

FACULTAD DE INGENIERÍA

Carrera de Ingeniería Industrial

“PROPUESTA DE MEJORA EN EL PRONÓSTICO DE
VENTAS PARA REDUCIR LOS COSTOS LOGÍSTICOS DE
UNA EMBOTELLADORA DE BEBIDAS NO ALCOHÓLICAS”

Tesis para optar el título profesional de:

Ingeniero Industrial

Autor:

Giancarlo Rene Callaoapaza Chavez

Asesor:

Ing. Oscar Alberto Goicochea Ramírez

Trujillo - Perú

2021



DEDICATORIA

A mi madre y a mi abuela, por su paciencia y confianza.

AGRADECIMIENTO

Al soporte de mi madre durante los tiempos del pregrado, sin su apoyo, llegar a este momento no sería posible.

A los buenos profesores que encontré durante el camino, y de quienes guardo el más grato recuerdo. De ellos aprendí que, mucho queda por hacer en la práctica de la profesión. Espero que este trabajo tenga algún mérito en ello.

A Gerardo Cailloma, profesor de las asignaturas de Lengua I y II. De él aprendimos que, más allá de nuestra disciplina específica, el acercamiento a las humanidades resulta fundamental para nuestro mundo social e interno.

Tabla de contenidos

DEDICATORIA.....	2
AGRADECIMIENTO	3
ÍNDICE DE TABLAS	5
ÍNDICE DE FIGURAS	7
ÍNDICE DE ECUACIONES.....	8
CAPÍTULO I. INTRODUCCIÓN	10
CAPÍTULO II. METODOLOGÍA	48
CAPÍTULO III. RESULTADOS	76
CAPÍTULO IV. DISCUSIÓN Y CONCLUSIONES	87
REFERENCIAS.....	91
ANEXOS.....	95

ÍNDICE DE TABLAS

Tabla 1. Estado de los indicadores de la gestión logística	13
Tabla 2. Causas de desviación del desabasto	14
Tabla 3. Causas de desviación del costo de mantener inventario	15
Tabla 4. Priorización de causas de desviación en el sistema logístico	15
Tabla 5. Operacionalización de variables	18
Tabla 6. Procedimiento de desarrollo de la tesis	49
Tabla 7. Categoría de productos de la empresa	49
Tabla 8. Canales de venta de la empresa	50
Tabla 9. Regiones de venta de la empresa	51
Tabla 10. Volumen de venta según su variabilidad	52
Tabla 11. Evaluación del modelamiento del pronóstico estadístico	55
Tabla 12. Evaluación de la estructura jerárquica del pronóstico	56
Tabla 13. Evaluación de la segmentación de productos y clientes	57
Tabla 14. Evaluación de la colaboración de la demanda	58
Tabla 15. Evaluación de la capacitación y entrenamiento del equipo de trabajo	59
Tabla 16. Estado de indicadores del planeamiento de la demanda	60
Tabla 17. Causas raíz de la inadecuada gestión de la demanda	61
Tabla 18. Priorización de las causas raíz de la inadecuada gestión de la demanda	61
Tabla 19. Matriz de indicadores	62
Tabla 20. Propuesta de mejora de pronósticos de venta	63
Tabla 21. Mejora en el modelamiento del pronóstico estadístico	66
Tabla 22. Mejora en la estructura jerárquica del pronóstico	66
Tabla 23. Programa de capacitación técnica en modelamiento predictivo	70
Tabla 24. Mejora en la capacitación y entrenamiento	72
Tabla 25. Costo del software requerido por la propuesta	73
Tabla 26. Costo del hardware requerido por la propuesta	73
Tabla 27. Costo de la contratación del especialista de modelamiento predictivo	73
Tabla 28. Costo de la capacitación y certificación en modelamiento predictivo	74
Tabla 29. Costo laboral de un especialista de modelamiento predictivo	74
Tabla 30. Presupuesto para capacitaciones técnicas anuales	75

Tabla 31. Mejora en la exactitud del pronóstico semanal por demand sensing	78
Tabla 32. Mejora propuesta de la matriz de indicadores	79
Tabla 33. Impacto de la propuesta en el desabasto y en las utilidades del año 2019	80
Tabla 34. Impacto de la propuesta en el costo de mantener inventario en el año 2019	82
Tabla 35. Impacto de la propuesta en el costo de transporte del año 2019	82
Tabla 36. Impacto de la propuesta en los costos logísticos y en las utilidades del 2019 ...	83
Tabla 37. Flujo de caja económico de la propuesta de mejora	85

ÍNDICE DE FIGURAS

Figura 1. Participación de bebidas no alcohólicas por regiones del mundo	10
Figura 2. Principales países latinoamericanos en bebidas no alcohólicas	11
Figura 3. Participación de mercado de bebidas no alcohólicas en el Perú	12
Figura 4. Árbol de desviaciones en el sistema logístico	14
Figura 5. Optimización logística	25
Figura 6. Diagrama de árbol jerárquico de tres niveles	29
Figura 7. Esquematación de una red neuronal de tres capas	41
Figura 8. Ejemplo de modelamiento por validación cruzada	47
Figura 9. Mapa de centros de distribución y plantas productoras	51
Figura 10. Diagrama de flujo del proceso de planeamiento de la demanda	53
Figura 11. Eventos que impactan en el histórico de venta	55
Figura 12. Propuesta de estructura jerárquica para la generación de pronósticos	64
Figura 13. Pipeline del pronóstico estadístico propuesto	65
Figura 14. Modelo de datos de la propuesta	67
Figura 15. Flujo de información propuesto entre los sistemas de planeamiento	67
Figura 16. Línea base de venta nacional	68
Figura 17. Pronósticos estadísticos para una gaseosa de 296 ml en un CD de Lima	69
Figura 18. Pronósticos estadísticos para el agua sin gas de 625 ml en un CD de Lima	69
Figura 19. Diagrama de Gantt del programa de capacitación	71
Figura 20. Evaluación del enfoque metodológico del planeamiento de la demanda	76
Figura 21. Mejora en la exactitud del nuevo pronóstico estadístico semanal	77
Figura 22. Mejora en la exactitud del nuevo pronóstico semanal por <i>demand sensing</i>	77
Figura 23. Mejora en la señal de rastreo del nuevo pronóstico estadístico semanal	78
Figura 24. Prospectiva del enfoque metodológico del planeamiento de la demanda	79
Figura 25. Mejora en el nivel de servicio por los nuevos pronósticos estadísticos	81
Figura 26. Mejora en los niveles de inventario por los nuevos pronósticos estadísticos ...	81
Figura 27. Realidad futura del sistema logístico	84
Figura 28. Sensibilidad del VPN de la propuesta	86

ÍNDICE DE ECUACIONES

Ecuación 1. Costo logístico	25
Ecuación 2. Costo de venta perdida	25
Ecuación 3. Costo de mantener inventario	25
Ecuación 4. Costo de almacenamiento	25
Ecuación 5. Costo de transporte	25
Ecuación 6. Relación inventario y pronóstico de ventas	26
Ecuación 7. Nivel de servicio o <i>fill rate</i>	26
Ecuación 8. Relación nivel de servicio y pronóstico de ventas	27
Ecuación 9. Enfoque <i>bottom-up</i>	29
Ecuación 10. Enfoque <i>top-down</i> basado en proporciones de pronósticos	30
Ecuación 11. <i>Bias</i>	32
Ecuación 12. WMAPE	32
Ecuación 13. Exactitud del pronóstico	33
Ecuación 14. Señal de rastreo	33
Ecuación 15. Valor agregado del pronóstico	33
Ecuación 16. Modelo AR	34
Ecuación 17. Modelo ARMA	34
Ecuación 18. Estimación de la magnitud de venta según Croston	37
Ecuación 19. Estimación del tiempo de interllegada según Croston	37
Ecuación 20. Pronóstico mediante el método Croston	38
Ecuación 21. Naive forecast sin estacionalidad	38
Ecuación 22. Naive forecast con estacionalidad	39
Ecuación 23. Expresión general de la regresión lineal múltiple	42
Ecuación 24. Suavización exponencial simple	44
Ecuación 25. Suavización exponencial con tendencia	44
Ecuación 26. Actualización de la suavización exponencial con tendencia	45
Ecuación 27. Suavización exponencial con tendencia y estacionalidad	45
Ecuación 28. Actualización de la suavización exponencial triple	45

RESUMEN

El presente trabajo tiene como objetivo determinar si la propuesta de mejora en el pronóstico de ventas permitirá reducir los costos logísticos de una embotelladora de bebidas no alcohólicas. Para ello se desarrolló un diagnóstico del enfoque metodológico, encontrando debilidades en el modelamiento y estructura del pronóstico, y en la formación técnica del grupo humano, derivando en una exactitud del pronóstico operativo y señal de rastreo fuera de objetivo. Posteriormente, se desarrolló una propuesta técnica de modelamiento en el lenguaje de programación R que, para su sostenimiento, necesita de la incorporación de un especialista de modelamiento predictivo, y de un programa de capacitación y certificación técnica. Finalmente, esta propuesta, con una inversión de 462.7 miles de soles, tiene como impacto la mejora de la exactitud del pronóstico operativo del actual 71.9% a entre 81.1% y 87.4%, con una señal de rastreo controlada, mejorando el nivel de servicio del actual 95.3% a entre 96.3% y 96.8%, reduciendo los inventarios entre 11.6% a 13.7%, y generando ahorros en el costo logístico de entre 20 y 20.8 millones de soles para el año 2019; siendo además una propuesta viable, al presentar a 4 años un VPN de 76.69 millones de soles, un PR de 0.2 meses y una relación B/C de 55.4.

Palabras clave: Pronóstico de ventas, Costos logísticos, R programming language.

CAPÍTULO I. INTRODUCCIÓN

1.1. Realidad problemática

La industria mundial de bebidas no alcohólicas en el año 2019 se dimensiona en los 838,350 millones de dólares americanos, con las regiones Asia-Pacífico y Norteamérica concentrando el 52% del mercado global (ver figura 1), configurándose Coca-Cola como líder en el mercado global con el 18.2% del volumen de ventas, duplicando en participación a PepsiCo (Euromonitor International, 2020).

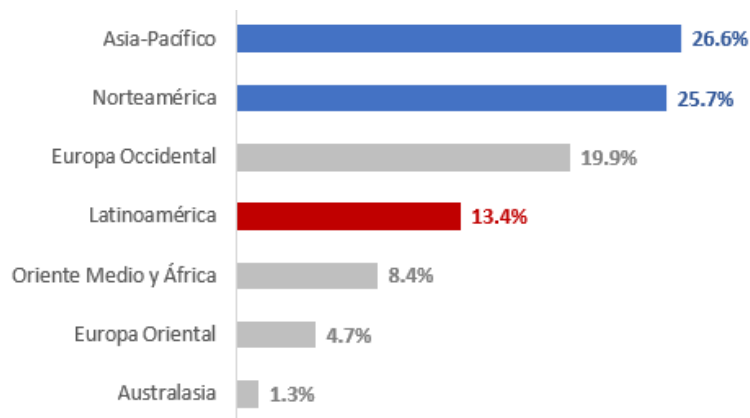


Figura 1. Participación de bebidas no alcohólicas por regiones del mundo.

Fuente: Euromonitor International, 2020.

A nivel de Latinoamérica, como la cuarta región del mundo en venta, el mercado se dimensiona en los 112,030 millones de dólares americanos, con México y Brasil concentrando el 61% (ver figura 2), configurándose Coca-Cola como líder de mercado con el 36.8% del volumen de ventas, cuatriplicando a PepsiCo. Adicionalmente, se observa que Perú, en este contexto, ocupa el séptimo lugar en participación con el 3.2% (Euromonitor Internacional, 2020).

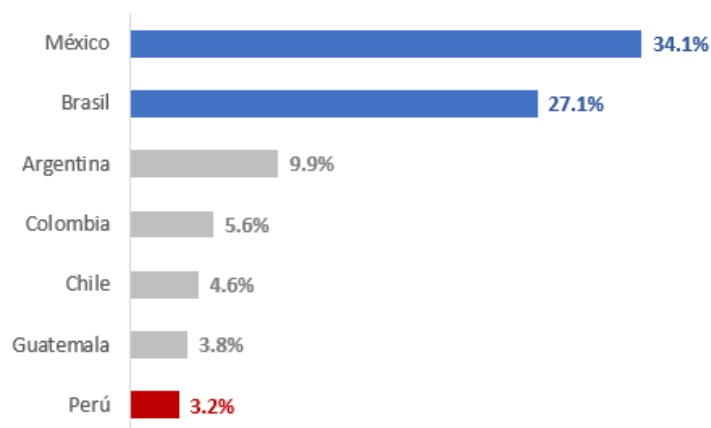


Figura 2. Principales países latinoamericanos en bebidas no alcohólicas.

Fuente: Euromonitor International, 2020.

A nivel peruano, la industria de bebidas no alcohólicas, dimensionada en los 3,636 millones de dólares americanos, se ubica en un entorno altamente complejo debido a su nivel competitivo, con más de 14 empresas integrando este sector (Euromonitor International, 2020), ocupando 4 de ellas una posición dominante (ver figura 3), donde además, a nivel político-legal, se cuenta con una normativa que afecta a la rentabilidad de la industria por el aumento del impuesto selectivo al consumo (ISC) dado en el año 2018 (Decreto Supremo N° 091-2018-EF, 2018), sumado a la normativa de etiquetado de advertencia de productos derivada de la Ley de Alimentación Saludable (Ley N° 30021, 2013). Adicionalmente, la crisis económica del país por la pandemia del COVID-19 (Banco Central de Reserva del Perú, 2020), con la caída del PBI, el empleo y el aumento de la pobreza, viene obligando a la búsqueda de una mayor eficiencia operativa, pero que a su vez permita contar con la flexibilidad necesaria en la cadena de suministros para reaccionar ante cambios abruptos en el comportamiento de la demanda del mercado.

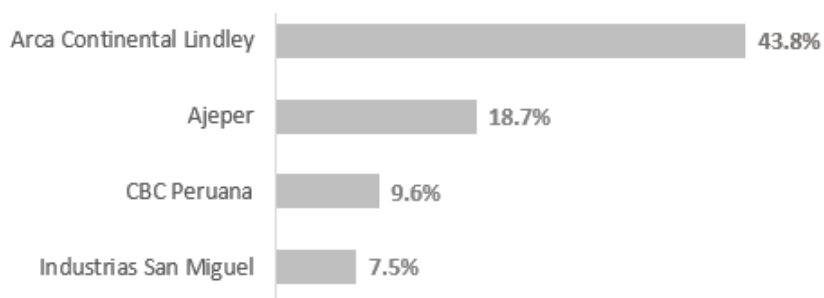


Figura 3. Participación de mercado de bebidas no alcohólicas en el Perú.

Fuente: Euromonitor International.

Bajo este escenario, la embotelladora de bebidas no alcohólicas objeto de este estudio, viene presentando problemas en la gestión de su cadena de suministros, ya que en el año 2019, con unas utilidades antes de impuestos (EBT) de 249.7 millones de soles, la empresa presentó un sobrecosto logístico neto de 14.7 millones de soles (5.9% del EBT) por desviaciones respecto a sus objetivos de costo de venta perdida y de mantener inventario, incumpliendo además el objetivo de nivel de servicio de 96% (ver tabla 1). Y al profundizar en la cadena de causas de estos efectos indeseados, empleando para ello el método efecto-causa-efecto (Goldratt, 1990), se logra evidenciar una inadecuada gestión de la demanda (ver figura 4) por desviaciones en sus principales indicadores: el de exactitud del pronóstico operativo y el de señal de rastreo, confirmando esta conclusión al valorizar las contribuciones de cada causa raíz al resultado de costo (ver tabla 2 y 3), obteniendo de esta forma que la inadecuada gestión de la demanda es la responsable del 76% de las desviaciones en el costo logístico, lo cual nos conduce a abordar este proceso para la mejora de todo sistema (ver tabla 4).

Tabla 1
Estado de los indicadores de la gestión logística

Indicador	Unidad de medida	Objetivo 2019 (a)	Resultado 2019 (b)	Variación (b) - (a)	Estado
Costo logístico	Soles/Caja unitaria	1.69	1.75	0.06	Fuera de objetivo
Costo de venta perdida	Soles/Caja unitaria	0.10	0.12	0.02	Fuera de objetivo
Costo de mantener inventario	Soles/Caja unitaria	0.26	0.30	0.04	Fuera de objetivo
Costo de transporte	Soles/Caja unitaria	1.09	1.09	-0.002	En objetivo
Costo de almacenamiento	Soles/Caja unitaria	0.24	0.24	-0.002	En objetivo
Nivel de servicio	%	96%	95.3%	-0.7%	Fuera de objetivo
Desabasto	%	1.3%	2.1%	0.8%	Fuera de objetivo
Rechazo y devolución de producto	%	2.7%	2.6%	-0.1%	En objetivo
Exactitud de inventario	%	99%	99.7%	0.7%	En objetivo
Cumplimiento de la política de inventario	%	80%	26.3%	-53.7%	Fuera de objetivo
Exactitud de pedido y carga	%	98%	98.7%	0.7%	En objetivo
Cumplimiento del plan de distribución secundaria	%	96%	97.8%	1.8%	En objetivo
Exactitud del plan de producción	%	90%	80.1%	-9.9%	Fuera de objetivo
Exactitud del plan de distribución primaria	%	90%	93.2%	3.2%	En objetivo
Nivel de servicio de materiales	%	98%	98.9%	0.9%	En objetivo
Nivel de servicio del proveedor de transporte	%	96%	97.7%	1.7%	En objetivo
Exactitud del pronóstico operativo	%	80%	71.9%	-8%	Fuera de objetivo
Señal de rastreo	Adimensional	[-2.5, 2.5]	-13.3	-10.8	Fuera de objetivo

Fuente: Elaboración propia.

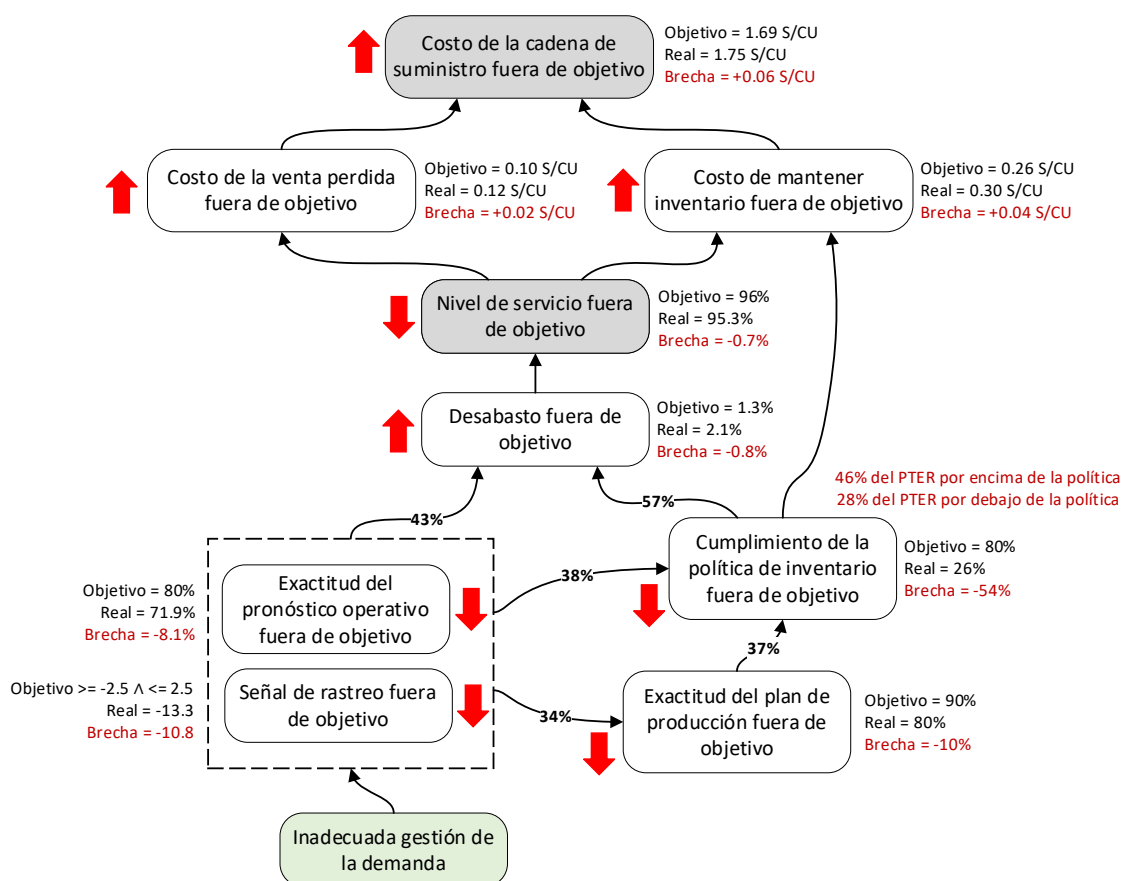


Figura 4. Árbol de desviaciones en el sistema logístico. Fuente: Elaboración propia.

Tabla 2
Causas de desviación del desabasto

Concepto	Valor (Soles)	Categoría
Desabasto real anual	15,172,812	
- Desabasto objetivo anual	9,542,522	
= Incremento en el desabasto	5,630,290	
(1) Desviaciones del plan de demanda	2,421,025	Gestión de la demanda
(2) Incumplimiento de la política de inventarios	3,209,265	
(2.1) Desviaciones del plan de demanda	1,219,521	Gestión de la demanda
(2.2) Factores internos de la distribución	802,316	Gestión de la distribución
(2.3) Desviaciones del plan de producción	1,187,428	
(2.3.1) Desviaciones del plan de demanda	403,726	Gestión de la demanda
(2.3.2) Factores internos de producción	783,703	Gestión de la producción

Fuente: Elaboración propia.

Tabla 3

Causas de desviación del costo de mantener inventario

Concepto	Valor (Soles)	Categoría
Costo real de mantener inventario	83,599,459	
- Costo objetivo de mantener inventario	73,247,317	
= Incremento de costo de mantener inventario	10,352,142	
(1) Incumplimiento de la política de inventarios	9,676,507	
(1.1) Desviaciones del plan de demanda	7,547,675	Gestión de la demanda
(1.2) Factores internos de producción	2,128,832	Gestión de la producción
(2) Desviaciones del desabasto	675,635	
(2.1) Desviaciones del plan de demanda	290,523	Gestión de la demanda
(2.2) Incumplimiento de la política de inventarios	385,112	
(2.2.1) Desviaciones del plan de demanda	146,342	Gestión de la demanda
(2.2.2) Factores internos de la distribución	96,278	Gestión de la distribución
(2.2.3) Desviaciones del plan de producción	142,491	
(2.2.3.1) Desviaciones del plan de deman	48,447	Gestión de la demanda
(2.2.3.2) Factores internos de producción	94,044	Gestión de la producción

Fuente: Elaboración propia.

Tabla 4

Priorización de causas de desviación en el sistema logístico

Categoría	Valor (Soles)	Peso (%)
Inadecuada gestión de la demanda	12,077,259	76%
Inadecuada gestión de la producción	3,006,578	19%
Inadecuada gestión de la distribución	898,594	6%
Desviación total	15,982,431	100%

Fuente: Elaboración propia.

1.2. Formulación del problema

¿La propuesta de mejora en el pronóstico de ventas permitirá reducir los costos logísticos de una embotelladora de bebidas no alcohólicas?

1.3. Objetivos

1.3.1. Objetivo general

- Determinar si la propuesta de mejora en el pronóstico de ventas permitirá reducir los costos logísticos de una embotelladora de bebidas no alcohólicas.

1.3.2. Objetivos específicos

- Evaluar el enfoque metodológico del proceso de planeamiento de la demanda de una embotelladora de bebidas no alcohólicas.
- Desarrollar una propuesta de mejora del pronóstico de ventas.
- Determinar las mejoras en el planeamiento de la demanda, nivel de servicio y nivel de inventario, derivadas de la propuesta planteada.
- Determinar la reducción del costo logístico derivado de la propuesta de mejora.
- Evaluar la viabilidad económica de la propuesta de mejora del pronóstico de ventas.

1.4. Hipótesis

La propuesta de mejora en el pronóstico de ventas permite reducir los costos logísticos de una embotelladora de bebidas no alcohólicas.

1.5. Justificación

1.5.1. Justificación práctica

Este trabajo permitirá desarrollar una propuesta que impacte de manera positiva a los resultados de costos logísticos de la embotelladora de bebidas no alcohólicas objeto de estudio, cuyos costos presentaron una desviación neta por

14.7 millones de soles en el año 2019, identificando como causa raíz a la inadecuada gestión de la demanda.

1.5.2. Justificación teórica

Este trabajo contribuirá, mediante la propuesta de mejora, con un modelo de aplicación que beneficiaría no solo a una empresa particular del sector de bebidas no alcohólicas, sino a cualquier empresa del sector industrial o de comercio en el que intervenga un pronóstico de ventas para el funcionamiento de su cadena de suministros.

1.5.3. Justificación metodológica

Este trabajo permitirá sentar un precedente en la aplicación del método de validación cruzada para evaluar la calidad de una propuesta de mejora en un pronóstico de ventas, en contraposición a lo visto en los antecedentes locales de esta investigación, además del uso en conjunto de diversas técnicas de pronóstico, empleando para ello el lenguaje de programación R para la automatización de procesos y la reproducibilidad de resultados.

1.6. Operacionalización de las variables

1.6.1. Variable independiente

La propuesta de mejora en el pronóstico de ventas.

1.6.2. Variable dependiente

Costos logísticos de una embotelladora de bebidas no alcohólicas.

En la tabla 5 se muestra los indicadores de las variables independientes y dependientes.

Tabla 5

Operacionalización de variables

OPERACIONALIZACIÓN DE VARIABLES					
Problema	Hipótesis	Variables	Definición	Indicador	Fórmula
¿La propuesta de mejora en el pronóstico de ventas permitirá reducir los costos logísticos de una embotelladora de bebidas no alcohólicas?	La propuesta de mejora en el pronóstico de ventas permite reducir los costos logísticos de una embotelladora de bebidas no alcohólicas.	Variable Independiente: Propuesta de mejora en el pronóstico de ventas.	El pronóstico de ventas es la proyección en el futuro de la demanda esperada según una serie de supuestos dados para su desarrollo.	Error porcentual absoluto medio ponderado (WMAPE)	$(\sum Venta - Pronóstico / \sum Venta) \times 100 [\%]$
				Exactitud del pronóstico (FA)	$1 - WMAPE [\%]$
				Señal de rastreo (TS)	$(\sum (Venta - Pronóstico) / \sum Venta - Pronóstico) \times N^\circ \text{ periodos}$ [Adimensional]
				Valor agregado del pronóstico (FVA)	$FA_{etapa 2} - FA_{etapa 1} [\%]$
				Bias	$((\sum Pronóstico / \sum Venta) - 1) \times 100 [\%]$
		Variable Dependiente: Costos logísticos de una embotelladora de bebidas no alcohólicas.	Los costos logísticos son aquellos que se asumen por almacenar, mantener y transportar inventario, incluyendo a la venta perdida.	Costo logístico (CL)	$(\text{Costo de venta perdida} + \text{Costo de mantener inventario} + \text{Costo de almacenamiento} + \text{Costo de transporte})$ [Soles/Caja unitaria]
		Reducción del costo logístico	$(\text{Costo inicial} - \text{Costo final})$ [Soles]		

Fuente: Elaboración propia.

1.7. Alcance

Se enmarca en el ámbito del campo de la ingeniería industrial en el área de gestión de la cadena de suministros, específicamente en el planeamiento de la demanda.

1.8. Antecedentes de la investigación

1.8.1. Antecedentes internacionales

Ma (2020), en su investigación titulada: “Human-machine teaming for intelligent demand-planning”, presenta como objetivo determinar cómo las diferentes estructuras de toma de decisiones que involucran una integración humano-computador influyen en el desempeño del planeamiento de la demanda, realizando para ello un experimento con grupos de tratamiento que involucran distintas estrategias de trabajo basadas en modelamiento con inteligencia artificial y ciertos grado de intervención humana, versus el grupo de control formado por el proceso tradicional de ajuste manual de pronósticos, concluyendo que, un modelo híbrido de inteligencia artificial con una adecuada intervención humana, mejora la exactitud de pronósticos operativos en 53%, los tácticos en 64%, y los niveles de inventario en 70%. La relevancia de esta investigación radica en que determina la necesidad de contar con un proceso de modelamiento predictivo sólido para alcanzar la excelencia operacional en la cadena de suministro.

Delfs (2018), en su investigación titulada: “Forecasting in the supply chain with machine learning techniques”, presenta como objetivo pronosticar las ventas futuras de sistemas de rayos X para la gestión de la cadena de suministros de Siemens Healthineers Advanced Therapies, realizando para

ello una comparativa entre el desempeño de técnicas tradicionales de modelamiento de pronósticos como la suavización exponencial y la autorregresión integrada de media móvil (ARIMA), versus técnicas de aprendizaje automático como redes neuronales y bosques aleatorios, concluyendo en que el mejor desempeño se obtiene mediante un modelo extendido de ARIMA, ARIMAX, que incluye variables explicativas, presentando un error porcentual absoluto promedio (MAPE) de 9.55%, mostrando en segundo lugar al método de ensamble (pronósticos combinados) con 11.94%. Esta investigación como tal proporciona una referencia importante sobre la aplicación práctica del modelo ARIMA y de los pronósticos combinados, técnicas que serán utilizadas en el presente trabajo.

Athanasopoulos, Ahmed y Hyndman (2007), en su investigación titulada: “Hierarchical forecasts for Australian domestic tourism”, presenta como objetivo comparar el desempeño del método propuesto de generación de pronósticos jerárquicos llamado “*top-down* basado en proporciones de pronósticos” respecto a las alternativas clásicas de generación de pronósticos jerárquicos (*bottom-up*, y *top-down* basado en proporciones históricas), concluyendo en que este nuevo método presenta los menores niveles de error de pronóstico medidos a través del MAPE. Esta investigación como tal proporciona uno de los principales enfoques de generación de pronósticos para el desarrollo del presente trabajo.

Hyndman y Khandakar (2008), en su investigación titulada: “Automatic time series forecasting: the forecast package for R”, presenta como objetivo desarrollar una librería en el lenguaje de programación R llamada “forecast” que permita automatizar la generación de pronósticos de series de tiempo basados en algoritmos de suavización exponencial y de ARIMA para su aplicación a serie de datos estacionales y no estacionales, realizando la selección de las mejores parametrizaciones mediante el criterio de información de Akaike (AIC), concluyendo en la publicación de la nueva librería en “The Comprehensive R Archive Network” (CRAN) para su uso por parte de toda la comunidad. Esta investigación proporciona una de las librerías principales empleadas en el presente trabajo desarrollado en su parte técnica con el lenguaje de programación R.

1.8.2. Antecedentes nacionales

Chávez, De la Cruz y Rodas (2018), en su investigación titulada: “Propuesta de mejora del rendimiento sobre la inversión de inventarios para una empresa de bebidas no alcohólicas”, presenta como objetivo “maximizar el retorno sobre el capital invertido en inventario para mantener el nivel de servicio deseado”, utilizando para ello el modelo RightStockTM desarrollado por Edward Frazelle, el cual se orienta a la optimización de referencias, pronósticos, tiempos de entrega y tamaños de lote productivos, concluyendo con un resultado favorable al obtenerse un valor presente neto (VPN) de 3.8 millones de soles en un periodo de 4 años, mostrando los primeros indicios sobre el potencial efecto de mejora del pronóstico de ventas en el sistema logístico de una embotelladora de bebidas no alcohólicas, aunque sin

profundizar en los detalles a nivel técnico, tema que el presente trabajo sí desarrollará.

Pérez, Montalvo y Carruitero (2016), en su investigación titulada: “Resideño del modelo de planificación y gestión de inventarios de productos terminados en una empresa de colchones”, presenta como objetivo “rediseñar el modelo actual de la gestión de inventarios de productos terminados con base al modelo RightStock™ propuesto por Edward Frazelle”, proponiendo para ello la reducción de inventarios basada en una mejora en el error de pronóstico de la demanda, concluyendo con un resultado favorable al obtenerse un VPN de 83,589 dólares americanos en un periodo de 4 años, ofreciendo detalles técnicos sobre el modelamiento de pronósticos que sustentan la mejora.

1.8.3. Antecedentes locales

Campos (2017), en su investigación titulada: “Implementación de un sistema logístico para reducir costos de inventarios en la Botica La Luz S.A.C. Trujillo 2017”, presenta como objetivo “implementar un sistema logístico para reducir costos de inventarios en la Botica La Luz S.A.C. en la ciudad de Trujillo – Perú para el año 2017”, mediante la segmentación del portafolio de productos considerando una clasificación ABC, proponiendo además una metodología de generación de pronósticos y de gestión de inventarios, sumado a la clasificación de proveedores, logrando como resultado la reducción de los costos de inventarios en un 20.34%, considerándose una investigación relevante por el enfoque metodológico tradicional en el

modelamiento de pronósticos, en contraposición al enfoque adoptado en el presente trabajo.

Paredes (2017), en su investigación titulada: “Gestión de inventarios y su influencia en los costos logísticos en la empresa Maderas del Norte S.R.L.”, presenta como objetivo “determinar cómo influye la gestión de inventarios en los costos logísticos de la empresa Maderas del Norte S.R.L.”, mediante el cálculo del costo logístico y el desarrollo del pronóstico de demanda, determinando el lote óptimo para hallar la relación entre los costos y la gestión de inventarios, encontrando un coeficiente de determinación (R^2) de 92.1%, indicando que la gestión de inventarios a través del lote óptimo está significativamente relacionada a los costos logísticos representando el 60.88% de los mismos, considerándose esta investigación relevante por el enfoque metodológico tradicional en el modelamiento de pronósticos, en contraposición al enfoque adoptado en el presente trabajo.

1.9. Limitaciones

- Falta de datos de ingresos brutos y descuentos valorizados entre los periodos de enero 2014 y abril 2016, imposibilitando incluir dentro del modelado por regresión lineal múltiple a esta franja de tiempo, optando en este caso por el periodo de mayo 2016 a fines del 2019.
- Falta de datos de temperaturas para todas las ciudades donde la empresa cuenta con presencia de un centro de distribución, debido a que el monitoreo actual considera

solo 11 localidades definidas como las principales, agrupando a más del 70% del volumen de venta.

1.10. Marco teórico

1.10.1. Logística de la cadena de suministro

Definida como el flujo de materiales, información, y dinero entre corporaciones, donde la cadena de suministro engloba la red de infraestructura, vehículos y sistemas de información, y la logística es lo que sucede dentro de la cadena de suministro como la gestión de inventarios, compras, transporte y almacenamiento (Frazelle, 2002).

1.10.1.1. Métricas de la gestión logística

A. Costo logístico (CL)

El costo logístico se define como la suma del costo de venta perdida (CVP), el costo de mantener inventario (CMI), el costo de almacenamiento (CA) y el costo de transporte (CT); lo que implica que la gestión logística se aboque a minimizar este costo total respetando restricciones definidas como el alcanzar un determinado nivel de servicio. En este sentido se realiza un *trade-off*, es decir, un balance entre el beneficio de contar con un mejor nivel de servicio reduciendo el costo de venta perdida, y los costos adicionales en inventario y transporte para hacer esto posible, tal y como se aprecia en la figura 5 (Frazelle, 2002).

Ecuación 1. Costo logístico

$$CL = (CVP + CPI + CA + CT) \times TCE$$

Ecuación 2. Costo de venta perdida

$$CVP = \frac{(1 - NS) \times Total \text{ cajas requeridas} \times MCU}{TCE}$$

MCU: Margen de contribución unitaria

Ecuación 3. Costo de mantener inventario

$$CMI = \frac{Inventario \text{ promedio} \times Tasa \text{ posesión inventarios}}{TCE}$$

Ecuación 4. Costo de almacenamiento

$$CA = \frac{Costo \text{ almacenamiento}}{TCE}$$

Ecuación 5. Costo de transporte

$$CT = \frac{Costo \text{ transporte materia prima} + T1 + T2}{TCE}$$

T1: Distribución primaria

T2: Distribución secundaria

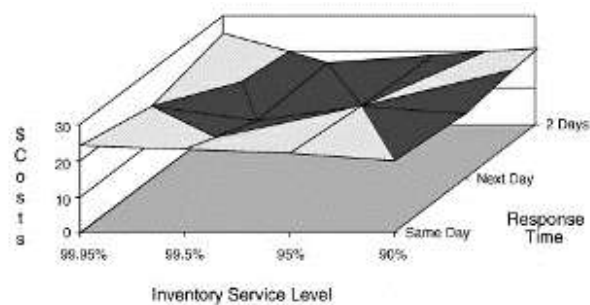


Figura 5. Optimización logística. Fuente: Frazelle, 2002.

Para determinar qué tanto influye el desempeño del pronóstico sobre los niveles de inventario valorizado, se empleará la regresión de la ecuación 6 obtenida a través de los datos históricos de la empresa, con validez demostrada por el estadístico F con valor p de 5.27×10^{-11} (< 0.05) y un coeficiente de determinación ajustado R^2 de 99.4%.

Ecuación 6. Relación inventario y pronóstico de ventas

$$Inv = 6.199 \times ventas[cu] + 1.25 \times 10^9 \times bias \\ - 4.418 \times 10^1 \times venta[cu] \times bias$$

B. Nivel de servicio (NS)

El nivel de servicio mide el desempeño de la disponibilidad del inventario expresado con el *unit fill rate*, comparando el total de cajas unitarias entregadas (TCE) respecto al total de las cajas unitarias requeridas (TCR) para el caso de la embotelladora de bebidas no alcohólicas objeto de estudio.

Ecuación 7. Nivel de servicio o *fill rate*

$$NS = TCE / TCR$$

Y de forma semejante al costo logístico, para determinar qué tanto influye el desempeño del pronóstico sobre el nivel de servicio, se empleará la regresión de la ecuación 8, obtenida mediante los resultados de proyectos de mejora de la empresa,

con validez demostrada por el estadístico F con valor p de 0.001808 (< 0.05) y un coeficiente de determinación ajustado R^2 de 78.7%.

Ecuación 8. Relación nivel de servicio y pronóstico de ventas

$$NS = 0.947 + 0.049 \times FA - 0.137 \times bias + 0.317 \times FA \times bias - 0.785 \times rech_{nc}$$

1.10.2. Pronóstico de ventas

Los pronósticos o *forecasts* son definidos como la proyección en el futuro de la demanda esperada, según una serie de supuestos dados para su desarrollo (Mentzer y Moon, 2005).

1.10.2.1. Demand sensing

Chase (2013) define al *demand sensing* como al uso de datos de ventas muy granulares para mejorar los pronósticos de demanda de corto plazo, impactando de forma positiva en la gestión de inventarios, y requiriendo a su vez de tres elementos claves:

- El uso de datos lo más próximos al cliente. Esto requiere la habilidad de coleccionar estos datos en todos los mercados, para entender quién compra qué producto y en qué cantidad.
- La medición del impacto de los programas de *demand shaping*. Esto refiere a la habilidad analítica de medir y determinar el impacto de las actividades que dan forma a la demanda, como

los precios promocionales, tácticas de venta, eventos de marketing y otros factores que aumentan la demanda.

- La reducción de la latencia o mínima latencia posible. Eso refiere a la habilidad de modelar los cambios en los pronósticos de manera más frecuente. Tradicionalmente, los pronósticos de demanda se realizaban mensualmente. El *demand sensing* requiere que la demanda sea modelada en periodos de tiempo más reducidos, en base semanal o incluso diaria, dependiendo de la frecuencia con que se reciba nueva información.

1.10.2.2. Generación de pronósticos jerárquicos

Athanasopoulos, Ahmed y Hyndman (2007) exponen que en la generación de pronósticos es frecuente encontrarnos ante una estructura jerárquica de series temporales, lo que implica disponer de distintos niveles de agregación según determinados atributos (país/región/ciudad, por ejemplo), por lo que tenemos que emplear algún método para conciliar los pronósticos en toda la estructura jerárquica. Para esquematizar esta idea, en la figura 6 contamos con un diagrama de árbol jerárquico de tres niveles con el nivel 0 definido en el total, y el primer nivel de desagregación en el nivel 1, continuando así hasta el nivel 2 para este ejemplo. Además, definiremos $\hat{Y}_{X,t,k,v}$ como el pronóstico en el tiempo t en el nivel k con versión v de series \hat{Y}_X que corresponden al nodo X en el árbol jerárquico, y utilizamos letras para denotar los nodos individuales.

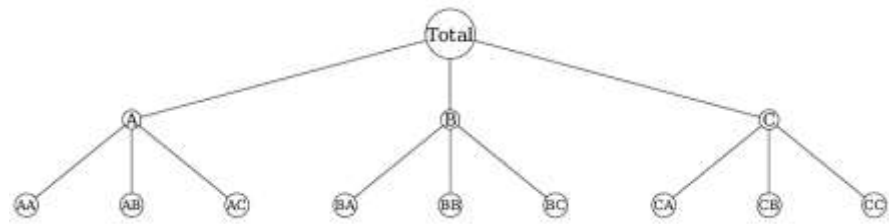


Figura 6. Diagrama de árbol jerárquico de tres niveles.

Fuente: Athanasopoulos, Ahmed y Hydman, 2007.

A. Enfoque bottom-up

La característica principal del enfoque *bottom-up* es la de obtener el pronóstico total de la jerarquía partiendo de la acumulación de pronósticos generados al mínimo nivel de detalle, con la ventaja de no perder información debido a un pronóstico generado a nivel agregado, pero con el problema de ser más desafiante al presentarse una alta variabilidad en el comportamiento de las series de tiempo.

Ecuación 9. Enfoque *bottom-up*

$$\hat{Y}_{A,t,1,0} = \sum_{X=AA}^{AC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{B,t,1,0} = \sum_{X=BA}^{BC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{C,t,1,0} = \sum_{X=CA}^{CC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{t,0} = \sum_{X=AA}^{CC} \hat{Y}_{X,t,2,0}$$

B. Enfoque top-down basado en proporciones de pronósticos

El enfoque *top-down* para generar un pronóstico jerárquico se construye de manera opuesta al *bottom-up*, es decir, se genera un pronóstico desde los más altos niveles de agregación, y se desagrega hacia los niveles inferiores empleando algún método, tradicionalmente mediante la proporciones de venta histórica, pero Athanasopoulos y Hyndman (2007) introdujeron la noción de desagregar los pronósticos mediante una serie de pronósticos generados a mayor detalle, es decir, mediante pronósticos multinivel, logrando de esta forma mejoras empíricas en exactitud, siendo adoptado en el transcurso de los años en plataformas como SAP IBP (Sankaran, Sasso, Kepczynski y Chiaraviglio, 2019).

Ecuación 10. Enfoque *top-down* basado en proporciones de pronósticos

$$\hat{Y}_{A,t,1,1} = \hat{Y}_{t,0} \times \hat{Y}_{A,t,1,0} / \sum_{X=A}^C \hat{Y}_{X,t,1,0}$$

$$\hat{Y}_{B,t,1,1} = \hat{Y}_{t,0} \times \hat{Y}_{B,t,1,0} / \sum_{X=A}^C \hat{Y}_{X,t,1,0}$$

$$\hat{Y}_{C,t,1,1} = \hat{Y}_{t,0} \times \hat{Y}_{C,t,1,0} / \sum_{X=A}^C \hat{Y}_{X,t,1,0}$$

$$\hat{Y}_{AA,t,2,1} = \hat{Y}_{A,t,1,1} \times \hat{Y}_{AA,t,2,0} / \sum_{X=AA}^{AC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{AB,t,2,1} = \hat{Y}_{A,t,1,1} \times \hat{Y}_{AB,t,2,0} / \sum_{X=AA}^{AC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{AC,t,2,1} = \hat{Y}_{A,t,1,1} \times \hat{Y}_{AC,t,2,0} / \sum_{X=AA}^{AC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{BA,t,2,1} = \hat{Y}_{B,t,1,1} \times \hat{Y}_{BA,t,2,0} / \sum_{X=BA}^{BC} \hat{Y}_{X,t,2,0}$$

∴ =

$$\hat{Y}_{CA,t,2,1} = \hat{Y}_{C,t,1,1} \times \hat{Y}_{CA,t,2,0} / \sum_{X=CA}^{CC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{CB,t,2,1} = \hat{Y}_{C,t,1,1} \times \hat{Y}_{CB,t,2,0} / \sum_{X=CA}^{CC} \hat{Y}_{X,t,2,0}$$

$$\hat{Y}_{CC,t,2,1} = \hat{Y}_{C,t,1,1} \times \hat{Y}_{CC,t,2,0} / \sum_{X=CA}^{CC} \hat{Y}_{X,t,2,0}$$

1.10.2.3. Métricas de desempeño del pronóstico

A. Bias

Es una métrica que evalúa el sesgo del pronóstico, es decir, si el pronóstico está sobreestimado mediante un *bias* positivo, o si se presenta subestimación por un *bias* negativo. Como tal es un indicador que no siempre se considera, pero que podría llegar a evidenciar signos de una influencia política negativa que contamina el proceso de generación de pronósticos (Gilliland, 2010).

Ecuación 11. *Bias*

$$Bias = \left[\left(\frac{\sum Pronóstico}{\sum Venta} \right) - 1 \right] \times 100$$

B. Error porcentual absoluto promedio ponderado

(WMAPE)

El WMAPE es una variación de la métrica clásica de error de pronóstico MAPE, que mide el error de pronóstico absoluto de manera porcentual, con la diferencia de que los errores porcentuales se promedian de manera ponderada considerando el peso relativo de la venta que representa cada registro, y no el promedio simple entre ellos. Métricas como MAPE o WMAPE indican la magnitud de los errores de pronóstico, pero no dan una señal sobre qué tan eficiente es la organización para alcanzar un determinado nivel de desempeño (Gilliland, 2010).

Ecuación 12. WMAPE

$$WMAPE = \left(\frac{\sum_{i=1}^n |Venta_i - Pronóstico_i|}{\sum_{i=1}^n Venta_i} \right) \times 100$$

C. Exactitud del pronóstico (FA)

La exactitud de pronóstico es el complemento porcentual de la métrica de error relativa considerada, en el caso de este trabajo el WMAPE.

Ecuación 13. Exactitud del pronóstico

$$FA = 1 - WMAPE$$

D. Señal de rastreo (TS)

La señal de rastreo es una métrica que evalúa el sesgo de un pronóstico, facilitando el monitoreo de su consistencia en el transcurso de un periodo de tiempo n que abarca t_1, \dots, t_n evidenciando una sobreestimación con un TS negativo o una subestimación con un TS positivo (Sankaran et al., 2019).

Ecuación 14. Señal de rastreo

$$TS = \left[\sum_{t=1}^n (Venta_t - Pronóstico_t) / \sum_{t=1}^n |Venta_t - Pronóstico_t| \right] \times n$$

E. Valor agregado del pronóstico (FVA)

El valor agregado del pronóstico es una métrica que evidencia el aporte concreto al desempeño del pronóstico por cada etapa, participante o método que interviene en su generación. Este indicador como tal ayuda a definir qué tan eficiente resulta este proceso y si los esfuerzos desplegados para mejorar la exactitud conducen a algún efecto positivo (Gilliland, 2010).

Ecuación 15. Valor agregado del pronóstico

$$FVA = FA_{etapa 2} - FA_{etapa 1}$$

1.10.2.4. Técnicas de pronóstico cuantitativas

A. Autorregresión integrada de media móvil (ARIMA)

Shmueli y Lichtendahl (2016) explican que un modelo ARIMA es aquel en el que se modela directamente la autocorrelación de los valores de la serie de tiempo y al mismo tiempo los errores de pronóstico, considerando tres componentes: AR, I y MA.

Este proceso inicia con un modelo $AR(p)$ que autocorrelaciona los valores de la serie de tiempo considerando un desfase $1, 2, \dots, p$.

Ecuación 16. Modelo AR

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \epsilon_t$$

Posteriormente se adiciona la autocorrelación de los errores de pronóstico (promedio móvil) hasta el desfase q obteniendo el modelo $ARMA(p, q)$.

Ecuación 17. Modelo ARMA

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Debido a que tanto el modelo AR como el modelo ARMA no pueden emplearse sobre series de tiempo con estacionalidad o

tendencia, ARIMA incorpora un paso preliminar que remueve la tendencia, la operación I (integrada) para la diferenciación, especificando el orden de esta mediante el parámetro d en $ARIMA(p, d, q)$.

En el lenguaje de programación R, la librería “forecast” contiene una implementación del modelo ARIMA que permite seleccionar de manera automática el mejor modelo, considerando incluso variantes como SARIMA, que incluye la estacionalidad o SARIMAX que, además de la estacionalidad, permite anexar otras series para emplearlas como predictores fuera de la serie principal.

```
1. ## ARIMA EN R ##
2.
3. #I. Carga de librería----
4. library(forecast)
5.
6. #II. Desarrollo de modelo básico----
7. modelo <- auto.arma(y)
8.
9. #III. Generación de pronóstico----
10. pronostico <- forecast(modelo, h = N°_periodos)
```

B. Intervalos de predicción

Un intervalo de predicción es un pronóstico por rangos con un nivel de incertidumbre asociado a este, por lo que resulta más informativo que los puntos de pronóstico o *point forecast* en solitario, al indicar el nivel de riesgo asociado al pronóstico. A manera de ejemplo, un punto de pronóstico 500, no nos dice

qué tan lejos el pronóstico podría ir arriba o por debajo de este número. En contraste, el intervalo de predicción de 95%: [460, 540], nos indica que estamos un 95% seguros que la venta caerá dentro de este intervalo (Shmueli y Lichtendahl, 2016).

En el lenguaje de programación R los distintos modelos como el de suavización exponencial o el de ARIMA ofrecen la posibilidad de generar intervalos de predicción mediante la configuración del parámetro *level*, y para el caso de la regresión lineal, a través del parámetro *interval*.

```
1. ## SUAVIZACIÓN EXPONENCIAL EN R ##
2.
3. #I. Carga de librería----
4. library(forecast)
5.
6. #II. Desarrollo de modelo básico----
7. modelo <- ets(y, model = "ZZZ")
8.
9. #II. Generación de pronóstico----
10. pronostico <- forecast(modelo, h = N°_periodos,
    level = %_intervalo_predicción)
```

```
1. ## REGRESIÓN LINEAL EN R ##
2.
3. #I. Desarrollo de modelo básico---
4. modelo <- lm(y ~ x1 + x2 + x3 +..., data = Set_de_datos)
5.
6. #II. Generación de pronóstico---
7. pronostico <- predict(modelo, newdata =
    Nuevo_set_de_datos, interval = "predict")
```

C. Método Croston

Hyndman y Athanasopoulos (2018) exponen que en la generación de pronósticos es frecuente trabajar con series que presentan intermitencia en su comportamiento, para lo cual se emplean métodos como el de Croston, que aborda el problema generando una estimación de la magnitud de la venta considerando el último valor no cero registrado q_i y a su vez una estimación del tiempo de interllegada considerando a a_i como el tiempo entre q_{i-1} y q_i , y aplicando sobre ambas estimaciones la suavización exponencial con un coeficiente α .

Ecuación 18. Estimación de la magnitud de venta según Croston

$$\hat{q}_{i+1|i} = (1 - \alpha)\hat{q}_{i|i-1} + \alpha q_i$$

Ecuación 19. Estimación del tiempo de interllegada según Croston

$$\hat{a}_{i+1|i} = (1 - \alpha)\hat{a}_{i|i-1} + \alpha a_i$$

Considerando que el coeficiente α puede asumir valores entre 0 y 1, además que j es el último tiempo en que se registró una venta positiva, y ante valores incrementales de h , podemos obtener el pronóstico para la demanda en el tiempo $T + h$.

Ecuación 20. Pronóstico mediante el método Croston

$$\hat{y}_{T+h|T} = \hat{q}_{j+1|j} / \hat{a}_{j+1|j}$$

En el lenguaje de programación R el paquete “tsintermittent” proporciona una implementación del método Croston que optimiza el parámetro α .

```
1. ## MÉTODO CROSTON EN R ##
2.
3. #I. Carga de librería---
4. library(tsintermittent)
5.
6. #II. Desarrollo de modelo básico---
7. modelo <- crost(y, h = N°_periodos)
8.
9. #III. Obtención de pronóstico----
10. pronóstico <- modelo$frc.out
```

D. Naive forecast

Shmueli y Lichtendahl (2016) exponen que, como una buena práctica dentro de la generación de pronósticos, se debe considerar el desempeño del *naive forecast*, es decir del pronóstico en el tiempo $t+k$ que adopte el último valor de la serie en el tiempo t .

Ecuación 21. Naive forecast sin estacionalidad

$$F_{t+k} = y_t$$

Para el caso de las series estacionales, se considera al *seasonal naive forecast* como el pronóstico en el tiempo $t+k$ del último valor de la serie que corresponda a la misma estación.

Ecuación 22. Naive forecast con estacionalidad

$$\begin{aligned}F_{t+1} &= y_{t-M+1} \\F_{t+2} &= y_{t-M+2} \\&\vdots \\F_{t+k} &= y_{t-M+k}\end{aligned}$$

La importancia del *naive forecast* radica en que se convierte en una línea base para la comparación con los diversos métodos a considerar, ya que al basarse en el supuesto de que la última información es la más relevante para un pronóstico, si este método supera en exactitud a otros modelos más sofisticados, el *naive forecast* debería primar en la selección de alternativas.

En el lenguaje de programación R el paquete “forecast” cuenta con una implementación del *naive forecast* para series tanto estacionales como no estacionales.

```
1. ## NAIVE FORECAST EN R ##
2.
3. #I. Carga de librería---
4. library(forecast)
5.
6. #II. Desarrollo de modelo básico---
7. modelo <- naive(y, h = N°_periodos) #para series no
   estacionales
8. modelo <- snaive(y, h = N°_periodos) #para series
   estacionales
9.
10. #III. Obtención de pronóstico---
11. pronostico <- modelo$mean
```

E. Pronósticos combinados

Al momento de seleccionar entre distintos modelos, la combinación de métodos de pronóstico nos puede llevar a mejorar la exactitud versus la sola elección de un método particular. Dicha combinación de manera frecuente se realiza mediante el ensamble, esto significa que, múltiples métodos de pronóstico se aplican sobre la misma serie de tiempo generando pronósticos independientes. Posteriormente estos pronósticos se ponderan entre sí para dar paso al pronóstico final, con la ventaja de poder captar diferentes aspectos de la serie de tiempo (Shmueli y Lichtendahl, 2016).

F. Redes neuronales

Shmueli y Lichtendahl (2016) explican que el algoritmo de redes neuronales se basa en la actividad biológica del cerebro, emulando la interconexión de las neuronas entre sí (ver figura 7), siendo capaces de aprender con la experiencia, por lo cual pueden llegar a consumir gran cantidad de recursos computacionales en su desarrollo. Como principal ventaja este algoritmo permite la captura de relaciones complejas entre los predictores y el resultado (variable respuesta) sin requerir muchos datos de entrada por parte del usuario, pero a su vez posee como desventaja que las relaciones que captura son difíciles de entender, considerándose este método como una caja negra que vincula los predictores y los resultados

mediante una secuencia de capas, donde en cada capa se realiza una operación sobre sus datos de entrada generando un resultado que posteriormente será procesado en la capa siguiente, continuando de esta forma hasta llegar a la última capa de la red. Las capas intermedias de la red (entre la primera y la última capa), se componen de nodos llamados variables derivadas que consisten en sumas ponderadas de datos de entrada sobre los que se aplica una función de activación, siendo una de las más comunes la función sigmoide.

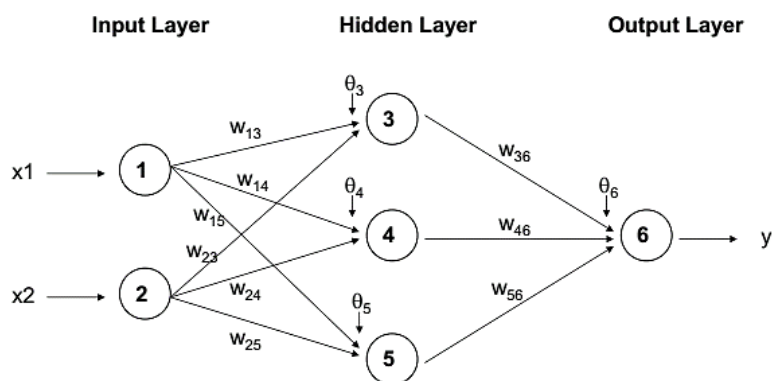


Figura 7. Esquematización de una red neuronal de tres capas.

Fuente: Shmueli y Lichtendahl, 2016.

En el lenguaje de programación R el paquete “forecast” proporciona una implementación del algoritmo de redes neuronales autorregresivas de una sola capa intermedia sobre la que pueden definirse el número de nodos o variables derivadas, otorgando además la posibilidad de anexar otras series para emplearlas como predictores en el modelo.

```

1. ## REDES NEURONALES EN R ##
2.
3. #I. Carga de librería----
4. library(forecast)
5.
6. #II. Desarrollo de modelo básico----
7. modelo <- nnetar(y)
8.
9. #III. Generación de pronóstico----
10. pronostico <- forecast(modelo, h = N°_periodos)

```

G. Regresión lineal múltiple

Los modelos de regresión múltiple se basan en el supuesto de que una serie de tiempo y mantiene una relación lineal con otras series de tiempo x_1, \dots, x_k considerando a la variable y como dependiente o resultante, y a las variables x como independientes o predictoras, aplicándose sobre estas últimas unos coeficientes β_1, \dots, β_k , que miden el efecto de cada predictor después de considerar el efecto de los demás predictores en el modelo (Hyndman y Athanasopoulos, 2018).

Ecuación 23. Expresión general de la regresión lineal múltiple

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t$$

Para la estimación de los coeficientes β empleamos el principio de mínimos cuadrados, eligiendo valores que permiten minimizar la sumatoria de errores al cuadrado.

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_k x_{k,t})^2$$

En el lenguaje de programación R contamos con la función de regresión lineal dentro del paquete “stats”, incorporada con la instalación predeterminada de R.

```
8. ## REGRESIÓN LINEAL EN R ##
9.
10. #I. Desarrollo de modelo básico---
11. modelo <- lm(y ~ x1 + x2 + x3 +..., data = Set_de_datos)
12.
13. #II. Generación de pronóstico---
14. pronostico <- predict(modelo, newdata =
    Nuevo_set_de_datos)
```

H. Suavización exponencial

Según Shmueli y Lichtendahl (2016), es una de las técnicas de generación de pronósticos más empleada en los entornos de negocios debido a su popularidad, flexibilidad, facilidad para automatizar, y al poco esfuerzo computacional que requiere.

La forma básica de esta técnica es la de suavización exponencial simple, empleándose para generar un pronóstico F_{t+1} para el tiempo $t + 1$ ponderando la información histórica más reciente y_t y la más antigua (y_{t-1}, y_{t-2} , etc.) mediante un coeficiente α cuyos valores pueden oscilar entre 0 y 1:

$$F_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots$$

O en su forma equivalente, donde se considera al error de pronóstico e_t y al pronóstico F_t en el periodo t .

Ecuación 24. Suavización exponencial simple

$$F_{t+1} = F_t + \alpha e_t$$

Cabe mencionar que la suavización exponencial simple se aplica sobre series de tiempo que no muestran tendencia ni estacionalidad, aplicándose para estos casos las variantes de suavización que veremos a continuación.

En la suavización exponencial para series con tendencia se asume que la tendencia no es global, ya que esta cambia con el paso del tiempo ante las actualizaciones realizadas al incorporarse más datos, expresando el pronóstico F_{t+k} para una serie con tendencia aditiva considerando valores incrementales de k , como la combinación del nivel L_t y el estimado de la tendencia T_t en el tiempo t .

Ecuación 25. Suavización exponencial con tendencia

$$F_{t+k} = L_t + kT_t$$

Para el caso de la actualización del nivel y tendencia se emplean coeficientes α y β que ponderan la información más reciente y antigua, similar al caso de la suavización exponencial simple.

Ecuación 26. Actualización de la suavización exponencial con tendencia

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

En la suavización exponencial con tendencia y estacionalidad o triple, se extiende el mismo concepto de combinación de componentes incluyendo en esta ocasión la estación S_{t+k} en el tiempo $t + k$ considerando que el periodo evaluado contiene M estaciones.

Ecuación 27. Suavización exponencial con tendencia y estacionalidad

$$F_{t+k} = (L_t + kT_t)S_{t+k-M}$$

Y de la misma forma que en casos previos, para el nivel, tendencia y estacionalidad se emplean coeficientes α , β y γ para las actualizaciones a medida que se incorpora nueva información disponible.

Ecuación 28. Actualización de componentes de la suavización exponencial triple

$$L_t = \alpha y_t / S_{t-M} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$S_t = \gamma(y_t / L_t) + (1 - \gamma)S_{t-M}$$

Ahora bien, para automatizar la selección entre las distintas variantes de suavización exponencial mencionadas incluyendo variantes como las de tendencia multiplicativa y amortiguada, el paquete “forecast” del lenguaje de programación R, facilita la selección de la variante de suavización y parametrización de coeficientes mediante el criterio de Akaike (AIC), el cual penaliza a un modelo por un pobre ajuste a los datos y por su nivel de complejidad.

```
11. ## SUAVIZACIÓN EXPONENCIAL EN R ##
12.
13. #I. Carga de librería----
14. library(forecast)
15.
16. #II. Desarrollo de modelo básico----
17. modelo <- ets(y, model = “ZZZ”)
18.
19. #II. Generación de pronóstico----
20. pronostico <- forecast(modelo, h = N°_periodos)
```

1.10.2.5. Validación cruzada

Un paso importante para evaluar la capacidad predictiva de los modelos de pronósticos se da mediante la validación cruzada, lo que implica dividir el conjunto total de datos en un segmento de entrenamiento y un segmento de validación (ver figura 8), puesto que la calidad de un modelo se determina evaluando su desempeño sobre datos no empleados para su desarrollo. Sobre este punto se debe considerar que, un modelo que se ajusta muy bien al segmento de entrenamiento, no necesariamente realizará un buen pronóstico; además, un ajuste perfecto puede siempre obtenerse mediante un

modelo que posea suficientes parámetros; y finalmente, el sobreajuste de un modelo a sus datos implica la no identificación de patrones sistemáticos de comportamiento (Hyndman y Athanasopoulos, 2018).

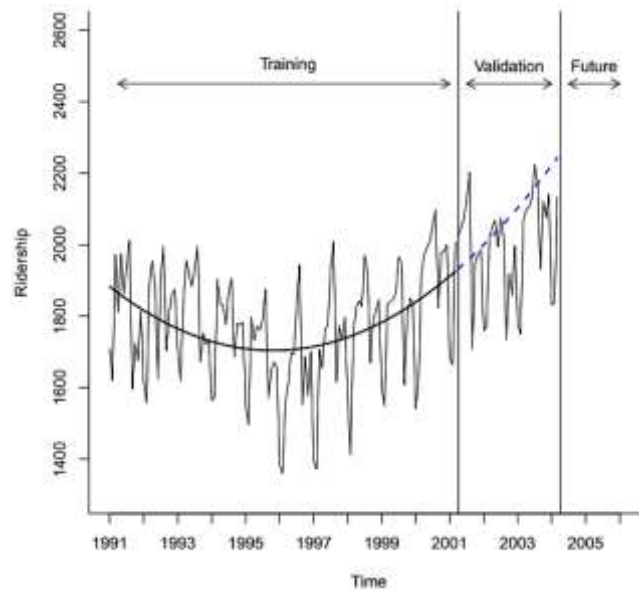


Figura 8. Ejemplo de modelamiento por validación cruzada.

Fuente: Shmueli y Lichtendahl, 2016.

Cabe resaltar que la validación cruzada, dentro del modelamiento estadístico, forma parte de una cultura denominada modelamiento algorítmico, la cual se centra en encontrar una buena solución basada en la capacidad predictiva del modelo sobre el segmento de validación, más allá de su nivel de interpretabilidad y del uso de pruebas de bondad de ajuste, siendo este enfoque empleado de manera intensiva en disciplinas como el aprendizaje automático (Breiman, 2001).

CAPÍTULO II. METODOLOGÍA

2.1. Tipo de investigación

Este trabajo explicativo se clasificará dentro de la investigación no experimental de tipo longitudinal. Es explicativo porque busca determinar el efecto que produciría la mejora del pronóstico de ventas sobre el costo logístico. Es no experimental porque el análisis se realizará *ex post-facto*, de manera retrospectiva, sin una manipulación directa de las variables como se da en los experimentos con grupos de tratamiento y de control. Es de tipo longitudinal, o evolutiva, por la recolección de datos en distintos periodos de tiempo formando series, necesarias para generar pronósticos según los diversos comportamientos observados (Hernández, Fernández y Baptista, 2014).

2.2. Población y muestra (Materiales, instrumentos y métodos)

La población está conformada por el registro total de los volúmenes de venta a nivel nacional expresados en cajas unitarias y que correspondan a la empresa objeto de estudio.

La muestra está conformada por el registro entre los años 2014 y 2019 de los volúmenes de venta a nivel nacional expresados en cajas unitarias y que correspondan a la empresa objeto de estudio. Esto implica seis años de venta, del año 2014 al 2018 para el desarrollo de modelos, y los del año 2019 para la validación de estos.

2.3. Procedimiento

El procedimiento de trabajo de la presente tesis está compuesto por tres etapas, según lo mostrado en la tabla 6.

Tabla 6
Procedimiento de desarrollo de la tesis

Etapa	Procedimiento
Diagnóstico	Se empleará el modelo de Mentzer y Moon (2005) para situar el proceso en su estado actual a nivel de enfoque metodológico.
Solución propuesta	Se desarrollará la propuesta de mejora de los indicadores del pronóstico con la finalidad de reducir los costos logísticos de la embotelladora de bebidas no alcohólicas.
Evaluación económica	Se evaluará la viabilidad de la propuesta mediante la elaboración de un flujo de caja económico considerando el valor presente neto, el periodo de recuperación, la relación beneficio/costo y el análisis de sensibilidad.

Fuente: Elaboración propia.

2.3.1. Diagnóstico del proceso de planeamiento de la demanda

2.3.1.1. Estructura de la demanda de la empresa

La empresa atiende a sus 340 mil clientes con un portafolio de más de 300 productos distribuidos en 6 categorías, con la categoría carbonatadas consolidando el 76% del volumen de venta nacional (ver tabla 7).

Tabla 7
Categoría de productos de la empresa

Categoría	Referencias	Volumen %
Carbonatadas	217	76%
Agua embotellada	60	20%
Jugos	48	3%
Bebidas deportivas	28	1%
Bebidas energéticas	4	0.04%
Té	2	0.03%

Fuente: Elaboración propia.

Con el propósito de establecer modelos de atención diferenciados para cada grupo de clientes, se realiza una segmentación en 4 canales, con el minorista tradicional concentrando el 61% de la venta, mostrando una variabilidad, medida con el coeficiente de variación, de 44%, muy por debajo del resto de canales que presentan variabilidades arriba del 80% (ver tabla 8).

Tabla 8
Canales de venta de la empresa

Canal	Volumen %	Variabilidad %
Minorista Tradicional	61%	44%
Minorista Moderno	17%	83%
Mayorista Tradicional	20%	83%
Mayorista Moderno	2%	106%

Fuente: Elaboración propia.

Además, la empresa cuenta con una cobertura nacional desde sus 65 centros de distribución (CD) y 6 plantas productoras (ver figura 9), capturando en Lima el 50% del volumen de venta, mostrando a su vez el más alto nivel de variabilidad (ver tabla 9).



Figura 9. Mapa de centros de distribución y plantas productoras.

Fuente: Elaboración propia.

Tabla 9

Regiones de venta de la empresa

Región	Volumen %	Variabilidad %
Lima	50%	69%
Sur	19%	50%
Norte	16%	57%
Centro Oriente	15%	59%

Fuente: Elaboración propia.

La mención de la variabilidad importa debido a que influye en el grado de predictibilidad de las series de tiempo, dificultando la planificación a medida que esta aumenta. De esta manera, con la clasificación XYZ

(Sankaran et al., 2019) encontramos que el 44% del volumen de venta presenta una variabilidad entre intermedia y alta (ver tabla 10).

Tabla 10

Volumen de venta según su variabilidad

Clase Variabilidad	Volumen %
X (Baja)	56%
Y (Intermedia)	26%
Z (Alta)	18%

Fuente: Elaboración propia.

2.3.1.2. Estructura del proceso de planeamiento

El planeamiento de la demanda en la empresa se realiza considerando dos horizontes: el táctico y el operativo. En el táctico, el objetivo es entregar a la compañía un pronóstico de ventas mensual y semanal consensuado en cajas unitarias (1 caja unitaria equivale a 24 botellas de 237 ml) y en cajas físicas (la caja física se define por cómo se arma cada producto para su entrega al cliente) a nivel CD, canal y referencia (SKU). En el caso del horizonte operativo, el objetivo es entregar un pronóstico de ventas semanal y diario consensuado en cajas físicas a nivel CD y SKU.

Ambos horizontes de planificación son secuenciales, por lo que, para planificar en el nivel operativo, se requiere del plan táctico como insumo. En este proceso de planeamiento, detallado en la figura 10, interviene no solo el equipo de planeamiento de la demanda,

conformado por siete personas, si no a su vez los equipos de marketing, comercial, producción y abastecimiento.

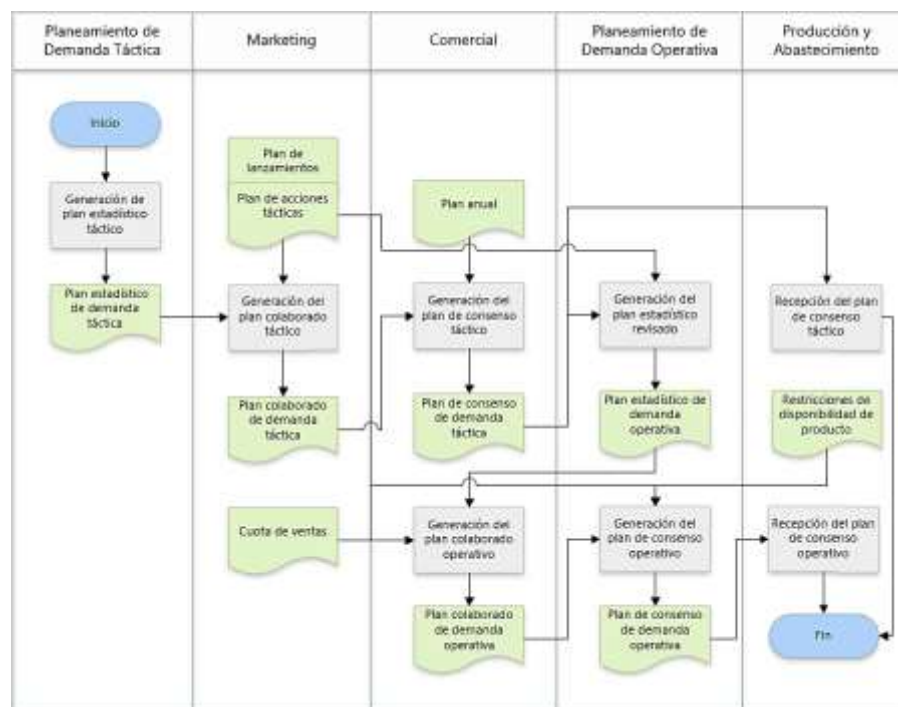


Figura 10. Diagrama de flujo del proceso de planeamiento de la demanda. Fuente: Elaboración propia.

2.3.1.3. Evaluación del enfoque metodológico del planeamiento

Tomando como referencia el modelo propuesto por Mentzer y Moon (2005), se dividirá la evaluación del enfoque metodológico en sus diversos aspectos clasificándolos en una de cuatro fases: incipiente, transición, desarrollo y excelencia.

La primera componente evalúa el modelamiento del pronóstico estadístico ubicándola en la fase de transición (ver la tabla 11), esto debido al uso exclusivo de métodos de series de tiempo, lo cual

imposibilita entender los diversos mecanismos que influyen sobre las ventas, tales como cambios en las políticas de precio y descuentos, además de la influencia de eventos como el fenómeno El Niño en el 2017 y el aumento del impuesto selectivo al consumo en el 2018 (ver figura 11), o la correlación existente con la venta total de la industria o con el consumo privado nacional, entre otras series económicas de interés, resultando en la incapacidad de dimensionar una línea base de volumen o cuantificar el impacto de la actividad promocional.

El segundo componente evalúa la estructura jerárquica del pronóstico ubicándola en la fase de transición (ver tabla 12) al contarse con una estructura rígida que genera pronósticos estadísticos solo al nivel CD/SKU/canal/semana, por la limitación que existe en la generación de pronósticos a cualquier nivel de la estructura del negocio, esto a causa del empleo de la caja física como unidad de medida de modelamiento, creando un problema de consistencia en la agregación, ya que, por ejemplo, una caja física de agua de 20 litros tiene 13 litros más que la caja física de 7 litros. Por esto, lo correcto sería emplear la caja unitaria por su semejanza al litro, permitiendo la generación de pronósticos agregados por categoría o por marca, facilitando luego a un adecuado *top-down* (desagregado en cascada de un pronóstico agregado), para luego conciliarlos con volúmenes generados mediante el enfoque *bottom-up* (partiendo del detalle a lo agregado).

Tabla 11

Evaluación del modelamiento del pronóstico estadístico

Fase	Concepto	Marcar (X)
Incipiente	No existe un análisis estadístico de la demanda histórica.	
Transición	Uso de modelos basados en series de tiempo para analizar los patrones de la demanda histórica.	X
Desarrollo	Uso de modelos basados en regresión y en series de tiempo considerando un enfoque clásico.	
Excelencia	Uso de modelos basados en regresión y series de tiempo considerando un enfoque clásico y de aprendizaje automático (redes neuronales, por ejemplo).	

Fuente: Elaboración propia.

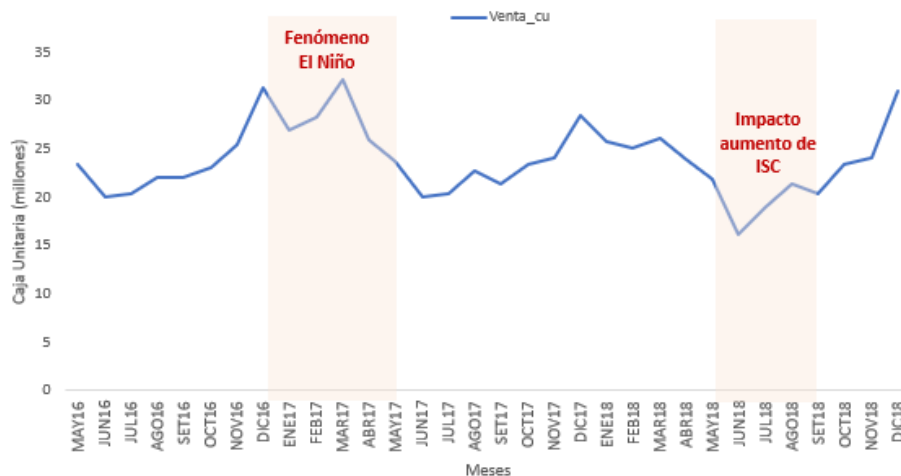


Figura 11. Eventos que impactan en el histórico de venta.

Fuente: Elaboración propia.

Tabla 12
Evaluación de la estructura jerárquica del pronóstico

Fase	Concepto	Marcar (X)
Incipiente	No existe una estructura jerárquica definida para generar pronósticos.	
Transición	Uso de enfoque <i>bottom-up</i> (del detalle a lo general) en la generación de pronósticos.	X
Desarrollo	Uso del enfoque <i>top-down</i> (de lo macro a lo micro) y del enfoque <i>bottom-up</i> (de lo micro a lo macro) sin conciliación en la generación de pronósticos.	
Excelencia	Uso de manera conjunta y conciliada de los <i>enfoques top-down</i> (de lo macro a lo micro) y <i>bottom-up</i> (de lo micro a lo macro) en la generación de pronósticos.	

Fuente: Elaboración propia.

La tercera componente evalúa la segmentación de productos y clientes ubicándola en la fase de desarrollo (ver tabla 13) al realizarse una clasificación de las combinaciones a pronosticar (CD/SKU/canal) considerando su importancia relativa a nivel de volumen de venta (ABC) y en cuanto a variabilidad (XYZ), y tomando esta clasificación en el modelamiento de los pronósticos estadísticos. Ahora bien, como punto de mejora se podría contar con una segmentación de las distintas series de tiempo según su comportamiento (estacional, tendencia, tendencia con estacionalidad, intermitente, etc.) de forma algorítmica.

Tabla 13

Evaluación de la segmentación de productos y clientes

Fase	Concepto	Marcar (X)
Incipiente	Los pronósticos de todos los productos son tratados (generados) de la misma manera.	
Transición	Uso de análisis ABC para segmentar la exactitud de pronósticos como soporte en el ajuste manual del pronóstico colaborado.	
Desarrollo	Uso del análisis ABC para segmentar la exactitud de pronósticos por niveles de importancia y como soporte a la generación de pronósticos estadísticos.	X
Excelencia	Uso de segmentación por patrones de comportamiento como soporte a la generación de pronósticos estadísticos en conjunto con el análisis ABC.	

Fuente: Elaboración propia.

La cuarta componente evalúa a la colaboración del plan de demanda ubicándola en la fase de desarrollo (ver tabla 14), al contarse en la actualidad con un proceso formal de integración de iniciativas de marketing en el proceso de generación de pronósticos, donde a su vez el equipo comercial dispone de la facultad de realizar ajustes manuales a los planes ya sean tácticos u operativos (ver figura 10). Ahora bien, como punto de mejora se podría extender la colaboración involucrando a determinados clientes externos clave mediante el CPFR (reabastecimiento, pronóstico y planeamiento colaborado, por sus siglas en inglés).

Tabla 14

Evaluación de la colaboración de la demanda

Fase	Concepto	Marcar (X)
Incipiente	No se cuenta con un proceso de colaboración de demanda.	
Transición	Se incorpora en el pronóstico las iniciativas promocionales y de marketing al pronóstico estadístico como <i>building blocks</i> .	
Desarrollo	Se incorpora información subjetiva del área de marketing y ventas en el pronóstico.	X
Excelencia	Dentro de la planificación de la demanda intervienen estrategias de integración entre cliente y proveedor mediante CPFR (<i>Collaborative Planning, Forecasting and Replenishment</i>)	

Fuente: Elaboración propia.

La última componente evalúa la capacitación y entrenamiento del equipo de planeamiento de la demanda ubicándola en la fase incipiente (ver tabla 15), al no contar con un grupo humano que disponga del conocimiento técnico necesario para realizar la función de modelamiento predictivo, además de no contar con un programa establecido por la empresa que aborde capacitaciones y certificaciones en la materia.

Tabla 15

Evaluación de la capacitación y entrenamiento del equipo de trabajo

Fase	Concepto	Marcar (X)
Incipiente	El personal de planeamiento de la demanda no cuenta con el conocimiento técnico ni recibe capacitación y entrenamiento en técnicas de pronóstico.	X
Transición	El personal de planeamiento de la demanda recibe capacitación y entrenamiento limitado en técnicas de pronóstico.	
Desarrollo	El personal de planeamiento de la demanda recibe de manera estructurada capacitación y entrenamiento en análisis cuantitativo y estadístico.	
Excelencia	El personal de planeamiento de la demanda recibe de manera estructurada capacitación y entrenamiento en análisis cuantitativo, estadístico, y en el entendimiento del entorno de negocio.	

Fuente: Elaboración propia.

Estas debilidades a nivel metodológico se evidencian a través del bajo desempeño en los indicadores del planeamiento de la demanda, con especial énfasis en la parte estadística semanal y mensual del proceso (ver tabla 16), con una escasa contribución del proceso de colaboración semanal, debido a un FVA operativo de 0.2%.

Tabla 16

Estado de indicadores del planeamiento de la demanda

Indicador	Unidad de medida	Objetivo 2019 (a)	Resultado 2019 (b)	Variación (b) - (a)	Estado
Exactitud estadística mensual	%	80%	75.6%	-4.4%	Fuera de objetivo
FVA táctico	%	2%	4%	2.0%	En objetivo
Exactitud del plan táctico	%	82%	79.6%	-2.4%	Fuera de objetivo
Exactitud estadística semanal	%	78%	71.7%	-6.3%	Fuera de objetivo
FVA operativo	%	2%	0.2%	-1.8%	Fuera de objetivo
Exactitud del plan operativo	%	80%	71.9%	-8.1%	Fuera de objetivo
Bias	%	[-4%, 4%]	7.2%	3.2%	Fuera de objetivo
Señal de rastreo	Adimensional	[-2.5, 2.5]	-13.3	-10.8	Fuera de objetivo

Fuente: Elaboración propia.

De esta manera podemos enumerar las causas raíz de la inadecuada gestión de la demanda con sus respectivas contribuciones (ver tabla 17), para priorizarlas y obtener que, el deficiente modelamiento de pronóstico, la deficiente estructura jerárquica de pronóstico y la falta de capacitación/entrenamiento técnico ocasionan el 89.7% de las pérdidas en el proceso (ver tabla 18), desarrollándose finalmente una matriz de indicadores (ver tabla 19).

Tabla 17
Causas raíz de la inadecuada gestión de la demanda

Causa raíz	Descripción	Pérdida exactitud operativa	Valorizado (Soles)	Peso %
CR-1	Deficiente modelamiento de pronóstico	4.7%	3,260,860	27.0%
CR-2	Deficiente estructura jerárquica de pronóstico	6.3%	4,324,809	35.8%
CR-3	La colaboración de la demanda no es efectiva	1.8%	1,242,232	10.3%
CR-4	Falta de capacitación /entrenamiento técnico	4.7%	3,249,358	26.9%
Total		17.5%	12,077,259	100.0%

Fuente: Elaboración propia.

Tabla 18
Priorización de las causas raíz de la inadecuada gestión de la demanda

Causa raíz	Descripción	Pérdida exactitud operativa	Valorizado (Soles)	Peso %	Acum. %
CR-2	Deficiente estructura jerárquica de pronóstico	6.3%	4,324,809	35.8%	35.8%
CR-1	Deficiente modelamiento de pronóstico	4.7%	3,260,860	27.0%	62.8%
CR-4	Falta de capacitación /entrenamiento técnico	4.7%	3,249,358	26.9%	89.7%
CR-3	La colaboración de la demanda no es efectiva	1.8%	1,242,232	10.3%	100.0%
Total		17.5%	12,077,259	100.0%	100.0%

Fuente: Elaboración propia.

Tabla 19
Matriz de indicadores

Causa raíz	Descripción	Indicador	Fórmula	Valor actual	Valor meta	Propuesta
CR-2	Deficiente estructura jerárquica de pronóstico	Señal de rastreo	$\frac{\sum(Venta - Pronóstico)}{\sum Venta - Pronóstico } \times N^{\circ} periodos$	-13.3	[-2.5, 2.5]	Mejora estructura jerárquica (tabla 20)
CR-1	Deficiente modelamiento de pronóstico	% exactitud estadística semanal	$\left(1 - \frac{\sum Venta - Pronóstico }{\sum Venta}\right) \times 100$	71.7%	78%	Mejora modelamiento (tabla 20)
CR-4	Falta de capacitación/entrenamiento técnico	% de trabajadores capacitados	$\frac{Trabajadores capacitados}{Total trabajadores} \times 100$	0%	100%	Mejora capacitación (tabla 20)

Fuente: Elaboración propia.

2.3.2. Diseño de la propuesta de mejora

Según lo evidenciado en el diagnóstico, el esfuerzo de mejora debe orientarse a las debilidades del proceso, en este caso, al modelamiento, estructura jerárquica y capacitación/entrenamiento en los pronósticos. De tal manera, la propuesta de mejora aborda los puntos detallados en la tabla 20.

Tabla 20

Propuesta de mejora de pronósticos de venta

Componente	Propuesta
Modelamiento del pronóstico estadístico	<ul style="list-style-type: none"> - La implementación del modelamiento causal, de eventos, y de correlación con otras series de tiempo, a través de la regresión lineal múltiple, logrando estimar la línea base de venta, y el impacto de la actividad promocional y de los cambios en la política de precios. - La implementación del modelamiento mediante ARIMA y redes neuronales, considerando variables explicativas en su diseño, para la generación de pronósticos diarios desarrollando el <i>demand sensing</i>.
Estructura jerárquica del pronóstico	<ul style="list-style-type: none"> - La implementación de pronósticos estadísticos por intervalos de predicción, ofreciendo el nivel de incertidumbre que acompaña a cada punto de pronóstico o <i>point forecast</i>, facilitando una mejora en la capacidad de respuesta de la cadena de suministro. - La implementación de pronósticos estadísticos multinivel (por capas) conciliando el enfoque “<i>top-down</i> basado en proporciones de pronósticos” y el enfoque <i>bottom-up</i> en los niveles táctico y operativo (ver figura 12), con la caja unitaria como unidad de medida empleada para la estimación del volumen de venta.
Capacitación y entrenamiento	<ul style="list-style-type: none"> - La incorporación al equipo de un especialista de modelamiento predictivo, sumado a la implementación de un programa de capacitación y entrenamiento en técnicas de pronóstico, considerando certificaciones en la materia.

Fuente: Elaboración propia.

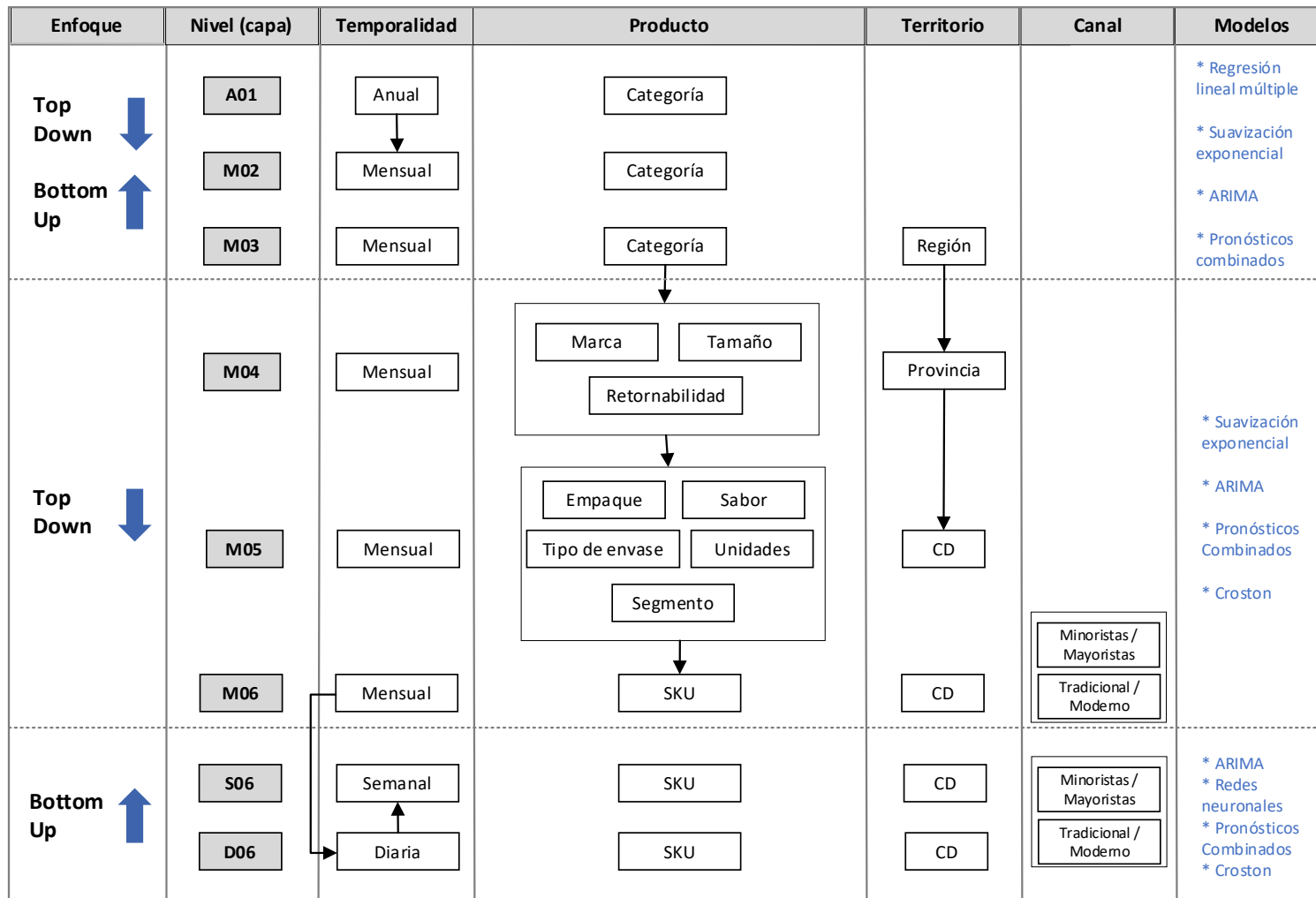


Figura 12. Propuesta de estructura jerárquica para la generación de pronósticos. Fuente: Elaboración propia.

En el desarrollo del presente trabajo, esta propuesta ha sido implementada según el *pipeline* mostrado en la figura 13, empleando el lenguaje de programación R, el entorno de desarrollo RStudio y el gestor de base de datos Microsoft SQL Server, todo esto bajo la administración del equipo de planeamiento, considerando los históricos de venta en cajas unitarias del año 2014 al 2018 para el entrenamiento de modelos y los del año 2019 para las validaciones, según especificaciones. En los anexos 1, 2, 3 y 4 se detallan los *scripts* para cada etapa del *pipeline* de modelamiento desarrollados para el nivel diario 06 (D06).



Figura 13. Pipeline del pronóstico estadístico propuesto.

Fuente: Elaboración propia.

Esta propuesta como tal permite alcanzar niveles de excelencia en las debilidades técnicas del proceso de planeamiento de la demanda, actualmente ubicadas en la fase de transición (ver tablas 21 y 22).

A su vez, la propuesta para su implementación, requiere del modelo de datos presentado en la figura 14, ya que mediante la consulta desarrollada en T-SQL (anexo 1), se extraerá los datos de Microsoft SQL Server para su procesamiento en RStudio, en un flujo de información integrado entre los distintos sistemas del proceso de planeamiento de la demanda mostrado en la figura 15.

Tabla 21
Mejora en el modelamiento del pronóstico estadístico

Fase	Concepto	Marcar (X)
Incipiente	No existe un análisis estadístico de la demanda histórica.	
Transición	Uso de modelos basados en series de tiempo para analizar los patrones de la demanda histórica.	
Desarrollo	Uso de modelos basados en regresión y en series de tiempo considerando un enfoque clásico.	
Excelencia	Uso de modelos basados en regresión y series de tiempo considerando un enfoque clásico y de aprendizaje automático.	X

Fuente: Elaboración propia.

Tabla 22
Mejora en la estructura jerárquica del pronóstico

Fase	Concepto	Marcar (X)
Incipiente	No existe una estructura jerárquica definida para generar pronósticos.	
Transición	Uso de enfoque <i>bottom-up</i> (del detalle a lo general) en la generación de pronósticos.	
Desarrollo	Uso del enfoque <i>top-down</i> (de lo macro a lo micro) y del enfoque <i>bottom-up</i> (de lo micro a lo macro) sin conciliación en la generación de pronósticos.	
Excelencia	Uso de manera conjunta y conciliada de los enfoques <i>top-down</i> (de lo macro a lo micro) y <i>bottom-up</i> (de lo micro a lo macro) en la generación de pronósticos.	X

Fuente: Elaboración propia.

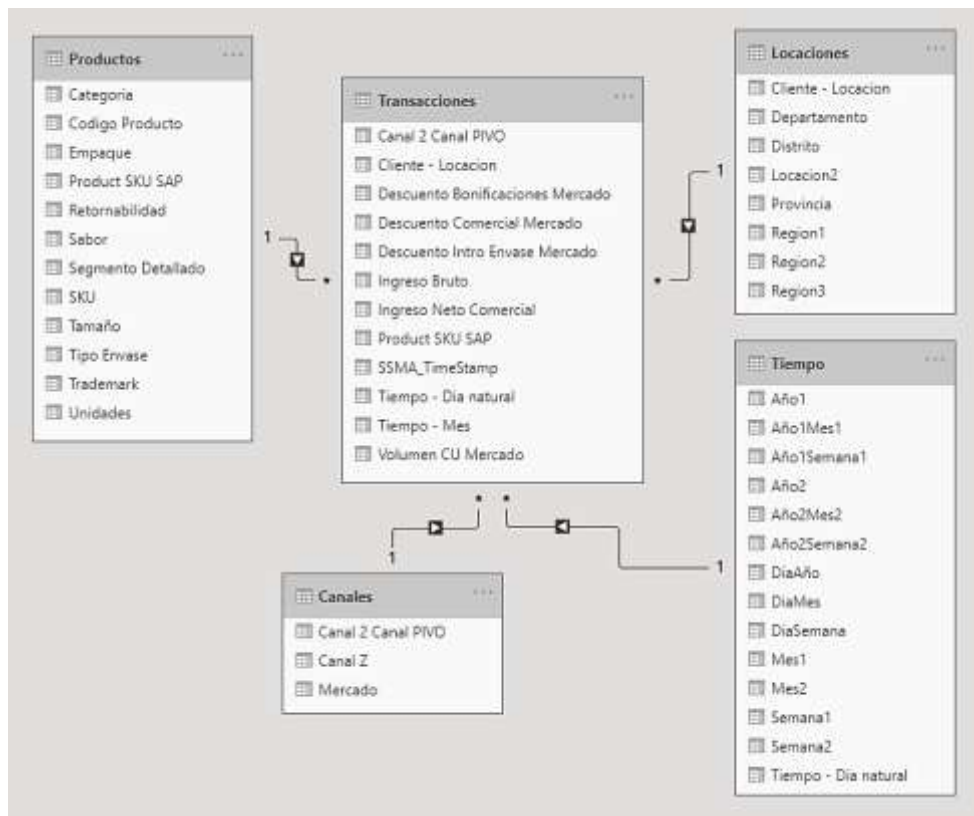


Figura 14. Modelo de datos de la propuesta.

Fuente: Elaboración propia.

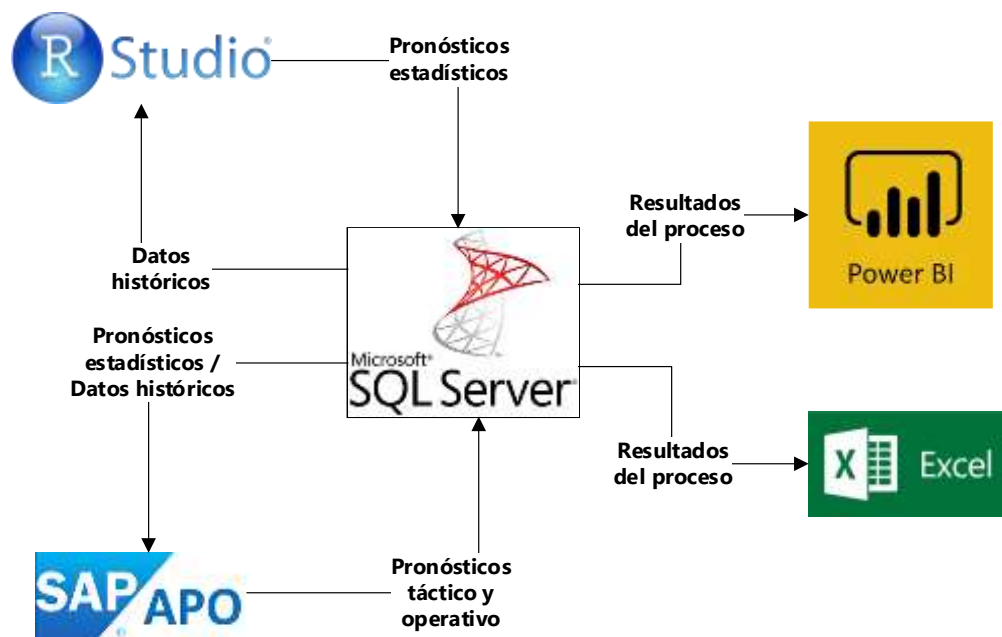


Figura 15. Flujo de información propuesto entre los sistemas de planeamiento.

Fuente: Elaboración propia.

Ahora bien, mediante esta propuesta se destaca que, la empresa dispondría de la capacidad de dimensionar su línea base de venta, discriminando los impactos de la actividad promocional y de eventos como el fenómeno El Niño o el aumento del ISC (ver figura 16), favoreciendo el desarrollo de iniciativas con marketing sobre una base de venta definida.

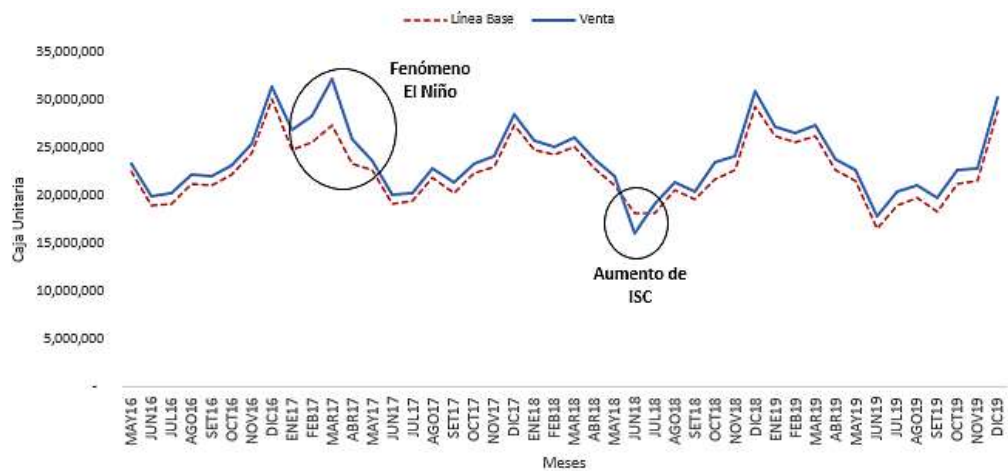


Figura 16. Línea base de venta nacional. Fuente: Elaboración propia.

Adicionalmente, esta metodología permitiría generar pronósticos estadísticos con intervalos de predicción, mostrando el nivel de incertidumbre de las estimaciones, como en el caso de una gaseosa de 296 ml en un CD de Lima (ver figura 17), donde las desviaciones registradas en el año 2019 son contenidas por el límite superior del pronóstico, siendo esta funcionalidad de suma importancia para la para la evaluación de las capacidades respuesta de la cadena de suministros.

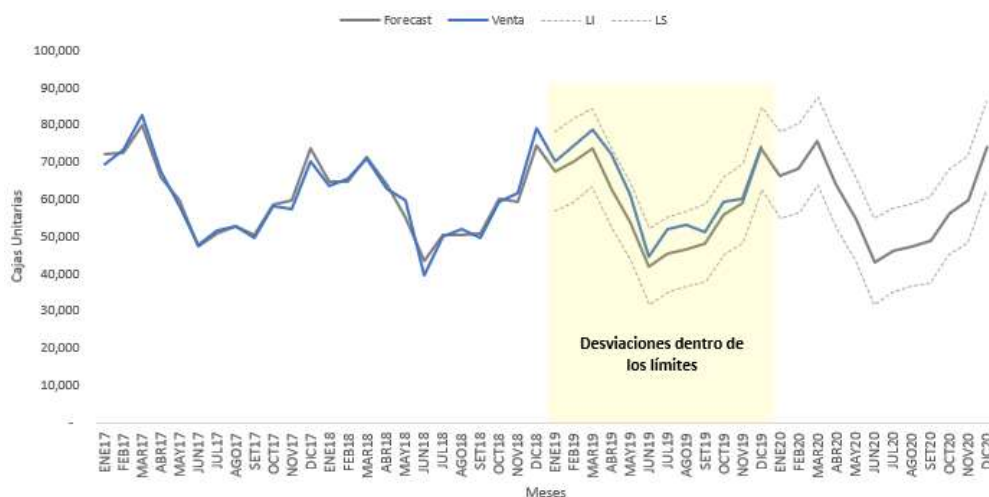


Figura 17. Pronósticos estadísticos para una gaseosa de 296 ml en un CD de Lima. Fuente: Elaboración propia.

Además, debido a la generación de pronósticos multinivel, se podría entregar una mejor estimación a nivel operativo diario y semanal, como en el caso del agua sin gas de 625 ml en un CD limeño (ver figura 18), donde se observa una mejora en exactitud del pronóstico por *demand sensing* respecto al pronóstico conciliado con el nivel táctico mediante el enfoque *top down* propuesto.

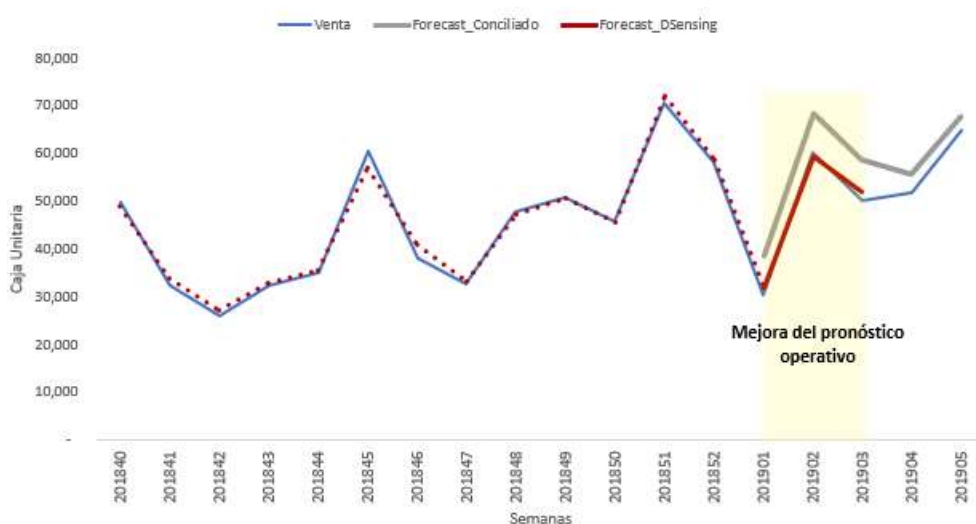


Figura 18. Pronósticos estadísticos para el agua sin gas de 625 ml en un CD de Lima. Fuente: Elaboración propia.

Finalmente, para que las mejoras técnicas se puedan mantener en el tiempo, se necesita de un equipo humano con las capacidades adecuadas, por lo tanto, se considera la contratación de un especialista de modelamiento predictivo con el rol y responsabilidades definidas en el anexo 5, y a su vez de la implementación de un programa de capacitaciones y certificaciones en modelamiento para todo el equipo (ver tabla 23) a realizarse durante el año 2021 y 2022 (ver figura 19).

De esta manera se alcanzaría un nivel de desarrollo en el aspecto de capacitación y entrenamiento del equipo encargado del planeamiento de la demanda, actualmente ubicado en una fase incipiente (ver tabla 24).

Tabla 23

Programa de capacitación técnica en modelamiento predictivo

Programa	Institución / Plataforma	Tipo
Applied Forecasting	University of Nicosia	Curso
Data Science	The Johns Hopkins University / Coursera	Especialización
Mathematics for Machine Learning	Imperial College London / Coursera	Especialización
Analytics: Essential Tools and Methods	Georgia Institute of Technology / edX	Micromaster
Statistics and Data Science	Massachusetts Institute of Technology / edX	Micromaster
IBF Certified Professional Forecaster (CPF)	Institute of Business Forecasting	Certificación
IBF Advanced Certified Professional Forecaster (ACPF)	Institute of Business Forecasting	Certificación

Fuente: Elaboración propia.

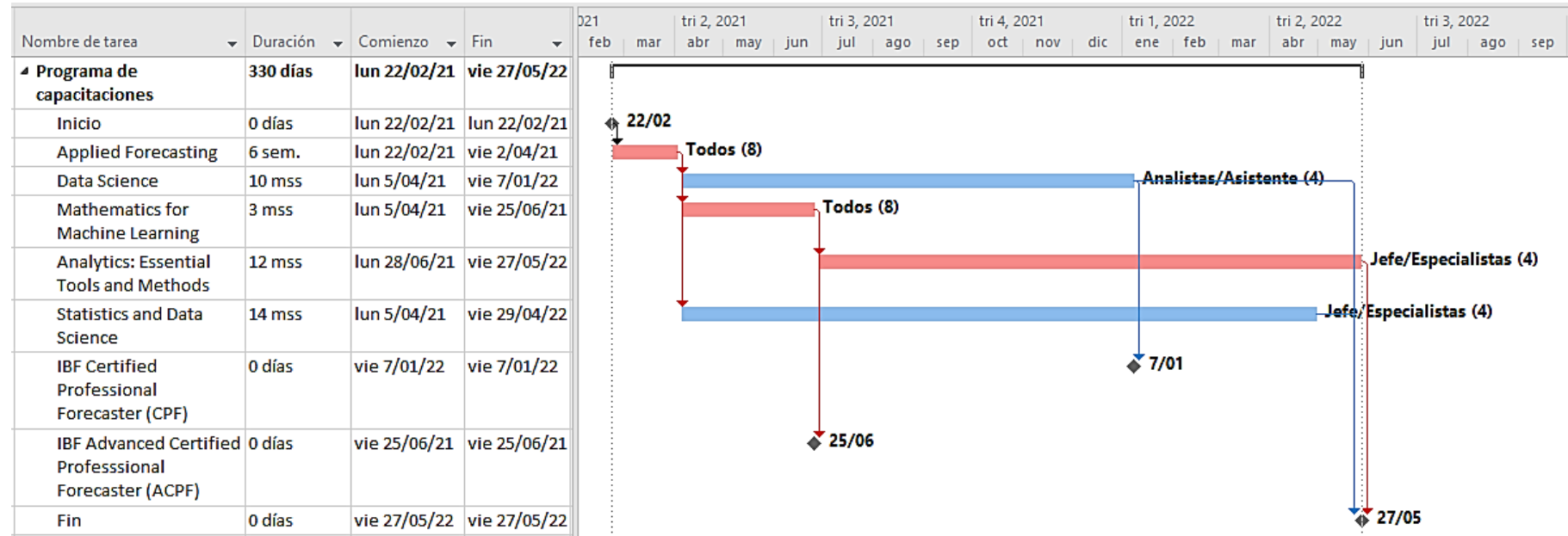


Figura 19. Diagrama de Gantt del programa de capacitación.

Fuente: Elaboración propia.

Tabla 24

Mejora en la capacitación y entrenamiento

Fase	Concepto	Marcar (X)
Incipiente	El personal de planeamiento de la demanda no cuenta con el conocimiento técnico ni recibe capacitación/entrenamiento en técnicas de pronóstico.	
Transición	El personal de planeamiento de la demanda recibe capacitación/entrenamiento limitado en técnicas de pronóstico.	
Desarrollo	El personal de planeamiento de la demanda recibe de manera estructurada capacitación/entrenamiento en análisis cuantitativo y estadístico.	X
Excelencia	El personal de planeamiento de la demanda recibe de manera estructurada capacitación/entrenamiento en análisis cuantitativo, estadístico, y en el entendimiento del entorno de negocio.	

Fuente: Elaboración propia.

2.3.3. Costos de la propuesta de mejora

Para evaluar la viabilidad de la propuesta, se ha considerado como inversión lo concerniente a la adquisición de software (ver tabla 25), hardware (ver tabla 26), la contratación del especialista de modelamiento predictivo (ver tabla 27) y la capacitación y certificación en modelamiento del equipo de planeamiento de la demanda (ver tabla 28), todo esto a desarrollarse durante el año 2021. Para el caso de los costos en dólares americanos se ha empleado el tipo de cambio de 3.6 soles por dólar.

Tabla 25
Costo del software requerido por la propuesta

Software	Costo unitario		Costo total	
	Dólares	Soles	Dólares	Soles
SQL Server Enterprise Edition x 1	13,748	49,493	13,748	49,493
Rstudio Pro x 8	995	3,582	7,960	28,656
Total			21,708	78,149

Fuente: Elaboración propia.

Tabla 26
Costo del hardware requerido por la propuesta

Hardware	Costo unitario		Costo total	
	Dólares	Soles	Dólares	Soles
Laptop Lenovo i7 (81YT001YLM) x 1	3,611	12,999	3,611	12,999
Laptop Lenovo i7 (81YT006MLM) x 5	2,694	9,699	13,471	48,495
Total			17,082	61,494

Fuente: Elaboración propia.

Tabla 27
Costo de la contratación del especialista de modelamiento predictivo

Concepto	Costo mensual	Costo total
	Soles	Soles
Remuneración	9,000	108,000
Sobrecosto laboral	4,440	53,276
Certificación IBF		7,650
Total		168,926

Fuente: Elaboración propia.

Tabla 28

Costo de la capacitación y certificación en modelamiento predictivo

Programa	Costo unitario		Costo total	
	Dólares	Soles	Dólares	Soles
Applied Forecasting x 8	1,839	6,620	14,712	52,963
Data Science x 4	539	1,940	2,156	7,762
Mathematics for Machine Learning x 8	147	529	1,176	4,234
Analytics: Essential Tools and Methods x 4	2,466	8,878	9,864	35,512
Statistics and Data Science x 4	1,495	5,381	5,979	21,524
IBF CPF x 7	1,275	4,590	8,925	32,130
Total			42,812	154,124

Fuente: Elaboración propia.

Además, dentro de los costos operativos a partir del año 2022 se han considerado los relacionado al software (ver tabla 25), el costo laboral del especialista de modelamiento predictivo (ver tabla 29), asignando un presupuesto para las capacitaciones técnicas del equipo de planeamiento de la demanda (ver tabla 30).

Tabla 29

Costo laboral de un especialista de modelamiento predictivo

Concepto	Costo mensual	Costo anual
	Soles	Soles
Remuneración	9,000	108,000
Sobrecosto laboral	4,440	53,276
Total		161,276

Fuente: Elaboración propia.

Tabla 30

Presupuesto para capacitaciones técnicas anuales

Programa	Costo total anual	
	Dólares	Soles
Cursos/Programas de modelamiento predictivo x 8	16,000	57,600
Congresos/Workshops x 8	4,000	14,400
Total	20,000	72,000

Fuente: Elaboración propia.

CAPÍTULO III. RESULTADOS

3.1. Estado actual del enfoque metodológico del planeamiento de la demanda

Del diagnóstico realizado en el capítulo II, se determina que los elementos de mayor énfasis técnico del proceso se ubican en una fase de transición e incipiente, mostrando por lo tanto un retraso en su desarrollo (ver figura 20).



Figura 20. Evaluación del enfoque metodológico del planeamiento de la demanda.

Fuente: Elaboración propia.

3.2. Mejora en el desempeño del planeamiento de la demanda

De la propuesta planteada, según los resultados de la validación cruzada, el nuevo pronóstico estadístico semanal alcanzaría el 81% de exactitud en su versión conciliada con la estimación mensual, con una mejora de 9% versus el estadístico actual y el plan operativo (ver figura 21). Y si a este resultado consideramos la funcionalidad de *demand sensing*, la exactitud del pronóstico estadístico se incrementaría al 87%, sobrepasando en 16% a los resultados actuales (ver figura 22).

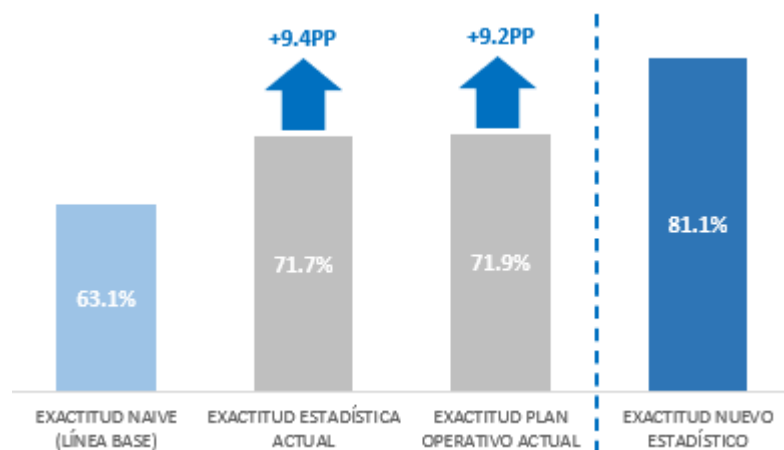


Figura 21. Mejora en la exactitud del nuevo pronóstico estadístico semanal.

Fuente: Elaboración propia.

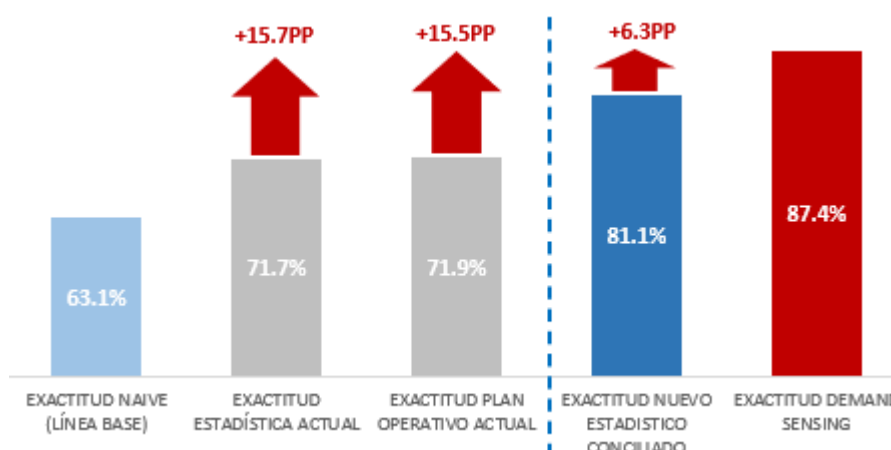


Figura 22. Mejora en la exactitud del nuevo pronóstico semanal por *demand sensing*.

Fuente: Elaboración propia.

Además, en las cuatro principales categorías de la empresa (99.9% del volumen), se podría alcanzar una exactitud del pronóstico superior al 85%, con la ayuda del *demand sensing* (ver tabla 31).

Tabla 31

Mejora en la exactitud del pronóstico semanal por demand sensing

Categoría	Volumen %	Exactitud Operativa Actual %	Exactitud Demand Sensing %	Mejora de Exactitud %
Carbonatadas	76%	74%	87%	13%
Agua embotellada	20%	69%	89%	20%
Jugos	3%	61%	89%	28%
Bebidas deportivas	1%	71%	86%	15%
Bebidas energéticas	0.04%	15%	69%	54%
Té	0.03%	18%	66%	48%

Fuente: Elaboración propia.

Ahora bien, a la par de las mejoras en exactitud de pronóstico, la implementación del nuevo modelamiento permitiría generar pronósticos con una señal de rastreo dentro de los límites de control (ver figura 23).

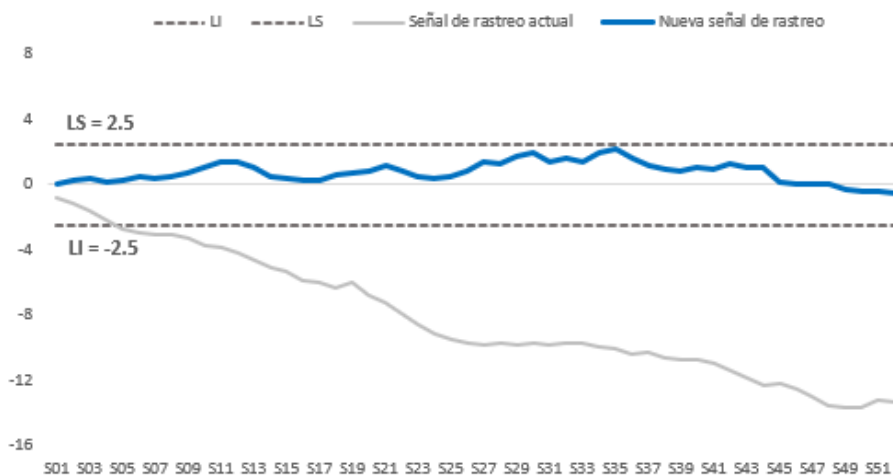


Figura 23. Mejora en la señal de rastreo del nuevo pronóstico estadístico semanal.

Fuente: Elaboración propia.

Y a su vez, el programa de capacitación propuesto involucrará al 100% de los miembros del equipo de planeamiento de la demanda (a los actuales siete miembros

del equipo más el especialista de modelamiento predictivo). De esta forma, lograremos que los indicadores ahora desviados (tabla 19), alcancen sus metas (tabla 32).

Tabla 32

Mejora propuesta de la matriz de indicadores

Causa raíz	Descripción	Indicador	Valor propuesto	Valor meta	Estado
CR-2	Deficiente estructura jerárquica de pronóstico	Señal de rastreo	[-0.5, 2.1]	[-2.5, 2.5]	En meta
CR-1	Deficiente modelamiento de pronóstico	% Exactitud estadística semanal	[81.1%, 87.4%]	78%	En meta
CR-4	Falta de capacitación/entrenamiento técnico	% de trabajadores capacitados	100%	100%	En meta

Fuente: Elaboración propia.

Finalmente, todas las mejoras descritas colocarían al planeamiento de la demanda de la empresa en niveles de excelencia y desarrollo en los componentes del enfoque metodológico rezagados en la actualidad (ver figura 24).

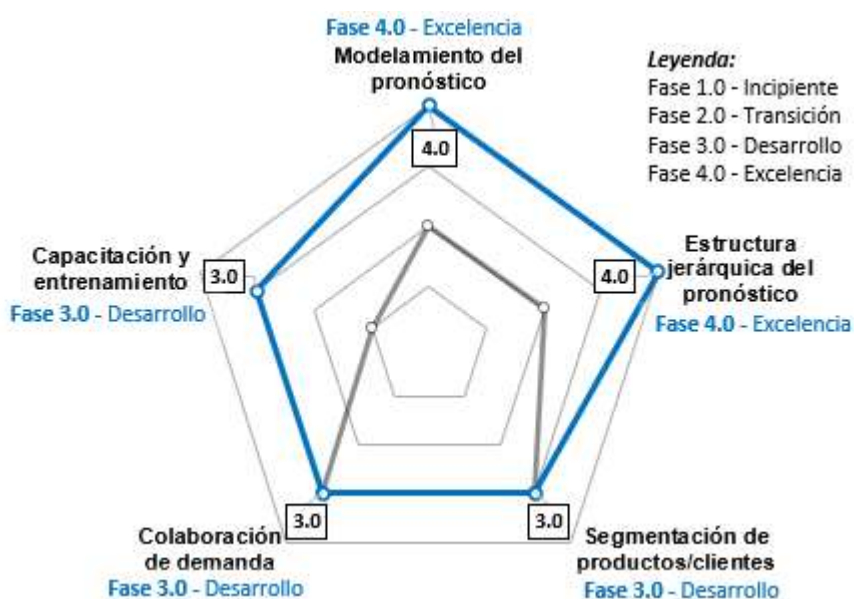


Figura 24. Prospectiva del enfoque metodológico del planeamiento de la demanda.

Fuente: Elaboración propia.

3.3. Impacto en el sistema logístico

La importancia real de las mejoras en la exactitud y sesgo del pronóstico radica en su impacto en los niveles de servicio e inventario de la empresa. De esta manera, considerando las ecuaciones 6 y 7, podemos indicar que, con esta propuesta de mejora, se alcanzaría entre 96.3% y 96.8% de nivel de servicio (ver figura 25), representando un incremento de venta de 3 a 4.3 millones de cajas unitarias en el 2019, con una generación de utilidades adicionales de entre 7.5 y 10.8 millones de soles (ver tabla 33).

Tabla 33

Impacto de la propuesta en el desabasto y en las utilidades del año 2019

Parámetros	Escenario Actual	Escenario Estadístico Conciliado	Escenario Demand Sensing
Exactitud de pronóstico [%]	71.9%	81.1%	87.4%
Bias [%]	7.2%	-2.2%	-0.9%
Rechazo_NC [%]	2.6%	2.6%	2.6%
Nivel de servicio [%]	95.3%	96.3%	96.8%
Desabasto [%]	2.1%	1.0%	0.6%
Pedidos [cu]	295,741,705	295,741,705	295,741,705
Desabasto [cu]	6,069,125	3,060,255	1,739,126
Abasto recuperado [cu]		3,008,870	4,329,999
Margen unitario [soles/cu]	2.5	2.5	2.5
Margen perdido por desabasto [soles]	15,182,235	7,656,773	4,351,301
Utilidad recuperada [soles]		7,525,461	10,830,934

Fuente: Elaboración propia.

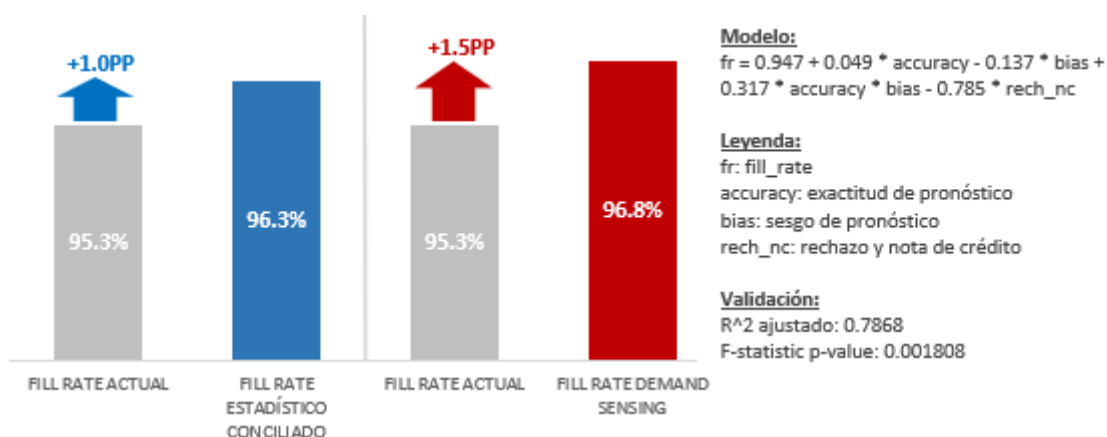


Figura 25. Mejora en el nivel de servicio por los nuevos pronósticos estadísticos.

Fuente: Elaboración propia.

Por otra parte, con estas mejoras los niveles de inventario de la empresa se reducirían entre 11.6% y 13.7% (ver figura 26), lo que implica un ahorro en el costo de mantener inventario de 9.7 a 11.4 millones de soles respecto al escenario actual (tabla 34).

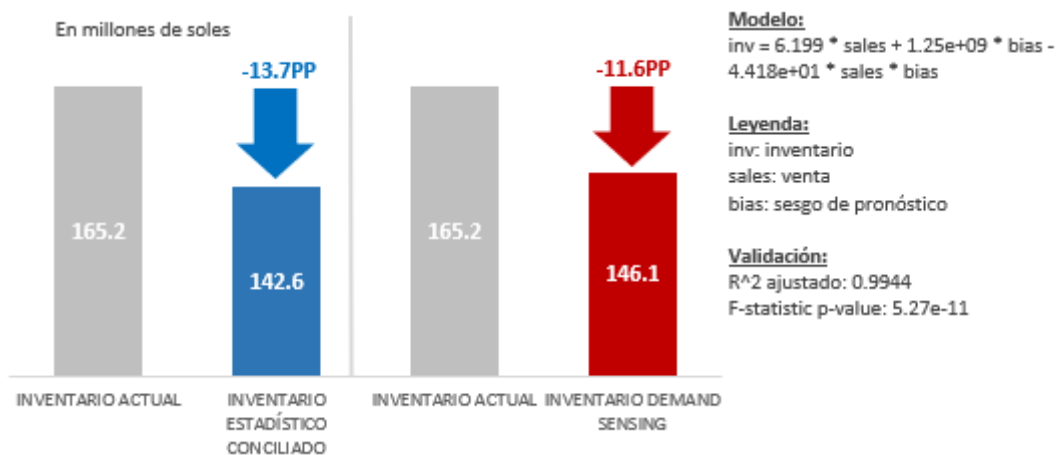


Figura 26. Mejora en los niveles de inventario por los nuevos pronósticos estadísticos.

Fuente: Elaboración propia.

Tabla 34

Impacto de la propuesta en el costo de mantener inventario en el año 2019

Parámetros	Escenario Actual	Escenario Estadístico Conciliado	Escenario Demand Sensing
Venta anual [cu]	281,871,419	284,880,289	286,201,418
Bias [%]	7.2%	-2.2%	-0.9%
Inventario PTER+MP [soles]	165,230,106	142,621,944	146,059,397
Costo de mantener inventario [soles/cu]	0.30	0.25	0.26
Costo de mantener inventario [soles]	83,599,459	72,160,683	73,898,729
Ahorro en costo de mantener inventario [soles]		11,438,776	9,700,730

Fuente: Elaboración propia.

Adicionalmente, con estas mejoras el costo de transporte de la empresa mostraría ahorros de 279 mil a 1.1 millones de soles (ver tabla 35), esto por la reducción de inventarios, a pesar del aumento del costo de distribución debido al incremento de venta ante un mejor nivel de servicio.

Tabla 35

Impacto de la propuesta en el costo de transporte del año 2019

Parámetros	Escenario Actual	Escenario Estadístico Conciliado	Escenario Demand Sensing
Venta anual [cu]	281,871,419	284,880,289	286,201,418
Inventario PTER+MP [soles]	165,230,106	142,621,944	146,059,397
Costo de transporte [soles/cu]	1.09	1.08	1.07
Costo de transporte [soles]	307,521,718	306,414,472	307,242,421
Ahorro en costo de transporte [soles]		1,107,246	279,297

Fuente: Elaboración propia.

Además, estos beneficios por recuperación del margen perdido, por ahorro de mantener inventario y por transporte, conducen a un ahorro del costo logístico de entre 20 y 20.8 millones de soles, incrementando las utilidades antes de impuestos entre 8% y 8.3% sobre el resultado del año 2019 (ver tabla 36).

Tabla 36

Impacto de la propuesta en los costos logísticos y en las utilidades del 2019

Parámetros	Escenario Actual	Escenario Estadístico Conciliado	Escenario Demand Sensing
Margen perdido por desabasto [soles]	15,182,235	7,656,773	4,351,301
Costo de mantener inventario [soles]	83,599,459	72,160,683	73,883,406
Costo de transporte [soles]	307,521,718	306,414,472	307,242,421
Costos logísticos relevantes [soles]	406,303,411	386,231,928	385,477,128
Ahorro del costo logístico [soles]		20,071,483	20,826,284
Utilidades antes de impuestos [soles]	249,664,000	269,735,483	270,490,284
Aumento de utilidades antes de impuestos [%]		8.0%	8.3%

Fuente: Elaboración propia.

De esta forma, la cadena de suministro de la empresa con un costo real de 1.75 soles por caja unitaria en el año 2019, hubiera alcanzado un costo entre 1.65 y 1.66 soles por caja unitaria con las mejoras planteadas en la gestión de la demanda (ver figura 27).

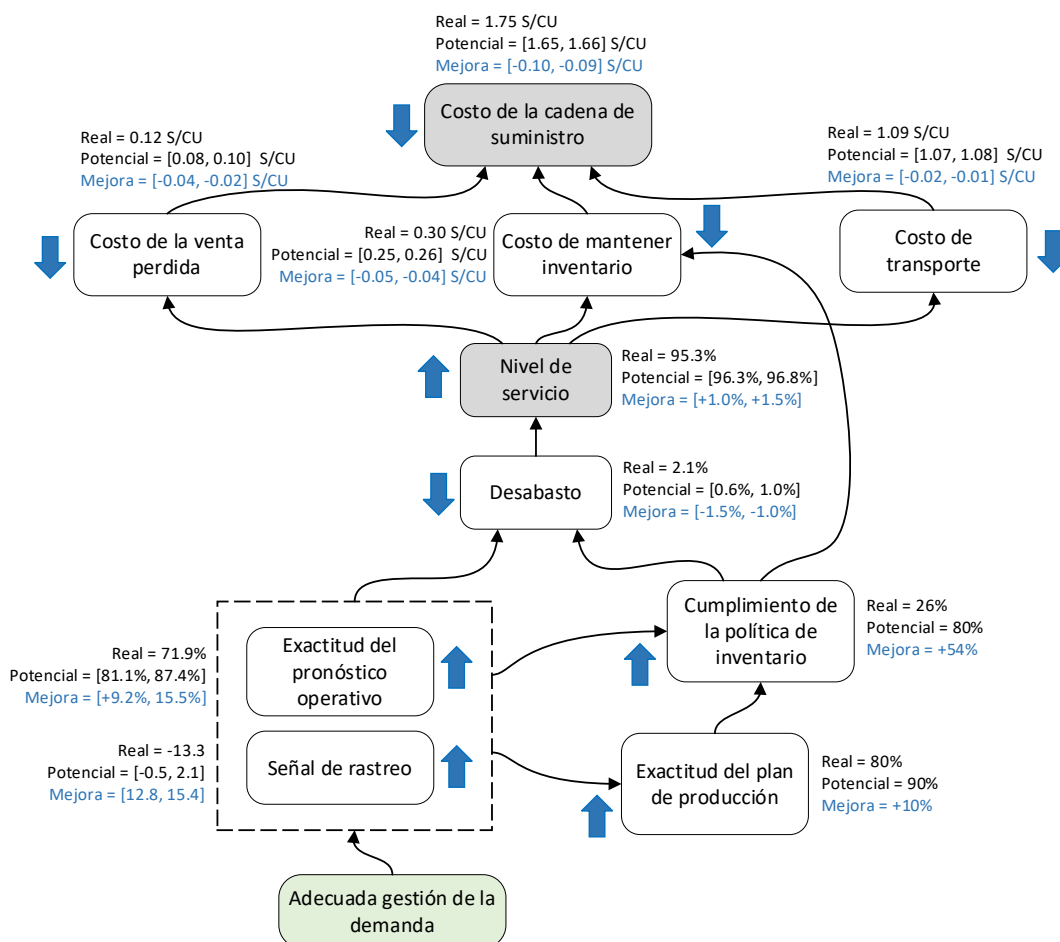


Figura 27. Realidad futura del sistema logístico. Fuente: Elaboración propia.

3.4. Flujo de caja económico

Para la elaboración del flujo de caja del proyecto (ver tabla 37) se ha considerado como datos de entrada las proyecciones de venta para el periodo 2022 – 2025, el costo de oportunidad de capital de 12% y un escenario conservador con costo logístico mejorado de 1.66 soles/caja unitaria; obteniendo de esta forma un valor presente neto del proyecto (VPN) de 76.69 millones de soles, a un periodo de recuperación de la inversión (PR) de 0.2 meses, con una relación beneficio/costo (B/C) de 55.4. Esto implica que el proyecto de mejora del planeamiento de la demanda sí es viable.

Tabla 37

Flujo de caja económico de la propuesta de mejora

Año	2021	2022	2023	2024	2025
A. Inversión [soles]	-462,694	0	0	0	0
A.1. Software [soles]	-78,149				
A.2. Hardware [soles]	-61,494				
A.3. Contratación especialista de modelamiento [soles]	-168,926				
A.4. Capacitación y certificación en modelamiento [soles]	-154,124				
B. Costos Operativos [soles]	0	-311,425	-311,425	-311,425	-311,425
B.1. Renovación software [soles]		-78,149	-78,149	-78,149	-78,149
B.2. Costo laboral especialista de modelamiento [soles]		-161,276	-161,276	-161,276	-161,276
B.3. Capacitaciones técnicas [soles]		-72,000	-72,000	-72,000	-72,000
C. Beneficios [soles]		24,480,592	25,347,241	26,249,628	27,294,967
C.1. Mejora del costo logístico [soles]		24,480,592	25,347,241	26,249,628	27,294,967
C.1.1. Proyecciones de venta [cajas unitarias]		274,000,000	283,700,000	293,800,000	305,500,000
C.1.2. Costo logístico propuesto [soles/caja unitaria]		1.66	1.66	1.66	1.66
C.1.3. Costo logístico actual [soles/caja unitaria]		1.75	1.75	1.75	1.75
D. Flujo de Caja Económico [soles]	-462,694	24,169,167	25,035,816	25,938,203	26,983,542

$$D = A + B + C$$

$$A = A.1 + A.2 + A.3 + A.4$$

$$B = B.1 + B.2 + B.3$$

$$C = C.1$$

$$C.1 = C.1.1 * (C.1.3 - C.1.2)$$

Costo de oportunidad de capital

12%

Valor Presente Neto (VPN)

76,686,148

 soles

Periodo de Recuperación de la Inversión (PR)

0.2

 meses

Ratio Beneficio/Costo (B/C)

55.4

Fuente: Elaboración propia.

3.5. Análisis de sensibilidad

Dentro del análisis de la viabilidad de la propuesta, es importante determinar la sensibilidad del resultado mediante la simulación de escenarios a producirse en la venta futura y en el costo logístico, de tal forma que pueda evaluarse las condiciones necesarias para que el proyecto pierda su viabilidad. De esta manera, según la figura 28, se observa que, para descartar el proyecto, el volumen de venta en el año 2022 debería reducirse en un 98% (de 274 a 4.9 millones de cajas unitarias), o debería perderse el 90% de la mejora en costo (de 1.66 a 1.74 soles/caja unitaria) y a su vez un 85% del volumen de venta estimado (de 274 a 40 millones de cajas unitarias).

	VPN	Costo logístico [soles/caja unitaria]									
	76,686,148	1.66	1.67	1.68	1.69	1.70	1.71	1.72	1.73	1.74	1.75
274,000,000	76,686,148	67,945,362	59,204,577	50,463,791	41,723,005	32,982,219	24,241,434	15,500,648	6,759,862	-1,980,923	
262,300,000	73,351,445	64,983,897	56,616,350	48,248,802	39,881,254	31,513,706	23,146,159	14,778,611	6,411,063	-1,956,485	
250,600,000	70,016,742	62,022,433	54,028,123	46,033,813	38,039,503	30,045,193	22,050,883	14,056,574	6,062,264	-1,932,046	
238,900,000	66,682,040	59,060,968	51,439,896	43,818,824	36,197,752	28,576,680	20,955,608	13,334,536	5,713,464	-1,907,608	
227,200,000	63,347,337	56,099,503	48,851,669	41,603,835	34,356,001	27,108,167	19,860,333	12,612,499	5,364,665	-1,883,169	
215,500,000	60,012,634	53,138,038	46,263,442	39,388,846	32,514,250	25,639,654	18,765,058	11,890,462	5,015,866	-1,858,730	
203,800,000	56,677,931	50,176,573	43,675,215	37,173,857	30,672,499	24,171,141	17,669,783	11,168,424	4,667,066	-1,834,292	
192,100,000	53,343,229	47,215,108	41,086,988	34,958,868	28,830,748	22,702,628	16,574,507	10,446,387	4,318,267	-1,809,853	
180,400,000	50,008,526	44,253,644	38,498,761	32,743,879	26,988,997	21,234,114	15,479,232	9,724,350	3,969,468	-1,785,415	
168,700,000	46,673,823	41,292,179	35,910,534	30,528,890	25,147,246	19,765,601	14,383,957	9,002,313	3,620,668	-1,760,976	
157,000,000	43,339,120	38,330,714	33,322,307	28,313,901	23,305,495	18,297,088	13,288,682	8,280,275	3,271,869	-1,736,537	
145,300,000	40,004,417	35,369,249	30,734,080	26,098,912	21,463,744	16,828,575	12,193,407	7,558,238	2,923,070	-1,712,099	
133,600,000	36,669,715	32,407,784	28,145,854	23,883,923	19,621,992	15,360,062	11,098,131	6,836,201	2,574,270	-1,687,660	
121,900,000	33,335,012	29,446,319	25,557,627	21,668,934	17,780,241	13,891,549	10,002,856	6,114,164	2,225,471	-1,663,222	
110,200,000	30,000,309	26,484,854	22,969,400	19,453,945	15,938,490	12,423,036	8,907,581	5,392,126	1,876,672	-1,638,783	
98,500,000	26,665,606	23,523,390	20,381,173	17,238,956	14,096,739	10,954,523	7,812,306	4,670,089	1,527,872	-1,614,344	
86,800,000	23,330,904	20,561,925	17,792,946	15,023,967	12,254,988	9,486,009	6,717,031	3,948,052	1,179,073	-1,589,906	
75,100,000	19,996,201	17,600,460	15,204,719	12,808,978	10,413,237	8,017,496	5,621,755	3,226,015	830,274	-1,565,467	
63,400,000	16,661,498	14,638,995	12,616,492	10,593,989	8,571,486	6,548,983	4,526,480	2,503,977	481,474	-1,541,029	
51,700,000	13,326,795	11,677,530	10,028,265	8,379,000	6,729,735	5,080,470	3,431,205	1,781,940	132,675	-1,516,590	
40,000,000	9,992,093	8,716,065	7,440,038	6,164,011	4,887,984	3,611,957	2,335,930	1,059,903	-216,124	-1,492,151	
28,300,000	6,657,390	5,754,601	4,851,811	3,949,022	3,046,233	2,143,444	1,240,655	337,865	-564,924	-1,467,713	
16,600,000	3,322,687	2,793,136	2,263,584	1,734,033	1,204,482	674,931	145,379	-384,172	-913,723	-1,443,274	
4,900,000	-12,016	-168,329	-324,642	-480,956	-637,269	-793,582	-949,896	-1,106,209	-1,262,522	-1,418,836	
0	-1,408,601	-1,408,601	-1,408,601	-1,408,601	-1,408,601	-1,408,601	-1,408,601	-1,408,601	-1,408,601	-1,408,601	

Figura 28. Sensibilidad del VPN de la propuesta. Fuente: Elaboración propia.

CAPÍTULO IV. DISCUSIÓN Y CONCLUSIONES

4.1 Discusión

El estado actual del planeamiento de la demanda evidencia una desatención de elementos centrales en cuanto al modelamiento de un pronóstico y a la formación requerida por el equipo humano a cargo de su generación, y tal como evidencia Ma (2020), se requiere de una integración adecuada entre las capacidades predictivas automáticas y el criterio humano informado, para lograr los mejores resultados en esta especialidad, explicando el porqué las debilidades presentes en el proceso generan una exactitud y sesgo del pronóstico fuera de los niveles requeridos por la empresa, conduciendo a un desempeño subóptimo de la cadena de suministros.

La propuesta de mejora en el planeamiento de la demanda nos muestra que sí es posible, a través de un modelamiento adecuado, el logro de niveles de exactitud y de sesgo de pronóstico según los objetivos establecidos, llegando incluso a superar lo esperado. Implementaciones como la generación de pronósticos multinivel, basado en el desarrollo de Athanasopoulos et al. (2007), la generación de pronósticos por *demand sensing* empleando técnicas como ARIMAX y pronósticos combinados, probados como efectivos en la práctica por Delf (2018), y las automatizaciones del pronóstico en el lenguaje de programación R desarrolladas por Hyndman y Khandakar (2008), necesarias para procesar los más de 12 millones de registros de venta involucrados en el desarrollo de esta propuesta, forman parte de un conjunto de prácticas consideradas en la actualidad como estándar, implicando una adopción natural en la empresa de contarse con el capital humano adecuado y los recursos tecnológicos requeridos para soportar este desarrollo.

Frente al alcance técnico desplegado en esta propuesta, trabajos como el de Chávez et al. (2018), carecen de sustento, al no proporcionar detalles sobre cómo se mejora en 10% el error de pronóstico, que luego se emplea como dato de entrada para obtener la mejora en el costo de mantener inventario. En Pérez et al. (2016), si bien se realiza una validación cruzada, se generan pronósticos a alto nivel de agregación, no a nivel SKU, por lo que no resultan pertinentes para la toma de decisiones en la cadena de suministros a nivel operativo, además de emplear solo un tipo de modelo. Caso similar se encuentra en Campos (2017), al considerar un modelamiento agregado a pesar de contar con 1,577 SKUS en el portafolio de productos, sin realizar validación cruzada alguna, empleando solo la suavización exponencial, sin considerar variables como precios y descuentos. En Paredes (2017) encontramos pronósticos generados por cada SKU, pero sin realizar validación cruzada, y dependiendo de un solo tipo de modelo. De estos antecedentes se desprende que, la relevancia del presente trabajo, es la de sentar las bases que permitan la adopción de las mejores prácticas en la generación de pronósticos para la cadena de suministros.

El impacto de la propuesta en el sistema logístico evidencia a su vez los beneficios de la mejora de la exactitud y del sesgo del pronóstico en indicadores clave como el nivel de servicio y el nivel de inventario, generando resultados positivos en cuanto a recuperación de margen perdido, costo de mantener inventario y de transporte, logrando un ahorro logístico de entre 20 y 20.8 millones de soles.

El flujo de caja económico nos muestra que la propuesta cuenta con un elevado nivel de VPN a un periodo de recuperación inferior a 1 mes, evidenciando el potencial que el modelamiento predictivo puede traer a las organizaciones siempre que se disponga

de los recursos y la formación adecuada del equipo humano, con el apoyo de la alta gerencia.

El análisis de sensibilidad evidencia que la decisión de realizar el proyecto es robusta ya que la empresa tendría que ver reducido su volumen de venta del año 2022 en 85% y la mejora del costo verse afectada en un 90%, para que de esta forma la decisión no resultará rentable, y este escenario extremo, como tal, no es factible si consideramos que, en el año 2020 con la pandemia del COVID-19, la venta anual solo sería un 13% menor a los 274 millones de cajas unitarias consideradas para el año 2022.

4.2 Conclusiones

- El enfoque metodológico actual del planeamiento de la demanda presenta debilidades en sus elementos de carácter técnico a nivel de modelamiento y estructura jerárquica del pronóstico, siendo fundamental en esta situación la escasa formación del equipo humano responsable de esta actividad. De esta forma se explica el porqué se presentan en la actualidad exactitudes de pronóstico que no superan el 72% a nivel operativo, encontrándose el proceso alejado del 80% requerido, sumado a un sesgo de pronóstico fuera de control por una recurrente sobreestimación, lo que conduce a mantener elevados niveles de inventario.
- La propuesta de mejora consiste en la implementación de una estructura jerárquica que permita la generación de pronósticos multinivel mediante la aplicación adecuada de los enfoques “*top-down* basado en proporciones de pronósticos” y del *bottom-up*, según la especificación planteada, para lo cual se requiere establecer a la caja unitaria como la unidad de medida para el modelamiento. A

su vez, se está incorporando nuevas técnicas como la regresión lineal múltiple, las redes neuronales, ARIMA y sus variantes, así como los pronósticos combinados, en un esquema donde también contaremos con el *demand sensing* para actualizar los pronósticos de una manera más frecuente y a un nivel de mayor granularidad, atendiendo las necesidades del nivel operativo, bajo el soporte del lenguaje de programación R para su desarrollo. Esto, por lo tanto, hace necesaria la incorporación al equipo de un especialista de modelamiento predictivo, además de la capacitación en técnicas de modelamiento a todo el grupo en general, para que la implementación y sus mejoras puedan mantenerse en el tiempo

- Los beneficios de la propuesta a nivel del proceso se evidencian en una mejora que llevaría a la exactitud operativa a entre 81.1% y 87.4%, con una señal de rastreo dentro de los límites de control especificados. De esta manera, el nivel de servicio mejoraría del actual 95.3% a entre 96.3% y 96.8%; y en cuanto al nivel de inventario, este se reduciría entre 11.6% y 13.7%, manteniéndose entre los 142 y 146 millones de soles.
- La reducción del costo logístico, derivado de las mejoras en nivel de servicio y la reducción de los niveles de inventario, alcanzaría entre 20 y 20.8 millones de soles para el año 2019, equivalente a un aumento de la utilidad neta antes de impuestos de entre 8% y 8.3%.
- La propuesta de mejora implica una inversión de 462.7 miles de soles, mostrando un VPN de 76.69 millones de soles, considerando un periodo de 4 años, con un PR de 0.2 meses y una relación B/C de 55.4, por lo que esta propuesta cuenta con la viabilidad necesaria, resultando una decisión robusta ya que, según el análisis de sensibilidad, se necesitaría de escenarios extremos para la pérdida de viabilidad.

REFERENCIAS

- Athanasopoulos, G., Ahmed, R., y Hyndman R.J. (2007). *Hierarchical forecasts for Australian domestic tourism*. Melbourne, Australia: Universidad de Monash. Recuperado de <https://www.monash.edu/business/ebs/research/publications/ebs/wp12-07.pdf>.
- Banco Central de Reserva del Perú (2020). *Reporte de inflación junio 2020*. Recuperado de <https://www.bcrp.gob.pe/docs/Publicaciones/Reporte-Inflacion/2020/junio/reportede-inflacion-junio-2020.pdf>.
- Breiman, L. (2001). Statistical modeling: the two cultures. *Statistical Science*, 16(3), 199-231. Recuperado de <https://projecteuclid.org/euclid.ss/1009213726>
- Campos, J. (2017). *Implementación de un sistema logístico para reducir costos de inventarios en la Botica La Luz S.A.C.* (tesis de pregrado). Universidad Privada del Norte, Trujillo, Perú.
- Chase, C. (2013). *Demand-driven forecasting: a structured approach to forecasting*. Nueva Jersey, Estados Unidos: John Wiley & Sons, Inc.
- Chávez, K., De La Cruz, L., & Rodas, S. (2018). *Propuesta de mejora del rendimiento sobre la inversión en inventarios para una empresa de bebidas no alcohólicas* (tesis de maestría). Universidad del Pacífico, Lima, Perú.

Decreto Supremo N° 091-2018-EF. Diario Oficial El Peruano, Lima, Perú, 9 de mayo de 2018.

Delfs, M. (2018). *Forecasting in the supply chain with machine learning techniques*. (tesis de maestría). Otto-Friedrich-Universität Bamberg, Bamberg, Alemania.

Euromonitor International (2020). *Market statistics about soft drinks and retail* [Data set]. Passport. Recuperado de <https://www.euromonitor.com>.

Frazelle, E. (2002). *Supply chain strategy: the logistics of supply chain management*. Nueva York, Estados Unidos: McGraw-Hill.

Gilliland, M. (2010). *The business forecasting deal: exposing myths, eliminating bad practices, providing practical solutions*. Nueva Jersey, Estados Unidos: John Wiley & Sons, Inc.

Goldratt, E. (1990). *Theory of constraints*. Great Barrington, Estados Unidos: The North River Press.

Hernández, R., Fernández, C., y Baptista M. (2014). *Metodología de la investigación*. México D.F., México: McGraw-Hill.

Hyndman, R.J., y Athanasopoulos, G. (2018). *Forecasting: principles and practice*. Melbourne, Australia: OText.

Hyndman, R.J., y Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistiscal Software*, 27(3), 1-22.

Ley N° 30021. Diario Oficial El Peruano, Lima, Perú, 17 de mayo de 2013.

Ma, Y. (2020). *Human-machine teaming for intelligent demand-planning*. (tesis de maestría). Massachusetts Institute of Technology, Cambridge, Estados Unidos de América.

Mentzer, J., y Moon, M. (2005). *Sales forecasting management: a demand management approach*. California, Estados Unidos: Sage Publications.

Paredes, J. (2017). *Gestión de inventarios y su influencia en los costos logísticos en la empresa Maderas del Norte S.R.L.* (tesis de pregrado). Universidad Privada del Norte, Trujillo, Perú.

Pérez, W., Montalvo, M., & Carruitero, W. (2016). *Rediseño del modelo de planificación y gestión de inventarios de productos terminados en una empresa de colchones* (tesis de maestría). Universidad del Pacífico, Lima, Perú.

Sankaran, G., Sasso, F., Kepczynski, R., y Chiaraviglio, A. (2019). *Improving forecasts with Integrated Business Planning: from short-term to long-term demand planning enabled by SAP IBP*. Cham, Suiza: Springer Nature Switzerland AG.

Shmueli, G., y Lichtendahl, K. (2016). *Practical time series forecasting with R: a hands-on guide*. Green Cove Springs, Estados Unidos: Axelrod Schnall Publishers.

ANEXOS

Anexo N° 1: Código en T-SQL para la extracción de datos: nivel D6

```

1. SELECT 1.Region3 as region3,
2.         1.Provincia as provincia,
3.         1.Locacion2 as locacion,
4.         c.[Canal Z] as canal,
5.         p.Categoria as categoria,
6.         p.Trademark as trademark,
7.         p.Retornabilidad as retornabilidad,
8.         p.Tamaño as tamaño,
9.         p.[Segmento Detallado] as segmento_detallado,
10.        p.Sabor as sabor,
11.        p.Empaque as empaque,
12.        p.Unidades as unidades,
13.        p.[Tipo Envase] as tipo_envase,
14.        p.[Codigo Producto] as cod_producto,
15.        p.SKU as producto,
16.        m.Año1 as año1,
17.        m.Mes1 as mes1,
18.        m.Año1Mes1 as año1mes1,
19.        m.Año2Mes2 as año2mes2,
20.        m.Año2Semana2 as año2semana2,
21.        m.DiaSemana as diasemana,
22.        m.DiaAño as dia_año,
23.        m.DiaMes as diames,
24.        m.[Tiempo - Dia natural] as fecha,
25.        Sum(emb.[Volumen CU Mercado]) as [volumen_cu_emb],
26.        Sum(emb.[Ingreso Bruto]) as [ingreso_bruto],
27.        Sum(emb.[Descuento Comercial Mercado]) as [descuento_comercial_mercado],
28.        Sum(emb.[Descuento Bonificaciones Mercado]) as [descuento_bonificaciones_mercado],
29.        ISNULL(Sum(emb.[Descuento Intro Envase Mercado]),0) as
    [descuento_intro_envase_mercado],
30.        Sum(emb.[Ingreso Neto Comercial]) as [ingreso_netocomercial]
31.
32. FROM    Transacciones emb
33. LEFT JOIN [EMB].[dbo].[Tiempo] m ON emb.[Tiempo - Dia natural] = m.[Tiempo - Dia natural]
34. LEFT JOIN [EMB].[dbo].[Productos] p ON emb.[Product SKU SAP] = p.[Product SKU SAP]
35. LEFT JOIN [EMB].[dbo].[Locaciones] l ON emb.[Cliente - Locacion] = l.[Cliente - Locacion]
36. INNER JOIN [EMB].[dbo].[Canales] c ON emb.[Canal 2 Canal PIVO] = c.[Canal 2 Canal PIVO]
37.
38. WHERE   c.Mercado = 'NACIONAL' AND
39.         p.Categoria != 'PACKAGED FOODS DAIRY' AND p.Categoria != 'RTD COFFEE' AND
40.         p.Categoria != 'ASIAN SPECIALITY DRINKS' AND p.Categoria != 'PACKAGED FOOD' AND
41.         m.Año1Mes1 >= 201701 AND m.Año1Mes1 <= 201903
42.
43. GROUP BY 1.Region3,
44.          1.Provincia,
45.          1.Locacion2,
46.          c.[Canal Z],
47.          p.Categoria,
48.          p.Trademark,
49.          p.Retornabilidad,
50.          p.Tamaño,
51.          p.[Segmento Detallado],
52.          p.Sabor,
53.          p.Empaque,
54.          p.Unidades,
55.          p.[Tipo Envase],
56.          p.[Codigo Producto],
57.          p.SKU,
58.          m.Año1,
59.          m.Mes1,
60.          m.Año1Mes1,
61.          m.Año2Mes2,
62.          m.Año2Semana2,
63.          m.DiaSemana,
64.          m.DiaAño,
65.          m.DiaMes,

```

```

66.         m.Año1Mes1Dia,
67.         m.[Tiempo - Dia natural]
68.
69. ORDER BY  l.Region3,
70.           l.Provincia,
71.           l.Locacion2,
72.           c.[Canal Z],
73.           p.Categoria,
74.           p.Trademark,
75.           p.Retornabilidad,
76.           p.Tamaño,
77.           p.[Segmento Detallado],
78.           p.Sabor,
79.           p.Empaque,
80.           p.Unidades,
81.           p.[Tipo Envase],
82.           p.[Codigo Producto],
83.           p.SKU,
84.           m.Año1 ASC,
85.           m.Mes1 ASC,
86.           m.Año2Semana2 ASC,
87.           m.DiaAño ASC

```

Anexo N° 2: Código en R para la preparación de datos: nivel D6

```

1.  ## PREPARACIÓN DE DATOS NIVEL 06 - DIARIO: SKU / CD / CANAL ##
2.
3.  #I. Carga de datos----
4.  data <- read.csv("D:/Daily_Level06_EMB.csv", encoding = "UTF-8")
5.  status_l06 <- read.csv("D:/Combo_Status.csv") #mensual
6.  status_temp <- read.csv("D:/Status_Temp.csv")
7.  borrar_l00 <- read.csv("D:/BorrarVenta100%.csv")
8.  borrar_prov <- read.csv("D:/BorrarVentaParcial.csv")
9.
10.
11. #II. Carga de librerías----
12. library(dplyr)
13.
14. #III. Preprocesamiento----
15. # Limpieza de datos
16. data$volumen_cu_emb[data$fecha %in% borrar_l00$fecha] <- 0
17. data$ingreso_bruto[data$fecha %in% borrar_l00$fecha] <- 0
18. data$descuento_comercial_mercado[data$fecha %in% borrar_l00$fecha] <- 0
19. data$descuento_bonificaciones_mercado[data$fecha %in% borrar_l00$fecha] <- 0
20. data$descuento_intro_envase_mercado[data$fecha %in% borrar_l00$fecha] <- 0
21. data$ingreso_netocomercial[data$fecha %in% borrar_l00$fecha] <- 0
22. data$volumen_cu_emb[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
23. data$ingreso_bruto[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
24. data$descuento_comercial_mercado[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
25. data$descuento_bonificaciones_mercado[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
26. data$descuento_intro_envase_mercado[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
27. data$ingreso_netocomercial[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
28.
29. data <- subset(data, volumen_cu_emb != 0)
30. colnames(data)[1] <- "region3"
31. data$fecha <- as.Date(data$fecha, format = "%d/%m/%Y")
32.
33. data$id <- paste(data$region3, data$provincia, data$locacion, data$canal, data$categoria,
data$trademark, data$retornabilidad, data$tamaño, data$segmento_detallado, data$sabor,
data$empaque, data$unidades, data$tipo_envase, data$cod_producto, data$producto, sep = " ")
34.

```



```

35. # Preparación de datos para la evaluación de combinaciones
36. status <- aggregate(data[,23], by=list(region3 = data$region3, provincia = data$provincia,
    locacion = data$locacion, canal = data$canal, categoria = data$categoria, trademark =
    data$trademark, retornabilidad = data$retornabilidad, tamaño = data$tamaño,
    segmento_detallado = data$segmento_detallado, sabor = data$sabor, empaque = data$empaque,
    unidades = data$unidades, tipo_envase = data$tipo_envase, cod_producto = data$cod_producto,
    producto = data$producto), FUN=sum, na.rm=TRUE)
37.
38. status$x <- NULL
39.
40. status <- status[order(status$region3, status$provincia, status$locacion, status$canal,
    status$categoria, status$trademark, status$retornabilidad, status$tamaño,
    status$segmento_detallado, status$sabor, status$empaque, status$unidades,
    status$tipo_envase, status$cod_producto, status$producto),]
41.
42. rownames(status) <- 1:nrow(status)
43.
44. status_l06 <- status_l06[, -c(16:58)]
45.
46. status <- left_join(status, status_l06, by = c("region3", "provincia", "locacion", "canal",
    "categoria", "trademark", "retornabilidad", "tamaño", "segmento_detallado", "sabor",
    "empaque", "unidades", "tipo_envase", "cod_producto", "producto"))
47.
48. # Definir valores iniciales de los clasificadores
49. status$activo_inactivo_l06[is.na(status$activo_inactivo_l06)] <- "INACTIVO"
50. status$fase_comb_l06[is.na(status$fase_comb_l06)] <- ""
51. status$tipo_modelo_l06[is.na(status$tipo_modelo_l06)] <- ""
52.
53. status$id <- paste(status$region3, status$provincia, status$locacion, status$canal,
    status$categoria, status$trademark, status$retornabilidad, status$tamaño,
    status$segmento_detallado, status$sabor, status$empaque, status$unidades,
    status$tipo_envase, status$cod_producto, status$producto, sep = " ")
54.
55. # Creación de variables adicionales
56. status$fecha.inicial <- as.Date("", format = "%d/%m/%Y")
57. status$fecha.final <- as.Date("", format = "%d/%m/%Y")
58. status$fecha.referencia <- as.Date("31/03/2019", format = "%d/%m/%Y")
59. status$var.fecha.fin <- 0
60. status$nper.real <- 0
61. status$nper.teo <- 0
62. status$var.nper <- 0
63. status$fecha.ini.2017 <- as.Date("", format = "%d/%m/%Y")
64. status$fecha.fin.2017 <- as.Date("", format = "%d/%m/%Y")
65. status$fecha.ini.2018 <- as.Date("", format = "%d/%m/%Y")
66. status$fecha.fin.2018 <- as.Date("", format = "%d/%m/%Y")
67. status$fecha.ini.2019 <- as.Date("", format = "%d/%m/%Y")
68. status$fecha.fin.2019 <- as.Date("", format = "%d/%m/%Y")
69.
70. status$nper.real2017 <- 0
71. status$nper.teo2017 <- 0
72. status$var.nper2017 <- 0
73. status$nper.real2018 <- 0
74. status$nper.teo2018 <- 0
75. status$var.nper2018 <- 0
76. status$nper.real2019 <- 0
77. status$nper.teo2019 <- 0
78. status$var.nper2019 <- 0
79. status$var.fecha.fin.2017 <- 0
80.
81. status$venta2018 <- 0
82. status$venta2019 <- 0
83. status$ing_bruto1718 <- 0
84. status$ing_bruto19 <- 0
85. status$desc_com1718 <- 0
86. status$desc_com19 <- 0
87. status$desc_bon1718 <- 0
88. status$desc_bon19 <- 0
89. status$desc_intro1718 <- 0
90. status$desc_intro19 <- 0
91.
92. status$fecha.ini.ult.seq <- as.Date("", format = "%d/%m/%Y")
93.

```

```

94. #IV. Clasificación de estado de las combinaciones----
95. for (i in 1:nrow(status)){
96.
97.   print(i)
98.   combo <- status$id[i]
99.
100.  x <- subset(data, id == combo)
101.
102.  minimo <- min(x$fecha)
103.  maximo <- max(x$fecha)
104.  venta2018 <- sum(subset(x, año1 == 2018)$volumen_cu_emb)
105.  venta2019 <- sum(subset(x, año1 == 2019)$volumen_cu_emb)
106.  ing_bruto1718 <- sum(subset(x, año1 > 2016 & año1 < 2019)$ingreso_bruto)
107.  ing_bruto19 <- sum(subset(x, año1 == 2019)$ingreso_bruto)
108.  desc_com1718 <- sum(subset(x, año1 > 2016 & año1 < 2019)$descuento_comercial_mercado)
109.  desc_com19 <- sum(subset(x, año1 == 2019)$descuento_comercial_mercado)
110.  desc_bon1718 <- sum(subset(x, año1 > 2016 & año1 <
2019)$descuento_bonificaciones_mercado)
111.  desc_bon19 <- sum(subset(x, año1 == 2019)$descuento_bonificaciones_mercado)
112.  desc_intro1718 <- sum(subset(x, año1 > 2016 & año1 <
2019)$descuento_intro_envases_mercado)
113.  desc_intro19 <- sum(subset(x, año1 == 2019)$descuento_intro_envases_mercado)
114.  nper.real <- nrow(x)
115.
116.  status[status$id == combo, 20] <- minimo
117.  status[status$id == combo, 21] <- maximo
118.  status[status$id == combo, 24] <- nper.real
119.  status[status$id == combo, 43] <- venta2018
120.  status[status$id == combo, 44] <- venta2019
121.  status[status$id == combo, 45] <- ing_bruto1718
122.  status[status$id == combo, 46] <- ing_bruto19
123.  status[status$id == combo, 47] <- desc_com1718
124.  status[status$id == combo, 48] <- desc_com19
125.  status[status$id == combo, 49] <- desc_bon1718
126.  status[status$id == combo, 50] <- desc_bon19
127.  status[status$id == combo, 51] <- desc_intro1718
128.  status[status$id == combo, 52] <- desc_intro19
129.
130.  y1 <- as.numeric(substr(minimo, start = 1, stop = 4))
131.  y2 <- as.numeric(substr(maximo, start = 1, stop = 4))
132.
133.  for (j in y1:y2){
134.
135.    nper <- nrow(subset(x, año1 == j))
136.
137.    if (nper == 0){
138.
139.      k <- j + 1
140.
141.      if (k<=y2){
142.
143.        min2 <- min(subset(x, año1 == k)$fecha)
144.
145.        status[status$id == combo, 20] <- min2
146.        status[status$id == combo, 53] <- min2
147.
148.        z <- subset(x, fecha >= min2)
149.        nper.ult.seq <- nrow(z)
150.
151.        status[status$id == combo, 24] <- nper.ult.seq
152.      }
153.    } else {
154.
155.      min1 <- min(subset(x, año1 == j)$fecha)
156.      max1 <- max(subset(x, año1 == j)$fecha)
157.
158.      n <- 27 + (j-2017)*2
159.      status[status$id == combo, n] <- min1
160.      n <- n + 1
161.      status[status$id == combo, n] <- max1
162.      m <- 33 + (j-2017)*3
163.

```

```

164. status[status$id == combo, m] <- nper
165. r <- 34 + (j-2017)*3
166. status[status$id == combo, r] <- max1-min1+1
167. s <- 35 + (j-2017)*3
168. status[status$id == combo, s] <- (max1-min1+1) - nper
169.
170. if(nper <= 2){
171.
172.   w <- j + 1
173.
174.   if (w <= min(y2,2018)){
175.
176.     min2 <- min(subset(x, año1 == w)$fecha)
177.
178.     status[status$id == combo, 20] <- min2
179.     status[status$id == combo, 53] <- min2
180.
181.     z <- subset(x, fecha >= min2)
182.     nper.ult.seq <- nrow(z)
183.
184.     status[status$id == combo, 24] <- nper.ult.seq
185.   }
186. }
187. }
188. }
189. }
190.
191. status$var.fecha.fin.2017 <- as.Date("31/12/2017", format = "%d/%m/%Y") -
  status$fecha.fin.2017
192. status$var.fecha.fin.2017[is.na(status$var.fecha.fin.2017)] <- 0
193. status$var.fecha.fin <- status$fecha.referencia - status$fecha.final
194.
195. status$activo_inactivo_l06_daily <- ""
196. status$activo_inactivo_l06_daily[status$locacion == "PLANTA RIMAC"] <- "INACTIVO"
197. status$activo_inactivo_l06_daily[status$venta2019 <= 1] <- "INACTIVO"
198. status$activo_inactivo_l06_daily[status$venta2018 <= 0] <- "INACTIVO"
199. status$activo_inactivo_l06_daily[status$activo_inactivo_l06_daily != "INACTIVO"] <-
  "ACTIVO"
200. status$activo_inactivo_l06_daily[status$activo_inactivo_l06_daily == "ACTIVO" &
  status$nper.real2019 <=10] <- "INACTIVO"
201. status$activo_inactivo_l06_daily[status$activo_inactivo_l06_daily == "ACTIVO" &
  status$nper.real2018 <=1] <- "INACTIVO"
202.
203. status$fecha.inicial[status$var.fecha.fin.2017>30 &
  status$activo_inactivo_l06_daily=="ACTIVO"] <-
  status$fecha.ini.2018[status$var.fecha.fin.2017>30 &
  status$activo_inactivo_l06_daily=="ACTIVO"]
204. status$nper.real[status$var.fecha.fin.2017>30 & status$activo_inactivo_l06_daily=="ACTIVO"]
  <- status$nper.real2018[status$var.fecha.fin.2017>30 &
  status$activo_inactivo_l06_daily=="ACTIVO"] +
  status$nper.real2019[status$var.fecha.fin.2017>30 &
  status$activo_inactivo_l06_daily=="ACTIVO"]
205. status$nper.teo <- status$fecha.final - status$fecha.inicial + 1
206. status$var.nper <- status$nper.teo - status$nper.real
207.
208. status$fase_comb_l06_daily <- ""
209. status$fase_comb_l06_daily[status$venta2018 == 0 & status$venta2019 > 0] <- "LANZAMIENTO"
210.
211. status$tipo_modelo_l06_daily <- ""
212. status$tipo_modelo_l06_daily[status$tipo_modelo_l06 == "CROSTON" &
  status$activo_inactivo_l06_daily == "ACTIVO"] <- "CROSTON"
213. status$tipo_modelo_l06_daily[status$tipo_modelo_l06 == "NAIVE" &
  status$activo_inactivo_l06_daily == "ACTIVO"] <- "CROSTON"
214. status$tipo_modelo_l06_daily[status$tipo_modelo_l06 == "ES_AR" &
  status$activo_inactivo_l06_daily == "ACTIVO"] <- "AR_NN"
215. status$tipo_modelo_l06_daily[status$tipo_modelo_l06 == "ES" &
  status$activo_inactivo_l06_daily == "ACTIVO"] <- "AR_NN"
216. status$tipo_modelo_l06_daily[status$tipo_modelo_l06_daily == "" &
  status$activo_inactivo_l06_daily == "ACTIVO" & status$fecha.fin.2018 >= "20/12/2018"] <-
  "AR_NN"

```

```

217.status$tipo_modelo_l06_daily[status$tipo_modelo_l06_daily == "" &
    status$activo_inactivo_l06_daily == "ACTIVO" & status$fecha.fin.2018 < "20/12/2018"] <-
    "CROSTON"
218.
219.status <- left_join(status, status_temp, by = c("provincia"))
220.status$data_temp[is.na(status$data_temp)] <- 0
221.
222.#V. Grabar archivo de estado de combinaciones----
223.write.csv(status, file = "D:/Combo_Status.csv", row.names = FALSE)
224.
225.#VI. Grabar archivo combinaciones válidos----
226.combo <- status
227.combo$combo_status <- ""
228.combo$combo_status[combo$venta2018 <= 1 & combo$venta2019 <= 1] <- "DESCARTAR COMBO"
229.combo$combo_status[combo$venta2019 <= 0] <- "DESCARTAR COMBO"
230.combo$combo_status[combo$combo_status != "DESCARTAR COMBO"] <- "MANTENER COMBO"
231.combo <- subset(combo, combo_status == "MANTENER COMBO")
232.write.csv(combo[,1:15], file = "D:/Daily Level06/Combo.csv", row.names = FALSE)

```

Anexo N° 3: Código en R para el entrenamiento y selección de modelos de pronósticos:

nivel D6

```

1. ## MODELAMIENTO NIVEL 06 - DIARIO: SKU / CD / CANAL ##
2.
3. #I. Carga de datos----
4. comb <- read.csv("D:/Combo_Tiempo.csv", encoding = "UTF-8")
5. data <- read.csv("D:/Daily_Level06_EMB.csv", encoding = "UTF-8")
6. status <- read.csv("D:/Combo_Status.csv")
7. outputlvl06 <- read.csv("D:/OutputLvl06_toDaily.csv")
8. temp <- read.csv("D:/temp6_c.csv")
9. no_lab <- read.csv("D:/No_Laborado.csv")
10. comb_redis <- read.csv("D:/Redis_File_Level06_toDaily.csv")
11. loc_19 <- read.csv("D:/locaciones19_status.csv")
12.
13. colnames(comb)[1] <- "region3"
14. colnames(data)[1] <- "region3"
15. status$fecha.inicial <- as.Date(status$fecha.inicial, format = "%Y-%m-%d")
16.
17. #II. Carga de librerías----
18. library(dplyr)
19. library(forecast)
20. library(tsintermittent)
21.
22. #III. Preprocesamiento----
23. borrar100 <- read.csv("D:/BorrarVenta100%.csv")
24. borrar_prov <- read.csv("D:/BorrarVentaParcial.csv")
25.
26. data$volumen_cu_emb[data$fecha %in% borrar100$fecha] <- 0
27. data$ingreso_bruto[data$fecha %in% borrar100$fecha] <- 0
28. data$descuento_comercial_mercado[data$fecha %in% borrar100$fecha] <- 0
29. data$descuento_bonificaciones_mercado[data$fecha %in% borrar100$fecha] <- 0
30. data$descuento_intro_envase_mercado[data$fecha %in% borrar100$fecha] <- 0
31. data$ingreso_netocomercial[data$fecha %in% borrar100$fecha] <- 0
32.
33. data$volumen_cu_emb[paste(data$fecha,data$provincia) %in%
    paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
34. data$ingreso_bruto[paste(data$fecha,data$provincia) %in%
    paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
35. data$descuento_comercial_mercado[paste(data$fecha,data$provincia) %in%
    paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
36. data$descuento_bonificaciones_mercado[paste(data$fecha,data$provincia) %in%
    paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
37. data$descuento_intro_envase_mercado[paste(data$fecha,data$provincia) %in%
    paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0

```

```

38. data$ingreso_netocomercial[paste(data$fecha,data$provincia) %in%
paste(borrar_prov$fecha,borrar_prov$provincia)] <- 0
39.
40. bd <- left_join(comb, data, by = c("region3", "provincia", "locacion", "canal",
"categoria", "trademark", "retornabilidad", "tamaño", "segmento_detallado", "sabor",
"empaque", "unidades", "tipo_envase", "cod_producto", "producto", "año1", "mes1",
"año1mes1", "año2mes2", "año2semana2", "diasemana", "dia_año", "diameter", "fecha"))
41.
42. bd$volumen_cu_emb[is.na(bd$volumen_cu_emb)] <- 0
43. bd$ingreso_bruto[is.na(bd$ingreso_bruto)] <- 0
44. bd$descuento_comercial_mercado[is.na(bd$descuento_comercial_mercado)] <- 0
45. bd$descuento_bonificaciones_mercado[is.na(bd$descuento_bonificaciones_mercado)] <- 0
46. bd$descuento_intro_envase_mercado[is.na(bd$descuento_intro_envase_mercado)] <- 0
47. bd$ingreso_netocomercial[is.na(bd$ingreso_netocomercial)] <- 0
48.
49. bd$precio_bruto <- bd$ingreso_bruto/bd$volumen_cu_emb
50. bd$precio_bruto[is.na(bd$precio_bruto)] <- 0
51. bd$precio_bruto[bd$precio_bruto == "Inf" | bd$precio_bruto == "-Inf"] <- 0
52.
53. bd <- left_join(bd, temp, by = c("provincia", "año1", "mes1", "año1mes1", "fecha"))
54. bd$temp6_c[is.na(bd$temp6_c)] <- 0
55.
56. bd <- left_join(bd, no_lab, by = c("region3", "provincia", "fecha"))
57. bd$fecha <- as.Date(bd$fecha, format = "%d/%m/%Y")
58.
59. # Grabar la BD previo a la segmentación
60. write.csv(bd, file = "D:/bd_daily_lvl06.csv", row.names = FALSE)
61.
62. # Segmentación de BD por regiones
63. # Repetir este proceso por cada región #
64. bd_sub <- subset(bd, region3 == "COSTA CENTRO")
65. write.csv(bd_sub, file = "D:/costa_centro_daily_lvl06.csv", row.names = FALSE)
66.
67. #IV. Inicia modelamiento----
68. # Repetir los siguientes pasos por cada región #
69.
70. # Carga de datos segmentados por región
71. bd <- read.csv("D:/costa_centro_daily_lvl06.csv")
72. status <- read.csv("D:/Combo_Status.csv")
73. outputlvl06 <- read.csv("D:/OutputLvl06_toDaily.csv")
74. status <- subset(status, region3 == "COSTA CENTRO")
75. outputlvl06 <- subset(outputlvl06, region3 == "COSTA CENTRO")
76. bd$fecha <- as.Date(bd$fecha, format = "%Y-%m-%d")
77. status$fecha.inicial <- as.Date(status$fecha.inicial, format = "%Y-%m-%d")
78.
79. bd$temp6_c[bd$no_laborado == 1] <- 0
80.
81. bd$id <- paste(bd$region3, bd$provincia, bd$locacion, bd$canal, bd$categoria, bd$trademark,
bd$retornabilidad, bd$tamaño, bd$segmento_detallado, bd$sabor, bd$empaque, bd$unidades,
bd$tipo_envase, bd$cod_producto, bd$producto, sep = " ")
82.
83. bd$predy_ar <- 0
84. bd$predy_nn <- 0
85. bd$predy_cr <- 0
86. bd$predy_pond <- 0
87. bd$predy_naive <- 0
88.
89. bd$predy_ar_li <- 0
90. bd$predy_ar_ls <- 0
91. bd$predy_nn_li <- 0
92. bd$predy_nn_ls <- 0
93. bd$predy_pond_li <- 0
94. bd$predy_pond_ls <- 0
95.
96. comb.activa <- subset(status, activo_inactivo_l06_daily == "ACTIVO")
97. rownames(comb.activa) <- 1:nrow(comb.activa)
98.
99. temp <- bd[-(1:nrow(bd)),]
100.
101. # Modelamiento estadístico
102. for(i in 1:nrow(comb.activa)){
103.

```

```

104. print(i)
105. combo <- comb.activa$id[i]
106. inicio <- comb.activa$fecha.inicial[i] #fecha inicial
107. añomes <- as.numeric(paste(format(inicio, "%Y"),format(inicio, "%m"), sep = ""))
    #año1mes1 inicial
108. model <- comb.activa$tipo_modelo_l06_daily[i] #tipo modelo
109. categoria <- comb.activa$categoria[i]
110. region <- comb.activa$region3[i]
111.
112. x <- subset(bd, id == combo & año1mes1 >= añomes)
113. x.ts <- ts(subset(x, año1 < 2019)$volumen_cu_emb, frequency = 7)
114. n.ts <- length(x.ts)
115. n.x <- nrow(x)
116.
117. f <- fourier(ts(x$volumen_cu_emb, frequency = 365), K = 5)
118.
119. xreg <- cbind(precio_bruto = x$precio_bruto, descuento_comercial =
    x$descuento_comercial_mercado, descuento_bonificaciones =
    x$descuento_bonificaciones_mercado, descuento_intro_envase =
    x$descuento_intro_envase_mercado, temp_c = x$temp6_c, no_laborado = x$no_laborado)
120.
121. xreg1 <- xreg[1:(n.x - 90),]
122.
123. desc_comercial <- sum(xreg1[,colnames(xreg1) %in% 'descuento_comercial'])
124. desc_bonificaciones <- sum(xreg1[,colnames(xreg1) %in% 'descuento_bonificaciones'])
125. desc_intro <- sum(xreg1[,colnames(xreg1) %in% 'descuento_intro_envase'])
126. temperatura <- sum(xreg1[,colnames(xreg1) %in% 'temp_c'])
127.
128. if (region %in% c("SIERRA SUR", "SELVA ALTA", "SELVA BAJA")){
129.   xreg <- xreg[,!colnames(xreg) %in% 'precio_bruto']
130. }
131.
132. if (desc_comercial <= 1){
133.   xreg <- xreg[,!colnames(xreg) %in% 'descuento_comercial']
134. }
135.
136. if (desc_bonificaciones <= 1){
137.   xreg <- xreg[,!colnames(xreg) %in% 'descuento_bonificaciones']
138. }
139.
140. if (desc_intro <= 2){
141.   xreg <- xreg[,!colnames(xreg) %in% 'descuento_intro_envase']
142. }
143.
144. if (temperatura <= 0 | region %in% c("SIERRA SUR", "COSTA SUR", "SIERRA CENTRO", "SIERRA
    NORTE")){
145.   xreg <- xreg[,!colnames(xreg) %in% 'temp_c']
146. }
147.
148. if (model == "AR_NN"){
149.
150.   #1. ARIMAX
151.   if (is.null(nrow(xreg)) == TRUE){
152.
153.     xreg.train <- xreg[1:(n.x - 90)]
154.     xreg.test <- xreg[(n.x - 90 + 1):n.x]
155.
156.   } else {
157.
158.     xreg.train <- xreg[1:(n.x - 90),]
159.     xreg.test <- xreg[(n.x - 90 + 1):n.x,]
160.
161.   }
162.
163. x.ar <- auto.arima(x.ts, xreg = xreg.train)
164. x.ar.pred <- forecast(x.ar, xreg = xreg.test, h = 90, level = 95)
165.
166. x$predy_ar[1:(n.x - 90)] <- x.ar.pred$fitted
167. x$predy_ar[(n.x - 90 + 1):n.x] <- x.ar.pred$mean
168. x$predy_ar[x$predy_ar <= 0] <- 0
169. x$predy_ar_li[(n.x - 90 + 1):n.x] <- x.ar.pred$lower
170. x$predy_ar_ls[(n.x - 90 + 1):n.x] <- x.ar.pred$upper

```



```

171.
172.   x$predy_ar[x$predy_ar < 0] <- 0
173.   x$predy_ar_li[x$predy_ar_li < 0] <- 0
174.   x$predy_ar_ls[x$predy_ar_ls < 0] <- 0
175.   x$predy_ar[x$no_laborado == 1] <- 0
176.   x$predy_ar_li[x$no_laborado == 1] <- 0
177.   x$predy_ar_ls[x$no_laborado == 1] <- 0
178.
179.
180.   #2. Redes Neuronales
181.   x.ts <- ts(subset(x, año1mes1 >= max(201801, año1mes1) & año1 < 2019)$volumen_cu_emb,
frequency = 7)
182.   n.ts <- length(x.ts)
183.
184.   if (region %in% c("SIERRA SUR", "COSTA CENTRO", "COSTA SUR")){
185.     xreg <- cbind(xreg, fourier = f)
186.   }
187.
188.
189.   if (is.null(nrow(xreg)) == TRUE){
190.
191.     xreg.train <- xreg[(n.x - 90 - n.ts + 1):(n.x - 90)]
192.     xreg.test <- xreg[(n.x - 90 + 1):n.x]
193.
194.   } else {
195.
196.     xreg.train <- xreg[(n.x - 90 - n.ts + 1):(n.x - 90),]
197.     xreg.test <- xreg[(n.x - 90 + 1):n.x,]
198.
199.   }
200.
201.   set.seed(34)
202.   x.nn <- nnetar(x.ts, xreg = xreg.train)
203.   x.nn.pred <- forecast(x.nn, xreg = xreg.test, h = 90)
204.
205.   x$predy_nn[(n.x - 90 - n.ts + 1):(n.x - 90)] <- x.nn.pred$fitted
206.   x$predy_nn[(n.x - 90 + 1):n.x] <- x.nn.pred$mean
207.   x$predy_nn[x$predy_nn < 0] <- 0
208.   x$predy_nn[x$no_laborado == 1] <- 0
209.   x$predy_nn[is.na(x$predy_nn)] <- 0
210.   x$predy_nn_li <- (x$predy_ar_li/x$predy_ar) * x$predy_nn
211.   x$predy_nn_ls <- (x$predy_ar_ls/x$predy_ar) * x$predy_nn
212.   x$predy_nn_li[x$predy_nn_li == "Inf" | x$predy_nn_li == "-Inf"] <- 0
213.   x$predy_nn_ls[x$predy_nn_ls == "Inf" | x$predy_nn_ls == "-Inf"] <- 0
214.   x$predy_nn_li[is.na(x$predy_nn_li)] <- 0
215.   x$predy_nn_ls[is.na(x$predy_nn_ls)] <- 0
216.
217.   #3. Pronóstico Combinado
218.   x$predy_pond <- (x$predy_ar + x$predy_nn)/2
219.   x$predy_pond[x$año1mes1 < max(201801, año1mes1) & año1 < 2019]$volumen_cu_emb,
año1mes1]
220.   x$predy_pond_li <- (x$predy_ar_li/x$predy_ar) * x$predy_pond
221.   x$predy_pond_ls <- (x$predy_ar_ls/x$predy_ar) * x$predy_pond
222.   x$predy_pond_li[x$predy_pond_li == "Inf" | x$predy_pond_li == "-Inf"] <- 0
223.   x$predy_pond_ls[x$predy_pond_ls == "Inf" | x$predy_pond_ls == "-Inf"] <- 0
224.   x$predy_pond_li[is.na(x$predy_pond_li)] <- 0
225.   x$predy_pond_ls[is.na(x$predy_pond_ls)] <- 0
226.
227.   #4. Naive
228.   x.ts <- ts(subset(x, año1mes1 >= max(201801, año1mes1) & año1 < 2019)$volumen_cu_emb,
frequency = 365)
229.   x.naive <- snaive(x.ts, h = 90)
230.   x$predy_naive[(n.x - 90 + 1):n.x] <- x.naive$mean
231.
232.
233.   # Acumulación en el temporal
234.   temp <- union_all(temp, x)
235.
236. } else if (model == "CROSTON") {
237.
238.   #1. Croston
239.   x.cr <- crost(x.ts, h = 90)

```

```

240. x$predy_cr[1:(n.x-90)] <- x.cr$frfc.in
241. x$predy_cr[(n.x-90+1):n.x] <- x.cr$frfc.out
242. x$predy_cr[is.na(x$predy_cr)] <- 0
243. x$predy_cr[x$predy_cr < 0] <- 0
244. x$predy_cr[x$no_laborado == 1] <- 0
245.
246. #2. ARIMA, Neural Networks y Combinado
247. # Asignar los resultados de croston a estas variables
248. x$predy_ar <- x$predy_cr
249. x$predy_nn <- x$predy_cr
250. x$predy_pond <- x$predy_cr
251.
252. #3. Naive
253. x.naive <- naive(x.ts, h = 90)
254. x$predy_naive[(n.x-90+1):n.x] <- x.naive$mean
255. x$predy_naive[x$no_laborado == 1] <- 0
256.
257. # Acumulación en el temporal
258. temp <- union_all(temp, x)
259.
260. }
261.}
262.
263.# Grabar resultados del segmento regional
264.write.csv(temp, file = "D:/output_level06_daily (COSTA CENTRO).csv", row.names = FALSE)
265.
266.# Reunificación de segmentos regionales
267.bd <- read.csv("D:/bd_daily_lvl06.csv")
268.outputlvl06 <- read.csv("D:/OutputLvl06_toDaily.csv")
269.status <- read.csv("D:/Combo_Status.csv")
270.
271.bd$temp6_c[bd$no_laborado == 1] <- 0
272.
273.bd$id <- paste.bd$region3, bd$provincia, bd$locacion, bd$canal, bd$categoria, bd$trademark,
bd$retornabilidad, bd$tamaño, bd$segmento_detalldado, bd$sabor, bd$empaquete, bd$unidades,
bd$tipo_envase, bd$cod_producto, bd$producto, sep = " ")
274.
275.bd$predy_ar <- 0
276.bd$predy_nn <- 0
277.bd$predy_cr <- 0
278.bd$predy_pond <- 0
279.bd$predy_naive <- 0
280.
281.bd$predy_ar_li <- 0
282.bd$predy_ar_ls <- 0
283.bd$predy_nn_li <- 0
284.bd$predy_nn_ls <- 0
285.bd$predy_pond_li <- 0
286.bd$predy_pond_ls <- 0
287.
288.bd_cn <- read.csv("D:/output_level06_daily (COSTA NORTE).csv")
289.bd_cc <- read.csv("D:/output_level06_daily (COSTA CENTRO).csv")
290.bd_cs <- read.csv("D:/output_level06_daily (COSTA SUR).csv")
291.bd_sn <- read.csv("D:/output_level06_daily (SIERRA NORTE).csv")
292.bd_sc <- read.csv("D:/output_level06_daily (SIERRA CENTRO).csv")
293.bd_ss <- read.csv("D:/output_level06_daily (SIERRA SUR).csv")
294.bd_sa <- read.csv("D:/output_level06_daily (SELVA ALTA).csv")
295.bd_sb <- read.csv("D:/output_level06_daily (SELVA BAJA).csv")
296.
297.temp <- bd_cn[-(1:nrow.bd_cn)],]
298.
299.temp <- union_all(temp, bd_cn)
300.temp <- union_all(temp, bd_cc)
301.temp <- union_all(temp, bd_cs)
302.temp <- union_all(temp, bd_sn)
303.temp <- union_all(temp, bd_sc)
304.temp <- union_all(temp, bd_ss)
305.temp <- union_all(temp, bd_sa)
306.temp <- union_all(temp, bd_sb)
307.
308.# Anexo de resultados temporales en la BD general
309.bd$predy_ar[match(paste(temp$id, temp$fecha), paste.bd$id, bd$fecha))] <- temp$predy_ar

```



```

310.bd$predy_nn[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <- temp$predy_nn
311.bd$predy_cr[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <- temp$predy_cr
312.bd$predy_pond[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <- temp$predy_pond
313.bd$predy_naive[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_naive
314.
315.bd$predy_ar_li[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_ar_li
316.bd$predy_ar_ls[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_ar_ls
317.bd$predy_nn_li[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_nn_li
318.bd$predy_nn_ls[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_nn_ls
319.bd$predy_pond_li[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_pond_li
320.bd$predy_pond_ls[match(paste(temp$id, temp$fecha), paste(bd$id, bd$fecha))] <-
temp$predy_pond_ls
321.
322.bd$predy_naive[is.na(bd$predy_naive)] <- 0
323.
324.bd$predy_ar[bd$predy_ar == "Inf" | bd$predy_ar == "-Inf"] <- 0
325.bd$predy_nn[bd$predy_nn == "Inf" | bd$predy_nn == "-Inf"] <- 0
326.bd$predy_cr[bd$predy_cr == "Inf" | bd$predy_cr == "-Inf"] <- 0
327.bd$predy_pond[bd$predy_pond == "Inf" | bd$predy_pond == "-Inf"] <- 0
328.
329.#V. Procesamiento modelado inicial----
330.# Errores absolutos
331.bd$error_abs_ar <- abs(bd$predy_ar - bd$volumen_cu_emb)
332.bd$error_abs_nn <- abs(bd$predy_nn - bd$volumen_cu_emb)
333.bd$error_abs_pond <- abs(bd$predy_pond - bd$volumen_cu_emb)
334.bd$error_abs_naive <- abs(bd$predy_naive - bd$volumen_cu_emb)
335.
336.# Peso de límites inferiores y superiores
337.bd$peso_ar_li <- bd$predy_ar_li/bd$predy_ar
338.bd$peso_ar_ls <- bd$predy_ar_ls/bd$predy_ar
339.bd$peso_nn_li <- bd$predy_nn_li/bd$predy_nn
340.bd$peso_nn_ls <- bd$predy_nn_ls/bd$predy_nn
341.bd$peso_pond_li <- bd$predy_pond_li/bd$predy_pond
342.bd$peso_pond_ls <- bd$predy_pond_ls/bd$predy_pond
343.
344.bd$peso_ar_li[is.na(bd$peso_ar_li) | bd$peso_ar_li == "Inf"] <- 0
345.bd$peso_ar_ls[is.na(bd$peso_ar_ls) | bd$peso_ar_ls == "Inf"] <- 0
346.bd$peso_nn_li[is.na(bd$peso_nn_li) | bd$peso_nn_li == "Inf"] <- 0
347.bd$peso_nn_ls[is.na(bd$peso_nn_ls) | bd$peso_nn_ls == "Inf"] <- 0
348.bd$peso_pond_li[is.na(bd$peso_pond_li) | bd$peso_pond_li == "Inf"] <- 0
349.bd$peso_pond_ls[is.na(bd$peso_pond_ls) | bd$peso_pond_ls == "Inf"] <- 0
350.
351.# Conciliación con el nivel 06 mensual
352.predy.total <- aggregate(bd[,35:38], by=list(region3 = bd$region3, provincia =
bd$provincia, locacion = bd$locacion, canal = bd$canal, categoria = bd$categoria, trademark
= bd$trademark, retornabilidad = bd$retornabilidad, tamaño = bd$tamaño, segmento_detallado
= bd$segmento_detallado, sabor = bd$sabor, empaque = bd$empaque, unidades = bd$unidades,
tipo_envase = bd$tipo_envase, cod_producto = bd$cod_producto, producto = bd$producto, año1
= bd$año1, mes1 = bd$mes1, año1mes1 = bd$año1mes1), FUN = sum, na.rm = TRUE)
353.
354.predy.total <- predy.total[,-21] #retirar el pronóstico croston
355.
356.colnames(predy.total)[19:21] <- c("predy_ar_total", "predy_nn_total", "predy_pond_total")
357.
358.bd <- left_join(bd, predy.total, by = c("region3", "provincia", "locacion", "categoria",
"canal", "trademark", "retornabilidad", "tamaño", "segmento_detallado", "sabor", "empaque",
"unidades", "tipo_envase", "cod_producto", "producto", "año1", "mes1", "año1mes1"))
359.
360.bd$peso_ar <- bd$predy_ar/bd$predy_ar_total
361.bd$peso_nn <- bd$predy_nn/bd$predy_nn_total
362.bd$peso_pond <- bd$predy_pond/bd$predy_pond_total
363.
364.bd$peso_ar[is.na(bd$peso_ar) | bd$peso_ar == "Inf"] <- 0
365.bd$peso_nn[is.na(bd$peso_nn) | bd$peso_nn == "Inf"] <- 0
366.bd$peso_pond[is.na(bd$peso_pond) | bd$peso_pond == "Inf"] <- 0
367.

```

```

368.bd <- left_join(bd, outputlv106, by = c("region3", "provincia", "locacion", "categoria",
      "canal", "trademark", "retornabilidad", "tamaño", "segmento_detallado", "sabor", "empaque",
      "unidades", "tipo_envase", "cod_producto", "producto", "año1", "mes1", "año1mes1"))
369.colnames(bd)[62:63] <- c("predy_lv106", "baseline_lv106")
370.bd$predy_lv106[is.na(bd$predy_lv106)] <- 0
371.
372.# Cálculo de resultados conciliados
373.bd$predy_ar_ajust <- bd$predy_lv106 * bd$peso_ar
374.bd$predy_nn_ajust <- bd$predy_lv106 * bd$peso_nn
375.bd$predy_pond_ajust <- bd$predy_lv106 * bd$peso_pond
376.
377.bd$error_abs_ar_ajust <- abs(bd$predy_ar_ajust - bd$volumen_cu_emb)
378.bd$error_abs_nn_ajust <- abs(bd$predy_nn_ajust - bd$volumen_cu_emb)
379.bd$error_abs_pond_ajust <- abs(bd$predy_pond_ajust - bd$volumen_cu_emb)
380.
381.bd$predy_ar_ajust_li <- bd$peso_ar_li * bd$predy_ar_ajust
382.bd$predy_ar_ajust_ls <- bd$peso_ar_ls * bd$predy_ar_ajust
383.bd$predy_nn_ajust_li <- bd$peso_nn_li * bd$predy_nn_ajust
384.bd$predy_nn_ajust_ls <- bd$peso_nn_ls * bd$predy_nn_ajust
385.bd$predy_pond_ajust_li <- bd$peso_pond_li * bd$predy_pond_ajust
386.bd$predy_pond_ajust_ls <- bd$peso_pond_ls * bd$predy_pond_ajust
387.
388.status_short <- status[,-(16:53)]
389.
390.bd <- left_join(bd, status_short[,1:17], by = c("region3", "provincia", "locacion",
      "canal", "categoria", "trademark", "retornabilidad", "tamaño", "segmento_detallado",
      "sabor", "empaque", "unidades", "tipo_envase", "cod_producto", "producto"))
391.
392.bd$activo_inactivo_l06_daily[is.na(bd$activo_inactivo_l06_daily)] <- "INACTIVO"
393.bd$fase_comb_l06_daily[is.na(bd$fase_comb_l06_daily)] <- ""
394.
395.# Eliminación de límites fuera de control----
396.bd_temp <- subset(bd, año1mes1 >= 201901)
397.bd_1901 <- aggregate(bd_temp[,c(38, 44:45)], by = list(id = bd_temp$id), FUN = sum, na.rm =
      TRUE)
398.bd_1901$peso_val <- bd_1901$predy_pond_ls/bd_1901$predy_pond
399.bd_1901$peso_val[is.na(bd_1901$peso_val)] <- 0
400.bd_1901$elim_ls <- 0
401.bd_1901$elim_ls[bd_1901$peso_val >= 4] <- 1
402.bd_1901 <- bd_1901[, -c(2:5)]
403.
404.bd <- left_join(bd, bd_1901, by = c("id"))
405.
406.bd$predy_ar_li[bd$elim_ls == 1] <- 0
407.bd$predy_ar_ls[bd$elim_ls == 1] <- 0
408.bd$predy_nn_li[bd$elim_ls == 1] <- 0
409.bd$predy_nn_ls[bd$elim_ls == 1] <- 0
410.bd$predy_pond_li[bd$elim_ls == 1] <- 0
411.bd$predy_pond_ls[bd$elim_ls == 1] <- 0
412.bd$predy_ar_ajust_li[bd$elim_ls == 1] <- 0
413.bd$predy_ar_ajust_ls[bd$elim_ls == 1] <- 0
414.bd$predy_nn_ajust_li[bd$elim_ls == 1] <- 0
415.bd$predy_nn_ajust_ls[bd$elim_ls == 1] <- 0
416.bd$predy_pond_ajust_li[bd$elim_ls == 1] <- 0
417.bd$predy_pond_ajust_ls[bd$elim_ls == 1] <- 0
418.bd$peso_ar_li[bd$elim_ls == 1] <- 0
419.bd$peso_ar_ls[bd$elim_ls == 1] <- 0
420.bd$peso_nn_li[bd$elim_ls == 1] <- 0
421.bd$peso_nn_ls[bd$elim_ls == 1] <- 0
422.bd$peso_pond_li[bd$elim_ls == 1] <- 0
423.bd$peso_pond_ls[bd$elim_ls == 1] <- 0
424.
425.bd$elim_ls <- NULL
426.
427.#VII. Redistribución de pronóstico de nuevos productos del nivel 06----
428.x_acum <- bd[1,]
429.x_acum$group <- 0
430.x_acum <- x_acum[-(1:nrow(bd)),]
431.
432.regions <- aggregate(comb_redis[,4], by = list(region = comb_redis$region3), FUN=sum,
      na.rm=TRUE)
433.regions <- regions[, -2]

```

```

434.
435. for (r in 1:length(regions)){
436.
437.   print(regions[r])
438.   comb_reg <- subset(comb_redis, region3 == regions[r])
439.
440.   for (i in 1:max(comb_reg$group)){
441.
442.     print(i)
443.     comb_actual <- subset(comb_reg, group == i)
444.
445.     x <- bd[1,]
446.     x$group <- 0
447.     x <- x[-(1:nrow(bd)),]
448.
449.     for (j in 1:nrow(comb_actual)){
450.
451.       temp_x <- subset(bd, as.character(region3) == comb_actual[j, 1] &
         as.character(cod_producto) == comb_actual[j, 2] & as.character(canal) == comb_actual[j, 3]
         & activo_inactivo_l06_daily == "INACTIVO" & año1 == 2019)
452.       temp_x$group <- i
453.
454.       x <- union_all(x, temp_x)
455.     }
456.
457.     x_acum <- union_all(x_acum, x)
458.
459.   }
460.}
461.
462. venta_2019 <- aggregate(x_acum[,25], by = list(region3 = x_acum$region3, provincia =
         x_acum$provincia, locacion = x_acum$locacion, canal = x_acum$canal, cod_producto =
         x_acum$cod_producto, group = x_acum$group, año1 = x_acum$año1, mes1 = x_acum$mes1, año1mes1
         = x_acum$año1mes1), FUN = sum, na.rm = TRUE)
463.
464. colnames(venta_2019)[10] <- "volumen_cu_emb_total"
465.
466. x_acum <- left_join(x_acum, venta_2019, by = c("region3", "provincia", "locacion", "canal",
         "cod_producto", "group", "año1", "mes1", "año1mes1"))
467.
468. x_acum$peso_redis <- x_acum$volumen_cu_emb/x_acum$volumen_cu_emb_total
469. x_acum$peso_redis[is.na(x_acum$peso_redis)] <- 0
470. x_acum$peso_redis[x_acum$peso_redis == "-Inf" | x_acum$peso_redis == "Inf"] <- 0
471.
472. x_acum$predy_ar <- x_acum$predy_lv106 * x_acum$peso_redis
473. x_acum$predy_nn <- x_acum$predy_lv106 * x_acum$peso_redis
474. x_acum$predy_pond <- x_acum$predy_lv106 * x_acum$peso_redis
475. x_acum$predy_ar_ajust <- x_acum$predy_lv106 * x_acum$peso_redis
476. x_acum$predy_nn_ajust <- x_acum$predy_lv106 * x_acum$peso_redis
477. x_acum$predy_pond_ajust <- x_acum$predy_lv106 * x_acum$peso_redis
478.
479. bd$predy_ar[match(paste(x_acum$id, x_acum$fecha), paste(bd$id, bd$fecha))] <-
         x_acum$predy_ar
480. bd$predy_nn[match(paste(x_acum$id, x_acum$fecha), paste(bd$id, bd$fecha))] <-
         x_acum$predy_nn
481. bd$predy_pond[match(paste(x_acum$id, x_acum$fecha), paste(bd$id, bd$fecha))] <-
         x_acum$predy_pond
482. bd$predy_ar_ajust[match(paste(x_acum$id, x_acum$fecha), paste(bd$id, bd$fecha))] <-
         x_acum$predy_ar_ajust
483. bd$predy_nn_ajust[match(paste(x_acum$id, x_acum$fecha), paste(bd$id, bd$fecha))] <-
         x_acum$predy_nn_ajust
484. bd$predy_pond_ajust[match(paste(x_acum$id, x_acum$fecha), paste(bd$id, bd$fecha))] <-
         x_acum$predy_pond_ajust
485.
486. bd$error_abs_ar <- abs(bd$predy_ar - bd$volumen_cu_emb)
487. bd$error_abs_nn <- abs(bd$predy_nn - bd$volumen_cu_emb)
488. bd$error_abs_pond <- abs(bd$predy_pond - bd$volumen_cu_emb)
489. bd$error_abs_ar_ajust <- abs(bd$predy_ar_ajust - bd$volumen_cu_emb)
490. bd$error_abs_nn_ajust <- abs(bd$predy_nn_ajust - bd$volumen_cu_emb)
491. bd$error_abs_pond_ajust <- abs(bd$predy_pond_ajust - bd$volumen_cu_emb)
492.
493. #VIII. MODELO DE SELECCIÓN FORECAST AJUSTADO AL VOLUMEN MENSUAL-----

```

```

494.# Agregación de resultados a nivel semanal
495.data_val <- subset(bd, año2semana2 >= 201901)
496.mt_semanal <- aggregate(data_val[,c(25, 64:66)], by=list(locacion = data_val$locacion,
  canal = data_val$canal, cod_producto = data_val$cod_producto, año2semana2 =
  data_val$año2semana2), FUN=sum, na.rm=TRUE)
497.
498.mt_semanal$error_abs_ar_ajust <- abs(mt_semanal$predy_ar_ajust - mt_semanal$volumen_cu_emb)
499.mt_semanal$error_abs_nn_ajust <- abs(mt_semanal$predy_nn_ajust - mt_semanal$volumen_cu_emb)
500.mt_semanal$error_abs_pond_ajust <- abs(mt_semanal$predy_pond_ajust -
  mt_semanal$volumen_cu_emb)
501.
502.# Desarrollo de la matriz de decisión
503.mt_decision <- aggregate(mt_semanal[,c(5:11)], by=list(locacion = mt_semanal$locacion,
  canal = mt_semanal$canal, cod_producto = mt_semanal$cod_producto), FUN=sum, na.rm=TRUE)
504.
505.mt_decision$wmape_ar_ajust <- mt_decision$error_abs_ar_ajust/mt_decision$volumen_cu_emb
506.mt_decision$wmape_nn_ajust <- mt_decision$error_abs_nn_ajust/mt_decision$volumen_cu_emb
507.mt_decision$wmape_pond_ajust <- mt_decision$error_abs_pond_ajust/mt_decision$volumen_cu_emb
508.mt_decision$modelo_elegido <- ""
509.
510.mt_decision <- mt_decision[order(mt_decision$locacion, mt_decision$canal,
  mt_decision$cod_producto),]
511.
512.mt_decision <- mt_decision[mt_decision$volumen_cu_emb!=0,]
513.mt_decision <- mt_decision[mt_decision$predy_pond_ajust!=0,]
514.
515.for (i in c(1:nrow(mt_decision))){
516.  print(i)
517.  wmape_ar <- mt_decision$wmape_ar_ajust[i]
518.  wmape_nn <- mt_decision$wmape_nn_ajust[i]
519.  wmape_pond <- mt_decision$wmape_pond_ajust[i]
520.
521.  value <- min(wmape_ar, wmape_nn, wmape_pond)
522.
523.  if(value == wmape_ar){
524.    mt_decision$modelo_elegido[i] <- "predy_ar_ajust"
525.
526.  } else if (value == wmape_nn){
527.    mt_decision$modelo_elegido[i] <- "predy_nn_ajust"
528.
529.  } else if (value == wmape_pond){
530.    mt_decision$modelo_elegido[i] <- "predy_pond_ajust"
531.  }
532.}
533.
534.# Asignación de modelos para resultado final
535.modelo.final <- mt_decision[,-(4:13)]
536.
537.bd$predy_final_ajust <- 0
538.bd$error_abs_final_ajust <- 0
539.bd$predy_final_ajust_li <- 0
540.bd$predy_final_ajust_ls <- 0
541.bd$baseline_ajust <- 0
542.
543.bd <- left_join(bd, modelo.final, by = c("locacion", "canal", "cod_producto"))
544.
545.bd_ar <- subset(bd, modelo_elegido == "predy_ar_ajust")
546.bd_nn <- subset(bd, modelo_elegido == "predy_nn_ajust")
547.bd_pond <- subset(bd, modelo_elegido == "predy_pond_ajust")
548.
549.bd$predy_final_ajust[match(paste.bd_ar$id, bd_ar$fecha), paste.bd$id, bd$fecha)] <-
  bd_ar$predy_ar_ajust
550.bd$predy_final_ajust_li[match(paste.bd_ar$id, bd_ar$fecha), paste.bd$id, bd$fecha)] <-
  bd_ar$predy_ar_ajust_li
551.bd$predy_final_ajust_ls[match(paste.bd_ar$id, bd_ar$fecha), paste.bd$id, bd$fecha)] <-
  bd_ar$predy_ar_ajust_ls
552.
553.bd$predy_final_ajust[match(paste.bd_nn$id, bd_nn$fecha), paste.bd$id, bd$fecha)] <-
  bd_nn$predy_nn_ajust
554.bd$predy_final_ajust_li[match(paste.bd_nn$id, bd_nn$fecha), paste.bd$id, bd$fecha)] <-
  bd_nn$predy_nn_ajust_li

```

```

555.bd$predy_final_ajust_ls[match(paste(bd_nn$id, bd_nn$fecha), paste(bd$id, bd$fecha))] <-
bd_nn$predy_nn_ajust_ls
556.
557.bd$predy_final_ajust[match(paste(bd_pond$id, bd_pond$fecha), paste(bd$id, bd$fecha))] <-
bd_pond$predy_pond_ajust
558.bd$predy_final_ajust_li[match(paste(bd_pond$id, bd_pond$fecha), paste(bd$id, bd$fecha))] <-
bd_pond$predy_pond_ajust_li
559.bd$predy_final_ajust_ls[match(paste(bd_pond$id, bd_pond$fecha), paste(bd$id, bd$fecha))] <-
bd_pond$predy_pond_ajust_ls
560.
561.bd$error_abs_final_ajust <- abs(bd$predy_final_ajust - bd$volumen_cu_emb)
562.bd$modelo_elegido <- NULL
563.
564.predy.total <- aggregate(bd[,78], by=list(locacion = bd$locacion, canal = bd$canal,
cod_producto = bd$cod_producto, año1mes1 = bd$año1mes1), FUN=sum, na.rm=TRUE)
565.
566.colnames(predy.total)[5] <- c("predy_final_ajust_total")
567.
568.bd <- left_join(bd, predy.total, by = c("locacion", "canal", "cod_producto", "año1mes1"))
569.
570.bd$baseline_ajust <- bd$baseline_lv106 * bd$predy_final_ajust / bd$predy_final_ajust_total
571.bd$baseline_ajust[is.na(bd$baseline_ajust)] <- 0
572.bd$baseline_ajust[bd$baseline_ajust == "Inf" | bd$baseline_ajust == "-Inf"] <- 0
573.
574.# Eliminación de límites fuera de control
575.bd$predy_final_ajust_ls[bd$categoria == "RTD TEA"] <- 0
576.bd$predy_final_ajust_li[bd$categoria == "RTD TEA"] <- 0
577.bd$predy_final_ajust_ls[bd$categoria == "CARBONATES" & bd$trademark == "MIXTOS"] <- 0
578.bd$predy_final_ajust_li[bd$categoria == "CARBONATES" & bd$trademark == "MIXTOS"] <- 0
579.
580.#IX. MODELO DE SELECCION FORECAST SENSADO CORTO PLAZO-----
581.# Agregación de resultados a nivel semanal
582.data_val <- subset(bd, año2semana2 >= 201902 & año2semana2 <= 201903)
583.mt_semanal <- aggregate(data_val[,c(25,35,36,38,64:66)], by=list(locacion =
data_val$locacion, canal = data_val$canal, cod_producto = data_val$cod_producto,
año2semana2 = data_val$año2semana2), FUN=sum, na.rm=TRUE)
584.
585.mt_semanal$error_abs_ar <- abs(mt_semanal$predy_ar - mt_semanal$volumen_cu_emb)
586.mt_semanal$error_abs_nn <- abs(mt_semanal$predy_nn - mt_semanal$volumen_cu_emb)
587.mt_semanal$error_abs_pond <- abs(mt_semanal$predy_pond - mt_semanal$volumen_cu_emb)
588.mt_semanal$error_abs_ar_ajust <- abs(mt_semanal$predy_ar_ajust - mt_semanal$volumen_cu_emb)
589.mt_semanal$error_abs_nn_ajust <- abs(mt_semanal$predy_nn_ajust - mt_semanal$volumen_cu_emb)
590.mt_semanal$error_abs_pond_ajust <- abs(mt_semanal$predy_pond_ajust -
mt_semanal$volumen_cu_emb)
591.
592.#Desarrollo de la matriz de decisión
593.mt_decision <- aggregate(mt_semanal[,c(5:17)], by=list(locacion = mt_semanal$locacion,
canal = mt_semanal$canal, cod_producto = mt_semanal$cod_producto), FUN=sum, na.rm=TRUE)
594.
595.mt_decision$wmape_ar <- mt_decision$error_abs_ar/mt_decision$volumen_cu_emb
596.mt_decision$wmape_nn <- mt_decision$error_abs_nn/mt_decision$volumen_cu_emb
597.mt_decision$wmape_pond <- mt_decision$error_abs_pond/mt_decision$volumen_cu_emb
598.mt_decision$wmape_ar_ajust <- mt_decision$error_abs_ar_ajust/mt_decision$volumen_cu_emb
599.mt_decision$wmape_nn_ajust <- mt_decision$error_abs_nn_ajust/mt_decision$volumen_cu_emb
600.mt_decision$wmape_pond_ajust <- mt_decision$error_abs_pond_ajust/mt_decision$volumen_cu_emb
601.mt_decision$modelo_elegido <- ""
602.
603.mt_decision <- mt_decision[order(mt_decision$locacion, mt_decision$canal,
mt_decision$cod_producto),]
604.
605.mt_decision <- mt_decision[mt_decision$volumen_cu_emb!=0,]
606.mt_decision <- mt_decision[mt_decision$predy_pond!=0,]
607.
608.for (i in c(1:nrow(mt_decision))){
609. print(i)
610. wmape_ar <- mt_decision$wmape_ar[i]
611. wmape_nn <- mt_decision$wmape_nn[i]
612. wmape_pond <- mt_decision$wmape_pond[i]
613. wmape_ar_ajust <- mt_decision$wmape_ar_ajust[i]
614. wmape_nn_ajust <- mt_decision$wmape_nn_ajust[i]
615. wmape_pond_ajust <- mt_decision$wmape_pond_ajust[i]
616.

```

```

617. value <- min(wmape_ar, wmape_nn, wmape_pond, wmape_ar_ajust, wmape_nn_ajust,
wmape_pond_ajust)
618.
619. if(value == wmape_ar){
620.   mt_decision$modelo_elegido[i] <- "predy_ar"
621.
622. } else if (value == wmape_nn){
623.   mt_decision$modelo_elegido[i] <- "predy_nn"
624.
625. } else if (value == wmape_pond){
626.   mt_decision$modelo_elegido[i] <- "predy_pond"
627.
628. } else if (value == wmape_ar_ajust){
629.   mt_decision$modelo_elegido[i] <- "predy_ar_ajust"
630.
631. } else if (value == wmape_nn_ajust){
632.   mt_decision$modelo_elegido[i] <- "predy_nn_ajust"
633.
634. } else if (value == wmape_pond_ajust){
635.   mt_decision$modelo_elegido[i] <- "predy_pond_ajust"
636.
637. }
638. }
639.
640. # Asignación de modelos para resultado final
641. modelo.final <- mt_decision[,-(4:22)]
642.
643. bd$predy_final_sensed <- 0
644. bd$error_abs_final_sensed <- 0
645. bd$predy_final_sensed_li <- 0
646. bd$predy_final_sensed_ls <- 0
647. bd$baseline_sensed <- 0
648.
649. bd <- left_join(bd, modelo.final, by = c("locacion", "canal", "cod_producto"))
650.
651. bd_ar <- subset(bd, modelo_elegido == "predy_ar")
652. bd_nn <- subset(bd, modelo_elegido == "predy_nn")
653. bd_pond <- subset(bd, modelo_elegido == "predy_pond")
654. bd_ar_ajust <- subset(bd, modelo_elegido == "predy_ar_ajust")
655. bd_nn_ajust <- subset(bd, modelo_elegido == "predy_nn_ajust")
656. bd_pond_ajust <- subset(bd, modelo_elegido == "predy_pond_ajust")
657.
658. bd$predy_final_sensed[match(paste(bd_ar$id, bd_ar$fecha), paste(bd$id, bd$fecha))] <-
bd_ar$predy_ar
659. bd$predy_final_sensed_li[match(paste(bd_ar$id, bd_ar$fecha), paste(bd$id, bd$fecha))] <-
bd_ar$predy_ar_li
660. bd$predy_final_sensed_ls[match(paste(bd_ar$id, bd_ar$fecha), paste(bd$id, bd$fecha))] <-
bd_ar$predy_ar_ls
661.
662. bd$predy_final_sensed[match(paste(bd_nn$id, bd_nn$fecha), paste(bd$id, bd$fecha))] <-
bd_nn$predy_nn
663. bd$predy_final_sensed_li[match(paste(bd_nn$id, bd_nn$fecha), paste(bd$id, bd$fecha))] <-
bd_nn$predy_nn_li
664. bd$predy_final_sensed_ls[match(paste(bd_nn$id, bd_nn$fecha), paste(bd$id, bd$fecha))] <-
bd_nn$predy_nn_ls
665.
666. bd$predy_final_sensed[match(paste(bd_pond$id, bd_pond$fecha), paste(bd$id, bd$fecha))] <-
bd_pond$predy_pond
667. bd$predy_final_sensed_li[match(paste(bd_pond$id, bd_pond$fecha), paste(bd$id, bd$fecha))]
<- bd_pond$predy_pond_li
668. bd$predy_final_sensed_ls[match(paste(bd_pond$id, bd_pond$fecha), paste(bd$id, bd$fecha))]
<- bd_pond$predy_pond_ls
669.
670. bd$predy_final_sensed[match(paste(bd_ar_ajust$id, bd_ar_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_ar_ajust$predy_ar_ajust
671. bd$predy_final_sensed_li[match(paste(bd_ar_ajust$id, bd_ar_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_ar_ajust$predy_ar_ajust_li
672. bd$predy_final_sensed_ls[match(paste(bd_ar_ajust$id, bd_ar_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_ar_ajust$predy_ar_ajust_ls
673.
674. bd$predy_final_sensed[match(paste(bd_nn_ajust$id, bd_nn_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_nn_ajust$predy_nn_ajust

```



```

675.bd$predy_final_sensed_li[match(paste(bd_nn_ajust$id, bd_nn_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_nn_ajust$predy_nn_ajust_li
676.bd$predy_final_sensed_ls[match(paste(bd_nn_ajust$id, bd_nn_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_nn_ajust$predy_nn_ajust_ls
677.
678.bd$predy_final_sensed[match(paste(bd_pond_ajust$id, bd_pond_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_pond_ajust$predy_pond_ajust
679.bd$predy_final_sensed_li[match(paste(bd_pond_ajust$id, bd_pond_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_pond_ajust$predy_pond_ajust_li
680.bd$predy_final_sensed_ls[match(paste(bd_pond_ajust$id, bd_pond_ajust$fecha), paste(bd$id,
bd$fecha))] <- bd_pond_ajust$predy_pond_ajust_ls
681.
682.bd$error_abs_final_sensed <- abs(bd$predy_final_sensed - bd$volumen_cu_emb)
683.bd$modelo_elegido <- NULL
684.
685.predy.total <- aggregate(bd[,84], by=list(locacion = bd$locacion, canal = bd$canal,
cod_producto = bd$cod_producto, año1mes1 = bd$año1mes1), FUN=sum, na.rm=TRUE)
686.
687.colnames(predy.total)[5] <- c("predy_final_sensed_total")
688.
689.bd <- left_join(bd, predy.total, by = c("locacion", "canal", "cod_producto", "año1mes1"))
690.
691.bd$baseline_sensed <- bd$baseline_lvl06 * bd$predy_final_sensed /
bd$predy_final_sensed_total
692.bd$baseline_sensed[is.na(bd$baseline_sensed)] <- 0
693.bd$baseline_sensed[bd$baseline_sensed == "Inf" | bd$baseline_ajust == "-Inf"] <- 0
694.
695.# Eliminación de límites fuera de control--
696.bd$predy_final_sensed_ls[bd$categoria == "RTD TEA"] <- 0
697.bd$predy_final_sensed_li[bd$categoria == "RTD TEA"] <- 0
698.bd$predy_final_sensed_ls[bd$categoria == "CARBONATES" & bd$trademark == "MIXTOS"] <- 0
699.bd$predy_final_sensed_li[bd$categoria == "CARBONATES" & bd$trademark == "MIXTOS"] <- 0
700.
701.# Anexo de estado de locaciones
702.bd <- left_join(bd, loc_19, by = c("locacion"))
703.bd$loc_status <- as.character(bd$loc_status)
704.bd$loc_status[is.na(bd$loc_status)] <- "INACTIVO"
705.
706.#X. Grabar archivo final----
707.bd$id <- NULL
708.write.csv(bd, file = "D:/output_level06_daily_Redis_Final(Full).csv", row.names = FALSE)
709.
710.# Versión abreviada
711.bd2 <- subset(bd, año2semana2 >= 201801)
712.
713.write.csv(bd2, file = "D:/output_level06_daily_Redis_Final(Short).csv", row.names = FALSE)

```

Anexo N° 4: Código en R para el despliegue de modelos de pronóstico: nivel D06

```

1. ## DESPLIEGUE DE MODELOS NIVEL 06 - DIARIO: SKU / CD / CANAL ##
2.
3. #I. Carga de datos----
4. bd <- read.csv("D:/output_level06_daily_Redis_Final(Full).csv")
5.
6. #II. Carga de librerías----
7. library(dplyr)
8.
9. #III. RESULTADOS A NIVEL LOCACION / SKU / DIA----
10. bd1 <- aggregate(bd[,c(25:30, 34:38, 63:65, 77, 79:81, 83, 85:87)], by=list(region3 =
bd$region3, provincia = bd$provincia, locacion = bd$locacion, categoria = bd$categoria,
trademark = bd$trademark, retornabilidad = bd$retornabilidad, tamaño = bd$tamaño,
segmento_detallado = bd$segmento_detallado, sabor = bd$sabor, empaque = bd$empaque,
unidades = bd$unidades, tipo_envase = bd$tipo_envase, cod_producto = bd$cod_producto,
producto = bd$producto, año1 = bd$año1, mes1 = bd$mes1, año1mes1 = bd$año1mes1, año2mes2 =
bd$año2mes2, año2semana2 = bd$año2semana2, diasemana = bd$diasemana, dia_año = bd$dia_año,
dames = bd$dames, fecha = bd$fecha, activo_inactivo_l06_daily =
bd$activo_inactivo_l06_daily, fase_comb_l06_daily = bd$fase_comb_l06_daily, loc_status =
bd$loc_status), FUN=sum, na.rm=TRUE)

```

```

11.
12. bd1 <- bd1[,-35] #borrar croston
13. bd1$error_abs_ar <- abs(bd1$predy_ar - bd1$volumen_cu_emb)
14. bd1$error_abs_nn <- abs(bd1$predy_nn - bd1$volumen_cu_emb)
15. bd1$error_abs_pond <- abs(bd1$predy_pond - bd1$volumen_cu_emb)
16. bd1$error_abs_naive <- abs(bd1$predy_naive - bd1$volumen_cu_emb)
17.
18. bd1$error_abs_ar_ajust <- abs(bd1$predy_ar_ajust - bd1$volumen_cu_emb)
19. bd1$error_abs_nn_ajust <- abs(bd1$predy_nn_ajust - bd1$volumen_cu_emb)
20. bd1$error_abs_pond_ajust <- abs(bd1$predy_pond_ajust - bd1$volumen_cu_emb)
21.
22. bd1$error_abs_final_ajust <- abs(bd1$predy_final_ajust - bd1$volumen_cu_emb)
23. bd1$error_abs_final_sensed <- abs(bd1$predy_final_sensed - bd1$volumen_cu_emb)
24.
25. # Grabar archivo de pronóstico
26. write.csv(bd1, file ="D:/output_level06_Daily_Final_LocSku.csv", row.names = FALSE)
27.
28. #IV. RESULTADOS A NIVEL SEMANA / LOCACION / SKU / CANAL----
29. bd2 <- aggregate(bd[,c(25:30, 34:38, 63:65, 77, 79:81, 83, 85:87)], by=list(region3 =
bd$region3, provincia = bd$provincia, locacion = bd$locacion, canal = bd$canal, categoria =
bd$categoria, trademark = bd$trademark, retornabilidad = bd$retornabilidad, tamaño =
bd$tamaño, segmento_detallado = bd$segmento_detallado, sabor = bd$sabor, empaque =
bd$empaque, unidades = bd$unidades, tipo_envase = bd$tipo_envase, cod_producto =
bd$cod_producto, producto = bd$producto, año2semana2 = bd$año2semana2,
activo_inactivo_l06_daily = bd$activo_inactivo_l06_daily, fase_comb_l06_daily =
bd$fase_comb_l06_daily, loc_status = bd$loc_status), FUN=sum, na.rm=TRUE)
30.
31. bd2 <- bd2[,-28] #borrar croston
32.
33. bd2$error_abs_ar <- abs(bd2$predy_ar - bd2$volumen_cu_emb)
34. bd2$error_abs_nn <- abs(bd2$predy_nn - bd2$volumen_cu_emb)
35. bd2$error_abs_pond <- abs(bd2$predy_pond - bd2$volumen_cu_emb)
36. bd2$error_abs_naive <- abs(bd2$predy_naive - bd2$volumen_cu_emb)
37.
38. bd2$error_abs_ar_ajust <- abs(bd2$predy_ar_ajust - bd2$volumen_cu_emb)
39. bd2$error_abs_nn_ajust <- abs(bd2$predy_nn_ajust - bd2$volumen_cu_emb)
40. bd2$error_abs_pond_ajust <- abs(bd2$predy_pond_ajust - bd2$volumen_cu_emb)
41.
42. bd2$error_abs_final_ajust <- abs(bd2$predy_final_ajust - bd2$volumen_cu_emb)
43. bd2$error_abs_final_sensed <- abs(bd2$predy_final_sensed - bd2$volumen_cu_emb)
44.
45. # Grabar archivo de pronóstico
46. write.csv(bd2, file ="D:/output_level06_Weekly_Final.csv", row.names = FALSE)
47.
48. #2. RESULTADOS A NIVEL SEMANA / LOCACION / SKU
49. bd3 <- aggregate(bd[,c(25:30, 34:38, 63:65, 77, 79:81, 83, 85:87)], by=list(region3 =
bd$region3, provincia = bd$provincia, locacion = bd$locacion, categoria = bd$categoria,
trademark = bd$trademark, retornabilidad = bd$retornabilidad, tamaño = bd$tamaño,
segmento_detallado = bd$segmento_detallado, sabor = bd$sabor, empaque = bd$empaque,
unidades = bd$unidades, tipo_envase = bd$tipo_envase, cod_producto = bd$cod_producto,
producto = bd$producto, año2semana2 = bd$año2semana2, activo_inactivo_l06_daily =
bd$activo_inactivo_l06_daily, fase_comb_l06_daily = bd$fase_comb_l06_daily, loc_status =
bd$loc_status), FUN=sum, na.rm=TRUE)
50.
51. bd3 <- bd3[,-27] #borrar croston
52.
53. bd3$error_abs_ar <- abs(bd3$predy_ar - bd3$volumen_cu_emb)
54. bd3$error_abs_nn <- abs(bd3$predy_nn - bd3$volumen_cu_emb)
55. bd3$error_abs_pond <- abs(bd3$predy_pond - bd3$volumen_cu_emb)
56. bd3$error_abs_naive <- abs(bd3$predy_naive - bd3$volumen_cu_emb)
57.
58. bd3$error_abs_ar_ajust <- abs(bd3$predy_ar_ajust - bd3$volumen_cu_emb)
59. bd3$error_abs_nn_ajust <- abs(bd3$predy_nn_ajust - bd3$volumen_cu_emb)
60. bd3$error_abs_pond_ajust <- abs(bd3$predy_pond_ajust - bd3$volumen_cu_emb)
61.
62. bd3$error_abs_final_ajust <- abs(bd3$predy_final_ajust - bd3$volumen_cu_emb)
63. bd3$error_abs_final_sensed <- abs(bd3$predy_final_sensed - bd3$volumen_cu_emb)
64.
65. # Grabar archivo de pronóstico
66. write.csv(bd3, file ="D:/output_level06_Weekly_Final_LocSku.csv", row.names = FALSE)

```


Anexo N° 5: Rol y responsabilidades del especialista de modelamiento predictivo

		----- Definición del rol -----
Especialista de Modelamiento Predictivo		Mantiene y mejora continuamente los procesos y sistemas de generación de pronósticos considerando las variables que determinan la demanda, con sus supuestos, análisis de históricos, tendencias de mercado, modelos estadísticos, y con foco a la colaboración y consenso con los diversos stakeholders.
Parámetro	Impacto en el negocio	Responsabilidades
Pronóstico y detección de demanda	- Desarrolla algoritmos estadísticos avanzados.	<ul style="list-style-type: none"> - Utiliza analítica avanzada para la captura y análisis de señales de demanda, transfórmandolas en insumos para los modelos de pronósticos. - Revisa y evalúa regularmente los pronósticos de series de tiempo y de regresión lineal simple. - Desarrolla modelos de regresión lineal múltiple junto al analista de planeamiento de la demanda y los stakeholders. - Desarrolla instrumentos y herramientas para monitorear y evaluar la exactitud de los modelos de pronóstico estadístico y cualitativo. - Asegura formas de incorporar la inteligencia de mercado más actual y exacta en la generación de pronósticos estadísticos. - Provee capacitación y entrenamiento a los analistas de planeamiento de la demanda en la identificación, monitoreo y mitigación de los problemas, y a su vez en la identificación y ejecución de mejores prácticas del proceso. - Define reglas de desagregación semanal que reflejen los patrones de mercado.
Dar forma a la demanda y desarrollar el plan integrado de negocios	- Desarrolla algoritmos para mejorar la colaboración de demanda y el planeamiento de escenarios.	<ul style="list-style-type: none"> - Lidera y promueve el uso del demand shaping mediante el uso de técnicas de análisis de escenarios 'what if' en la generación de pronósticos. - Integra en el planeamiento y en la generación de pronósticos el enfoque de gestión basada en excepciones como una forma de mejorar la utilización de los recursos de la compañía.
Acuerdos y comunicación	- Acuerda planes de mejora. Acuerda planes de aprendizaje y desarrollo.	<ul style="list-style-type: none"> - Llega a un consenso sobre la forma de medir el desempeño del pronóstico y de vincularlo a los planes integrados de negocios y a los objetivos. - Establece y monitorea un conjunto integrado de indicadores de gestión que abarca la demanda y el suministro a cada nivel de agregación del pronóstico. - Desarrolla instrumentos para medir la efectividad, eficiencia y adherencia al proceso. - Lidera a la organización a introducir pronósticos diferenciados, alejando a la compañía de una visión "una talla se ajusta para todos". - Asegura que las mejoras del proceso sean comunicadas.
Gestión de datos, sistemas y mejora continua	- Capacita, entrena y lidera el intercambio de conocimiento. Asegura una adherencia global del proceso.	<ul style="list-style-type: none"> - Identifica y monitorea las fortalezas y debilidades de los diferentes modelos estadísticos usados en la generación de pronósticos. - Coordina con los analistas de planeamiento de la demanda la realización de análisis de causa raíz y la planificación de acciones correctivas para entender la exactitud de pronósticos y sus sesgos. - Provee análisis científicos detallados de los errores de pronósticos, tendencias, y causas raíces para los stakeholders locales, regionales y globales. - Propone algún uso potencial de ARIMAX o de otros modelos estadísticos avanzados en la generación de pronósticos. - Monitorea las mejores prácticas de generación de pronósticos y el uso de tecnología. - Lidera el desarrollo e intercambio del conocimiento. Administra un wiki interno, foros de discusión y las inducciones del proceso. - Lidera a la organización hacia la digitalización de los procesos de generación de pronósticos. - Regularmente prepara análisis personalizados sobre los resultados del pronóstico como un insumo para el establecimiento de objetivos de error de pronóstico. - Comunica a la gerencia las desviaciones más importantes en resultados, adherencia al proceso y calidad de datos. - Desarrolla técnicas de generación de pronósticos ex-post para mejorar los algoritmos empleados. - Utiliza técnicas de minería de datos para producir hallazgos útiles al proceso. - Mantiene la documentación actualizada de los procesos. - Lidera el enfoque basado en demanda y el uso de datos de consumo sindicados de Nielsen o de otra fuente si es aplicable. - Lidera la consistencia del proceso y de los datos en toda la organización.

Fuente: Sankaran *et al.*, 2019.