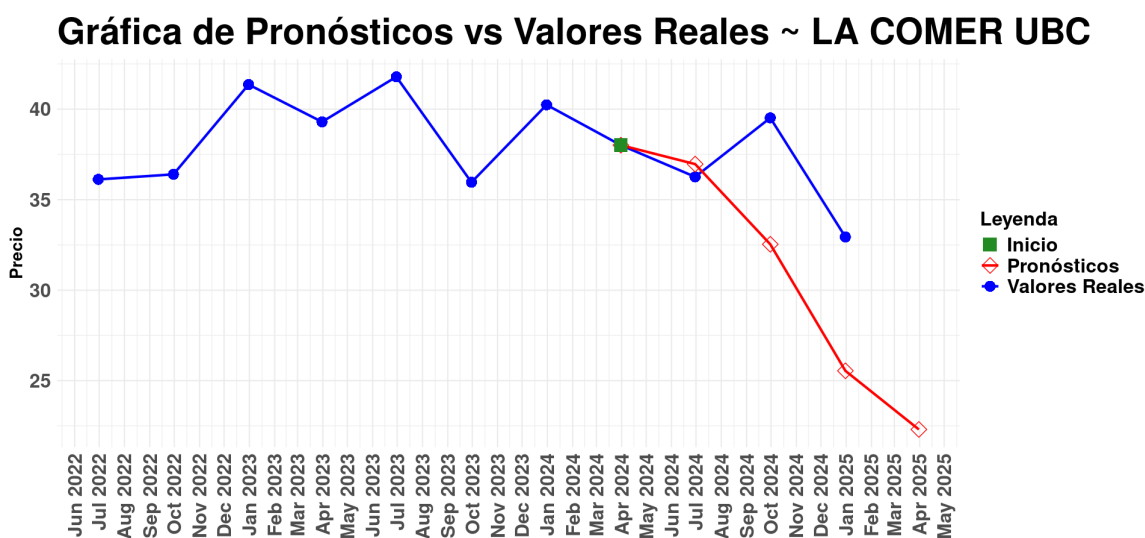
Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	36.96	36.26	0.70		
30/09/2024	32.54	39.52	6.98	7.11	65.39 %
31/12/2024	25.54	32.94	7.40*		
31/03/2025	22.30	-	-		

Nota: Los errores absolutos marcados con un * rebasan el error permitido.

Fuente: Elaboración propia.

Para este caso específico, los pronósticos realizados hasta el momento no todos se encuentran dentro del intervalo de error permitido, con errores absolutos que oscilan entre .70 y 7.40, uno de los pronósticos fuera del error permitido RMSE de 7.11. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales, aunque en un período más amplio los pronósticos no resultan tan precisos. A continuación, se muestra la Figura 3.26, que presenta la dirección de los precios de LA COMER UBC para observar la tendencia que toma.

Figura 3.26 *Dirección de los precios de LA COMER UBC*



Fuente: Elaboración propia.

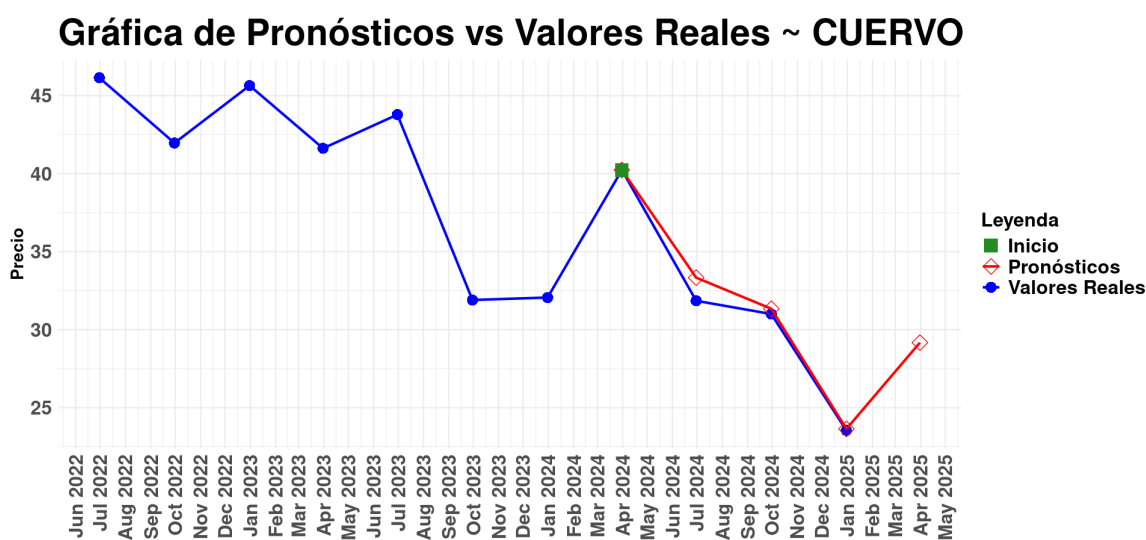
Tabla 3.22 *Pronóstico vs Valores Reales de CUERVO*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	33.33	31.85	1.48	4.46	84.87%
30/09/2024	31.34	31	0.34		
31/12/2024	23.64	23.54	0.10		
31/03/2025	29.17	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre .10 y 1.48, dentro del error permitido RMSE 4.46. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.27, que presenta la dirección de los precios de CUERVO para observar la tendencia que toma.

Figura 3.27 *Dirección de los precios de CUERVO*



Fuente: Elaboración propia.

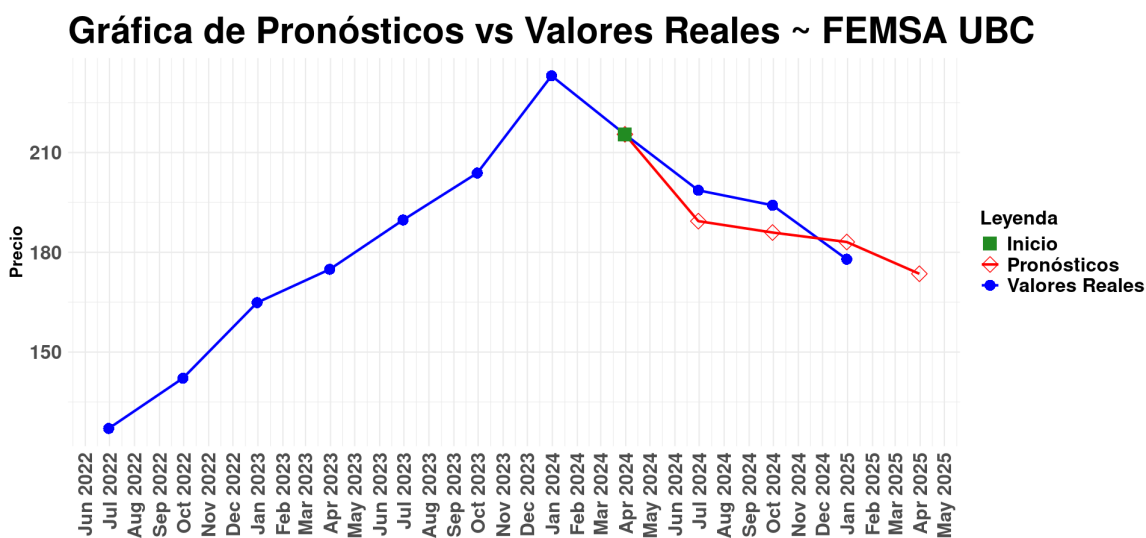
Tabla 3.23 *Pronóstico vs Valores Reales de FEMSA UBC*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	189.36	198.63	9.27	19.20	60.44 %
30/09/2024	185.94	194.13	8.19		
31/12/2024	183.10	177.91	5.19		
31/03/2025	173.54	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 5.29 y 9.27, dentro del error permitido RMSE 19.20. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.28, que presenta la dirección de los precios de FEMSA UBC para observar la tendencia que toma.

Figura 3.28 *Dirección de los precios de FEMSA UBC*



Fuente: Elaboración propia.

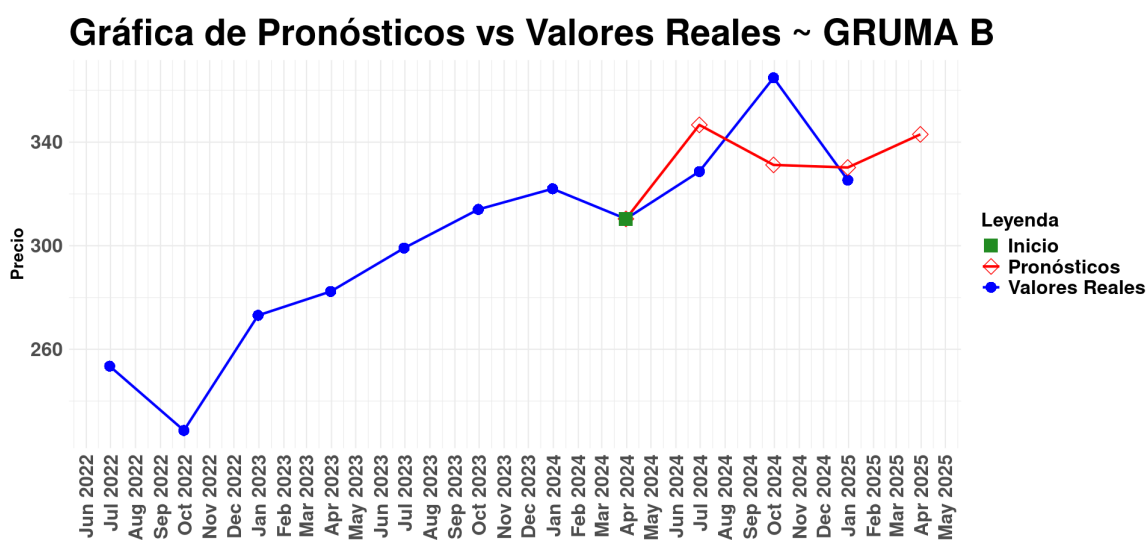
Tabla 3.24 *Pronóstico vs Valores Reales de GRUMA B*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	346.66	328.66	18.00	36.35	41.66 %
30/09/2024	331.21	364.86	33.65		
31/12/2024	330.21	325.36	4.85		
31/03/2025	343.01	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 4.85 y 33.65, dentro del error permitido RMSE 36.35. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.29, que presenta la dirección de los precios de GRUMA B para observar la tendencia que toma.

Figura 3.29 *Dirección de los precios de GRUMA B*



Fuente: Elaboración propia.

Tabla 3.25 *Pronóstico vs Valores Reales de HERDEZ*

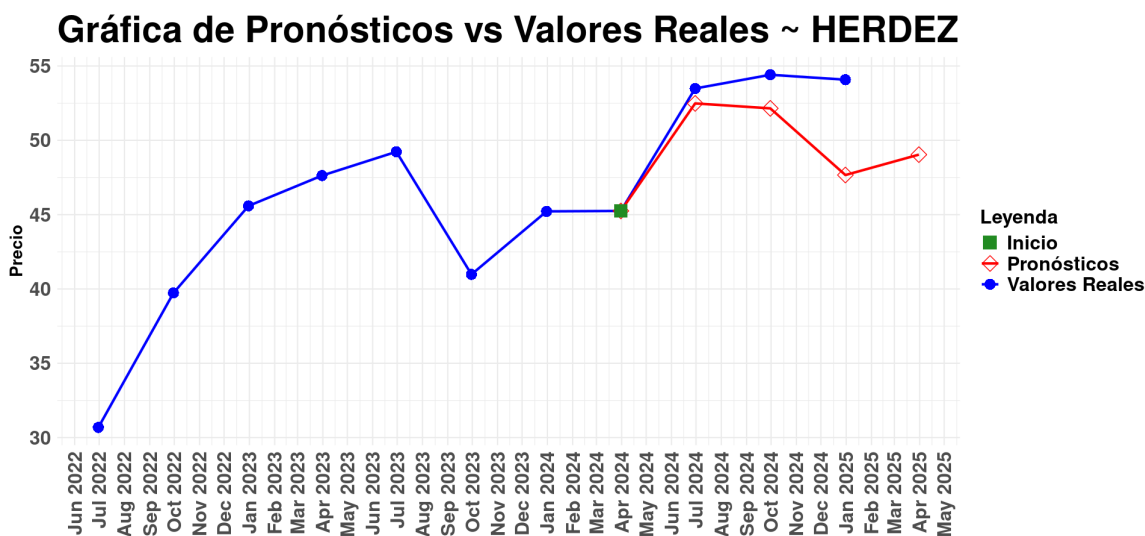
Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	52.48	53.49	1.01	4.94	51.96 %
30/09/2024	52.15	54.41	2.26		
31/12/2024	47.67	54.08	6.41*		
31/03/2025	49.04	-	-		

Nota: Los errores absolutos marcados con un * rebasan el error permitido.

Fuente: Elaboración propia.

Para este otro caso, los pronósticos realizados hasta el momento no todos se encuentran dentro del intervalo de error permitido, con errores absolutos que oscilan entre 1.01 y 6.41, uno de los pronósticos fuera del error permitido RMSE de 4.94. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales, aunque en un período más amplio los pronósticos no resultan tan precisos. A continuación, se muestra la Figura 3.30, que presenta la dirección de los precios de HERDEZ para observar la tendencia que toma.

Figura 3.30 *Dirección de los precios de HERDEZ*



Fuente: Elaboración propia.

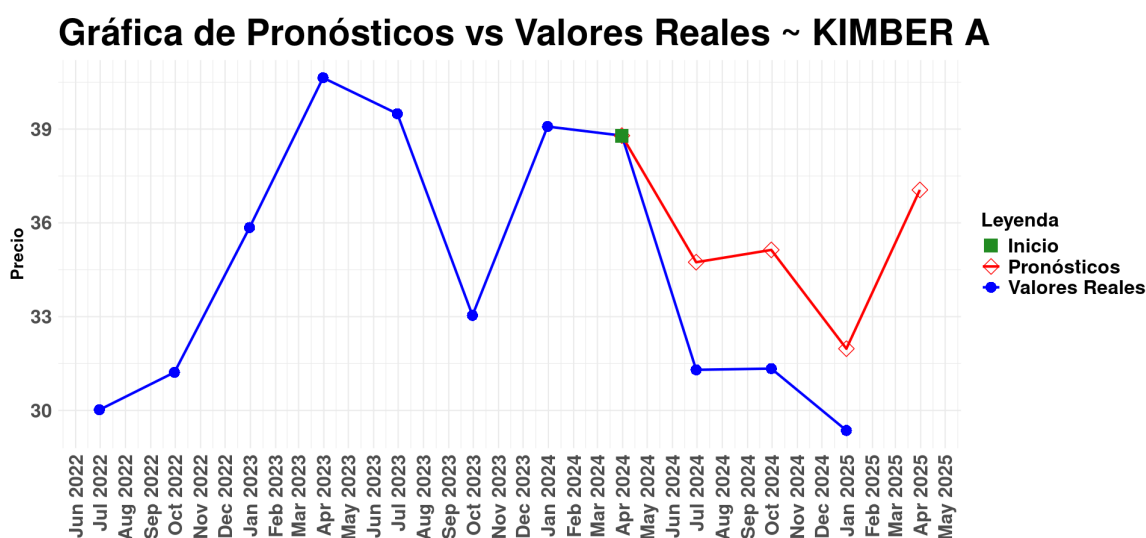
Tabla 3.26 *Pronóstico vs Valores Reales de KIMBER A*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	34.74	31.30	3.44	4.93	19.61 %
30/09/2024	35.14	31.34	3.80		
31/12/2024	31.98	29.36	2.62		
31/03/2025	37.05	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 3.44 y 3.80, dentro del error permitido RMSE 4.93. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.31, que presenta la dirección de los precios de KIMBER A para observar la tendencia que toma.

Figura 3.31 *Dirección de los precios de KIMBER A*



Fuente: Elaboración propia.

Tabla 3.27 *Pronóstico vs Valores Reales de KOFUBL*

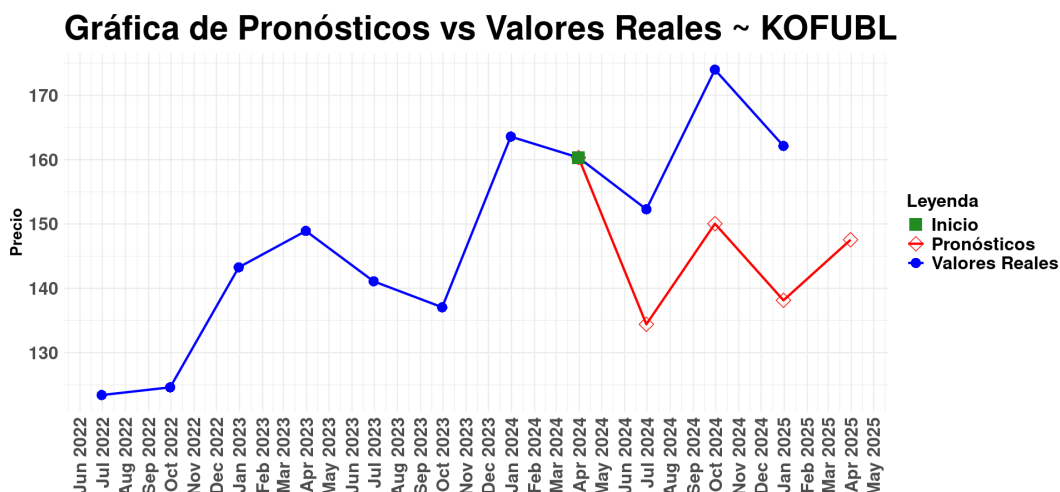
Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	134.42	152.28	17.86		
30/09/2024	150.04	173.98	23.94*	21.28	40.79 %
31/12/2024	138.14	162.13	23.99*		
31/03/2025	147.55	-	-		

Nota: Los errores absolutos marcados con un * rebasan el error permitido.

Fuente: Elaboración propia.

Este modelo es el que presenta más inconsistencias en los pronósticos realizados hasta el momento, solo uno se encuentra dentro del intervalo de error permitido; los errores absolutos oscilan entre 17.86 y 23.99. Dos de los pronósticos están fuera del error permitido RMSE de 21.28. Esto indica que el modelo tiene un mal desempeño en la predicción de los valores reales, en un período más amplio los pronósticos no resultan nada precisos. A continuación, se muestra la Figura 3.32, que presenta la dirección de los precios de KOFUBL para observar la tendencia que toma.

Figura 3.32 *Dirección de los precios de KOFUBL*



Fuente: Elaboración propia.

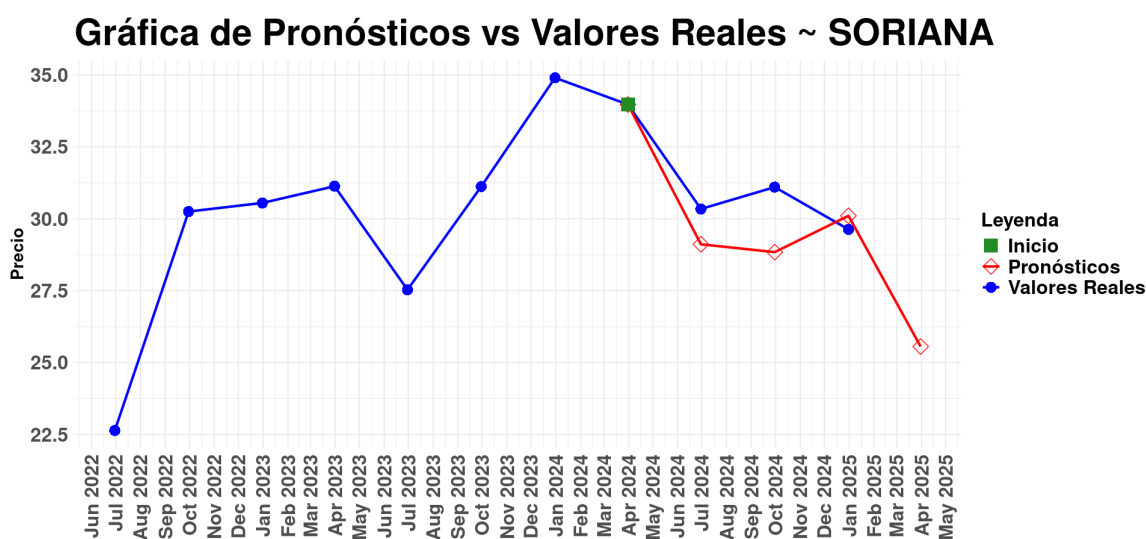
Tabla 3.28 *Pronóstico vs Valores Reales de SORIANA*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	29.11	30.34	1.23	5.30	76.86 %
30/09/2024	28.84	31.10	2.26		
31/12/2024	30.10	29.63	0.47		
31/03/2025	25.56	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 0.47 y 2.26, dentro del error permitido RMSE 5.30. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.33, que presenta la dirección de los precios de SORIANA para observar la tendencia que toma.

Figura 3.33 *Dirección de los precios de SORIANA*



Fuente: Elaboración propia.

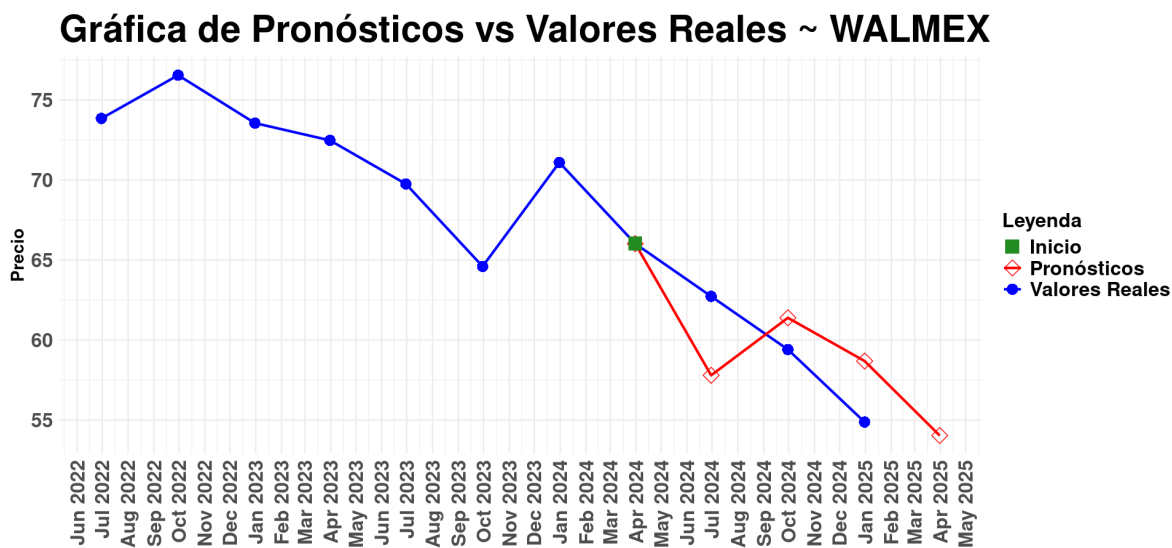
Tabla 3.29 *Pronóstico vs Valores Reales de WALMEX*

Fecha	Pronóstico	Valor real	Error Absoluto	RMSE con Intervalo (Error Permitido)	R^2
30/06/2024	57.80	62.73	4.93	9.38	50.34%
30/09/2024	61.39	59.41	1.98		
31/12/2024	58.69	54.88	3.81		
31/03/2025	54.04	-	-		

Fuente: Elaboración propia.

Los pronósticos realizados hasta el momento se cumplen dentro del intervalo de error permitido, con errores absolutos que oscilan entre 1.98 y 4.93, dentro del error permitido RMSE 9.38. Esto indica que el modelo tiene un desempeño aceptable en la predicción de los valores reales. A continuación, se muestra la Figura 3.34, que presenta la dirección de los precios de WALMEX para observar la tendencia que toma.

Figura 3.34 *Dirección de los precios de WALMEX*



Fuente: Elaboración propia.

Es importante recordar que el objetivo de esta investigación es realizar pronósticos para los precios de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, mediante el método estadístico de Bosques Aleatorios, utilizando como variables predictoras las razones financieras. Este enfoque permite extraer conocimiento y valor de los datos financieros, proporcionando una herramienta robusta para la toma de decisiones en el ámbito de inversiones.

A lo largo de este capítulo, se ha llevado a cabo un análisis exhaustivo de los resultados obtenidos a partir del modelo de bosques aleatorios. Se identificaron las variables predictoras más relevantes, se evitó la redundancia mediante un análisis de correlación y se estimaron los mejores modelos para cada acción del índice. Además, se aplicó un intervalo de confianza al RMSE para cuantificar la incertidumbre asociada a los pronósticos, lo que permite tener una visión más precisa y realista del error del modelo.

En el siguiente capítulo, se presentarán las conclusiones finales de esta investigación, así como las recomendaciones y posibles líneas de trabajo futuro derivadas de los hallazgos obtenidos.

Conclusiones

El presente trabajo ha abordado la elaboración de pronósticos para los precios de las acciones que conforman el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico, utilizando el método estadístico de Bosques Aleatorios y variables predictoras basadas en razones financieras. Los resultados obtenidos han permitido evaluar la eficacia del modelo en la predicción de los precios de las acciones, así como identificar las variables más relevantes y los casos en los que el modelo presenta limitaciones. A continuación, se presentan las conclusiones finales, destacando los hallazgos clave, las implicaciones para los inversionistas y las recomendaciones para futuras investigaciones.

El modelo ha demostrado un desempeño general satisfactorio en la mayoría de las acciones analizadas, con coeficientes de determinación (R^2) que oscilan entre 19.61 % y 88.26 %. Esto indica que, en términos generales, el método de Bosques Aleatorios es capaz de capturar una proporción significativa de la variabilidad en los precios de las acciones del sector. Acciones como CHEDRAUI (88.26 %) y CUERVO (84.87 %) destacan por su alto grado de ajuste, lo que sugiere que el modelo es particularmente efectivo para predecir su comportamiento. Sin embargo, en casos como KIMBER A (19.61 %) y KOFUBL (40.79 %), el modelo presenta limitaciones, lo que refleja la complejidad de predecir el comportamiento de ciertas acciones basándose únicamente en razones financieras.

Variables de Mayor Relevancia

La Relación del Precio con las Ganancias (R_P_E) ha emergido como la variable más influyente en la mayoría de los modelos, dominando de manera efectiva la predicción del precio de las acciones. Esta variable, que refleja la valoración del mercado en relación con las ganancias de la empresa, ha demostrado ser un indicador clave para anticipar movimientos en los precios. Otras variables importantes incluyen la Razón Corriente, que mide la liquidez de la empresa, y el Margen de Utilidad, que indica su capacidad para generar ganancias. Estas variables, en conjunto, proporcionan una visión integral de la salud financiera de las empresas y su potencial de crecimiento, lo que las convierte en herramientas fundamentales para los inversionistas.

Casos Fallidos y Errores en los Pronósticos

A pesar del buen desempeño general, el modelo ha presentado errores significativos en tres acciones: LA COMER UBC, HERDEZ y KOFUBL. En estos casos, los errores absolutos han superado el intervalo permitido, lo que sugiere que el modelo no ha logrado capturar completamente los factores que influyen en sus precios. A continuación, se analizan estos casos con mayor detalle:

■ LA COMER UBC:

- El pronóstico para el 31/12/2024 presentó un error absoluto de 7.40, superando el intervalo permitido. Este error podría estar relacionado con factores específicos de la empresa, como cambios en su estrategia de expansión o fluctuaciones en la demanda de sus productos, que no están reflejados en las razones financieras utilizadas.

■ HERDEZ:

- El pronóstico para el 31/12/2024 mostró un error absoluto de 6.41, también por encima del intervalo permitido. Este fallo podría deberse a la influencia de factores externos, como la volatilidad en los precios de los insumos agrícolas o cambios en las preferencias de los consumidores, que no son capturados por el modelo.
- **KOFUBL:**
- Los pronósticos para el 30/09/2024 y el 31/12/2024 presentaron errores absolutos de 23.94 y 23.99, respectivamente, ambos superando el intervalo permitido. Estos errores podrían estar asociados con la alta sensibilidad de la acción a factores macroeconómicos, como la tasa de interés o el tipo de cambio, que no están incluidos en el modelo actual.

Estos casos resaltan la importancia de complementar los modelos predictivos con información adicional, especialmente en acciones que son más sensibles a factores externos o que presentan un comportamiento atípico.

Se ha observado que, a medida que el horizonte temporal del pronóstico se extiende, la precisión del modelo tiende a disminuir. Este fenómeno es particularmente evidente en las proyecciones para el primer trimestre de 2025, donde los errores absolutos aumentan en comparación con los pronósticos a corto plazo. Esto sugiere que, aunque el modelo es efectivo para predecir el comportamiento de las acciones en el corto plazo, su capacidad predictiva se ve limitada en horizontes temporales más largos, donde la incertidumbre y la influencia de factores externos se incrementan.

El índice S&P/BMV Sector Selecto de Bienes de Consumo Básico ha mostrado una tendencia general al alza en el período analizado, reflejando la resiliencia del sector ante las fluctuaciones económicas. Este comportamiento es consistente con la naturaleza defensiva del sector, que tiende a mantener su desempeño incluso en contextos económicos adversos. Sin embargo, es importante destacar que algunas acciones dentro del índice, como KOFUBL y GRUMA B,

han presentado una mayor volatilidad, lo que sugiere que no todas las empresas del sector responden de la misma manera a los cambios en el entorno económico.

Los resultados de este estudio tienen implicaciones significativas para los inversionistas interesados en el índice S&P/BMV Sector Selecto de Bienes de Consumo Básico. Los pronósticos proporcionan una herramienta valiosa para la toma de decisiones, permitiendo a los inversionistas anticipar movimientos en los precios de las acciones y ajustar sus estrategias de inversión en consecuencia. Sin embargo, es crucial que los inversionistas consideren las limitaciones del modelo, especialmente en acciones donde los pronósticos han fallado, y complementen el análisis con información adicional sobre el contexto económico y sectorial.

Recomendaciones y Líneas de Trabajo Futuro

- **Inclusión de Variables Macroeconómicas:** Se recomienda incorporar variables como el PIB, la tasa de interés y la inflación en futuros modelos para mejorar la precisión de los pronósticos, especialmente en horizontes temporales más largos.
- **Análisis de Eventos Específicos:** Estudiar el impacto de eventos regulatorios, cambios en la demanda del consumidor y otros factores específicos del sector podrían ayudar a explicar los errores en acciones como LA COMER UBC y KOFUBL.
- **Exploración de Modelos Alternativos:** Probar técnicas de aprendizaje automático más avanzadas, como redes neuronales o modelos híbridos, podría mejorar la capacidad predictiva del modelo.
- **Monitoreo Continuo:** Actualizar periódicamente los modelos con datos recientes y ajustar las variables predictoras en función de cambios en el entorno económico y financiero.

En conclusión, este estudio subraya la importancia de los pronósticos en la gestión de inversiones y destaca el potencial de los modelos basados en razones financieras para predecir el

comportamiento de las acciones del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico. A lo largo del trabajo, se cumplió con el objetivo planteado, al generar pronósticos mediante el modelo de Bosques Aleatorios utilizando razones financieras como variables predictoras. Asimismo, se verificó el cumplimiento de la hipótesis, al comprobar que dicho modelo mejora la precisión de los pronósticos, permitiendo contrastar las estimaciones con los valores reales y evaluar su grado de exactitud.

No obstante, también se resalta la necesidad de un enfoque más integral que incluya factores macroeconómicos junto con las razones financieras, con el fin de mejorar aún más la precisión y confiabilidad de las predicciones. Estas mejoras no solo beneficiarían a los inversionistas en su toma de decisiones, sino que también contribuirían a un mercado más eficiente y transparente.

Anexo

Anexo 3.1

Código en R, que muestra la aplicación para cada uno de los modelos de los componentes del índice S&P/BMV Sector Selecto de Bienes de Consumo Básico.

```
library(rsample)
library(randomForest)
library(ranger)
library(caret)
library(ggplot2)
library(tidyr)
library(broom)
library(dplyr)
library(magrittr)
library(reshape2)

#Base datos, contiene cada accion y sus variables
Componentes=data_frame(AC ,BIMBOA ,CHEDRAUI ,COMERUBC ,
                       CUERVO ,FEMSA ,GRUMAB ,HERDEZ ,
                       KIMBERA .KOFUBL ,SORIANA .WALMEX)
cor_matrix <- cor(Componentes$ACCION) #Nombre de la accion
cor_matrix
```

```
##### Analisis de VARIABLES #####

heatmap_correlation <- function(data, variables) {
  data_selected <- data[, variables]

  corr_matrix <- cor(data_selected, use = "complete.obs")
  corr_melted <- melt(corr_matrix)

  # Creamos un mapa de calor
  ggplot(data = corr_melted, aes(x = Var1, y = Var2, fill = value)) +
    geom_tile() +
    scale_fill_gradient2(low = "red", high = "blue", mid = "white",
                        midpoint = 0, limit = c(-1, 1),
                        space = "Lab", name = "Correlacion") +
    theme_minimal() +
    theme(
      axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1,
                                  face = "bold", size = 15),
      axis.text.y = element_text(face = "bold", size = 15),
      plot.title = element_text(face = "bold", size = 20)
    ) +
    labs(x = "", y = "", title = "Mapa de Correlaciones de las
          variables ~ ACCION") #Nombre de la accion
}

#Mapa de Correlacion de las todas las Variables de razones
finacieras

selected_variables <- c("Margen_Utilidad", "Razon_Deuda",
```

```

"Razon_Corriente",      "ROA", "Prueba_ cida ", "R_P_E", "R_P_B",
"Rendimiento_x_Div", "Crecimiento_Ganancias")

heatmap_correlation(Componentes$ACCION, selected_variables)

#Variables elegidas posterior al analisis de correlacion

ACCION1<- Componentes$ACCION[, c("Precio_ACC", "Margen_Utilidad",
"Razon_Deuda", "Razon_Corriente", "ROA", "R_P_E",
"R_P_B", "Rendimiento_x_Div", "Crecimiento_Ganancias")]

##### Modelo de Bosque Aleatorio#####

set.seed(123)
ames_split <- initial_split(ACCION1, prop = .7)
ames_train <- training(ames_split)
ames_test  <- testing(ames_split)

#Primer modelo de bosques aleatorios
m1 <- randomForest(
  formula = Precio_ACC ~ .,
  data    = ames_train
)
m1
plot(m1,
     main = "Modelo 1 ~ ACCION",      #Nombre de la accion
     col = "blue",
     lwd = 5,
     cex.axis = 2,
     cex.lab = 2,

```

```
cex.main = 3)

# Numero de rboles con menor MSE
which.min(m1$mse)

# RMSE de este random forest optimo
sqrt(m1$mse[which.min(m1$mse)])

##### Ajuste mtry= el n mero de variables a muestrear aleatoriamente
      como candidatas en cada divisi n

features <- setdiff(names(ames_train), "Precio_ACC")

set.seed(123)
m2 <- tuneRF(
  x= ames_train[features], # Conjunto de caracter sticas (variables
    independientes)
  y= ames_train$Precio_ACC, # Variable dependiente (respuesta)
  ntreeTry= 500,           # N mero de rboles a probar en cada
    ajuste de 'mtry'
  mtryStart= 1,           # Valor inicial de 'mtry'
  stepFactor= .91,       # Factor por el cual se ajusta 'mtry' en
    cada iteraci n
  improve= 0.0001,      # Criterio m nimo de mejora en el error
    para continuar la b squeda
  trace= FALSE           # No mostrar el progreso en tiempo real
)

# Convertir los resultados de tuneRF en un data frame para ggplot
m2_df <- as.data.frame(m2)
```

```

# Crear la grafica con ggplot2
ggplot(m2_df, aes(x = mtry, y = OOBError)) +
  geom_line(size = 3, color = "blue") +
  geom_point(size = 10, color = "red") +
  labs(
    title = "Optimizaci n de mtry ~ ACCION", #Nombre de la acci n
    x = "mtry",
    y = "OOB Error"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 30, face = "bold", hjust = 0.5),
    axis.title.x = element_text(size = 27, face = "bold"),
    axis.title.y = element_text(size = 27, face = "bold"),
    axis.text.x = element_text(size = 25, face = "bold"),
    axis.text.y = element_text(size = 25, face = "bold"),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  )
##### Busqueda en rejilla de hiperparametros #####

hyper_grid <- expand.grid(
  mtry= seq(1, 8, by = 1),      # Valores de 'mtry' a probar
  node_size= seq(1,10, by = 1), # Valores de 'node_size' a probar
  sampe_size= c(.55, .632, .70, .80, .85,.90), # Fracciones del
  conjunto de datos
  OOB_RMSE= 0                  # Inicializa el valor de OOB_RMSE
  para almacenar los resultados
)

```

```

# Numero de combinaciones totales
nrow(hyper_grid)

for(i in 1:nrow(hyper_grid)) {
  # Aplicacion de combinaciones
  model <- ranger(
    formula      = Precio_ACC ~ .,
    data         = ames_train,
    num.trees    = 500,
    mtry         = hyper_grid$mtry[i],
    min.node.size = hyper_grid$node_size[i],
    sample.fraction = hyper_grid$sampe_size[i],
    seed        = 123
  )

  # OOB ERROR
  hyper_grid$OOB_RMSE[i] <- sqrt(model$prediction.error)
}

hyper_grid %>%
  dplyr::arrange(OOB_RMSE) %>%
  head(10) #Mostrar la mejores 10 combinaciones

#Eleccion del Mejor Modelo

OOB_RMSE <- vector(mode = "numeric", length = 100)

for(i in seq_along(OOB_RMSE)) {
  optimal_BS <- ranger(
    formula      = Precio_ACC ~ .,

```

```
data          = ames_train,
num.trees     = 500,
mtry          = 3,           #Var a respecto a la acci n
min.node.size = 2,           #Var a respecto a la acci n
sample.fraction = .70,      #Var a respecto a la acci n
importance    = 'impurity'
)

OOB_RMSE[i] <- sqrt(optimal_BS$prediction.error)
}

##### Histograma de la distribucion de las OOB_RMSE

hist(OOB_RMSE, breaks = 20, col = "blue", border = "black",
     main = "Distribuci n de OOB_RMSE ~ ACCION", xlab = "OOB_RMSE",
     ylab = "Frecuencia",
     cex.main = 3, cex.lab = 2.2, cex.axis =2, font.main = 2, font.
     lab = 2.5, font.axis = 2.5)

# Vector de importancia
importancia_variables <- optimal_BS1$importanceSD
print(importancia_variables)

importancia_df <- data.frame(
  Variable = names(importancia_variables),
  Importancia = as.numeric(importancia_variables)
)
```

```
# Ordenar por importancia las variables
importancia_df <- importancia_df %>%
  arrange(desc(Importancia)) %>%
  top_n(8, Importancia) #El n mero var a dependiendo de la acci n

# Grafico

ggplot(importancia_df, aes(x = reorder(Variable, Importancia), y =
  Importancia)) +
  geom_col(fill = "blue") +
  coord_flip() +
  labs(
    title = "Variables con mayor importancia SD ~ ACCION", #Nombre de
      la accion
    x = "",
    y = NULL
  ) +
  theme(
    plot.title = element_text(size = 30, face = "bold"),
    axis.text.x = element_text(size = 25, face = "bold"),
    axis.text.y = element_text(size = 25, face = "bold"),
    axis.title.x = element_text(size = 18, face = "bold")
  )

##### Pronosticos

PRONOSTICO<- predict(optimal_BS, ames_test)
head(PRONOSTICO)
```

```
##### Intervalo de confianza al 95% #####
# Datos necesarios
rmse <- 10 # RMSE del modelo
n <- nrow(ames_train) # Numero de observaciones en el conjunto de
  entrenamiento
alpha <- 0.05 # Nivel de significancia (95% de confianza)

# Valores criticos de la distribucion chi-cuadrado
chi2_lower <- qchisq(1 - alpha/2, df = n - 1)
chi2_upper <- qchisq(alpha/2, df = n - 1)

# Intervalo de confianza para la varianza
var_lower <- (n - 1) * rmse^2 / chi2_lower
var_upper <- (n - 1) * rmse^2 / chi2_upper

# Intervalo de confianza para el RMSE
rmse_lower <- sqrt(var_lower)
rmse_upper <- sqrt(var_upper)

# Mostrar el intervalo de confianza
IC_95 <- c(rmse_lower, rmse_upper)
print(IC_95)
```

Fuente: Elaboración propia

Bibliografía

- Bekaert, G., & Engstrom, E. (2017). The Economic Impact of Political Uncertainty on Financial Markets. *Journal of Financial Economics*, 126(3), 559-582.
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281-305. <http://www.jmlr.org/papers/v13/bergstra12a.html>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2018). Package ‘randomforest’. *University of California, Berkeley: Berkeley, CA, USA*, 81, 1-29.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth Brooks/Cole Advanced Books Software.
- Brigham, E. F., & Ehrhardt, M. C. (2021). *Financial Management: Theory Practice* (16th). Cengage Learning.
- Cárdenas, J. C., & Cruz, E. (2017). Impacto de la elección de Donald Trump en el mercado accionario mexicano. *Revista Mexicana de Economía y Finanzas*, 12(4), 255-271.
- Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd). Duxbury.
- CONCANACO. (2024). Consumo e inversión dan soporte al PIB en el inicio de año [Consultado en enero de 2025.]. <https://www.concanaco.com.mx/comercio/interesdecomercio/consumo-e-inversion-dan-soporte-al-pib-en-el-inicio-de-ano>

- Damodaran, A. (2012). *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset* (3rd). John Wiley & Sons.
- Domingos, P. (2012). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Graham, B., & Dodd, D. L. (2009). *Security Analysis* (6th). McGraw-Hill.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.
- Guide, U. B. A. R. P. (2023). Random Forests [Consultado el [fecha de acceso]]. https://uc-r.github.io/random_forests
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- INEGI. (2023). Indicadores de consumo privado en México. *Instituto Nacional de Estadística y Geografía*.
- Infobae, M. (2024). El principal indicador de la BMV registró alza de 3.24% al cierre de este 4 de junio. *Infobae*. <https://www.infobae.com/mexico/2024/06/04/el-principal-indicador-de-la-bmv-registro-alza-de-324-al-cierre-de-este-4-de-junio/>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 2(12), 1137-1143.
- Krugman, P. (1997). *The Age of Diminished Expectations*. MIT Press.

- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Moreno-Brid, J. C., & Paunovic, I. (2018). Mexico's Financial Markets and the Impact of U.S. Midterm Elections and Trade Uncertainty. *Journal of International Economics*, *115*, 98-112.
- País, E. (2024). El consumo privado en México muestra signos de desaceleración. *El País*.
- Ramelli, S., & Wagner, A. F. (2020). Feverish Stock Price Reactions to COVID-19. *The Review of Corporate Finance Studies*, *9*(3), 622-655.
- Research, B. (2014). El consumo interno como motor de la economía mexicana. *BBVA Research Economic Outlook*.
- Rodríguez, M., & Vargas, P. (2020). Impact of the COVID-19 Pandemic on the Mexican Stock Market: A Sectoral Perspective. *Journal of Emerging Markets*, *55*(4), 102-119.
- Ross, S. A., Westerfield, R. W., & Jaffe, J. (2019). *Corporate Finance* (12th). McGraw-Hill Education.
- Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press.
- S&P Dow Jones Indices. (2024a). *S&P/BMV Consumer Staples Select Sector Index* [Datos al 30 de agosto de 2024]. <https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-bmv-indices.pdf>
- S&P Dow Jones Indices. (2024b). *S&P/BMV IPC: Índice de Precios y Cotizaciones* [Recuperado de <https://www.spglobal.com/spdji/es/>].
- Zhou, J., Wang, S., & Zhang, W. (2021). Machine Learning in Financial Forecasting: Recent Advances and Practical Applications. *Journal of Financial Data Science*, *3*(2), 23-40.