MODELO PREDICTIVO DEL RIESGO DE ABASTECIMIENTO EN LA CADENA DE SUMINISTRO BASADO EN TÉCNICAS DE MACHINE LEARNING

Mauro Steban García Largo
Código 9027802

Proyecto Aplicado para optar al título de
Magíster en Ciencia de Datos

Directora
PhD. Isabel Cristina García Arboleda

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, MAYO 2026

Tabla de contenido

1. Introducción

La creciente volatilidad de los mercados globales ha impulsado una transformación significativa en la gestión logística mundial, imponiendo desafíos cada vez más complejos a la planificación del abastecimiento [1]. En industrias como la alimentaria, donde la disponibilidad de insumos y la demanda de productos presentan una alta variabilidad, la capacidad de anticipar interrupciones en la cadena de suministro se ha convertido en un factor determinante para la competitividad empresarial [2]. La ciencia de datos surgió como una herramienta estratégica para afrontar dichas variaciones, al permitir el análisis de grandes volúmenes de información histórica y la generación de modelos predictivos que coadyuven a la toma de decisiones informadas [3].

La organización objeto de estudio es una multinacional dedicada a la producción de alimentos con para el consumo humano. En el momento de la investigación, el abastecimiento de materiales de empaque, insumos e ingredientes se gestionaba mediante el modelo *Demand Driven Material Requirements Planning* (*DDMRP*), orientado a minimizar inventarios y operar bajo el enfoque de justo a tiempo [4]. Si bien este esquema había demostrado ser efectivo en condiciones de mercado estables, su naturaleza predominantemente reactiva resultaba insuficiente frente a escenarios de alta incertidumbre logística y fluctuaciones significativas en la demanda [5]. Esta situación había derivado en sobrecostos, retrasos en las entregas e incumplimientos con clientes, lo que evidenciaba la necesidad de transitar hacia enfoques de gestión proactiva del riesgo de abastecimiento.

El proyecto tuvo como propósito desarrollar un modelo predictivo de riesgo de abastecimiento basado en técnicas de aprendizaje automático (*Machine Learning*), con el fin de anticipar posibles interrupciones en el suministro y apoyar la toma de decisiones estratégicas en el área de compras y planeación. Para ello, la investigación contempló la identificación de las variables críticas que inciden en el riesgo asociado a proveedores, la comparación de diversos algoritmos de predicción para seleccionar el de mejor desempeño, y la integración de los resultados en tableros de control interactivos que faciliten su interpretación por parte de los equipos operativos.

La metodología empleada se fundamentó en el enfoque *CRISP-DM*, un marco estructurado ampliamente utilizado en proyectos de análisis de datos [6]. Se trabajó con datos históricos extraídos de los sistemas internos de la empresa, tales como el *ERP* y las bases de gestión de compras, que contenían información sobre tiempos de entrega, cumplimiento de órdenes, niveles de inventario y consumo de materiales. A través de este conjunto de datos, se buscó construir y evaluar modelos predictivos que permitieran generar alertas tempranas frente a eventos de desabastecimiento, contribuyendo así a fortalecer la resiliencia de la cadena de suministro y a la optimización de los costos operativos.

Es importante resaltar que, si bien el análisis predictivo del riesgo de abastecimiento ha sido abordado desde distintas perspectivas en la literatura, la naturaleza específica de las cadenas de suministro en la industria alimentaria plantea desafíos adicionales que requieren enfoques adaptados a sus particularidades [7]. El impacto de factores externos, como la variabilidad en los precios de materias primas, las restricciones logísticas internacionales y los cambios en la demanda del consumidor, sugiere que el estudio del riesgo de abastecimiento debe considerar no solo datos históricos internos, sino también variables contextuales que influyen en el comportamiento de los proveedores [8]. En este sentido, los desarrollos futuros en el modelado predictivo podrían beneficiarse de la incorporación de fuentes de datos complementarias, como indicadores macroeconómicos o métricas de desempeño sectorial, que amplíen la capacidad explicativa de los modelos.

2. Contextualización del Problema

2.1 Definición del Problema

La gestión del riesgo de abastecimiento constituye uno de los desafíos más relevantes en la administración de cadenas de suministro contemporáneas, particularmente en industrias donde la continuidad operativa depende de la disponibilidad oportuna de materiales e insumos [7]. En un entorno globalizado, caracterizado por la volatilidad de los mercados, la incertidumbre logística y la creciente interdependencia entre actores de la cadena, las organizaciones se enfrentan a una exposición cada vez mayor a eventos de desabastecimiento que pueden comprometer su eficiencia productiva y su capacidad de respuesta ante la demanda [1].

Tradicionalmente, la planificación del abastecimiento se ha apoyado en metodologías como el *Material Requirements Planning* (*MRP*) y su evolución hacia el *Demand Driven Material Requirements Planning (DDMRP*), las cuales han demostrado ser efectivas para optimizar niveles de inventario en condiciones de mercado relativamente estables [4]. No obstante, estos enfoques presentan limitaciones significativas cuando se aplican en escenarios de alta variabilidad, dado que su naturaleza dificulta la anticipación de disrupciones en el suministro [5]. Esta situación ha impulsado la necesidad de complementar dichas metodologías con herramientas analíticas que permitan predecir y gestionar de forma proactiva los riesgos asociados al abastecimiento.

La ciencia de datos y, en particular, las técnicas de aprendizaje automático (*Machine Learning*) han emergido como alternativas prometedoras para abordar esta problemática. Diversos estudios han demostrado que los modelos predictivos basados en permiten identificar patrones en datos históricos de la cadena de suministro, anticipar eventos críticos y generar alertas tempranas que contribuyen a la toma de decisiones informadas [1] [3]. No obstante, la aplicación de estas técnicas en contextos específicos, como la industria alimentaria, requiere una adaptación cuidadosa a las particularidades del sector, lo que justifica el desarrollo de investigaciones aplicadas que evalúen su efectividad en entornos reales [2].

2.2 Planteamiento del Problema

La organización objeto de estudio es una empresa multinacional dedicada a la producción de alimentos para el consumo humano. En el momento de la investigación, el abastecimiento de materiales de empaque, insumos e ingredientes se gestionaba mediante el modelo *DDMRP*, orientado a minimizar inventarios y operar bajo el enfoque de justo a tiempo [4]. Si bien este esquema había permitido alcanzar niveles de servicio aceptables en condiciones de mercado estables, su desempeño se veía comprometido frente a escenarios de alta volatilidad en los precios de materias primas, variabilidad logística internacional y fluctuaciones significativas en la demanda.

Esta situación había generado consecuencias operativas concretas: retrasos en las entregas de proveedores, problemas de calidad en los materiales recibidos, restricciones en la capacidad de almacenamiento interno y cambios inesperados en la planificación de la producción. En conjunto, estos factores habían derivado en sobrecostos, incumplimientos con clientes y una degradación progresiva de la eficiencia productiva [9]. A pesar de contar con sistemas de monitoreo de inventarios integrados en el *ERP* corporativo, la organización carecía de mecanismos que permitieran anticipar los riesgos de suministro de forma sistemática. La toma de decisiones en el área de compras seguía siendo predominantemente reactiva, lo que incrementaba la exposición a fallos críticos en la cadena de abastecimiento.

Desde la perspectiva de la ciencia de datos, la organización disponía de un volumen significativo de información histórica almacenada en sus sistemas internos, incluyendo datos sobre niveles de inventario, tiempos de entrega, desempeño de proveedores, órdenes de compra y consumo de materiales. Esta base de datos ofrecía la posibilidad de analizar patrones, identificar factores de riesgo y construir modelos analíticos que permitieran comprender y anticipar eventos de desabastecimiento [8]. Dicha información no era utilizada de forma estructurada para apoyar decisiones estratégicas. La ausencia de mecanismos basados en analítica predictiva representaba una brecha crítica en la gestión de riesgos de suministro, especialmente en un entorno caracterizado por la incertidumbre y la interdependencia global [7].

Se identificó la necesidad de desarrollar un modelo predictivo basado en técnicas de *Machine Learning* que, mediante los datos históricos disponibles, permitiera anticipar y gestionar de forma proactiva los riesgos asociados al suministro de materiales críticos. Este modelo debía no solo generar predicciones confiables, sino también integrarse en herramientas de visualización accesibles para los equipos de compras y planeación, de manera que sus resultados pudieran traducirse en acciones operativas concretas.

2.3 Formulación del Problema

Desde la aparición y consolidación de las metodologías de planificación basadas en demanda, las soluciones de abastecimiento se han expandido constantemente. La evidencia empírica sugiere que incluso los enfoques más avanzados, como el *DDMRP*, presentan limitaciones para anticipar disrupciones en entornos de alta volatilidad [5]. Al mismo tiempo, la creciente disponibilidad de datos transaccionales en los sistemas empresariales ha abierto la posibilidad de aplicar técnicas de *Machine Learning* para complementar estos enfoques, generando capacidades predictivas que permitan una gestión proactiva del riesgo de abastecimiento [1] [3].

Pregunta general: ¿Cuál es el modelo predictivo basado en técnicas de *Machine Learning* más efectivo para anticipar los riesgos de desabastecimiento de materiales de empaque, insumos e ingredientes en la empresa caso de estudio?

2.3.1 Sistematización

- ¿Cuáles son las variables más relevantes para explicar el nivel de riesgo de un proveedor?
- ¿Qué técnica de *Machine Learning* ofrece un rendimiento óptimo al predecir el nivel de riesgo?
- ¿Cómo pueden utilizarse los resultados del modelo para apoyar las decisiones del área de abastecimiento?

3. Objetivos

3.1 Objetivo General

Implementar un modelo predictivo basado en técnicas de *Machine Learning* que permita anticipar los riesgos de desabastecimiento de materiales de empaque, insumos e ingredientes, mediante el análisis de variables operativas.

3.2 Objetivos Específicos

- Identificar y analizar las variables con mayor relevancia para explicar el nivel de riesgo de desabastecimiento asociado a los proveedores, considerando factores como tiempos de entrega, cumplimiento de órdenes, niveles de inventario y volatilidad de la demanda
- Evaluar cuál es el mejor modelo predictivo con base en su rendimiento evidenciado a través de las métricas de desempeño
- Diseñar un tablero de control interactivo que consolide los resultados del modelo seleccionado y genere alertas tempranas sobre productos en riesgo, facilitando su interpretación y uso por parte de los equipos de compras y planeación.

4. Marco de Referencia

El desarrollo de un sistema de pronóstico de demanda para ingredientes industriales requiere combinar conocimiento de dos dominios: la gestión de la cadena de suministro, que aporta el contexto operativo y las restricciones del problema, y el aprendizaje automático, que provee las herramientas analíticas para el modelado. Las secciones siguientes cubren ambos dominios.

4.1 Cadena de Suministro y Abastecimiento

La cadena de suministro abarca la planificación y gestión de todas las actividades asociadas a la adquisición, abastecimiento y transformación de materiales, así como la coordinación logística entre las partes que la conforman [10]. El abastecimiento, en particular, comprende el ciclo completo desde que se origina un requerimiento de materiales hasta su cumplimiento mediante la entrega del bien o servicio [11].

Una gestión eficiente de la cadena de suministro requiere considerar la planificación integral del flujo de materiales, la evaluación de riesgos y beneficios, la calidad de la información disponible y el relacionamiento con proveedores y clientes [10]. Los cambios en la logística mundial han incrementado la complejidad de estas cadenas, especialmente por la variedad de proveedores disponibles, las variaciones inesperadas en la demanda y la necesidad de sincronización entre los actores [12].

El análisis de grandes volúmenes de datos mediante técnicas de inteligencia artificial permite identificar tendencias en el comportamiento de la demanda y del suministro que serían difíciles de detectar con métodos manuales [13]. Esta capacidad de análisis ha motivado aplicarlo al ámbito del abastecimiento, especialmente en industrias donde la disponibilidad oportuna de materiales condiciona la continuidad operativa.

4.2 Gestión de Compras

La gestión de compras se define como el proceso mediante el cual las organizaciones establecen sus necesidades de adquisición de productos y servicios, e identifican, evalúan y seleccionan entre

proveedores alternativos [14]. Su función principal consiste en obtener los materiales adecuados con la calidad correcta, en la cantidad correcta, al precio correcto y del proveedor correcto, en el momento y lugar oportunos [15].

En la organización objeto de estudio, la función de compras cobra especial importancia porque la adquisición de materiales de empaque, insumos e ingredientes se realiza a través de una red diversa de proveedores nacionales e internacionales. Esta diversidad, si bien amplía las opciones de suministro, incrementa la exposición a riesgos derivados de la variabilidad en los tiempos de entrega, las fluctuaciones en los precios de materias primas y las restricciones logísticas del comercio internacional [16]. Poder anticipar disrupciones en el proceso de compras es, por tanto, un factor clave para la estabilidad operativa.

4.3 Gestión de Inventarios

Un inventario es la existencia de bienes mantenidos para su uso o venta en el futuro [17]. La gestión de inventarios consiste en garantizar la disponibilidad de esos bienes cuando se requieren, definiendo políticas de reabastecimiento que permitan un movimiento constante de los materiales almacenados. Uno de los retos más frecuentes es el desbalance entre oferta y demanda: exceso de lo que no se consume y agotados de lo que sí se requiere [16].

Las políticas de inventarios definen la frecuencia de revisión, los puntos de reorden y las cantidades a solicitar [18]. La organización estudiada opera bajo el enfoque *DDMRP*, que establece puntos de desacoplamiento estratégicos y buffers dinámicos para absorber la variabilidad de la demanda y del suministro [4]. La metodología *DDMRP* presenta limitaciones frente a escenarios de alta volatilidad [5], lo que motivó la búsqueda de herramientas predictivas que lo complementen con base en datos históricos.

4.4 Gestión de Proveedores

Los proveedores son aquellos actores de la cadena de suministro que poseen un bien, servicio o conocimiento que pueden suministrar para atender los requerimientos de un cliente. La relación con los proveedores es una pieza clave: a través esta interacción es posible construir confianza y compromiso, generando beneficios tangibles para la gestión de la cadena [19].

Al ser el proveedor un eslabón central, se requieren mecanismos formales que habiliten una relación estructurada y medible [20]. La evaluación del desempeño de los proveedores, es decir, cumplimiento de entregas, calidad de los materiales, consistencia en los tiempos de respuesta, sirve como insumo para anticipar riesgos en el abastecimiento.

4.5 Riesgo de Desabastecimiento

El riesgo de desabastecimiento se define como la probabilidad de que una organización no pueda disponer oportunamente de los materiales, insumos o servicios necesarios para mantener la continuidad de sus operaciones productivas [20]. Sus causas incluyen la variabilidad en los tiempos de entrega de los proveedores, las fluctuaciones imprevistas en la demanda, las restricciones logísticas internacionales, los cambios en los precios de materias primas y la concentración excesiva del suministro en pocos proveedores [16].

Las disrupciones en el suministro generan consecuencias como la degradación de la eficiencia productiva, el incremento en los costos operativos, los cambios inesperados en la planificación y el incumplimiento con los clientes [9]. Las empresas que experimentan disrupciones frecuentes acumulan pérdidas financieras directas y pierden capacidad de respuesta ante la demanda del mercado [7][1].

En la industria alimentaria, la perecibilidad de los insumos, la dependencia de proveedores internacionales para ingredientes especializados y la estacionalidad de ciertas materias primas intensifican la exposición al desabastecimiento [2]. Estas condiciones hacen necesario el desarrollo de herramientas analíticas capaces de anticipar riesgos mediante modelos predictivos basados en datos históricos de la cadena de suministro.

En este trabajo, el riesgo de desabastecimiento se aborda a través del pronóstico de la demanda de ingredientes: cuando la demanda muestra señales claras de variaciones en el consumo, el sistema genera una alerta. Este enfoque permite pasar de una gestión reactiva a una proactiva, anticipando las necesidades de compra con base en la trayectoria de la demanda.

4.6 Fundamentos del Aprendizaje Automático

El aprendizaje automático (*Machine Learning*) agrupa algoritmos que aprenden patrones a partir de datos sin ser programados explícitamente para cada caso. El problema abordado en esta investigación (pronosticar la demanda futura de cada ingrediente y detectar tendencias que alerten sobre desequilibrios en el abastecimiento) se apoya en dos paradigmas: el aprendizaje supervisado, que estima cantidades de demanda mediante series históricas, y el aprendizaje no supervisado, que segmenta el portafolio de ingredientes mediante técnicas de agrupamiento.

4.6.1 Modelos de Pronóstico de Demanda

Se seleccionaron tres algoritmos con paradigmas distintos entre sí: un modelo de descomposición temporal, uno de *ensemble* no lineal y una línea base regularizada, para que la comparación abarcara enfoques con supuestos estructurales diferentes y no solo variantes de un mismo método [29].

Tabla 1. Modelos de pronóstico para la predicción de demanda de ingredientes

Modelo	Descripción
Prophet	Modelo *aditivo* de pronóstico desarrollado por Taylor y Letham [41]. Descompone la serie en tres componentes (tendencia, estacionalidad y efecto de eventos especiales) estimados mediante inferencia bayesiana a través de *Stan*. La formulación $y(t) = g(t) + s(t) + h(t) + \varepsilon$ captura estacionalidad semanal y anual mediante series de *Fourier*, y proporciona intervalos de credibilidad nativos que cuantifican la incertidumbre de las predicciones [41].
XGBoost (*Extreme Gradient Boosting*)	Algoritmo de *ensemble* basado en *gradient boosting*: construye secuencialmente árboles de decisión donde cada árbol corrige los residuos del anterior. Su función objetivo combina pérdida empírica con regularización L1 y L2 sobre los pesos de las hojas, lo que controla el sobreajuste sin poda externa. Aquí se emplea como regresor (no como clasificador), recibiendo como variables predictoras los rezagos de la serie y estadísticos de ventana deslizante [27].
Ridge Regression	Variante de la regresión lineal con penalización L2 sobre los coeficientes ($L = \|\|y - X\beta\|\|^2 + \alpha \cdot \|\|\beta\|\|^2$). Cuando se tabulan series de tiempo mediante rezagos, las columnas presentan multicolinealidad por construcción; la penalización L2 estabiliza la inversión de $X^\top X$ y produce coeficientes más robustos. Cumple la función de línea base contra la cual se mide si la complejidad de los otros modelos aporta mejoras reales [58].

Fuente: elaboración propia.

4.6.2 Modelos de Agrupamiento No Supervisado

Los algoritmos de aprendizaje no supervisado detectan agrupaciones naturales en datos que no están previamente etiquetados [28]. Antes de entrenar los modelos de pronóstico, se segmentó el portafolio de ingredientes para identificar grupos con patrones de demanda similares y poder asignar a cada grupo la familia de modelos más adecuada.

Se compararon dos algoritmos con supuestos geométricos opuestos:

Tabla 2. Modelos de agrupamiento seleccionados

Modelo	Descripción
K-Means	Algoritmo de partición que divide los datos en k grupos mutuamente excluyentes, minimizando la inercia (suma de distancias cuadradas al centroide más cercano). Asume clusters convexos de tamaño similar y requiere que k se fije de antemano. Computacionalmente eficiente, se usa ampliamente como referencia en tareas de segmentación [29].
DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*)	Algoritmo basado en densidad que identifica clusters como regiones de alta densidad separadas por regiones de baja densidad. No requiere especificar el número de grupos a priori, puede descubrir agrupaciones no convexas y etiqueta como ruido los puntos que no pertenecen a ninguna región densa, lo que lo hace más tolerante ante valores atípicos. El parámetro eps (r*adi*o de vecindad) puede estimarse automáticamente mediante el método del codo en la curva de k-distancias [51].

Fuente: elaboración propia.

4.6.3 Métodos de Ensamble

Los métodos de ensamble (*ensemble* learning) combinan múltiples modelos para obtener predicciones superiores a las de cualquier modelo individual [30]. Dos familias son relevantes para este trabajo:

Bagging (*Bootstrap Aggregating*) busca reducir la varianza. Algoritmos como *Random Forest* construyen múltiples árboles de decisión de forma independiente sobre subconjuntos aleatorios de los datos y agregan sus predicciones, lo que mitiga el sobreajuste típico de los árboles individuales [30].

Boosting busca reducir tanto el sesgo como la varianza. *XGBoost* construye el modelo de forma secuencial, donde cada nuevo árbol corrige los errores del anterior. La combinación de

regularización L1 y L2 con submuestreo de filas y columnas introduce un efecto de *bagging* implícito dentro del proceso de *boosting*, lo que mitiga el sobreajuste sin sacrificar la capacidad para capturar relaciones no lineales [26].

El torneo de pronóstico*s* implementado en esta tesis aprovecha ambas familias: *XGBoost* (*boosting*) compite directamente contra *Prophet* y *Ridge*, lo que permite determinar empíricamente si la corrección iterativa de errores aporta ventajas sobre los otros dos par*adi*gmas para cada ingrediente del portafolio.

4.6.4 Métricas de Evaluación de los Modelos

Se emplearon dos conjuntos de métricas: uno para los modelos de pronóstico y otro para los de agrupamiento, dado que el pronóstico produce valores continuos cuyo error se mide contra observaciones reales, mientras que el agrupamiento produce particiones evaluadas por cohesión y separación interna [32].

Tabla 3. Métricas de evaluación para los modelos de pronóstico de demanda

Métrica	Descripción	Rol en el sistema
RMSE (*Root Mean Squared Error*)	Raíz cuadrada del promedio de los errores al cuadrado entre valores predichos y reales. Penaliza errores grandes de forma cuadrática y se expresa en las mismas unidades que la variable original.	Métrica de selección del torneo. El modelo con menor *RMSE* promedio en validación *walk-forward* se declara ganador para cada producto [34].
MASE (*Mean Absolute Scaled Error*)	Cociente entre el *MAE* del modelo y el *MAE* de un pronóstico *naive* ($\hat{y}_t = y_{t-1}$). Independiente de la escala y simétrico ante errores positivos y negativos. Valores < 1 indican que el modelo supera al *naive*; ≥ 1, que no aporta valor sobre la regla más simple.	Métrica de diagnóstico. No participa en la selección del ganador, pero identifica productos cuyo modelo no supera al *naive* [61].
WMAPE (*Weighted Mean Absolute Percentage Error*)	Error porcentual absoluto ponderado por la magnitud real de cada observación. Asigna mayor peso a los periodos de alta demanda y evita la inestabilidad del MAPE cuando los valores reales son cercanos a cero.	Métrica de validación del pronóstico final. Permite comparar el desempeño entre productos con escalas de demanda diferentes [64].
MAE (*Mean Absolute Error*)	Promedio de los errores absolutos entre predicción y valor real, expresado en	Complemento operativo del *WMAPE*. Proporciona al planificador una

	unidades físicas (kilogramos, litros, unidades).	lectura directa del error promedio para dimensionar márgenes de seguridad en las órdenes de compra.

Fuente: elaboración propia.

Tabla 4. Métricas de evaluación para los modelos de agrupamiento

Métrica	Descripción	Criterio
Coeficiente de Silueta (*Silhouette Score*)	Mide la cohesión intra-cluster y la separación inter-cluster para cada observación. Varía de –1 a 1; valores cercanos a 1 indican agrupamientos bien definidos.	Mayor es mejor. Se usa además como desempate cuando los algoritmos empatan en el torneo [35].
Índice de *Davies-Bouldin* (DBI)	Evalúa la razón entre la dispersión interna de cada cluster y la distancia entre centroides. Un valor más bajo indica clusters densos y bien separados.	Menor es mejor [36].
Índice de *Calinski-Harabasz* (CHI)	Razón de la varianza inter-cluster sobre la varianza intra-cluster. Valores altos indican clusters compactos y bien separados entre sí.	Mayor es mejor [35].

Fuente: elaboración propia.

4.7 Demanda Intermitente y Clasificación de Patrones de Consumo

En las cadenas de suministro de la industria alimentaria es frecuente encontrar ingredientes cuyo consumo no ocurre de manera continua ni regular: algunos se demandan todos los días, otros solo ciertos meses, y otros presentan pedidos esporádicos de volúmenes muy variables. Esta heterogeneidad tiene implicaciones directas para la selección de métodos de pronóstico, dado que los algoritmos de suavizamiento exponencial estándar asumen demanda continua y pierden efectividad cuando la serie contiene periodos prolongados con demanda nula [48].

4.7.1 Esquema de Clasificación Syntetos-Boylan-Croston (SBC)

Syntetos y Boylan [48] propusieron un esquema de clasificación que se ha convertido en el referente estándar para categorizar patrones de demanda intermitente [48]. Opera sobre dos ejes: el *ADI* (*Average Demand Interval*), que mide el intervalo medio entre periodos con demanda positiva, y el CV^2 (coeficiente de variación al cuadrado de los tamaños de pedido). El cruce de

ambos ejes con los umbrales de la literatura (ADI = 1.32, CV^2 = 0.49) define cuatro cuadrantes:

Tabla 5. Cuadrantes de clasificación *SBC*

***ADI* / CV^2**	**$CV^2 < 0.49$**	**$CV^2 \geq 0.49$**
$ADI < 1.32$ (frecuente)	***Smooth***: demanda regular y predecible. Adecuada para suavizamiento exponencial.	***Erratic***: frecuente, pero con volúmenes variables. Requiere modelos que capturen variabilidad en el tamaño de los pedidos.
$ADI \geq 1.32$ (esporádica)	***Intermittent***: esporádica, pero con volúmenes estables cuando ocurre. Candidata a métodos de Croston.	***Lumpy***: esporádica y con volúmenes variables. La categoría más difícil de pronosticar; requiere métodos especializados como *SBA* o *TSB* [49].

Fuente: adaptado de Syntetos y Boylan en [48]

Clasificar el portafolio antes de entrenar los modelos permite saber de antemano qué proporción de ingredientes se ajusta al supuesto de demanda continua y qué proporción requiere tratamiento diferenciado. Esa información guía la selección de modelos y la interpretación de los resultados del torneo.

4.7.2 Exponente de Hurst: Memoria de Largo Plazo

El exponente de Hurst, introducido por Harold Hurst en 1951 para el estudio de niveles del río Nilo, cuantifica la dependencia de largo alcance en una serie temporal [50]. Se estima mediante el método del rango reescalado (R/S): se calcula el rango acumulado de las desviaciones respecto a la media, normalizado por la desviación estándar, para bloques de tamaño creciente, y la pendiente de la regresión log(R/S) contra log(n) da la estimación de H.

Valores de H superiores a 0.5 indican persistencia; valores inferiores a 0.5, antipersistencia; y valores cercanos a 0.5, comportamiento aleatorio. Este indicador aporta una dimensión que el CV y la intermitencia no capturan la estructura de dependencia temporal, y puede orientar la selección hacia modelos capaces de capturar tendencias sostenidas [70].

4.7.3 Entropía de Shannon

La entropía de Shannon, proveniente de la teoría de la información, mide la uniformidad de una distribución [47]. Aplicada a la serie de demanda normalizada de un ingrediente, una entropía alta indica consumo distribuido de manera homogénea en el tiempo; una entropía baja, concentración en pocos periodos. Dos series pueden tener el mismo CV pero distribuciones temporales muy diferentes, lo que hace de la entropía un complemento útil para la caracterización del portafolio.

4.8 Tratamiento de Datos y Validación en Series de Tiempo

Los datos transaccionales de la industria alimentaria presentan desafíos específicos para el análisis de series de tiempo: heterogeneidad de las fuentes, convenciones contables de signo invertido, diversidad de codificaciones y valores atípicos generados por pedidos extraordinarios o errores de digitación. Los métodos que se describen a continuación abordan estos desafíos.

4.8.1 Estadísticos Robustos: la Desviación Absoluta Mediana y el Identificador de Hampel

El método más difundido para detectar valores atípicos en distribuciones univariantes es el filtro de Tukey, basado en el rango intercuartílico (*IQR*). El *IQR* tiene un punto de quiebre del 25%: si más del 25% de las observaciones están contaminadas, el estimador se invalida [43]. En series de demanda de ingredientes, donde es frecuente encontrar múltiples valores extremos simultáneos, esta limitación puede producir enmascaramiento: los *outliers* inflan el *IQR* y pasan inadvertidos [44].

La Desviación Absoluta Mediana (*MAD*), definida como *MAD* = mediana (|xi – mediana(x)|), alcanza un punto de quiebre del 50%, el máximo posible para un estimador de escala. El identificador de Hampel clasifica como atípica toda observación cuya desviación respecto a la mediana supere un múltiplo t de la *MAD* normalizada (factor de corrección k = 1.4826 para consistencia bajo normalidad). Para series con alta asimetría y *outliers* múltiples, este método es más adecuado que el filtro de Tukey [45].

4.8.2 Validación Walk-Forward para Series de Tiempo

Evaluar modelos predictivos en series de tiempo requiere respetar la estructura temporal de los datos. La validación cruzada estándar con particiones aleatorias (*k-fold*) puede incurrir en fuga de información (*data leakage*) al utilizar datos futuros para entrenar, lo que produce estimaciones de error artificialmente optimistas [55].

La validación *walk-forward* evita este problema: cada fold de test contiene exclusivamente observaciones posteriores a las del fold de entrenamiento. En cada iteración, el modelo se entrena con todos los datos hasta el punto de corte y se evalúa sobre los días siguientes. El error reportado es el promedio del *RMSE* de cada fold, lo que reduce la sensibilidad a la elección de una ventana particular de evaluación [60].

4.8.3 Detección de Tendencias: Test de Mann-Kendall y Pendiente de Sen

El test de Mann-Kendall es un test no paramétrico para tendencia monotónica [65]. Al basarse en la concordancia de rangos y no en los valores numéricos, tolera violaciones del supuesto de normalidad y es resistente a valores atípicos. La alternativa más simple, la cual consiste en ajustar una regresión lineal y evaluar la pendiente, asume normalidad de los residuos y es sensible a *outliers*: un único periodo con demanda anormalmente alta puede generar una pendiente artificialmente positiva y disparar una alerta espuria.

La magnitud de la tendencia se cuantifica con la pendiente de Sen, definida como la mediana de todas las pendientes calculadas entre pares de puntos [65]. Al usar la mediana en lugar de la media, hereda la robustez del test de Mann-Kendall frente a valores extremos.

4.9 Metodología *CRISP-DM*

Se adoptó la metodología *CRISP-DM*, un enfoque estructurado para proyectos de ciencia de datos que divide el ciclo de vida en seis fases iterativas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación [6][38].

La naturaleza cíclica de *CRISP-DM* permite retroceder a fases anteriores cuando los resultados de

una etapa posterior lo requieren. En la práctica, el *pipeline* de este proyecto iteró varias veces entre la ingeniería de características, la segmentación del portafolio y el entrenamiento de modelos antes de converger a la arquitectura final.

4.10 Antecedentes

La tabla siguiente resume investigaciones relacionadas con el pronóstico de demanda y la predicción de riesgos en cadenas de suministro mediante *Machine Learning*.

Tabla 6. Principales estudios relacionados con la predicción del riesgo de abastecimiento

Estudio	Contribución
Modelo para la predicción de compra de materia prima en la gestión de inventario para producción en Pymes del sector retail. Salazar Vera y Antúnez Cueva [72]	Desarrollaron un modelo predictivo para la compra de materias primas en Pymes del sector retail, evaluando *Random Forest*, *Gradient Boost* y *Decision Tree*. Seleccionaron Regresión Lineal *Elastic Net* con base en *MAE*, *RMSE* y R^2. Validaron la viabilidad de utilizar ML para optimizar la gestión de inventarios y confirmaron la pertinencia de incorporar datos de riesgo y tiempos de entrega como variables predictoras [72].
Deconstructing Risk Factors for Predicting Risk Assessment in Supply Chains Using *Machine Learning*. Burstein y Zuckerman [73]	Categorizaron los tipos de riesgo que afectan la cadena de suministro y entrenaron un modelo de red neuronal con datos de encuestas a auditores, obteniendo R = 0.951 y *RMSE* = 0.00244. Su aporte fue la estructuración de una matriz de categorización de factores de riesgo [73].
Predicting supply chain risks using *machine learning*: The trade-off between performance and interpretability. Baryannis, Dani y Antoniou [74]	Evaluaron modelos de ML para la predicción de riesgos en cadena de suministro, con foco en el equilibrio entre rendimiento e interpretabilidad. Concluyeron que la selección del modelo no debe basarse solo en la máxima precisión, sino también en que los usuarios finales puedan comprender los resultados [74].

Fuente: elaboración propia.

Estos antecedentes confirmaron que es viable construir modelos predictivos a partir de datos transaccionales internos [20], que la categorización de factores de riesgo es un insumo útil para el modelado [21] y que la interpretabilidad del modelo afecta directamente su adopción por los equipos operativos [24].

A diferencia de los trabajos reseñados, esta tesis se enfoca en el pronóstico de demanda de

ingredientes, utiliza datos transaccionales reales de alto volumen de una empresa del sector alimentario colombiano y combina agrupamiento no supervisado con un torneo de modelos que selecciona automáticamente el algoritmo óptimo para cada ingrediente.

5. Determinación de los Factores Críticos Que Explican el Riesgo de Desabastecimiento de Proveedores

5.1 Comprensión del Negocio

El primer paso para cumplir este objetivo fue la comprensión de la lógica y dinámica del modelo de negocio de la empresa. Este análisis resultó fundamental, dado que el sector alimenticio en Colombia ha presentado dinámicas de volatilidad no observadas anteriormente, lo que impacta directamente en la cadena de suministro [7]. La organización objeto de estudio opera bajo un esquema de múltiples plantas productivas dedicadas a la manufactura de alimentos, donde la disponibilidad oportuna de materiales de empaque, insumos e ingredientes no constituye meramente un requisito logístico, sino un factor determinante para la continuidad operativa.

Con base en este contexto, el desarrollo del objetivo se estructuró en dos actividades principales: la identificación y categorización de los factores teóricos que contribuyen al desabastecimiento y la cuantificación del impacto de dichos factores en la cadena de suministro de la empresa. A continuación, se detalla el desarrollo de cada actividad.

5.2 Identificación y Categorización de los Factores Que Contribuyen al Desabastecimiento

Diversas organizaciones han reconocido la necesidad de proteger sus cadenas de suministro ante interrupciones que generan consecuencias operativas y financieras significativas. Sin embargo, las soluciones tradicionalmente implementadas, como el incremento del nivel de inventario o la diversificación de fuentes de suministro, no siempre resultan suficientes para atender las variaciones del entorno global [25]. Aunque los responsables de la gestión de abastecimiento comprenden los beneficios de mantener un suministro continuo, los esfuerzos orientados a mitigar los riesgos de manera estructurada han sido limitados, situación que ha comenzado a transformarse al evidenciarse el impacto financiero que genera desproteger esta parte de la cadena [25].

La cadena de suministro generalmente comprende un gran número de productos que son adquiridos, manufacturados o almacenados en múltiples ubicaciones, lo que dificulta el control

integral de todos los ítems. Esta complejidad puede generar en una reducción de la eficiencia en la gestión diaria del riesgo de retrasos y fluctuaciones, incrementando la probabilidad de interrupciones. Teniendo en cuenta lo anterior, considerar los riesgos de manera sistemática se convierte en un factor clave que apunta a la eficiencia operacional [25].

Con base en la revisión de la literatura, se identificaron los principales factores que constituyen potenciales interruptores de la cadena de suministro [26]. La Tabla 7 presenta una categorización de estos factores, agrupados por su naturaleza y ámbito de afectación.

Tabla 7. Categorización de los factores de riesgo en la cadena de suministro

Categoría de riesgo	Descripción
Concentración de fuentes de suministro	Se refiere a la dependencia de un número reducido de proveedores o ubicaciones geográficas para el abastecimiento de materiales críticos. La concentración del suministro incrementa la vulnerabilidad ante interrupciones asociadas a bloqueos de puertos marítimos, restricciones en vías terrestres o pérdida de mercancía por desastres naturales. La distribución geográfica de las fuentes de suministro constituye una estrategia de mitigación frente a este riesgo [26].
Riesgos económicos	Factores como la especulación en el mercado de *commodities*, la volatilidad cambiaria y las fluctuaciones en los precios de materias primas generan un panorama incierto que puede causar variaciones significativas en el costo de los productos, incrementos en los costos logísticos e imposición de aranceles, entre otras consecuencias de alto impacto financiero [26].
Riesgos organizacionales	La falta de claridad en la segregación de roles dentro de la cadena de suministro, deficiencias en la planificación del abastecimiento y descoordinación entre los actores internos pueden generar interrupciones cuando las actividades secuenciales no se ejecutan conforme a lo planificado [26].
Riesgos laborales	Asociados a las actividades humanas dentro de la cadena, incluyen errores transaccionales, problemas de calidad derivados de la operación manual y la probabilidad de huelgas o paros laborales que pueden detener parcial o totalmente las operaciones [26].
Riesgos relacionales	Dentro de la gestión interna y de proveedores, la falta de confianza o comunicación entre socios comerciales puede limitar la capacidad de acceder a información oportuna y fiable, o derivar en el deterioro de relaciones con proveedores estratégicos, comprometiendo la estabilidad del suministro [26].

Riesgos de suministro	La dependencia de proveedores únicos o altamente especializados genera un riesgo de no poder acceder a alternativas en los tiempos y formas requeridos para garantizar la continuidad de las operaciones. Este riesgo se intensifica cuando los materiales presentan especificaciones técnicas que limitan la sustituibilidad [26].
Riesgos de manufactura y producción	Comprenden la baja capacidad instalada, la obsolescencia de equipos y la baja fiabilidad de los procesos productivos internos, los cuales pueden generar cuellos de botella que agravan el efecto de las interrupciones en el suministro [26].
Riesgos de inventario	Se relacionan con la obsolescencia de materiales almacenados y la variabilidad en los niveles de stock, que pueden derivar tanto en excesos como en rupturas de inventario. Una gestión inadecuada de las políticas de reabastecimiento incrementa la exposición al desabastecimiento [26].
Riesgos de demanda y mercado	Incluyen la volatilidad en la demanda del consumidor final, el efecto látigo (*Bullwhip Effect*) que amplifica las variaciones a lo largo de la cadena, y la potencial pérdida de cuota de mercado derivada de incumplimientos reiterados [26].
Riesgos logísticos	Abarcan las restricciones en rutas de transporte, deficiencias en la infraestructura vial o portuaria, controles fronterizos y bloqueos que dificultan el flujo oportuno de materiales desde los proveedores hasta las plantas de producción [26].
Riesgos tecnológicos y de TIC	Comprenden los fallos en los sistemas de información que soportan la gestión de la cadena de suministro, los ciberataques que pueden comprometer datos transaccionales y las deficiencias en la integración entre plataformas tecnológicas de los diferentes actores [26].
Riesgos ambientales y regulatorios	Engloban los desastres naturales, las pandemias, los conflictos geopolíticos, los cambios en las políticas comerciales y las modificaciones en el marco regulatorio que pueden alterar de manera súbita las condiciones del entorno en el que opera la cadena de suministro [26].

Fuente: elaboración propia.

5.3 Cuantificación del Impacto Asociado al Desabastecimiento en la Cadena de Suministro

La gestión de la cadena de suministro en la industria de alimentos, bajo un ecosistema de alta complejidad caracterizado por la volatilidad de los *commodities*, la rigurosidad en los estándares de calidad fisicoquímica y la necesidad de sincronización precisa en los procesos de manufactura. Para la organización objeto de estudio, que opera bajo un esquema de múltiples plantas productivas, la disponibilidad oportuna de insumos e ingredientes constituye un requisito fundamental para la

preservación de la continuidad operativa.

El desabastecimiento, entendido técnicamente como la ruptura de stock o la incapacidad de satisfacer la demanda interna de producción en el momento requerido, genera una disrupción que se propaga a lo largo de la cadena de abastecimiento. En cadenas de suministro maduras, el impacto de una ruptura de stock no es lineal, sino exponencial, propagándose aguas abajo hacia el cliente final y generando el fenómeno conocido como efecto látigo (*Bullwhip Effect*) [39]. En el contexto de la organización objeto de estudio, la ausencia de un ingrediente minoritario tiene el potencial de detener líneas completas de refinación o empaque, magnificando el costo del insumo faltante por un factor que puede superar su valor nominal, debido al costo de oportunidad y los costos operativos fijos que son absorbidos por un volumen reducido de producto terminado durante la afectación [27].

5.3.1 Impacto Operativo: Degradación de la Eficiencia General de los Equipos (*OEE*)

Desde una perspectiva operativa, el desabastecimiento en cualquiera de las plantas productivas impacta directamente el indicador de Eficiencia General de los Equipos. El proceso de manufactura es por defecto continuo o semicontinuo, lo que implica que la interrupción del flujo productivo por falta de materiales no constituye simplemente una pausa, sino una alteración termodinámica del sistema.

5.3.2 Impacto Financiero: Costos Directos, Ocultos y de Oportunidad

La cuantificación financiera del desabastecimiento trasciende el valor de compra del material no suministrado. Para efectos de esta investigación, el impacto económico se categoriza en tres niveles de afectación:

Los costos directos y de sustitución se generan cuando, ante la ausencia de un insumo homologado, la urgencia obliga a la adquisición de sustitutos en el mercado *spot*, frecuentemente a precios superiores que erosionan el margen bruto del producto terminado. A estos se suman los costos logísticos extraordinarios, como fletes aéreos o servicios de mensajería urgente, necesarios para mitigar el tiempo de entrega [39].

En segundo lugar, los costos de parada y capacidad ociosa comprenden el costo de la mano de obra directa inactiva y la depreciación de maquinaria durante el tiempo de inactividad forzosa. En la industria de lípidos, donde los márgenes suelen ser dependientes del volumen, la pérdida de economías de escala por paradas no programadas afecta de manera severa la estructura de costos unitarios [27].

En tercer lugar, el lucro cesante y las penalizaciones representan el impacto más significativo desde la perspectiva del nivel de servicio al cliente. Dado que la organización objeto de estudio opera predominantemente en el sector *B2B* (*Business-to-Business*), un desabastecimiento interno se traduce directamente en un incumplimiento de entrega a clientes industriales. El costo de perder un cliente industrial debido a la falta de fiabilidad en el suministro resulta elevado, dado que no solo implica la pérdida de la venta actual, sino el decremento del valor de vida del cliente (*Customer Lifetime Value*), cuyo impacto acumulado puede superar con creces el costo del material faltante [40].

La cuantificación del impacto del desabastecimiento justifica la necesidad de desarrollar herramientas predictivas que permitan anticipar estos eventos y activar mecanismos de mitigación antes de que las consecuencias operativas y financieras se materialicen.

6. Preparación y Análisis Exploratorio de Datos

Siguiendo las fases de comprensión y preparación de los datos establecidas en la metodología *CRISP-DM* [6], en esta sección se describe el proceso de recolección, depuración y exploración del conjunto de datos utilizado para la construcción del modelo predictivo. Este proceso incluyó: Revisión de la estructura de los datos, Detección de valores atípicos o nulos, Identificación de relaciones entre variables y la generación de características derivadas, con el fin de garantizar la calidad de la información y descubrir patrones relevantes para el modelado posterior.

6.1 Recolección y Delimitación de los Datos

La organización proporcionó registros históricos de consumo de productos, que se extiende desde enero de 2022 hasta Octubre 2025. La base de datos original contiene un total de 28.374 registros diarios, correspondientes a 122 ítems.

Para efectos de la construcción del modelo predictivo, el alcance del estudio se delimitó a la subcategoría de Ingredientes e Insumos. Esta selección responde a criterios estratégicos de riesgo, dados los retos logísticos asociados a su adquisición y los tiempos de entrega prolongados por parte de los proveedores. Asimismo, esta subcategoría se consolida como la más representativa del portafolio, concentrando el 21% de los consumos totales de la compañía. Geográficamente, los datos reflejan la dinámica de operación en las plantas ubicadas en tres ciudades de Colombia.

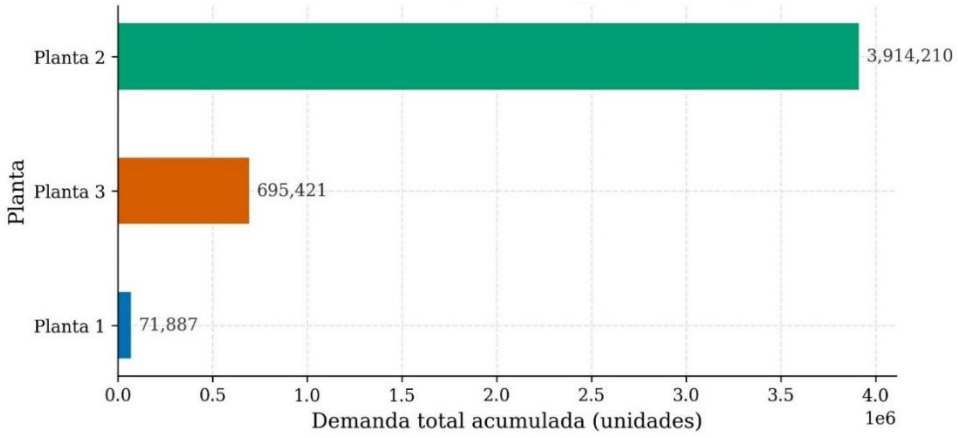

Figura 1. Demanda acumulada de ingredientes por planta de producción

La distribución de registros entre las tres plantas muestra una concentración marcada en la planta de mayor operación, seguida de dos instalaciones con volúmenes menores, pero con diversidad suficiente de ingredientes. Esta asimetría motiva el tratamiento independiente de cada planta durante la generación de pronósticos, dado que factores como la estacionalidad local y la política de abastecimiento pueden diferir entre sedes.

6.2 Caracterización del Conjunto de Datos

El conjunto de datos depurado para el análisis contiene registros transaccionales de consumos de la organización. La unidad de análisis primaria es el movimiento de inventario, definido por la fecha de transacción y la identificación del ítem, integrando atributos jerárquicos como la categoría y bodega en la que se registran los consumos.

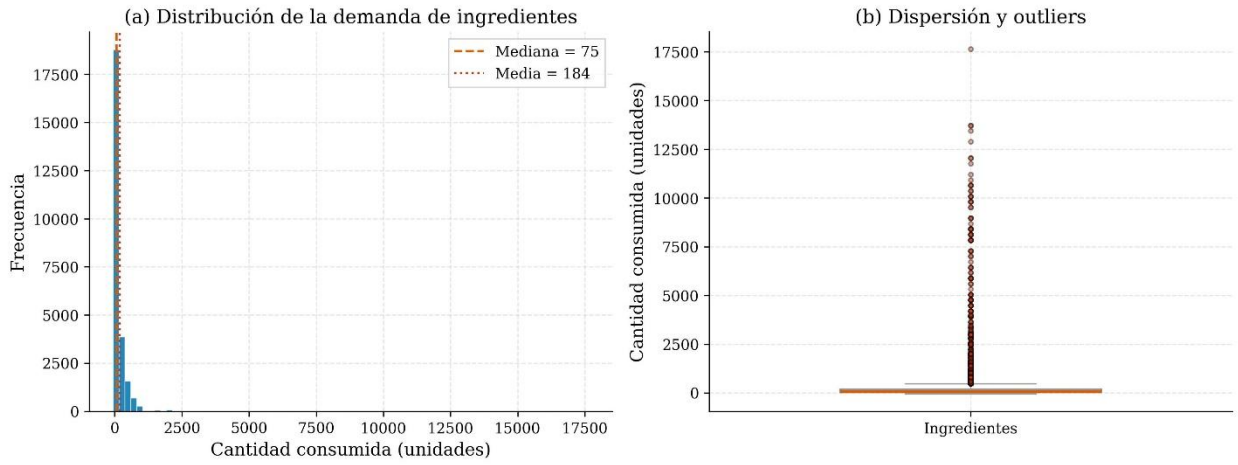

Figura 2. Distribución general de la demanda de ingredientes

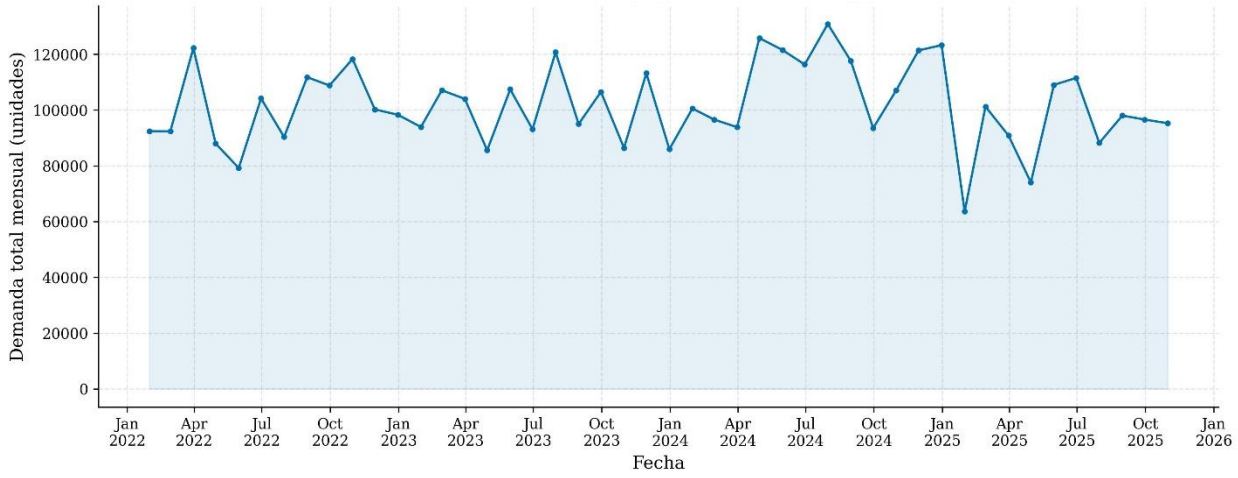

Figura 3. Serie temporal de consumo de ingredientes

La representación del consumo agregado a lo largo del horizonte de análisis permite identificar la dinámica temporal de la demanda. La serie muestra un comportamiento sostenido sin interrupciones prolongadas en el registro. Se aprecian patrones recurrentes compatibles con una estacionalidad de tipo anual, asociada al ciclo productivo de la organización, así como variaciones de mayor frecuencia atribuibles a la programación semanal de la producción.

Tabla 8. Variables del conjunto de datos

Variable	Descripción	Tipo de dato
LOTE	Lote del producto	Character
BODEGA_DESTINO	Bodega destino del movimiento	Character
FECHA_SENCILLA	Fecha en formato Siglo-Año	Numeric
PLANTA	Planta de operación	Character
BODEGA_ORIGEN	Bodega origen del movimiento	Character
FECHA_COMPLETA	Fecha en formato Siglo-Año-Mes-Día	Character
COD_PRODUCTO	Código del producto en el *ERP*	Character
DESC_PRODUCTO	Descripción del producto	Character

TIPO_MOVIMIENTO	Tipo de movimiento (traslado, consumo)	Character
CANTIDAD	Cantidad del movimiento	Numeric
COSTO	Costo del movimiento	Numeric
GENERADOR_MOVIMIENTO	Autorizador del movimiento en el *ERP*	Character
COD_SUBCATEGORÍA_PROD	Código de la subcategoría del producto	Character
COD_FAMILIA_PRODUCTO	Código de la familia del producto	Character
UNIDAD_MEDIDA	Unidad de medida	Character
COSTO_UNITARIO	Costo Unitario del producto	Numeric
FECHA_DATETIME	Fecha en formato Datetime	DateTime

Fuente: elaboración propia.

6.3 Preparación y Limpieza de Datos

Antes de alimentar los modelos, los datos requieren limpieza y transformación. En series de demanda de ingredientes, esta etapa es compleja por la heterogeneidad de las fuentes transaccionales, los signos invertidos en movimientos contables, la diversidad de codificaciones de planta y los valores atípicos generados por pedidos extraordinarios o errores de digitación [42]. El *pipeline* implementado se compone de siete etapas:

(I) Filtrado por subcategoría de producto

(II) Homologación de códigos de planta

(III) Normalización de cantidades con signo negativo

(IV) Cálculo del costo unitario

(V) Conversión y validación de fechas

(VI) Detección y eliminación de valores atípicos mediante el identificador de Hampel basado en la Desviación Absoluta Mediana (*MAD*)

(VII) Descarte de columnas innecesarias.

Filtrado por Subcategoría de Producto

El conjunto de datos original proviene de un sistema transaccional que registra el consumo de múltiples categorías de insumos: ingredientes, materiales de empaque, repuestos, suministros de limpieza, entre otros. Dado que el alcance de este estudio se limita a la predicción de la demanda de ingredientes, se aplicó un filtro sobre el campo *COD_SUBCATEGORIA_PRODUCTO,* reteniendo exclusivamente aquellos registros cuya subcategoría corresponde a *INGREDIENTE*. Este paso es necesario no solo por razones de delimitación temática, sino también porque la inclusión de categorías con dinámicas de consumo estructuralmente diferentes, como los materiales de empaque, que presentan patrones estacionales distintos y unidades de medida heterogéneas, lo que introduciría ruido y sesgo en los modelos posteriores, comprometiendo tanto la estimación de parámetros como la interpretabilidad de los resultados.

Homologación de Códigos de Planta

La empresa objeto de estudio opera tres plantas de producción distribuidas en el territorio colombiano. En el sistema transaccional de origen, cada planta se identifica mediante un código numérico interno cuya semántica no resulta intuitiva para el analista ni para los modelos que requieran segmentación geográfica. Por ello, se implementó una tabla de equivalencias que mapea cada código numérico a un nombre descriptivo de la planta.

Normalización de Cantidades Negativas

En los sistemas *ERP* utilizados en la industria alimentaria, los movimientos de devolución, ajuste de inventario o reversión contable se registran frecuentemente con signo negativo tanto en la cantidad como en el costo asociado. Desde la perspectiva del modelado de demanda, estos valores negativos no representan una demanda real negativa, sino una convención contable que debe ser normalizada antes de cualquier análisis cuantitativo. Se verificó si el valor mínimo de la columna de cantidad presentaba signo negativo; en caso afirmativo, se invirtió el signo de las columnas de cantidad y costo simultáneamente, preservando así la coherencia entre ambas magnitudes.

Cálculo del Costo Unitario

A partir de las columnas de costo total y cantidad, se derivó una nueva variable denominada costo

unitario, definida como el cociente entre el costo y la cantidad de cada transacción. Esta variable resulta de interés analítico dado que permite detectar anomalías de precio a nivel transaccional; por ejemplo, compras a proveedores con precios atípicamente elevados o errores de facturación y, además, constituye un insumo potencial para modelos que incorporen la dimensión económica de la demanda.

Para evitar errores de división por cero, se implementó una condición lógica mediante la función np.where: cuando la cantidad es igual a cero, el costo unitario se asigna como cero. Esta convención, resulta apropiada dado que un registro con cantidad nula carece de significado económico unitario y será, en la mayoría de los escenarios, descartado en etapas posteriores del análisis.

Conversión y Validación de Fechas

El campo de fecha en el sistema de origen se almacena como un valor numérico que requiere una transformación en dos pasos para su correcta interpretación temporal. En primer lugar, se aplicó un desplazamiento aritmético al valor numérico crudo, el cual compensa la diferencia entre la época de referencia utilizada por el sistema transaccional y la época estándar de los sistemas de fecha. En segundo lugar, el valor resultante se convierte a un objeto de fecha-hora (datetime) empleando el formato de fecha especificado en la configuración del proyecto.

Aquellos registros cuya conversión de fecha arroja un valor nulo, ya sea por datos corruptos, campos vacíos o valores numéricos fuera del rango válido, se eliminan del conjunto de datos. Se eliminaron estos registros porque la dimensión temporal es el eje del análisis de series de tiempo; un registro sin fecha válida no puede ubicarse en la secuencia cronológica y, por tanto, carece de utilidad para los modelos predictivos.

Detección y Eliminación de Valores Atípicos

La presencia de valores atípicos en las series de demanda de ingredientes industriales, pueden originarse por pedidos extraordinarios no recurrentes, errores de digitación en el sistema *ERP*, promociones puntuales de productos terminados que generan picos de consumo, o bien por la

consolidación de órdenes de compra de varios periodos en una sola transacción.

Método seleccionado: identificador de Hampel frente al filtro de Tukey

En la práctica del análisis estadístico, el método más difundido para la detección de valores atípicos en distribuciones univariantes es el filtro de Tukey, basado en el rango intercuartílico (*IQR*) [46]. Este método define como atípica toda observación que se sitúe por debajo de $Q_1 - k \times IQR$ o por encima de $Q_3 + k \times IQR$, donde k suele fijarse en 1.5 para valores atípicos moderados o en 3 para valores extremos. No obstante, el *IQR* posee un punto de quiebre del 25 %, lo que significa que la presencia de un 25 % o más de observaciones contaminadas puede invalidar completamente el estimador [43]. En series de demanda de ingredientes, donde es frecuente encontrar múltiples valores extremos simultáneos, esta limitación puede conducir al fenómeno de enmascaramiento (*masking effect*), en el que *outliers* múltiples inflan artificialmente el *IQR* y pasan inadvertidos [44].

Debido a lo anteriormente descrito, se optó por el identificador de Hampel, que emplea como estimador de escala la Desviación Absoluta Mediana (*MAD*) normalizada. La *MAD* posee un punto de quiebre del 50 %, el máximo alcanzable por un estimador de escala, lo que la convierte en un estadístico altamente resistente a la contaminación de la muestra [45]. Formalmente, dado un conjunto de observaciones $x_1, x_2, ..., x_n$, la *MAD* se define como:

$$MAD = \text{mediana}(|x_i - \text{mediana}(x)|) \quad (4)$$

Para que la *MAD* sea un estimador consistente de la desviación estándar bajo el supuesto de normalidad, se aplica un factor de corrección $k = 1/\Phi^{-1}(0.75) \approx 1.4826$, de modo que la *MAD* normalizada se expresa como:

$$MAD_{\text{normalizada}} = 1.4826 \times MAD \quad (5)$$

El identificador de Hampel clasifica como valor atípico toda observación x_i que satisfaga:

$$|x_i - \text{mediana}(x)| > t \times MAD_{\text{normalizada}} \quad (6)$$

donde t es el umbral de decisión. En este trabajo se empleó t = 3, equivalente a la regla de 3-sigma bajo distribución normal, lo que implica una probabilidad de clasificación incorrecta de aproximadamente el 0.3 % para datos gaussianos. La elección de este umbral busca un equilibrio

entre sensibilidad y especificidad: un valor menor incrementaría la proporción de falsos positivos, mientras que un valor mayor permitiría que *outliers* genuinos permanecieran en el conjunto de datos.

Aplicación del identificador Hampel segmentada por producto

La detección de valores atípicos se realizó de manera segmentada por código de producto (COD_PRODUCTO), mediante una operación *groupby-transform*. Esta decisión responde al hecho de que cada ingrediente presenta una distribución de demanda propia, con niveles de centralidad y dispersión característicos. Un valor que resulta atípico para un ingrediente de bajo consumo puede ser perfectamente normal para un insumo de alta rotación [71]. Aplicar un único criterio global, sin segmentación, habría conducido a la eliminación errónea de registros válidos en productos de alta demanda y a la retención de *outliers* en productos de baja demanda.

Adicionalmente, se estableció un tamaño mínimo de grupo (*min_group_size_for_outlier*) por debajo del cual no se aplica la detección de *outliers*. Este parámetro responde a un principio de estadística, el cual afirma que, con pocas observaciones, la estimación de la mediana y la *MAD* carece de la robustez necesaria para tomar decisiones confiables de descarte.

Mecanismo de respaldo: fallback al rango intercuartílico

Existe un caso particular en el que la *MAD* normalizada adopta el valor cero: cuando más del 50% de las observaciones del grupo son idénticas, la mediana de las desviaciones absolutas es nula [43]. Esta situación es usual en series de demanda, donde ciertos ingredientes se consumen en cantidades estandarizadas. Ante este escenario, la aplicación directa del criterio de Hampel resulta imposible por la indeterminación del cociente. Como mecanismo de respaldo, se implementó un filtro basado en el rango intercuartílico con un multiplicador configurable (por defecto, $k = 3$), que permite detectar aquellos valores que se desvían considerablemente de la distribución central aun cuando esta es altamente concentrada [46]. Si tanto la *MAD* como el *IQR* son iguales a cero, indicando una serie completamente constante, todos los registros del grupo se consideran válidos, dado que no existe dispersión que justifique la eliminación de ninguna observación.

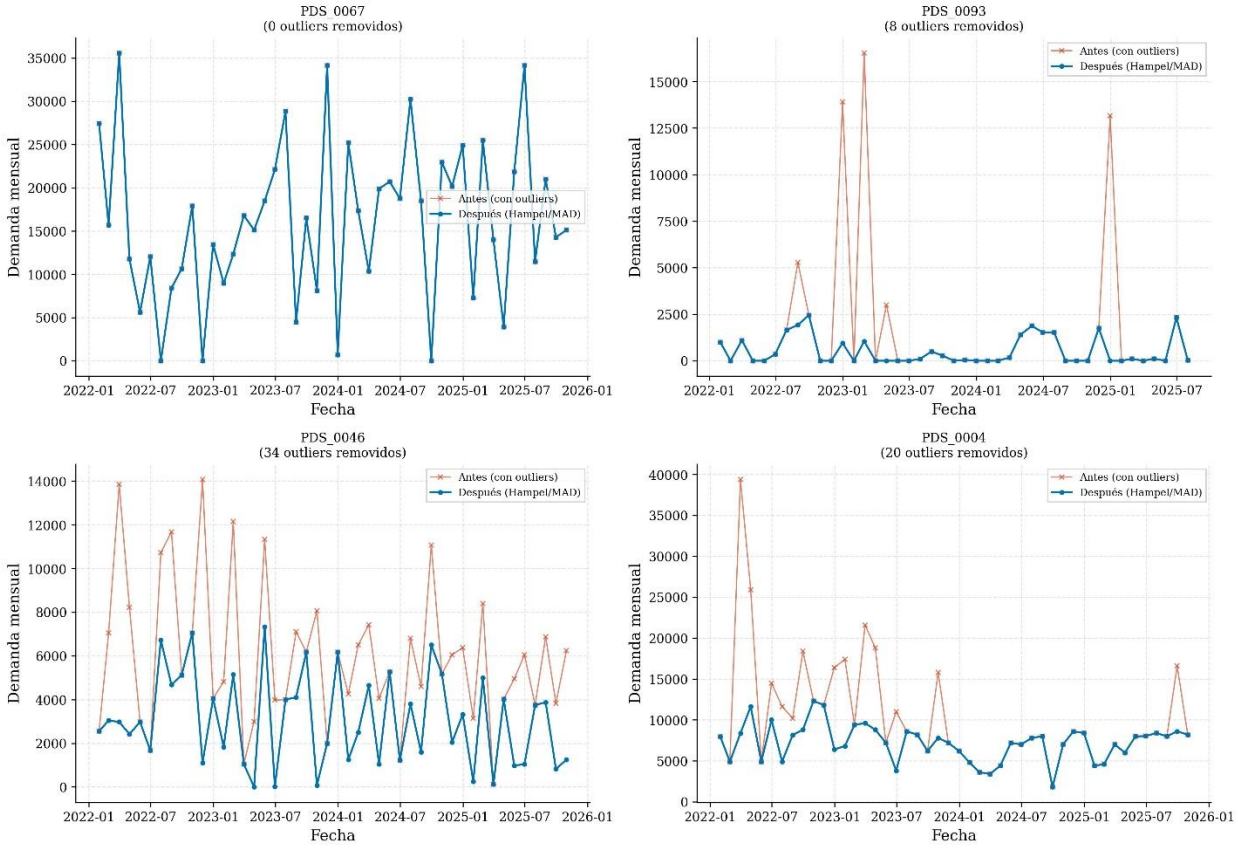

Figura 4. Impacto de la limpieza Hampel/MAD en series de demanda seleccionadas

El esquema resume la lógica del mecanismo de detección en dos niveles: el identificador de Hampel opera como criterio principal, y el filtro basado en el rango intercuartílico actúa como respaldo en aquellos grupos donde la *MAD* normalizada se anula. Esta arquitectura escalonada garantiza la aplicabilidad del proceso incluso en series altamente concentradas, situación frecuente en consumos estandarizados de insumos industriales.

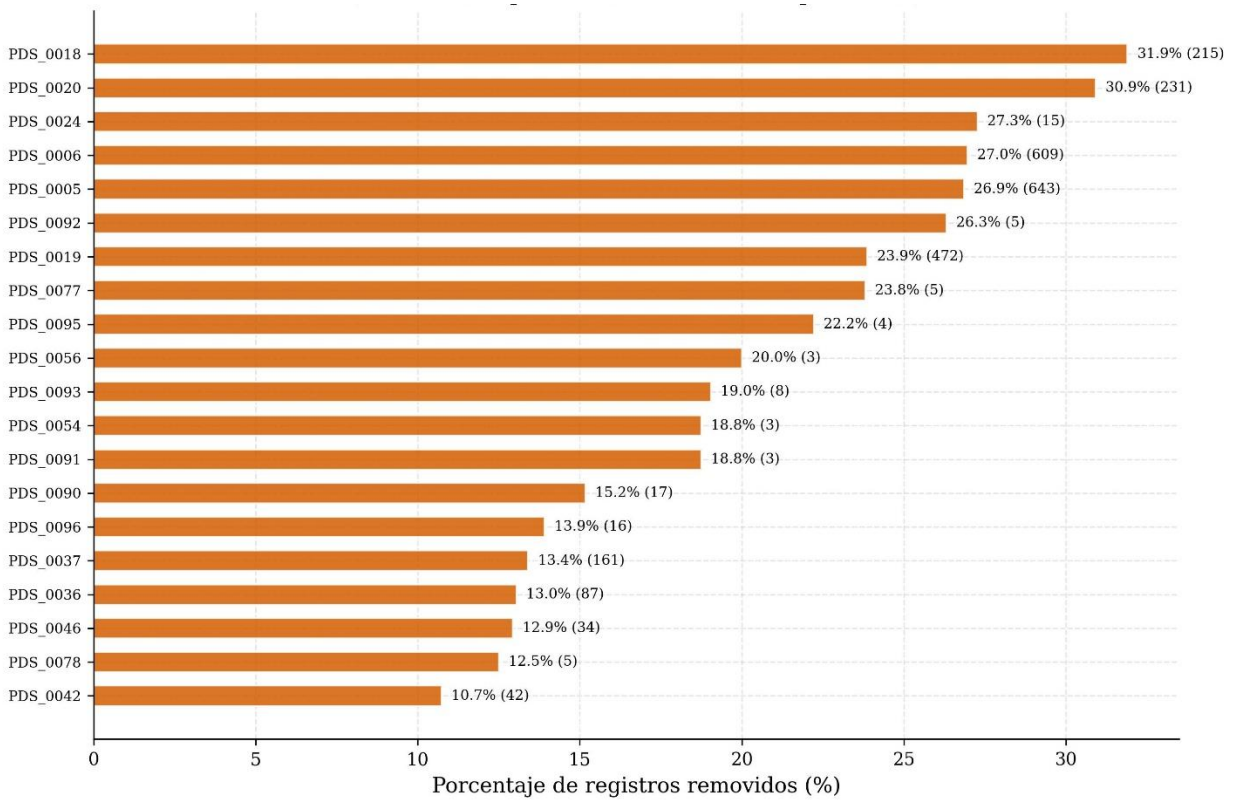

Figura 5. Top 20 ingredientes con mayor proporción de outliers eliminados

Efecto de la eliminación de nulos iniciales sobre la representatividad de las series cortas

En el conjunto de datos analizado, los nulos iniciales surgen de manera estructural cuando un ingrediente no registra movimientos al comienzo de la ventana temporal de análisis, ya sea porque fue incorporado al portafolio con posterioridad a la fecha de corte inicial del estudio, o porque sus registros más tempranos fueron eliminados durante las etapas de validación de fechas o cantidades descritas en secciones previas. Al construir las series con un origen temporal común para todos los productos, estos períodos de inactividad inicial se materializan como valores ausentes en los primeros puntos de la secuencia cronológica y son excluidos antes de cualquier estimación estadística.

El impacto de esta exclusión es asimétrico respecto a la longitud efectiva de la serie. Para un ingrediente con 200 observaciones históricas, la remoción de 10 nulos iniciales representa una pérdida del 5 % de la información disponible, con un efecto marginal sobre la estimación de la mediana o la MAD. En contraste, una serie de 25 observaciones que pierde 10 nulos iniciales ve

reducida su longitud efectiva en un 40 %, situación en la que los estimadores robustos pierden precisión: la MAD requiere un volumen mínimo de datos para que su punto de quiebre del 50 % opere con garantías estadísticas prácticas [43], y con muestras reducidas la varianza del estimador se incrementa de manera considerable [45].

Para atenuar este riesgo, el pipeline incorporó dos salvaguardas. La primera es el parámetro min_group_size_for_outlier, que desactiva la detección de outliers en series cuya longitud no supera el umbral definido, preservando íntegramente los registros de los ingredientes con historial corto y evitando descartes carentes de respaldo estadístico. La segunda consiste en la exclusión de las filas iniciales durante la ingeniería de características —allí donde las ventanas deslizantes de rezago no disponen de observaciones suficientes—, lo que garantiza la integridad vectorial de los predictores sin alterar el historial disponible. No obstante, debe reconocerse que los ingredientes con series muy cortas presentan, incluso con estas salvaguardas, una representatividad estadística inherentemente limitada: la escasa longitud de su historial restringe la capacidad del modelo para capturar patrones estacionales o ciclos de consumo recurrentes, y sus métricas de pronóstico deben interpretarse como estimaciones orientativas antes que como predicciones de alta confianza.

Efecto de la eliminación de nulos iniciales sobre la representatividad de las series cortas

En el conjunto de datos analizado, los nulos iniciales surgen de manera estructural cuando un ingrediente no registra movimientos al comienzo de la ventana temporal de análisis, ya sea porque fue incorporado al portafolio con posterioridad a la fecha de corte inicial del estudio, o porque sus registros más tempranos fueron eliminados durante las etapas de validación de fechas o cantidades descritas en secciones previas. Al construir las series con un origen temporal común para todos los productos, estos períodos de inactividad inicial se materializan como valores ausentes en los primeros puntos de la secuencia cronológica y son excluidos antes de cualquier estimación estadística.

El impacto de esta exclusión es asimétrico respecto a la longitud efectiva de la serie. Para un ingrediente con 200 observaciones históricas, la remoción de 10 nulos iniciales representa una pérdida del 5 % de la información disponible, con un efecto marginal sobre la estimación de la mediana o la MAD. En contraste, una serie de 25 observaciones que pierde 10 nulos iniciales ve

reducida su longitud efectiva en un 40 %, situación en la que los estimadores robustos pierden precisión; la MAD requiere un volumen mínimo de datos para que su punto de quiebre del 50 % opere con garantías estadísticas prácticas [43], y con muestras reducidas la varianza del estimador se incrementa de manera considerable [45].

Para disminuir este riesgo, el pipeline incorporó dos controles. El primero es el parámetro min_group_size_for_outlier, que desactiva la detección de outliers en series cuya longitud no supera el umbral definido, preservando íntegramente los registros de los ingredientes con historial corto y evitando descartes carentes de respaldo estadístico. La segunda consiste en la exclusión de las filas iniciales durante la ingeniería de características, y es allí donde las ventanas deslizantes de rezago no disponen de observaciones suficientes, lo que garantiza la integridad vectorial de los predictores sin alterar el historial disponible. No obstante, debe reconocerse que los ingredientes con series muy cortas presentan, incluso con estos controles, una representatividad estadística inherentemente limitada, puesto que la escasa longitud de su historial restringe la capacidad del modelo para capturar patrones estacionales o ciclos de consumo recurrentes, y sus métricas de pronóstico deben interpretarse como estimaciones orientativas antes que como predicciones de alta confianza.

Tabla 9. Comparación entre el identificador de Hampel y el filtro de Tukey

Criterio	Identificador de Hampel (*MAD*)	Filtro de Tukey (*IQR*)
Estimador de escala	*MAD* normalizada (k = 1.4826)	Rango intercuartílico (Q3 – Q1)
Punto de quiebre	50 %	25 %
Resistencia al masking	Alta: tolera hasta 50 % de contaminación	Moderada: vulnerable con >25 % de contaminación
Supuesto distribucional	Robusto ante asimetría	Mejor con distribuciones simétricas
Caso de uso ideal	Series con alta asimetría y *outliers* múltiples	Series con distribuciones aproximadamente normales
Referencia	Davies y Gather [44]; Rousseeuw y Croux [43]	Tukey [46]

Fuente: elaboración propia

Descarte de Columnas Innecesarias

Como etapa final del *pipeline* de preprocesamiento, se eliminaron del conjunto de datos aquellas columnas que no aportan información relevante para las fases de modelado y análisis predictivo. La lista de columnas a descartar se gestiona de manera centralizada a través del archivo de configuración del proyecto, lo que permite modificar el conjunto de variables retenidas sin alterar el código fuente.

La decisión de qué columnas descartar se fundamentó en un análisis previo de relevancia: se eliminaron campos identificadores internos del *ERP* (claves primarias, códigos auxiliares) y variables redundantes que no aportan capacidad predictiva, pero que incrementan la dimensionalidad del conjunto de datos y, con ello, el riesgo de sobreajuste.

Tabla 10. Resumen del *pipeline* de preprocesamiento

Paso	Operación	Justificación	Impacto
1	Filtrado por subcategoría	Delimitar el análisis al dominio de ingredientes	Reducción dimensional significativa
2	Homologación de plantas	Mejorar legibilidad e interpretabilidad	Nombres descriptivos
3	Normalización de cantidades	Corregir convenciones contables de signo	Eliminación de valores negativos
4	Cálculo de costo unitario	Derivar variable económica de análisis	Nueva variable derivada
5	Conversión de fechas	Habilitar el análisis temporal	Eliminación de fechas inválidas
6	Eliminación de *outliers* (Hampel/*MAD*)	Robustez frente a contaminación asimétrica	Remoción de extremos por producto
7	Descarte de columnas	Reducir dimensionalidad y ruido	Conjunto de datos final limpio

Fuente: elaboración propia.

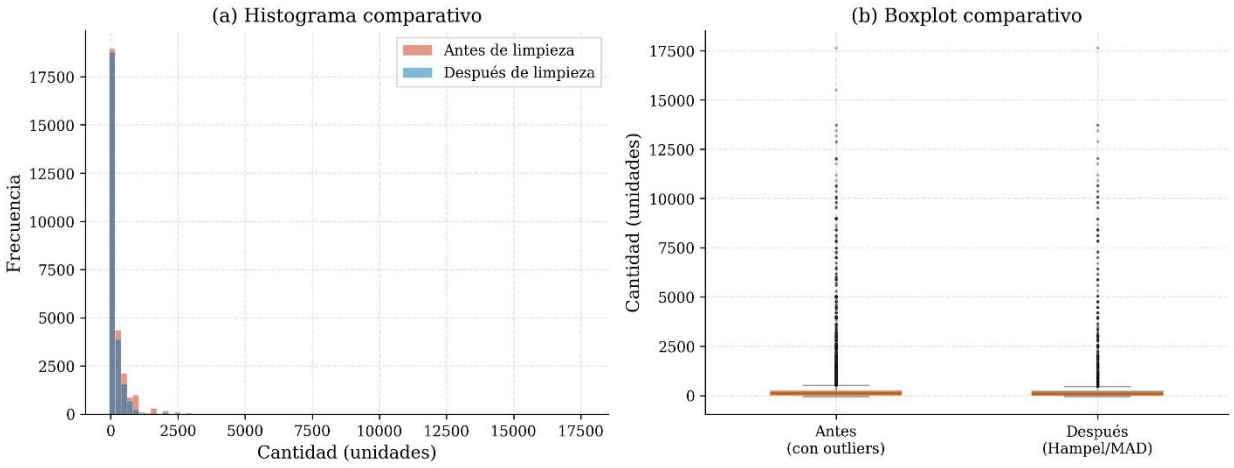

Figura 6. Impacto de la limpieza de datos

6.4 *Feature Engineering* y Agrupamiento de Productos

Tras haber sometido los datos transaccionales al proceso de limpieza y transformación descrito en el capítulo anterior, la información resultante se encuentra en condiciones de ser sintetizada en un conjunto de indicadores que caractericen el comportamiento temporal de la demanda de cada ingrediente. Esta etapa, conocida en la literatura de ciencia de datos como *feature engineering* o ingeniería de características, hace parte de la construcción de modelos predictivos, ya que las variables de entrada inciden de manera directa en la capacidad discriminativa y la generalización de los algoritmos de aprendizaje automático [47].

Para este abarcar capitulo, se generan dos fases; la extracción de métricas de comportamiento temporal para cada producto y el proceso de agrupamiento (*clustering*) mediante un torneo competitivo entre los algoritmos *K-Means* y *DBSCAN*, evaluado con tres indicadores de validación interna.

Extracción de Características por Producto

Para construir el espacio de características, las transacciones diarias se agregaron a nivel mensual mediante una tabla pivote que cruza el eje temporal con cada código de producto. Los meses sin transacciones se imputaron con valor cero, lo que refleja la ausencia real de demanda y resulta

indispensable para el cómputo correcto de métricas de intermitencia. Sobre esta serie mensual, que se emplea exclusivamente en esta etapa de caracterización y agrupamiento; el pipeline de modelado predictivo de la sección 6.5 retoma la serie diaria original, se calcularon las siguientes variables para cada producto:

Tabla 11. Características extraídas por producto

Característica	Definición y justificación
Volatilidad (CV)	Coeficiente de variación de la serie mensual completa (σ/μ). Cuantifica la dispersión relativa de la demanda respecto a su nivel medio. Valores altos indican productos con demanda errática.
Intermitencia	Proporción de meses con demanda igual a cero. Discrimina productos de consumo continuo (intermitencia ≈ 0) de aquellos con demanda esporádica (intermitencia → 1).
Entropía de Shannon	Calculada sobre la distribución normalizada de los valores positivos. Mide la uniformidad del patrón de consumo: entropía alta indica demanda distribuida de manera homogénea en el tiempo; entropía baja sugiere concentración en pocos periodos.
ADI	*Average Demand Interval* [48]. Intervalo medio entre periodos con demanda positiva: *ADI* = n / n_positivos. Eje vertical del esquema de clasificación *SBC*.
CV^2	Coeficiente de variación al cuadrado, calculado exclusivamente sobre periodos con demanda positiva. Eje horizontal del esquema *SBC*. Captura la variabilidad del tamaño de los pedidos cuando estos ocurren.
Exponente de Hurst	Estimado mediante el método R/S (rescaled range). Cuantifica la memoria de largo plazo: $H > 0.5$ indica persistencia (tendencias sostenidas); $H < 0.5$ indica antipersistencia (reversión a la media); $H \approx 0.5$ indica comportamiento aleatorio.
Clasificación *SBC*	Etiqueta categórica derivada del cruce *ADI*/CV^2 según los umbrales estándar de la literatura (*ADI* = 1.32, CV^2 = 0.49). Cuatro categorías: *Smooth*, *Intermittent*, *Erratic*, *Lumpy*.
Meses de historia	Número total de meses con datos disponibles. Controla la confiabilidad de las métricas calculadas.

Fuente: elaboración propia.

Clasificación de la Demanda Según el Esquema Syntetos-Boylan-Croston

El esquema de clasificación *SBC*, propuesto por Syntetos y Boylan [48], constituye el marco de referencia estándar en la gestión de inventarios para categorizar patrones de demanda intermitente. Su relevancia radica en que el tipo de patrón de demanda determina la familia de métodos de pronóstico más apropiada: los métodos de suavizamiento exponencial resultan adecuados para demanda *Smooth*, mientras que la demanda *Lumpy* requiere métodos especializados como el de Croston o su variante *SBA* [49].

La clasificación opera sobre dos ejes: el *ADI*, que mide la frecuencia de ocurrencia de la demanda, y el CV^2, que mide la variabilidad del tamaño de los pedidos cuando estos ocurren. El cruce de ambos ejes con los umbrales estándar (*ADI* = 1.32 y CV^2 = 0.49) define cuatro cuadrantes:

Tabla 12. Cuadrantes de clasificación *SBC*

***ADI* / CV^2**	**$CV^2 < 0.49$**	**$CV^2 \geq 0.49$**
ADI < 1.32 (frecuente)	*Smooth*: demanda regular y predecible	*Erratic*: frecuente pero con volúmenes variables
ADI ≥ 1.32 (esporádica)	*Intermittent*: esporádica pero con volúmenes estables	*Lumpy*: esporádica y con volúmenes variables

Fuente: adaptado de Syntetos y Boylan en [48]

Agrupamiento de Productos: Torneo K-Means Versus DBSCAN

Una vez construido el espacio de características, se procedió a agrupar los productos con el fin de identificar segmentos con patrones de demanda homogéneos. La segmentación es un paso previo al modelado predictivo que permite asignar a cada grupo la familia de modelos más apropiada, reducir la dimensionalidad del problema y mejorar la interpretabilidad de los resultados. Para ello, se implementó un mecanismo de selección automática entre dos modelos: *K-Means*, que asume clusters convexos de tamaño similar, y *DBSCAN*, que puede identificar agrupaciones de forma arbitraria y detectar puntos de ruido [51].

Estandarización previa

Antes de la ejecución de los algoritmos, las variables numéricas se estandarizaron mediante *StandardScaler*, operación indispensable dado que tanto *K-Means* como *DBSCAN* utilizan la distancia euclidiana como métrica de proximidad. Sin estandarización, las variables con mayor escala dominarían el cálculo de distancias, distorsionando la estructura de los agrupamientos.

K-Means

K-Means particiona las observaciones en k grupos minimizando la inercia. El número de clusters k se especificó desde el archivo de configuración del proyecto. El algoritmo se ejecutó con diez inicializaciones aleatorias para mitigar la sensibilidad a las condiciones iniciales, reteniendo la solución con menor inercia.

DBSCAN

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) identifica clusters como regiones de alta densidad separadas por regiones de baja densidad, lo que le confiere la capacidad de descubrir agrupaciones no convexas y etiquetar como ruido aquellos puntos que no pertenecen a ninguna región densa. El parámetro crítico eps (radio de vecindad) se estimó de manera automática mediante el método del codo en la curva de k-distancias [52]: se calculó la distancia al k-ésimo vecino más cercano para cada punto, se ordenó de forma descendente y se identificó el punto de máxima curvatura. Gracias a este valor inicial, se evaluaron 20 candidatos en un rango acotado, seleccionando el eps que maximizó el índice Silhouette.

Sistema de evaluación competitiva

La selección del algoritmo ganador se realizó mediante un sistema de puntos basado en tres indicadores de validación interna, calculados sin recurrir a etiquetas externas:

Tabla 13. Indicadores de validación interna del agrupamiento

Indicador	Interpretación	Criterio
Silhouette Score	Mide la cohesión intra-cluster y la separación	Mayor es mejor

	inter-cluster. Rango [–1, 1].	
Davies-Bouldin Index	Promedio de la razón entre dispersión intra-cluster y separación entre centroides.	Menor es mejor
Calinski-Harabasz Index	Razón de la varianza inter-cluster sobre la varianza intra-cluster.	Mayor es mejor

Fuente: elaboración propia.

Cada indicador asigna un punto al algoritmo que obtiene el mejor valor, de modo que el puntaje máximo posible es de tres puntos. Adicionalmente, se incorporó una penalización para *DBSCAN*: si el porcentaje de puntos clasificados como ruido excede un umbral configurable, se descuenta un punto de su marcador, lo que desincentiva soluciones que dejan una fracción excesiva de productos sin asignación. En caso de empate, el desempate se resuelve por el *Silhouette Score*, por ser el indicador más ampliamente aceptado en la literatura de validación de clusters [35]. Los puntos de ruido del modelo *DBSCAN*, cuando este resulta ganador, se reasignan al cluster más cercano por distancia euclidiana al centroide, garantizando la asignación completa de todos los productos.

6.5 Fundamentos de los Modelos Predictivos y Preparación de Datos para el Modelado

En la sección precedente se describió la construcción del espacio de características que permite segmentar los ingredientes según su patrón de demanda. Una vez definidos los grupos de productos, el paso siguiente consiste en seleccionar y configurar los modelos de pronóstico que se aplicarán a cada segmento. Se exponen los fundamentos teóricos de los tres modelos implementados (*Prophet*, *XGBoost* y *Ridge Regression*), justifica las decisiones de diseño adoptadas en cada caso y describe el mecanismo de preparación de datos que alimenta a los modelos de aprendizaje automático.

La elección de esta triada de modelos responde a un criterio de diversidad metodológica deliberado: *Prophet* opera como un modelo aditivo de descomposición que captura tendencia y estacionalidad de forma explícita; *XGBoost* representa la familia de métodos de *ensemble* basados en *gradient boosting*; y *Ridge Regression* constituye una línea base regularizada que permite evaluar si la complejidad adicional de los modelos no lineales se traduce en mejoras predictivas significativas [29].

Preparación de Datos para los Modelos de Aprendizaje Automático

A diferencia de la etapa de caracterización descrita en la sección 6.4, que agrega la demanda a nivel mensual para el cómputo de indicadores de producto y su agrupamiento, este pipeline trabaja sobre la serie diaria de cantidad consumida. Los modelos XGBoost y Ridge Regression operan sobre datos tabulares de tipo supervisado, donde cada observación consiste en un vector de variables predictoras (*features*) y una variable objetivo. Dado que la información de partida es una serie temporal univariante, fue necesario transformarla en un conjunto de *features* mediante dos estrategias complementarias: la generación de variables de rezago (*lags*) y la derivación de estadísticos de ventana deslizante (*rolling features*).

Variables de rezago

Para cada observación en el instante *t*, se generaron columnas lag_1, lag_2, ..., lag_k que contienen los valores de la demanda en los instantes *t–1*, *t–2*, ..., *t–k*, donde *k* es el número de rezagos definido en la configuración del proyecto. Esta transformación convierte el problema de pronóstico de series de tiempo en un problema de regresión estándar, permitiendo aplicar cualquier algoritmo supervisado. El número de rezagos se seleccionó de manera que capture al menos un ciclo completo de la estacionalidad de la demanda.

Estadísticos de ventana deslizante

Además de los rezagos puros, se derivaron tres variables adicionales que sintetizan el comportamiento reciente de la serie: la media móvil de 7 días, la media móvil de 14 días y la desviación estándar móvil de 7 días. Estas *rolling features* capturan respectivamente, el nivel local de la demanda a corto y mediano plazo y la volatilidad reciente. Para evitar la fuga de información (*data leakage*), todas las ventanas se calcularon a partir del valor en *t–1*, de modo que ninguna observación futura participa en la construcción de las variables predictoras [55].

Tabla 14. Variables predictoras para los modelos tabulares

Variable	Definición
lag_1, ..., lag_k	Valor de la demanda en los instantes t–1 hasta t–k.

roll_mean_7	Media de la demanda en los 7 días anteriores a t (calculada desde t–1). Captura el nivel local de corto plazo.
roll_mean_14	Media de la demanda en los 14 días anteriores a t. Captura el nivel local de mediano plazo.
roll_std_7	Desviación estándar de los 7 días anteriores a t. Cuantifica la volatilidad reciente de la demanda.

Fuente: elaboración propia.

Las primeras filas del conjunto resultante, cuyas ventanas deslizantes no disponen de suficientes observaciones precedentes, se eliminaron automáticamente. Este descarte es necesario para garantizar que todas las observaciones de entrenamiento cuenten con vectores de *features* completos y no introduzcan valores nulos en los algoritmos.

Modelo Prophet

Prophet es un modelo aditivo de pronóstico de series de tiempo desarrollado por Taylor y Letham [41], diseñado para capturar tres componentes: tendencia, estacionalidad y efecto de días festivos. Su formulación general se expresa como:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

donde *g(t)* modela la tendencia no periódica (crecimiento lineal o logístico con puntos de cambio), *s(t)* captura la estacionalidad mediante series de *Fourier*, *h(t)* incorpora el efecto de eventos especiales y ε_t es el término de error. La estimación se realiza por inferencia bayesiana a través de *Stan*, lo que permite cuantificar la incertidumbre de las predicciones mediante intervalos de credibilidad.

Se activaron las estacionalidades anual y semanal, desactivando la estacionalidad diaria. La estacionalidad semanal es relevante en cuanto a consumos industriales, donde los días de producción (lunes a viernes) y los paros de fin de semana generan un patrón cíclico de siete días. El parámetro *changepoint_prior_scale* controla la flexibilidad de la tendencia: valores bajos (< 0.05) producen tendencias suaves que no reaccionan ante fluctuaciones transitorias, mientras que valores altos (> 0.5) permiten al modelo adaptarse a cambios bruscos de nivel.

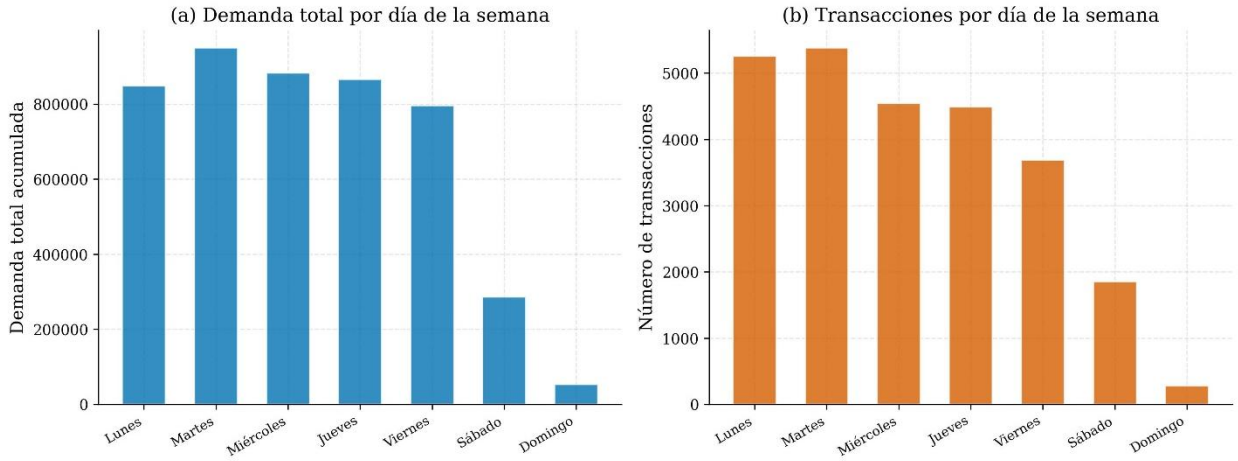

Figura 7. Patrón semanal de consumo de ingredientes

Modelo XGBoost

XGBoost (*Extreme Gradient Boosting*) es un algoritmo de *ensemble* basado en el principio de *gradient boosting*, que construye secuencialmente un conjunto de árboles de decisión donde cada árbol corrige los residuos del anterior [27]. Su función objetivo combina un término de pérdida empírica con un término de regularización:

$$L = \Sigma\, l(y_i, \hat{y}_i) + \Sigma\, \Omega(f_k) \qquad (2)$$

donde el primer término mide el error de predicción y el segundo penaliza la complejidad de los árboles individuales. El término de regularización Ω incluye una penalización sobre el número de hojas ($\gamma \cdot T$) y una penalización L2 sobre los pesos de las hojas ($\lambda \cdot ||w||^2$), lo que permite controlar el sobreajuste sin recurrir a técnicas externas de poda. La elección de L2 sobre L1 responde a que la penalización cuadrática es diferenciable en todo su dominio, condición que el algoritmo aprovecha para derivar analíticamente la ganancia óptima de cada división y el peso óptimo de cada hoja; una penalización L1 introduciría una singularidad en cero que impediría este cálculo cerrado [27].

Tabla 15. Hiperparámetros principales de *XGBoost*

Hiperparámetro	Rol

n_estimators	Número de árboles secuenciales. Mayor cantidad mejora el ajuste pero incrementa el riesgo de sobreajuste y el costo computacional.
max_depth	Profundidad máxima de cada árbol. Limita la capacidad de capturar interacciones de orden alto entre *features*.
learning_rate (η)	Tasa de encogimiento aplicada a cada árbol. Valores bajos requieren más árboles pero producen soluciones más robustas.
subsample	Fracción de observaciones muestreadas para entrenar cada árbol. Introduce estocasticidad que reduce el sobreajuste.
colsample_bytree	Fracción de *features* muestreadas por árbol. Complementa el efecto de *subsample* al diversificar las particiones.
reg_alpha (α)	Penalización L1 sobre los pesos de las hojas. Promueve soluciones dispersas (pesos nulos).
reg_lambda (λ)	Penalización L2 sobre los pesos de las hojas. Contrae los coeficientes y estabiliza el modelo.

Fuente: elaboración propia a partir de la documentación de XGBoost.

La combinación de submuestreo de filas (*subsample*) y de columnas (*colsample_bytree*) introduce un efecto de *bagging* implícito dentro del proceso de *boosting*, lo que mitiga el sobreajuste sin sacrificar la capacidad del modelo para capturar relaciones no lineales entre los rezagos y los estadísticos de ventana deslizante.

Modelo Ridge Regression

Como línea base de referencia se implementó *Ridge Regression,* una variante de la regresión lineal que incorpora una penalización L2 sobre la magnitud de los coeficientes [58]. Su función objetivo se define como:

$$L_{Ridge} = ||y - X\beta||^2 + \alpha \cdot ||\beta||^2 \quad (3)$$

donde α es el hiperparámetro de regularización que controla la intensidad de la penalización. Cuando α = 0, el modelo se reduce a la regresión lineal por mínimos cuadrados ordinarios (*OLS*); a medida que α crece, los coeficientes se contraen hacia cero, reduciendo la varianza del estimador al costo de introducir un sesgo controlado.

Justificación frente a la regresión lineal sin regularización

La decisión de emplear *Ridge* en lugar de *OLS* se fundamenta en la estructura de las variables predictoras. Cuando se tabula una serie de tiempo mediante rezagos, las columnas lag_1, lag_2, ..., lag_k presentan, por construcción, una correlación serial elevada, es decir, multicolinealidad. Bajo este escenario, la matriz X^TX se encuentra próxima a la singularidad, lo que produce estimaciones *OLS* con varianzas desproporcionadamente grandes y alta sensibilidad a perturbaciones en los datos. La penalización L2 de *Ridge* añade el término $\alpha \cdot I$ a la diagonal de X^TX, estabilizando la inversión de la matriz y produciendo coeficientes más robustos [58]. La penalización cuadrática actúa encogiendo los coeficientes hacia cero de manera proporcional a su magnitud: los predictores altamente correlacionados, que en *OLS* recibirían pesos grandes y de signos opuestos para compensarse mutuamente, quedan restringidos a valores moderados. Este encogimiento introduce un sesgo controlado a cambio de una reducción sustancial en la varianza de las estimaciones, un intercambio que resulta favorable en presencia de multicolinealidad severa, donde la varianza sin regularización domina el error cuadrático medio total [58].

Ajuste de Hiperparámetros

La calibración de hiperparámetros define el balance sesgo-varianza de cada modelo y, por tanto, su capacidad de generalización. Dado que el torneo opera de forma autónoma sobre cientos de ingredientes con distribuciones de demanda heterogéneas, la estrategia de búsqueda debía mantener un costo computacional acotado sin comprometer la validez de la selección.

Para XGBoost se aplicó búsqueda exhaustiva sobre una grilla predefinida (grid search) integrada con validación cruzada walk-forward. Cada configuración fue evaluada con el mismo esquema de n_splits folds descrito en la sección de validación, garantizando que la optimización ocurriera íntegramente dentro de la ventana de entrenamiento de cada producto, sin contaminación de los datos de prueba. El criterio de selección fue el RMSE promedio sobre los folds interiores; la configuración ganadora fue reentrenada sobre el historial completo antes de producir las predicciones finales. Los rangos explorados fueron la profundidad del árbol (max_depth $\in \{3, 5, 7\}$), tasa de aprendizaje (learning_rate $\in \{0.05, 0.1\}$), submuestreo de observaciones y columnas (subsample, colsample_bytree $\in \{0.6, 0.8, 1.0\}$) y penalización L2 (reg_lambda $\in \{0.1, 1.0, 10.0\}$); El parámetro n_estimators se fijó en un valor alto para que la búsqueda operara sobre la complejidad estructural de los árboles.

Para Ridge, la búsqueda del parámetro α se realizó con la implementación RidgeCV de scikit-learn, que ejecuta validación cruzada eficiente sobre una grilla logarítmica (α ∈ {0.01, 0.1, 1, 10, 100}) con costo computacional equivalente a un único entrenamiento [58]. El rango cubre desde regularización mínima hasta contracción pronunciada de coeficientes, permitiendo que el modelo ajuste la intensidad de la penalización según la estructura de multicolinealidad propia de cada ingrediente.

Prophet, cuya API no es compatible de forma eficiente con el bucle walk-forward en ejecuciones masivas, fue ajustado mediante un único split cronológico. El único hiperparámetro optimizado fue changepoint_prior_scale (rango: {0.001, 0.05, 0.3, 0.5}), que cubre el espectro de tendencias rígidas a altamente adaptativas [41]. Los demás componentes (estacionalidades anual y semanal activas, diaria desactivada) se fijaron a priori con base en el conocimiento del ciclo productivo industrial documentado en la Figura 7, dado que su configuración responde a la estructura del patrón de consumo de la empresa y no a características estadísticas de cada serie individual.

Comparación de los Modelos Seleccionados

Tabla 16. Características comparativas de los modelos implementados

Criterio	*Prophet*	*XGBoost*	*Ridge*
Paradigma	Aditivo bayesiano con descomposición explícita	Ensemble de *gradient boosting* con árboles	Regresión lineal con penalización L2
Captura de no linealidad	Limitada (tendencia segmentada + *Fourier*)	Alta (particiones recursivas)	Nula (relación lineal)
Estacionalidad	Explícita (series de *Fourier*)	Implícita (a través de *lags* y *rolling features*)	Implícita (a través de *lags* y *rolling features*)
Regularización	Priors bayesianos	L1 + L2 + submuestreo	L2 (α)
Incertidumbre	Intervalos de credibilidad nativos	No nativa (requiere técnicas adicionales)	No nativa
Rol en el estudio	Modelo de descomposición temporal	Modelo no lineal principal	Línea base regularizada

Fuente: elaboración propia.

Evaluación Competitiva y Selección de Modelos

El mecanismo de evaluación competitiva que permite identificar de manera automática y para cada ingrediente es explicado a continuación; este mecanismo se denomina torneo de pronósticos y se sustenta en dos pilares: una estrategia de validación cruzada adaptada a series de tiempo y un conjunto de métricas de error que cuantifican la precisión absoluta como la precisión relativa frente a un pronóstico sin de referencia.

Validación Cruzada Walk-Forward

La evaluación del rendimiento de modelos predictivos en series de tiempo requiere un tratamiento particular que respete la estructura temporal de los datos. A diferencia de la validación cruzada estándar con particiones aleatorias, que puede incurrir en fuga de información al utilizar datos futuros para entrenar, la validación *walk-forward*, también denominada validación con ventana expansiva, garantiza que cada fold de test contenga exclusivamente observaciones posteriores a las del fold de entrenamiento [55].

Se empleó la implementación *TimeSeriesSplit* de scikit-learn, configurada con *n* particiones (*n_splits*) y un tamaño de ventana de test fijo. En cada fold, el modelo se entrena con todos los datos disponibles hasta el punto de corte y se evalúa sobre los días siguientes. El error reportado es el promedio aritmético del *RMSE* obtenido en cada fold, lo que produce una estimación más estable y menos sensible a la elección arbitraria de una ventana particular de evaluación [60].

Justificación frente al *split* único

La alternativa más simple, un único *split* cronológico en proporción 80/20, adolece de una limitación fundamental: el error resultante refleja el desempeño del modelo en una ventana temporal específica, que puede ser representativa de un periodo favorable o desfavorable por azar. Si la ventana de test coincide con un periodo de demanda estable, el error será artificialmente bajo; si coincide con un evento atípico, será artificialmente alto. Al promediar múltiples folds, se mitiga

esta varianza y se obtiene una estimación más confiable de la capacidad de generalización del modelo. Esta mejora es especialmente relevante en series de demanda de ingredientes industriales, donde la presencia de estacionalidad, promociones y paros productivos introduce heterogeneidad temporal significativa.

Métricas de Evaluación

***RMSE*: métrica de selección**

La Raíz del Error Cuadrático Medio (*RMSE*) se adoptó como métrica principal de selección. Su formulación se expresa como:

$$RMSE = \sqrt{[\ (1/n) \cdot \Sigma\ (y_i - \hat{y}_i)^2\ ]} \quad (7)$$

donde y_i representa el valor real e $\hat{y}_i$ la predicción del modelo. El *RMSE* penaliza de manera cuadrática los errores grandes, lo que resulta deseable en el abastecimiento de ingredientes: un error de pronóstico excesivo en un periodo crítico puede derivar en desabastecimiento de línea de producción o en sobrestock con riesgo de vencimiento. La métrica se encuentra en las mismas unidades que la variable original (kilogramos, litros, unidades), lo que facilita su interpretación operativa.

***MASE*: métrica de diagnóstico**

Como complemento, se incorporó el Error Absoluto Escalado Medio (*MASE*), propuesto por Hyndman y Koehler [61] como solución a las limitaciones del MAPE en presencia de valores cercanos a cero. El *MASE* normaliza el *MAE* del modelo por el *MAE* de un pronóstico ingenuo (*naive forecast*) de referencia:

$$MASE = MAE_{\text{modelo}} \ / \ MAE_{naive} \quad (8)$$

donde el denominador corresponde al *MAE* de un pronóstico que repite el último valor observado ($\hat{y}_t = y_{t-1}$). El *MASE* posee tres propiedades que justifican su inclusión: es independiente de la escala de la variable, es simétrica ante errores positivos y negativos, y proporciona un punto de referencia intuitivo: un valor inferior a 1 indica que el modelo supera al *naive*; un valor igual o superior a 1 constituye una señal de advertencia de que el modelo no aporta valor predictivo por encima de la regla más simple posible.

Tabla 17. Métricas empleadas en el torneo de pronósticos

Métrica	Rol	Interpretación clave
RMSE	Selección del modelo ganador (promedio *walk-forward*)	Menor es mejor. En unidades de la variable. Penaliza errores grandes.
MASE	Diagnóstico complementario	< 1: supera al *naive*. ≥ 1: el modelo no aporta valor sobre la regla más simple.

Fuente: elaboración propia a partir de Hyndman y Koehler [61].

Diseño del Torneo de Pronósticos

El torneo opera de manera independiente para cada producto que cumple el requisito mínimo de historia, definido como el número de meses necesarios para alimentar los folds de validación cruzada más los rezagos requeridos por los modelos tabulares. Los productos con historia insuficiente se asignan a un pronóstico por promedio simple como mecanismo de respaldo.

Flujo de evaluación por producto

Para cada ingrediente calificado, la serie diaria de cantidad consumida se agrega a nivel diario mediante una suma, rellenando con ceros los días sin transacciones.

Tabla 18. Estrategia de evaluación por modelo

Modelo	Estrategia de validación	Justificación
Prophet	Split único cronológico (entrenamiento hasta T–d, test los últimos d días)	La API de *Prophet* no es compatible de forma eficiente con el bucle *walk-forward* cuando se evalúan cientos de productos.
XGBoost	*Walk-forward* con *n_splits* folds (*TimeSeriesSplit*). *RMSE* = promedio de folds.	La instanciación y el entrenamiento de *XGBoost* son computacionalmente ligeros, permitiendo múltiples folds sin penalización significativa.
Ridge	*Walk-forward* con *n_splits* folds (*TimeSeriesSplit*). *RMSE* = promedio de folds.	Mismo razonamiento. Además, *Ridge* es el modelo más rápido de entrenar, lo que hace viable un mayor número de folds.

Fuente: elaboración propia.

Selección del modelo ganador

El modelo ganador es aquel que obtiene el menor *RMSE* promedio entre los tres candidatos. En caso de que todos los modelos arrojen un *RMSE* infinito, se asigna *Ridge Regression* como modelo por defecto, dada su estabilidad numérica inherente. Adicionalmente, las predicciones de todos los modelos se restringen al dominio no negativo mediante la operación max(ŷ, 0), debido a que una demanda negativa carece de sentido físico en el dominio de abastecimiento de ingredientes.

Diagnóstico posterior mediante *MASE*

Una vez declarado el ganador para cada producto, se calcula su *MASE* sobre el último fold de evaluación. Este valor no participa en la decisión del torneo, que se rige exclusivamente por el *RMSE*, sino que funciona como indicador de diagnóstico a nivel de portafolio. Al finalizar el torneo masivo, el sistema reporta el porcentaje de productos cuyo *MASE* es igual o superior a 1, lo que permite al analista identificar aquellos ingredientes para los cuales ninguno de los tres modelos supera al pronóstico *naive*. Esto sugiere que dichos productos podrían beneficiarse de enfoques alternativos, como los métodos de Croston para demanda intermitente, o de la incorporación de variables exógenas que los modelos univariantes no capturan.

Ejecución Masiva y Criterios de Calificación

El torneo se ejecuta de manera masiva sobre todos los productos cuyo historial supera un umbral mínimo de meses (*min_history_months*), definido en el archivo de configuración. Este filtro de calificación cumple una función estadística: garantiza que los datos disponibles sean suficientes para alimentar al menos *n_splits* + 1 particiones de la validación cruzada, más los rezagos requeridos por los modelos tabulares. Sin este requisito mínimo, la validación *walk-forward* produciría folds de entrenamiento demasiado pequeños, con riesgo de sobreajuste y estimaciones de error poco confiables.

El proceso registra el tiempo de ejecución acumulado y reporta el avance cada 50 productos, lo que permite monitorizar el desempeño computacional y detectar productos cuya evaluación consume un tiempo desproporcionado, típicamente aquellos con series largas y modelos *Prophet* con

múltiples puntos de cambio. Al finalizar, se genera un resumen que incluye la distribución de modelos ganadores por frecuencia, el *MASE* promedio del portafolio y el porcentaje de productos que no superan al *naive*.

7. Generación de Pronósticos y Alertas de Tendencia

Una vez que el torneo de modelos ha asignado a cada ingrediente el algoritmo con menor *RMSE*, el sistema procede a generar los pronósticos de demanda futura que alimentarán las decisiones de abastecimiento. Por otra parte, un pronóstico puntual resulta insuficiente para la gestión operativa de compras, puesto que el planificador de abastecimiento necesita conocer no solo la demanda esperada, sino también el rango de valores plausibles y la dirección de la tendencia subyacente. En este capítulo se describen las tres capas que completan el sistema de pronóstico:

(I) la generación de predicciones puntuales con corrección de sesgo

(II) la construcción de intervalos de predicción empíricos al 80 % y 95 % de cobertura nominal

(III) el análisis de tendencia mediante el test de Mann-Kendall con estimación de la pendiente de Sen.

Estrategia de Predicción

El horizonte de pronóstico abarca un número configurable de días futuros. La estrategia para generar predicciones a múltiples pasos difiere según la naturaleza del modelo ganador:

Tabla 19. Estrategia de predicción según modelo

Modelo	Estrategia
Prophet	*Prophet* genera directamente el horizonte completo a partir de su descomposición aditiva de tendencia y estacionalidad, sin necesidad de retroalimentación.
XGBoost / *Ridge*	Cada predicción ŷ(t) se incorpora como lag para predecir ŷ(t+1). Los *rolling features* se recalculan en cada paso incluyendo las predicciones previas. Esto preserva la coherencia del vector de *features* a lo largo del horizonte.

Fuente: elaboración propia.

La predicción recursiva introduce un desafío el cual es la acumulación de error. Dado que cada paso utiliza como entrada las predicciones de pasos anteriores, los errores se propagan y amplifican conforme avanza el horizonte [62]. Este efecto motivó la implementación de un mecanismo de corrección de sesgo que se describe a continuación.

Corrección de Sesgo

Antes de emitir los pronósticos finales, se estima y corrige el sesgo sistemático del modelo. El sesgo se define como la media de los residuos ($\hat{y} - y$) observados durante el periodo de *backtest*, un segmento de validación al final de la serie histórica. Formalmente:

$$\text{sesgo} = (1/n) \cdot \Sigma\, (\hat{y}_i - y_i) \quad (9)$$

El pronóstico corregido se obtiene sustrayendo este sesgo medio de cada predicción futura: $\hat{y}^*_t = \max(\hat{y}_t - \text{sesgo}, 0)$. La operación $\max(\cdot, 0)$ garantiza la no negatividad del resultado, condición necesaria en el dominio de demanda de ingredientes. Si bien esta corrección es la más elemental dentro de la familia de métodos de recalibración, es efectiva para eliminar desviaciones sistemáticas y constituye una barrera como la predicción conformal o el bootstrap residual [63].

Intervalos de Predicción Empíricos

Para cuantificar la incertidumbre asociada a cada predicción, se construyeron intervalos de predicción empíricos a dos niveles de cobertura nominal: 80 % y 95 %. El procedimiento se fundamenta en los cuantiles de la distribución de residuos del *backtest*:

$$PI_\alpha = [\, \hat{y}^* + Q_{\alpha/2}(e)\,,\ \hat{y}^* + Q_{1-\alpha/2}(e)\,] \quad (10)$$

donde Q denota la función cuantil y *e* es el vector de residuos del *backtest*. Para el intervalo al 80 %, se emplean los cuantiles 10 y 90 de los residuos; para el intervalo al 95 %, los cuantiles 2.5 y 97.5.

Este enfoque opera bajo la hipótesis de que la distribución de los errores futuros será similar a la observada en el *backtest* (estacionariedad de los residuos). Aunque esta hipótesis puede no sostenerse ante cambios estructurales en la demanda, constituye una primera aproximación razonable y ampliamente utilizada en la práctica industrial [1]. En el caso particular de *Prophet*, el modelo proporciona intervalos de credibilidad bayesianos nativos, los cuales se adoptan como intervalo al 95% por su mayor fundamentación probabilística, conservando los intervalos empíricos para el nivel del 80%.

Métricas de Validación del Pronóstico

WMAPE

El Error Porcentual Absoluto Medio Ponderado (*WMAPE*) se calculó como métrica principal de validación del *backtest*. A diferencia del MAPE clásico, que promedia los errores porcentuales individuales y es inestable cuando los valores reales son cercanos a cero, el *WMAPE* pondera cada error por su magnitud real:

$$WMAPE = [\ \Sigma\ |y_i - \hat{y}_i|\ ]\ /\ \Sigma\ y_i \times 100 \quad (11)$$

Esta formulación asigna mayor peso a los periodos de alta demanda, que son precisamente los más críticos desde la perspectiva del abastecimiento y evita las distorsiones que el MAPE introduce en series con demanda intermitente [64].

MAE

Como complemento, se reporta el Error Absoluto Medio (*MAE*) en unidades físicas de la variable. Mientras el *WMAPE* permite comparaciones relativas entre productos con escalas diferentes, el *MAE* proporciona una lectura directa del error promedio en términos operativos, información necesaria para dimensionar los márgenes de seguridad en las órdenes de compra.

Análisis de Tendencia Mediante Test de Mann-Kendall y Pendiente de Sen

La última capa del sistema de pronóstico consiste en la detección automática de tendencias en la serie pronosticada, con el fin de activar alertas de abastecimiento cuando la demanda proyectada muestra un comportamiento creciente o decreciente. Para ello se implementó el test de Mann-Kendall, un test no paramétrico de hipótesis para tendencia monotónica [4], acompañado del estimador de la pendiente de Sen [65].

Estimador de la pendiente de Sen

La magnitud de la tendencia se cuantifica mediante la pendiente de Sen, definida como la mediana de todas las pendientes calculadas entre pares de puntos:

$$S_{Sen} = \text{mediana}\ \{\ (y_j - y_i)\ /\ (j - i)\ \}\quad \forall\ i < j \quad (12)$$

Al emplear la mediana en lugar de la media, este estimador hereda la robustez del test de Mann-

Kendall frente a valores extremos. El cambio porcentual total del periodo se calcula como el producto de la pendiente de Sen por el número de observaciones, normalizado por la media del pronóstico.

Regla de activación de alertas

La alerta de abastecimiento se dispara únicamente cuando se cumplen simultáneamente dos condiciones: el p-valor del test de Mann-Kendall es inferior al nivel de significancia configurado (por defecto, $\alpha = 0.10$) y el cambio porcentual acumulado supera un umbral mínimo. La conjunción de ambos criterios evita dos tipos de falsas alarmas, es decir, tendencias estadísticamente significativas, pero de magnitud irrelevante, y variaciones porcentuales grandes, pero estadísticamente indistinguibles del azar.

Tabla 20. Clasificación de alertas de tendencia

Etiqueta	Condición
ALERTA_ALZA	p-valor < α AND cambio porcentual > +umbral. Demanda creciente: revisar capacidad de abastecimiento.
ALERTA_BAJA	p-valor < α AND cambio porcentual < –umbral. Demanda decreciente: riesgo de sobrestock o vencimiento.
NORMAL	p-valor ≥ α OR |cambio porcentual| ≤ umbral. Demanda estable o sin tendencia significativa.

Fuente: elaboración propia.

8. Despliegue Operativo: Dashboard

En esta fase final de la metodología *CRISP-DM* correspondiente al despliegue, es el punto donde los resultados de los modelos de pronósticos migran a una herramienta utilizable por equipos que toman decisiones. Se construyó un tablero web interactivo que consume las salidas del *pipeline* y las presenta organizadas por planta, por categoría de riesgo y por tipo de comportamiento de la demanda.

El *dashboard* únicamente transmite la salida de los modelos, más no agrega una capa predictiva adicional. Esta segregación de la parte analítica ejecutada en Python y la presentación de la información, corresponde a una consideración práctica, puesto que las personas que deben consumir la información no deben ejecutar scripts ni tener conocimiento en Python.

8.1 Diseño y Estructura del Módulo

El módulo de Alertas y Pronósticos, contiene la tabla operativa. Cada fila corresponde a un producto y muestra su código, descripción, planta, clúster, tipo de alerta, cambio porcentual del periodo y modelo asignado. Al expandir una fila se despliega el gráfico de pronóstico a 30 / 60 / 90 días para ese producto de acuerdo con lo requerido, con la predicción puntual y las bandas de incertidumbre al 80% y 95%.

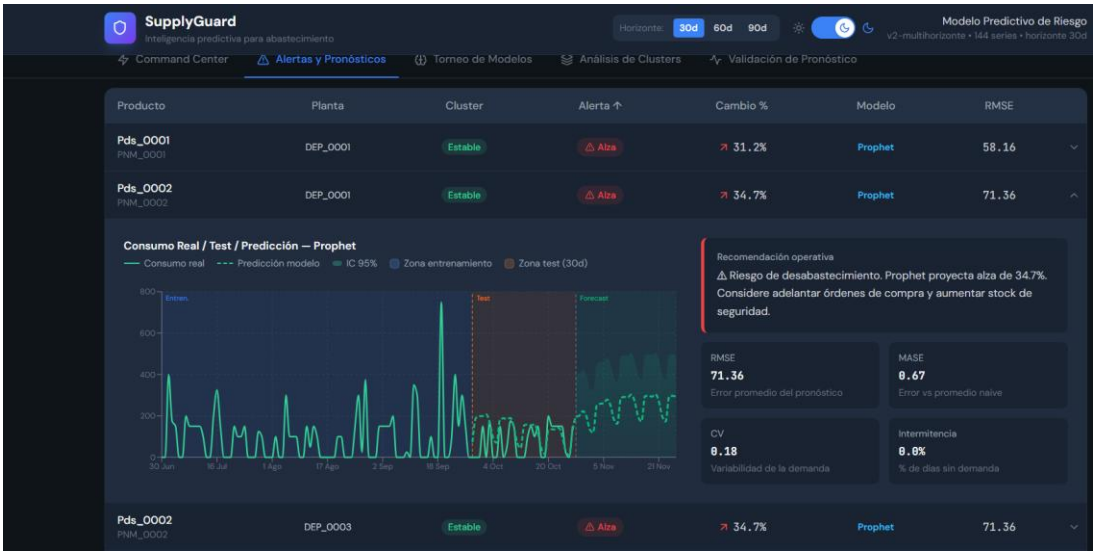

Figura 8. Módulo Alertas y Pronóstico*s del tablero*

9. Resultados y Discusión

Los resultados obtenidos en cada etapa del *pipeline* de pronóstico de demanda son presentados en esta sección, desde la caracterización del conjunto de datos hasta las alertas de tendencia generadas por el sistema.

9.1 Descripción del Conjunto de Datos Procesado

La distribución temporal evidencia que el volumen de transacciones mensuales se mantiene relativamente estable a lo largo del periodo de análisis, con picos acotados que pueden asociarse a ajustes de inventario o campañas de producción concretas. La ausencia de vacíos prolongados en la serie respalda la validez del conjunto para ejercicios de pronóstico de mediano plazo y facilita la construcción de ventanas de entrenamiento homogéneas en la validación cruzada.

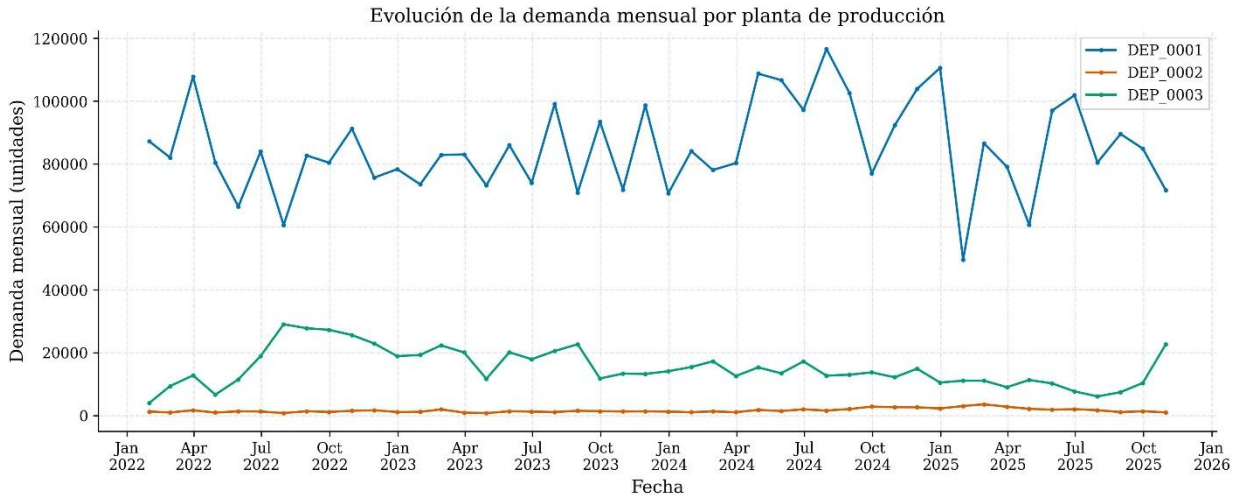

Figura 9. Evolución de la demanda mensual por planta de producción

La distribución de registros entre las tres plantas es heterogénea, lo que refleja las diferencias en volumen de operación y diversidad del portafolio manejado por cada instalación. Esta asimetría se consideró al momento de diseñar la validación *walk-forward*, dado que una planta con mayor volumen transaccional aporta, en promedio, series con mayor densidad de observaciones, por tanto, con mayor posibilidad de superar los requisitos mínimos de ventana para el entrenamiento de los modelos.

9.2 Resultados de la Ingeniería de Características y la Clasificación *SBC*

La extracción de métricas de comportamiento temporal reveló una diversidad significativa en los patrones de demanda del portafolio de ingredientes. La Tabla 21 resume los estadísticos descriptivos de las seis características numéricas calculadas para los 122 productos.

Tabla 21. Estadísticos descriptivos de las características por producto

Característica	Media	D.E.	Mín.	Máx.	Mediana	Rango
Volatilidad (CV)	1.726	1.791	0.131	6.782	1.075	6.651
Intermitencia	0.401	0.375	0.000	0.978	0.283	0.978
Entropía	2.780	1.142	0.000	3.820	3.312	3.820
ADI	5.752	10.529	1.000	46.000	1.353	45.000
CV^2	0.322	0.346	0.000	2.180	0.199	2.180
Hurst	0.671	0.175	0.495	1.000	0.665	0.505

Fuente: elaboración propia a partir de los datos procesados.

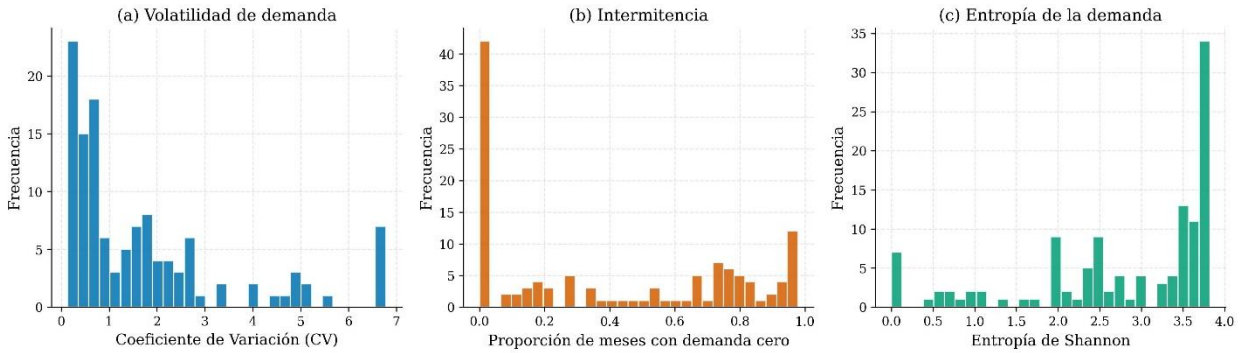

Figura 10. Distribuciones de features de comportamiento temporal

El coeficiente de variación promedio de 1.726 muestra una alta heterogeneidad en la variabilidad de la demanda entre productos, con ingredientes que oscilan entre patrones casi constantes y demandas extremadamente erráticas. La intermitencia promedio de 0.401 indica que, en promedio, el 40.1 % de los meses presentan demanda nula, lo que confirma la prevalencia de patrones intermitentes los ítems analizados. El exponente de Hurst promedio de 0.671 revela la presencia de

persistencia en la mayoría de las series, lo que indica que las tendencias de la demanda tienden a perpetuarse en el tiempo.

Clasificación *SBC*

La aplicación del esquema de *SBC* permitió clasificar el portafolio en cuatro categorías: 54 productos *Smooth* (44.3 %), 47 *Intermittent* (38.5 %), 19 *Lumpy* (15.6 %) y 2 *Erratic* (1.6 %). Que la categoría *Smooth* represente menos de la mitad del portafolio indica que mayoría de los ingredientes (55.7 %) presenta algún grado de intermitencia o variabilidad atípica en sus pedidos, lo que limita la aplicación de métodos de suavizamiento exponencial estándar y justifica la implementación de un sistema de selección automática de modelos.

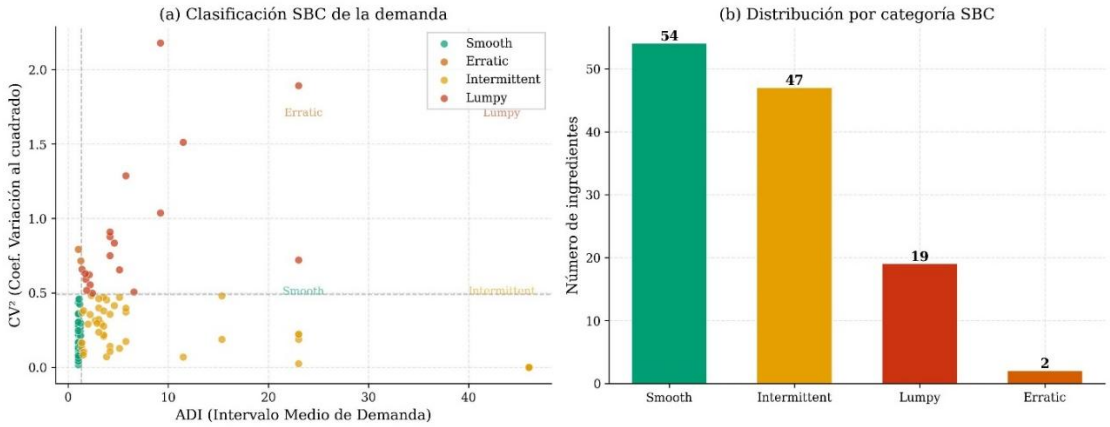

Figura 11. Distribución del portafolio según la clasificación SBC

9.3 Resultados del Agrupamiento

El torneo de algoritmos de *clustering* fue ganado por *DBSCAN*, que identificó dos clusters: el Cluster 0 con 114 productos (93.4 %) y el Cluster 1 con 8 productos (6.6 %). La distribución marcadamente asimétrica es debido a la naturaleza del modelo *DBSCAN* el cual agrupa en el clúster principal la masa de productos con densidad similar y aísla en clústers secundarios aquellos con perfiles atípicos.

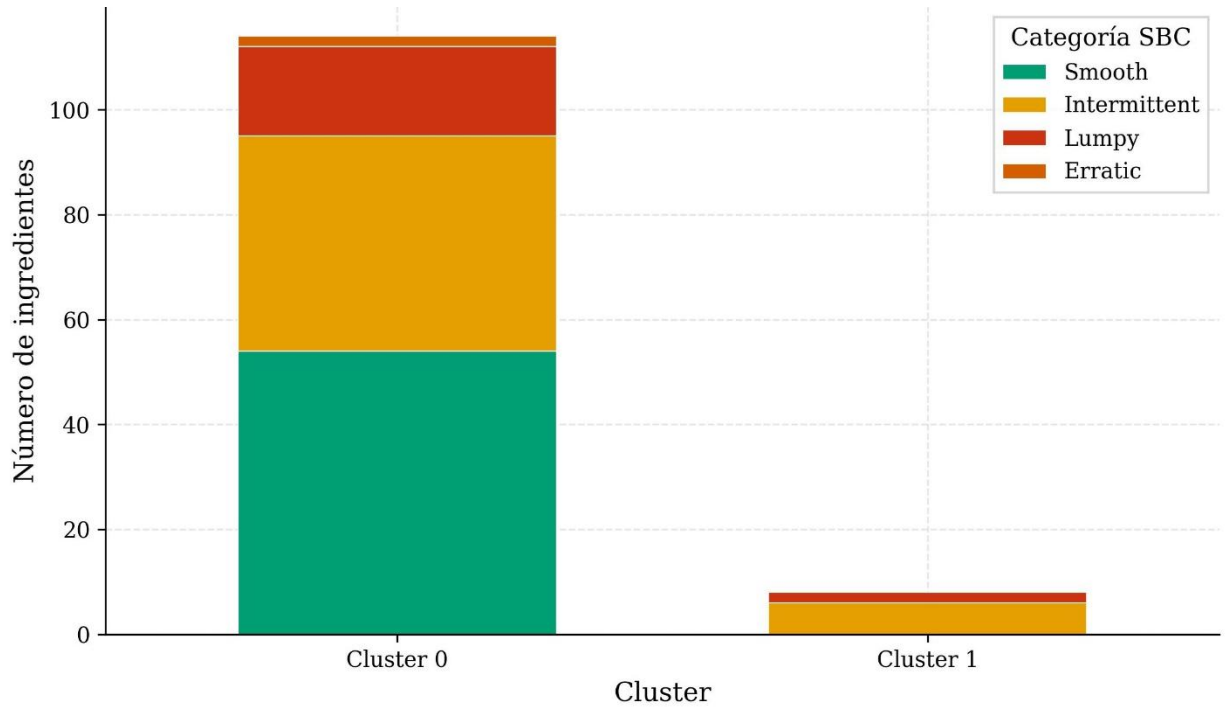

Figura 12. Caracterización comparativa de los clústers

La comparación de perfiles entre clústers evidencia que la segmentación capturó diferencias sustantivas en las dimensiones de intermitencia, volatilidad y memoria de largo plazo. El contraste más claro se observa en el exponente de Hurst, que pasa de 0.683 en el Cluster 0 a 0.500 en el Cluster 1, transición que marca el paso de series con persistencia a series cercanas al comportamiento aleatorio.

El Cluster 1 se compone exclusivamente de productos con demanda intermitente o *lumpy*, con un exponente de Hurst promedio de 0.500. En contraste, el Cluster 0 presenta un Hurst promedio de 0.683, lo que refleja series con memoria de largo plazo y con mayor potencial para el modelado predictivo.

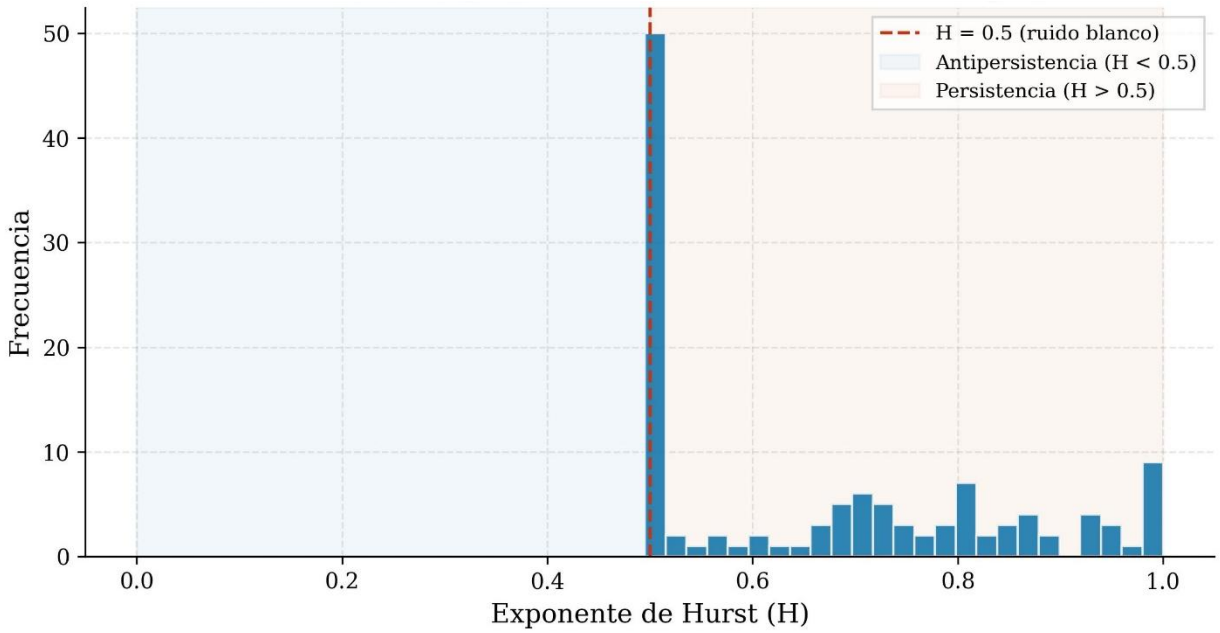

Figura 13. Distribución de productos entre el Cluster 0 y el Cluster 1

La asimetría entre los dos clusters está asociado al criterio de densidad empleado por *DBSCAN*, que aísla los productos con perfiles atípicos en un conglomerado minoritario en lugar de forzarlos dentro de una estructura mayor. Esta división a nivel operativo favorece, ya que separa el grupo mayoritario de productos con estructura temporal aprovechable de aquellos cuya demanda exige un tratamiento diferenciado o métodos de pronóstico alternativos.

9.4 Resultados del Torneo de Modelos

El torneo de pronósticos evaluó los tres modelos (*Prophet*, *XGBoost* y *Ridge Regression*) para cada uno de los 122 productos. A continuación, se detallan los resultados del torneo de modelos en la Figura 14:

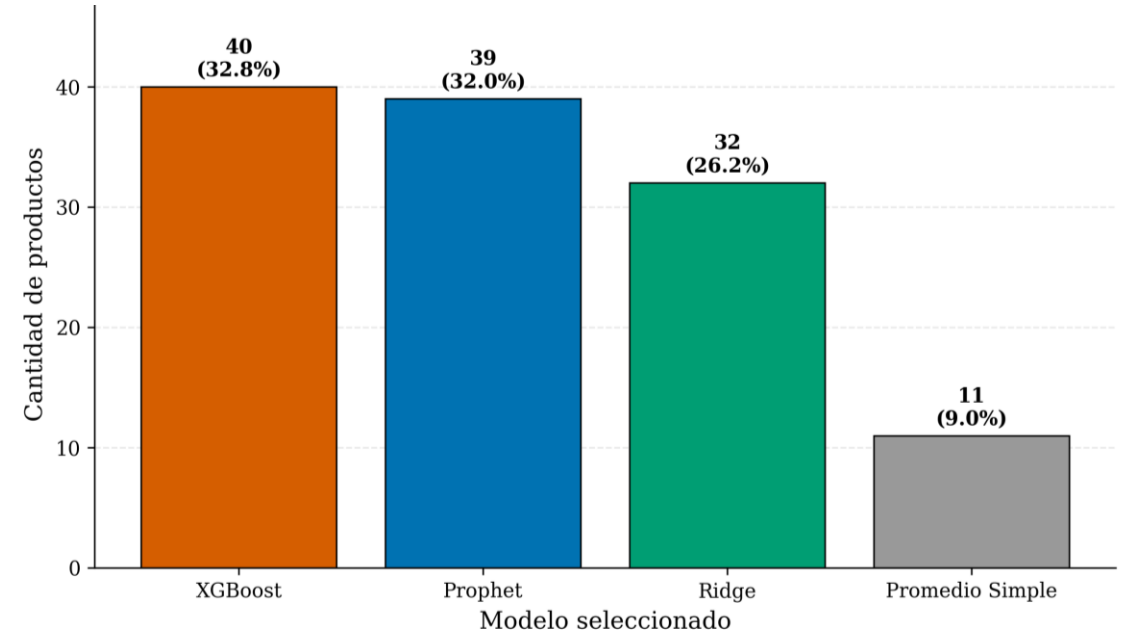

Figura 14. Distribución de modelos ganadores en el torneo

La distribución de modelos ganadores es equilibrada entre los tres candidatos principales: *XGBoost* fue seleccionado para 40 productos (33.6 %), *Prophet* para 39 (32.0 %) y *Ridge* para 32 (25.4 %). Los 11 productos restantes (9.0 %) fueron asignados al pronóstico por promedio simple por no cumplir el requisito mínimo de historia. Este equilibrio valida una de las premisas iniciales y el motivo principal por el que se ejecutaron 3 modelos de *forecast*: no existe un modelo superior para todos los casos, y la selección automática por producto aporta valor frente a la imposición de un único algoritmo para todos los productos.

Tabla 22. Métricas de Rendimiento por Modelo

Modelo	Productos	Proporción	*RMSE* mediano	*MASE* mediano
XGBoost	40	33.6 %	5.63	1.161
Prophet	39	32.0 %	28.04	1.173
Ridge	32	25.4 %	17.80	1.106
Promedio Simple	11	9.0 %	—	—

Fuente: elaboración propia

XGBoost obtuvo la mediana de *RMSE* más baja (5.63), es decir, para los cuales resultó ganador tienden a presentar errores de magnitud menor. *Ridge* se ubicó en una posición media (17.80) y *Prophet* registró la mediana más alta (28.04), lo que puede atribuirse a que fue seleccionado preferentemente para productos con demanda de mayor volumen absoluto, donde el error en unidades es naturalmente mayor. En términos de *MASE*, las tres medianas resultan muy cercanas (*Ridge* 1.106, *XGBoost* 1.161, *Prophet* 1.173), lo que confirma que la diferencia en *RMSE* refleja sobre todo la escala de la demanda y no una superioridad de un modelo sobre otro.

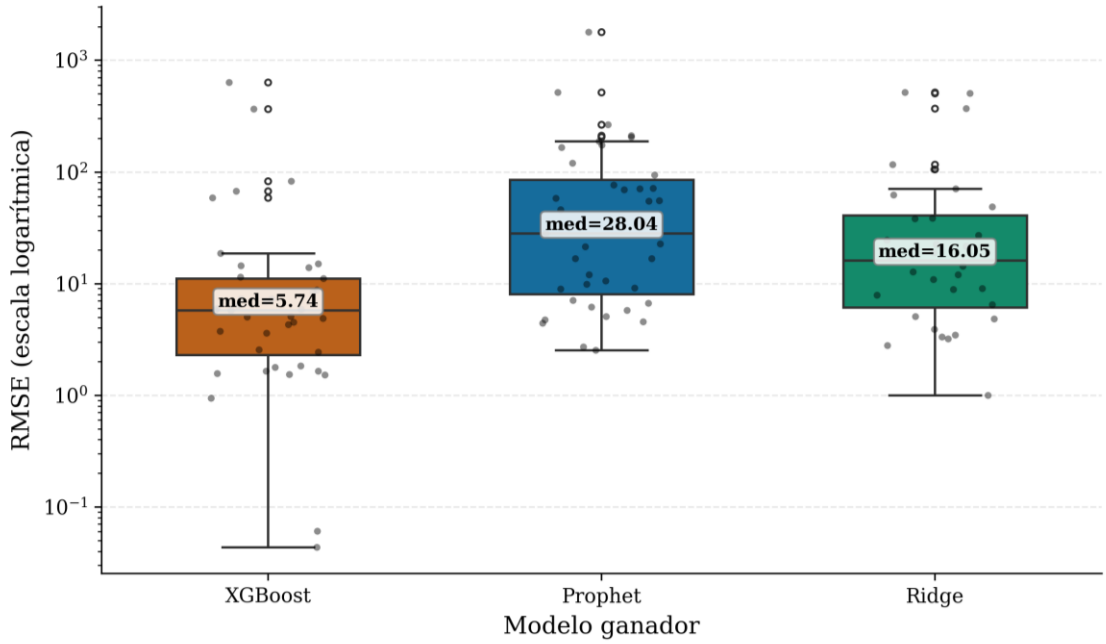

Figura 15. Distribución del RMSE por modelo ganador

La dispersión del *RMSE* dentro de cada modelo revela asimetrías marcadas, puesto que por una parte *XGBoost* concentra la mayoría de sus casos en magnitudes bajas de error, mientras que *Prophet* y *Ridge* presentan colas más largas hacia valores elevados. Este comportamiento es consistente con la naturaleza de cada algoritmo; el *gradient boosting* se ajusta mejor a productos con rezagos informativos y demanda acotada, mientras que *Prophet* tiende a ganar en series de mayor volumen donde el error absoluto crece proporcionalmente al nivel de la demanda. La

comparación directa entre medianas, por tanto, debe leerse en conjunto con la escala de cada producto pronosticado.

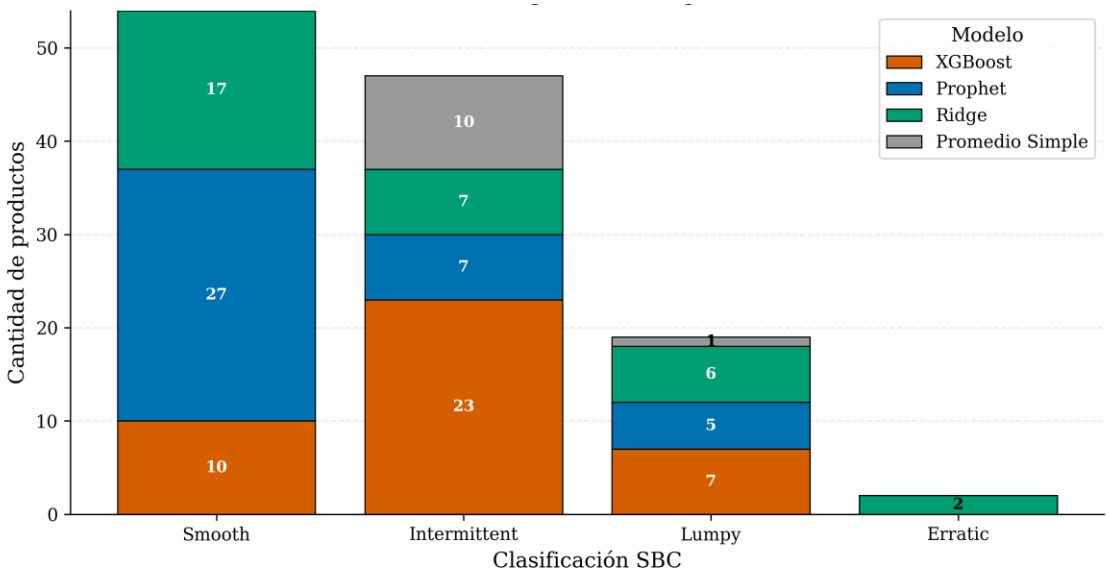

Figura 16. Relación entre la clasificación SBC y el modelo ganador

La Figura 16 revela la Relación entre clasificación *SBC* y modelo ganador, mostrando una relación entre la demanda y el modelo óptimo. *Prophet* domina en la categoría *Smooth* (27 de 54, 50.0 %), lo que resulta coherente con su diseño en el cual, su descomposición aditiva de tendencia y estacionalidad es más efectiva cuando la serie presenta regularidad temporal. *XGBoost*, en cambio, predomina en la categoría *Intermittent* (23 de 47, 48.9 %), donde su capacidad para capturar relaciones no lineales entre los rezagos le permite adaptarse a patrones de demanda esporádicos. *Ridge* se distribuye de manera más uniforme, con presencia relevante tanto en productos *Smooth* como *Intermittent*, lo que sugiere que la regularización L2 ofrece una solución robusta cuando los modelos más complejos no logran una ventaja.

9.5 Resultados de los Pronósticos

Se ilustra en la Figura 17 el comportamiento de la demanda y pronósticos para un item de referencia que es el más representativo a nivel de volúmenes de consumo para la organización. La gráfica se compone del consumo histórico, un periodo de validación de 60 días y, a partir de allí, el pronósico

en horizontes de 30, 60 y 90 días.

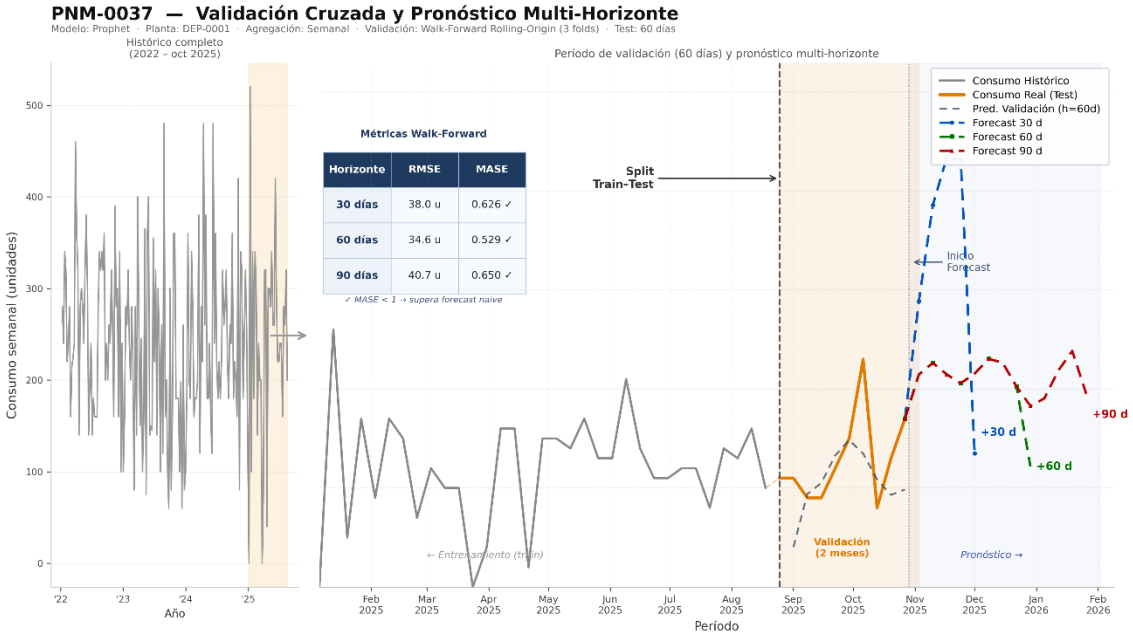

Figura 17. Pronósticos de demanda

Intervalos de predicción

Los intervalos de predicción empíricos se calcularon a partir de los cuantiles de los residuos del *backtest*. El ancho promedio del intervalo al 80 % fue de 49.11 unidades y el del intervalo al 95 % de 93.32 unidades. Para los productos pronosticados con *Prophet*, los intervalos resultaron más amplios, dado que este modelo genera intervalos bayesianos nativos que incorporan la incertidumbre completa del proceso generador. Para *XGBoost* y *Ridge*, los intervalos empíricos resultaron más estrech4s, lo que refleja tanto la menor varianza de sus predicciones como una posible subestimación de la incertidumbre inherente al método de cuantiles residuales. Esta diferencia sugiere que, en futuras iteraciones, convendría explorar técnicas de predicción conformal para los modelos de aprendizaje automático.

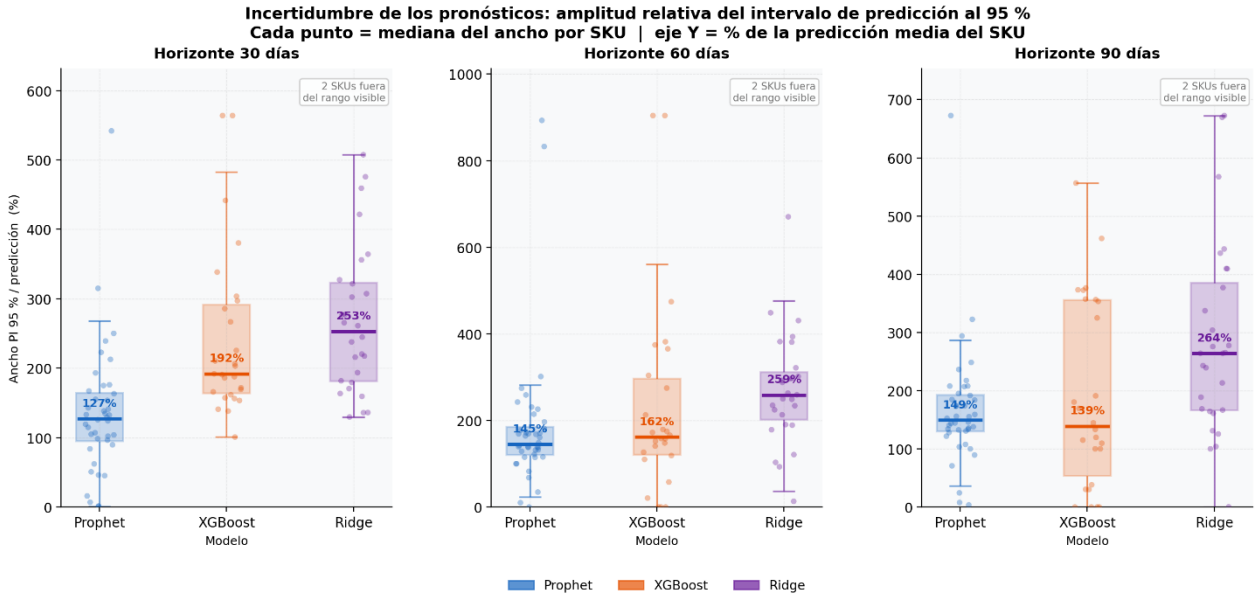

Figura 18. Amplitud de intervalos de predicción

9.6 Resultados del Análisis de Tendencias

El test de Mann-Kendall se aplicó a las series pronosticadas de los 122 productos evaluados. De estos, 66 items (54 %) presentaron tendencias estadísticamente significativas al nivel $\alpha = 0.10$. Sin embargo, la doble condición de significancia estadística y magnitud de cambio produjo la activación de alertas para 56 productos (45.9 %).

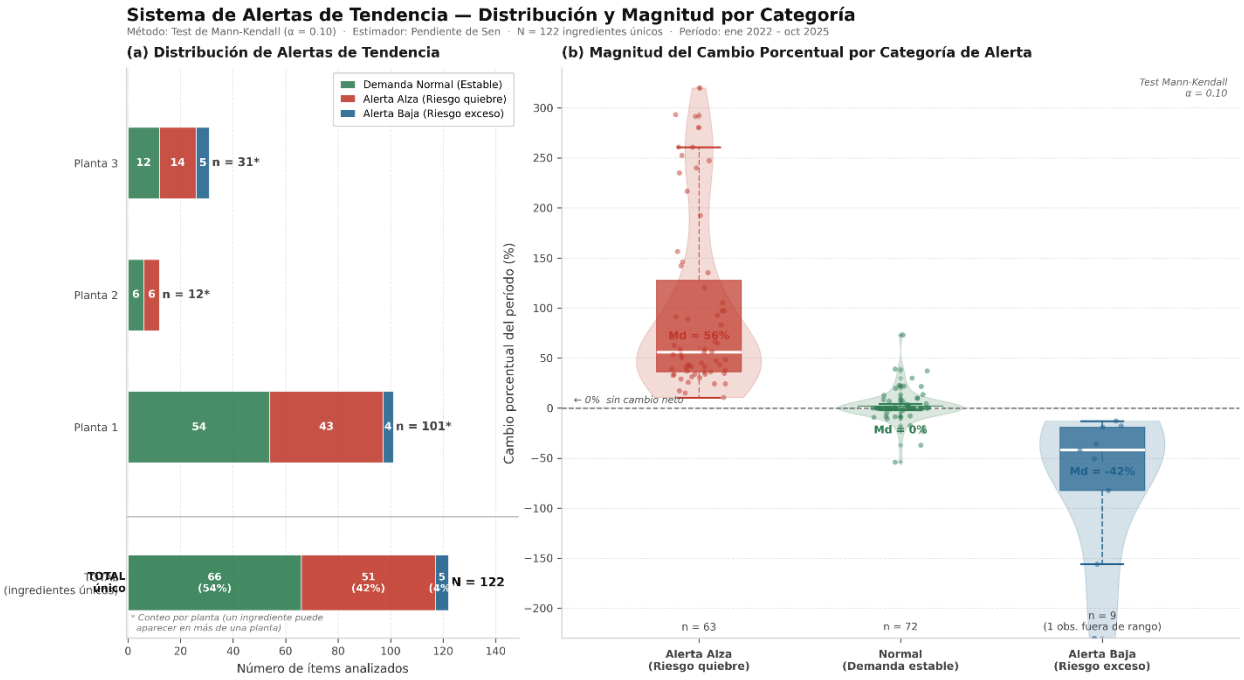

Figura 19. Sistema de alertas de abastecimiento

10. Conclusiones

El presente trabajo de grado abordó el diseño, la implementación y la evaluación de un sistema integral de pronóstico de demanda de ingredientes para una empresa del sector alimentario con tres plantas de producción en Colombia. A lo largo de los capítulos se describió *un pipeline* que comprende desde la limpieza de datos transaccionales hasta la generación de alertas de tendencia, pasando por la ingeniería de características, la segmentación de productos, la selección automática de modelos y la cuantificación de incertidumbre.

Diversidad de patrones de demanda y la pertinencia de la segmentación

La aplicación del esquema de clasificación de *SBC* reveló que únicamente el 44.3 % de los ingredientes del portafolio presenta demanda *Smooth*; el 55.7 % restante se distribuye entre patrones *Intermittent* (38.5 %), *Lumpy* (15.6 %) y *Erratic* (1.6 %). A través de esta segmentación de la demanda se apoya el proceso de gestión del abastecimiento de la empresa, puesto que evidencia que la mayoría de los ingredientes no se ajustan al supuesto de demanda continua y regular que generalmente es atendida mediante los métodos de planificación tradicionales empleados en la industria. La diversidad observada justifica la implementación de un sistema de selección automática de modelos y deja en desventaja la adopción de un único método de pronóstico para la totalidad del portafolio.

Selección automática de modelos

El torneo de modelos entre *Prophet*, *XGBoost* y *Ridge Regression* produjo una distribución de ganadores notablemente equilibrada: *XGBoost* fue seleccionado como modelo óptimo para el 33.8 % de los productos, *Prophet* para el 32.0 % y *Ridge* para el 26,2 %. Este resultado proporciona evidencia empírica del principio *No Free Lunch* que afirma que ningún algoritmo dominó a los demás en todos los escenarios. La validación cruzada *walk-forward*, al promediar el *RMSE* sobre múltiples folds temporales, ofreció estimaciones de error más estables que las que habría producido un único *split* cronológico, reduciendo el riesgo de seleccionar un modelo que se beneficiara artificialmente de una ventana de evaluación favorable.

Modelo Optimo según la naturaleza de la demanda

Se identificó una asociación entre la clasificación *SBC* y el modelo ganador; por una parte, prophet predominó en productos con demanda *Smooth* (50.0 % de los casos en esta categoría), mientras que *XGBoost* dominó en la categoría *Intermittent* (48.9 %). Este patrón es coherente con los fundamentos de cada algoritmo: la descomposición aditiva de *Prophet* es más efectiva cuando existen estacionalidades regulares, mientras que la capacidad de *XGBoost* para capturar relaciones no lineales entre los rezagos le permite adaptarse a patrones esporádicos. *Ridge Regression* se distribuyó de forma más uniforme, confirmando su rol como línea base robusta cuando los modelos más complejos no logran una ventaja decisiva. Esta asociación sugiere que, en una evolución futura del sistema, la clasificación *SBC* podría utilizarse como criterio de pre-asignación de familias de modelos, lo que reduciría el costo computacional del torneo sin sacrificar calidad predictiva.

Robustez estadística del *pipeline* de preprocesamiento

La elección del identificador de Hampel basado en la Desviación Absoluta Mediana con punto de quiebre del 50 % como método de detección de *outliers*, demostró ser más apropiada que el filtro de Tukey convencional para series de demanda con alta asimetría y múltiples valores extremos simultáneos. La aplicación segmentada por producto, junto con el mecanismo de respaldo al *IQR* para series constantes, garantizó que el tratamiento de *outliers* se adaptara a las particularidades de cada ingrediente sin requerir intervención manual. Así mismo, la parametrización centralizada del *pipeline*, aseguró la reproducibilidad completa.

Cuantificación de incertidumbre y las alertas de tendencia

La incorporación de intervalos de predicción empíricos al 80 % y 95 % y del análisis de tendencia mediante el test de Mann-Kendall con la pendiente de Sen es un avance respecto al pronóstico puntual convencional. El sistema de alertas identificó 51 ingredientes con tendencia creciente significativa y 5 con tendencia decreciente, proporcionando información accionable para la planificación de compras.

Estimación del impacto económico por reducción del riesgo de desabastecimiento

El beneficio económico del sistema se traduce en dos ventajas concretas para la operación. La primera es la anticipación: el sistema detectó que 56 ingredientes muestran tendencias que, si no se atienden a tiempo, pueden derivar en compras urgentes a precios más altos que los negociados, en paradas de producción por falta de materiales, o en incumplimientos de entrega que dañan las relaciones comerciales de largo plazo con los clientes. Al recibir la alerta con suficiente antelación, el equipo de compras puede actuar antes de que cualquiera de esos costos se materialice.

La segunda ventaja es el manejo más inteligente del inventario de seguridad, es decir, en lugar de aplicar el mismo colchón de stock a todos los ingredientes por igual, los rangos de predicción generados permiten asignar una reserva proporcional al riesgo real de cada uno, liberando capital inmovilizado en los productos de demanda estable y concentrándolo donde realmente se necesita. Para estimar el ahorro en términos monetarios sería necesario combinar estas predicciones con los costos de compra por ingrediente, los tiempos de parada históricos por desabastecimiento y los márgenes de los productos terminados afectados, un análisis que se recomienda adelantar durante la fase de implementación operativa del sistema.

11. Trabajos Futuros

A partir de los resultados obtenidos y las limitaciones identificadas, se proponen las siguientes líneas de desarrollo que podrían fortalecer la capacidad predictiva y la aplicabilidad operativa del sistema.

Incorporación de modelos especializados en demanda intermitente

La línea de trabajo de mayor prioridad consiste en incluir en el torneo de modelos las familias diseñadas específicamente para demanda intermitente el método de Croston, su variante con corrección de sesgo *SBA* y el método *TSB*. Estos modelos descomponen la demanda en dos procesos independientes y han demostrado ventajas sobre los métodos de propósito general en las categorías *Intermittent* y *Lumpy* del esquema *SBC*. Su incorporación permitiría abordar directamente la limitación evidenciada por el diagnóstico *MASE*, mejorando la precisión del pronóstico para el 55.7 % del portafolio que presenta patrones no regulares.

Mejora de la cuantificación de incertidumbre mediante predicción conformal

Los intervalos de predicción empíricos basados en cuantiles residuales, si bien constituyen una primera aproximación útil, no ofrecen garantías formales de cobertura. La predicción conformal, podría presentar una mejora sustancial, especialmente para los modelos *XGBoost* y *Ridge*, cuyos intervalos resultaron más estrechos que los de *Prophet*. La implementación de este enfoque requeriría un conjunto de calibración adicional, pero no modificaría la arquitectura del *pipeline* existente.

Incorporación de variables exógenas

El sistema actual opera exclusivamente con información univariante (la serie histórica de demanda de cada ingrediente). La incorporación de variables exógenas, como los planes de producción, los calendarios de mantenimiento de planta, la estacionalidad de los productos terminados que consumen cada ingrediente, indicadores macroeconómicos o el precio del ingrediente, podría capturar señales causales que los modelos univariantes no detectan.

Integración con el sistema *ERP* y automatización del ciclo de compras

El sistema actual genera pronósticos y alertas como archivos de salida que deben ser interpretados manualmente por el planificador de abastecimiento. Una mejora de alto impacto operativo consistiría en integrar el *pipeline* directamente con el módulo de abastecimiento del *ERP* de la empresa, de modo que los pronósticos se traduzcan automáticamente en pedidos sugeridos que consideren los tiempos de entrega del proveedor (*lead time*), los niveles de inventario en tiempo real, los mínimos de orden y las restricciones de capacidad de almacenamiento. Esta integración cerraría el ciclo entre la predicción y la acción, transformando el sistema de una herramienta analítica a un componente operativo del proceso de abastecimiento.

Referencias Bibliográficas

[1] M. Yang, M. K. Lim, Y. Qu, D. Ni y Z. Xiao, "Supply chain risk management with machine learning technology: A literature review and future research directions," *Computers & Industrial Engineering*, vol. 175, art. 108859, 2023.

[2] A. Aljohani, "Predictive Analytics and Machine Learning for Real-Time Supply Chain Risk Mitigation and Agility," *Sustainability*, vol. 15, n.º 20, art. 15088, oct. 2023.

[3] M. Nassif, M. A. Harfouche, D. Mukherjee y N. Tandon, "AI in Supply Chain Risk Assessment: A Systematic Literature Review and Bibliometric Analysis," arXiv preprint, arXiv:2401.10895, ene. 2024.

[4] C. Ptak y C. Smith, *Demand Driven Material Requirements Planning (DDMRP)*, 3.ª ed. Industrial Press, 2016.

[5] R. Miclo, F. Fontanili, M. Lauras, J. Lamothe y B. Milian, "MRP vs. Demand-Driven MRP: Towards an objective comparison," en *IFIP International Conference on Advances in Production Management Systems (APMS)*, Iguassu Falls, Brasil, pp. 590–597, 2016.

[6] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer y R. Wirth, "*CRISP-DM 1.0: Step-by-step data mining guide*," CRISP-DM Consortium, 2000.

[7] K. Katsaliaki, P. Galetsi y S. Kumar, "Supply chain disruptions and resilience: A major review and future research agenda," *Annals of Operations Research*, vol. 319, pp. 965–1002, 2022.

[8] A. Brintrup, E. Kosasih, P. Schaffer, G. Zheng, G. Demirel y B. L. MacCarthy, "Digital Supply Chain Surveillance Using Artificial Intelligence: Definitions, Opportunities and Risks," *International Journal of Production Research*, vol. 62, n.º 13, pp. 4674–4695, 2023.

[9] Z. Hamidu, F. O. Boachie-Mensah y K. Issau, "Supply chain resilience and performance of manufacturing firms: role of supply chain disruption," *Journal of Manufacturing Technology Management*, vol. 34, n.º 3, pp. 361–382, 2023.

[10] Y. Y. Alahmad, “The relationship between supply chain management practices and supply chain performance in Saudi Arabian firms,” *Amer. J. Ind. Bus. Manage.*, vol. 11, no. 1, pp.

42–59, 2021.

[11] J. Parra Peña, Y. A. Niño Villamizar y M. Suárez Serrano, "Reflexiones en torno a la logística de aprovisionamiento: Antecedentes y tendencias," *Ingeniería*, vol. 27, no. 2, 2022.

[12] J. Á. Alvarado, *La Inteligencia Artificial en la Administración de Empresas*. 2023.

[13] M. Gutiérrez y E. G. Polo, "*Inteligencia artificial dentro de la cadena de suministros*," Fundación Universitaria del Área Andina, Bogotá, Colombia, 2023.

[14] F. E. Webster y Y. Wind, *Organizational Buying Behavior*. Englewood Cliffs, NJ, EE. UU.: Prentice-Hall, 1972.

[15] A. van Weele, *Purchasing Control: Performance Measurement and Evaluation of the Industrial Purchasing Function*. Groningen, Países Bajos: Wolters-Noordhoff, 1984.

[16] C. J. Vidal, *Introducción a la gestión de inventarios*. Cali, Colombia: Universidad del Valle, Facultad de Ingeniería, 2006.

[17] J. L. Montes, *UF0476 - Gestión de inventarios*. Madrid, España: Editorial Elearning, S.L., 2014.

[18] V. Gutiérrez y C. J. Vidal, "Modelos de gestión de inventarios en cadenas de abastecimiento: Revisión de la literatura," *Rev. Fac. Ing. Univ. Antioquia*, no. 43, pp. 134–149, 2008.

[19] J. Yang, "The determinants of supply chain alliance performance: An empirical study," *Int. J. Prod. Res.*, vol. 47, no. 4, pp. 1055–1069, 2009.

[20] K. Petersen, R. Handfield y P. Cousins, "Buyer dependency and relational capital formation: The mediating effects of socialization," *J. Supply Chain Manage.*, vol. 44, no. 4, pp. 453–465, 2008.

[21] P. Cousins, R. Handfield, B. Lawson y K. Petersen, "Creating supply chain relational capital: The impact of formal and informal socialization processes," *J. Oper. Manage.*, vol. 24, no. 6, pp. 851–863, 2006.

[22] E. Rojas, “Glosario de los seis términos básicos del Machine Learning”, *MuyComputerPro*, 7 de feb. de 2018.

[23] V. Muñoz Jaramillo, “*Evaluación de modelos de Machine Learning para la predicción de crímenes en la ciudad de Medellín*,” Tesis de Maestría, Univ. Nac. de Colombia, Medellín, Colombia, 2021.

[24] J. Mueller y L. Massaron, *Machine Learning for Dummies*. Hoboken, NJ, EE. UU.: John Wiley & Sons, Inc., 2016.

[25] D. C. Montgomery, E. A. Peck y G. G. Vining, *Introduction to Linear Regression Analysis*, 6.ª ed. John Wiley & Sons, 2021.

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, n.º 1, pp. 5–32, 2001.

[27] T. Chen y C. Guestrin, "XGBoost: A Scalable Tree Boosting System," en *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[28] C. M. Bishop, *Pattern Recognition and Machine Learning*. Nueva York, NY, EE. UU.: Springer, 2007.

[29] T. Hastie, R. Tibshirani y J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2.ª ed. Springer, 2009.

[30] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ, EE. UU.: Princeton University Press, 1994.

[32] S. Raschka, “Model evaluation, model selection, and algorithm selection in machine learning,” arXiv preprint arXiv:1811.12808, 2020.

[33] M. Sokolova y G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, n.º 4, pp. 427–437, 2009.

[34] T. Chai y R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, n.º 3, pp. 1247–1250, 2014.

[35] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.

[36] D. L. Davies y D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, n.º 2, pp. 224–227, 1979.

[37] R. J. Hyndman y G. Athanasopoulos, *Forecasting: Principles and Practice*, 2.ª ed. Melbourne, Australia: OTexts, 2018.

[38] R. Wirth y J. Hipp, "CRISP-DM: Towards a standard process model for data mining," en *Proc. 4th Int. Conf. Pract. Appl. Knowl. Discovery Data Mining*, Manchester, Reino Unido, 2000.

[39] S. Chopra y P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, 6.ª ed. Pearson, 2016.

[40] M. Christopher, *Logistics & Supply Chain Management*, 4.ª ed. Pearson, 2011.

[41] S. J. Taylor y B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37-45, 2018.

[42] S. García, J. Luengo y F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham: Springer, 2015.

[43] P. J. Rousseeuw y C. Croux, "Alternatives to the median absolute deviation," *J. Amer. Statist. Assoc.*, vol. 88, n.º 424, pp. 1273–1283, 1993.

[44] L. Davies y U. Gather, "The identification of multiple outliers," *J. Amer. Statist. Assoc.*, vol. 88, n.º 423, pp. 782–792, 1993.

[45] F. R. Hampel, "The influence curve and its role in robust estimation," *J. Amer. Statist. Assoc.*, vol. 69, n.º 346, pp. 383–393, 1974.

[46] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.

[47] H. Zheng y J. Wu, "Feature engineering for machine learning," en *Principles and Practice*,

Cham: Springer, 2018.

[48] A. A. Syntetos y J. E. Boylan, “The accuracy of intermittent demand estimates,” *Int. J. Forecast.*, vol. 21, n.º 2, pp. 303–314, 2005.

[49] J. E. Boylan y A. A. Syntetos, “Forecasting for inventory management of service parts,” en *Complex System Maintenance Handbook*, London: Springer, 2008, pp. 479–506.

[50] H. E. Hurst, “Long-term storage capacity of reservoirs,” *Transactions of the ASCE*, vol. 116, pp. 770–799, 1951.

[51] M. Ester, H.-P. Kriegel, J. Sander y X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” en *Proc. KDD*, Portland, OR, 1996, pp. 226–231.

[52] J. Sander, M. Ester, H.-P. Kriegel y X. Xu, “Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications,” *Data Min. Knowl. Discov.*, vol. 2, n.º 2, pp. 169–194, 1998.

[55] C. Bergmeir y J. M. Benítez, “On the use of cross-validation for time series predictor evaluation,” *Inf. Sci.*, vol. 191, pp. 192–213, 2012.

[58] A. E. Hoerl y R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, n.º 1, pp. 55–67, 1970.

[60] R. J. Hyndman y G. Athanasopoulos, *Forecasting: Principles and Practice*, 3.ª ed. Melbourne: OTexts, 2021.

[61] R. J. Hyndman y A. B. Koehler, “Another look at measures of forecast accuracy,” *Int. J. Forecast.*, vol. 22, n.º 4, pp. 679–688, 2006.

[62] S. Makridakis, S. C. Wheelwright y R. J. Hyndman, *Forecasting: Methods and Applications*, 3.ª ed. New York: Wiley, 1998.

[63] V. Vovk, A. Gammerman y G. Shafer, *Algorithmic Learning in a Random World*. New York: Springer, 2005.

[64] S. Kolassa, “Evaluating predictive accuracy of intermittent demand forecasts,” *Int. J. Forecast.*, vol. 32, n.º 4, pp. 1163–1166, 2016.

[65] H. B. Mann, “Nonparametric tests against trend,” *Econometrica*, vol. 13, n.º 3, pp. 245–259, 1945.

[66] P. K. Sen, “Estimates of the regression coefficient based on Kendall’s tau,” *J. Amer. Statist. Assoc.*, vol. 63, n.º 324, pp. 1379–1389, 1968.

[68] D. H. Wolpert y W. G. Macready, “No free lunch theorems for optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 1, n.º 1, pp. 67–82, 1997.

[70] J. Beran, *Statistics for Long-Memory Processes*. Nueva York, NY, EE. UU.: Chapman & Hall, 1994.

[71] J. Han, M. Kamber y J. Pei, *Data Mining: Concepts and Techniques*, 3.ª ed. Waltham, MA, EE. UU.: Morgan Kaufmann, 2011.

[72] J. D. Salazar Vera y J. C. Antúnez Cueva, "*Modelo para la predicción de compra de materia prima en la gestión de inventario para producción en Pymes del sector retail aplicando machine learning*," Tesis de pregrado, Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Perú, 2024.

[73] G. Burstein y I. Zuckerman, "Deconstructing risk factors for predicting risk assessment in supply chains using machine learning," *Journal of Risk and Financial Management*, vol. 16, n.º 2, art. 97, 2023.

[74] G. Baryannis, S. Dani y G. Antoniou, "Predicting supply chain risks using machine learning: The trade-off between performance and interpretability," *Future Generation Computer Systems*, vol. 101, pp. 993–1004, 2019.