

## REGISTRO DE TRABAJO DE GRADO

FECHA	1	06	2025
-------	---	----	------

### DATOS DE LOS ESTUDIANTES

NOMBRES: JOSE IGNACIO						APELLIDOS: ARGUELLO TORRES	
TIPO IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO:	
CORREO INSTITUCIONAL: jarguellt@ucentral.edu.co						TELÉFONO: 3106497042	

NOMBRES: JAVIER IGNACIO						APELLIDOS: ACEVEDO QUIROS	
TIPO IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO:	
CORREO INSTITUCIONAL: jacevedoq@ucentral.edu.co						TELÉFONO: 3012702625	

NOMBRES: WILLIAN LEONARDO						APELLIDOS: LOPEZ GARCIA	
TIPO IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO: 1016039277	
CORREO INSTITUCIONAL: wlopezg@ucentral.edu.co						TELÉFONO: 3152993165	

### MODALIDAD DE TRABAJO DE GRADO

I. Modalidad de investigación:	II. Modalidad de profundización:
a. Proyecto final con resultados de nuevo conocimiento	a. Trabajo monográfico X
	b. Pasantía nacional o internacional

### Línea de profundización

**AVAL DEL DOCENTE DIRECTOR**

NOMBRES:	DEPARTAMENTO:
CORREO INSTITUCIONAL:	TELÉFONO-EXT. : +

**COMPONENTES**

**1. TÍTULO DEL TRABAJO DE GRADO O PASANTÍA**

**Modelo analítico de la planeación de la demanda en el sector manufacturero de PVC a partir de datos de distribución mayorista (sell-in) y consumo final (sell-out)**

**2. INTRODUCCIÓN Y JUSTIFICACIÓN**

Los pronósticos son fundamentales para la toma de decisiones en las organizaciones, ya que son la base para la planeación estratégica y la correcta gestión de recursos, presupuestos, costos y producción. A lo largo del tiempo, se han desarrollado varios métodos de pronóstico, como la suavización y el análisis de regresión, pero estos enfrentan limitaciones debido a las relaciones no lineales y la independencia entre variables. Es crucial actualizar las herramientas de pronóstico con los avances tecnológicos para obtener resultados más precisos, lo cual permite ajustar aspectos como precios y promociones, mejorando la rentabilidad. El aprendizaje de máquinas (Machine Learning), una rama de la inteligencia artificial, se presenta como una alternativa a los métodos tradicionales, al ofrecer pronósticos dinámicos y adaptables a cambios del mercado. Este enfoque tiene un alto potencial de generar valor significativo en áreas como marketing, ventas y cadena de suministro. En países desarrollados, como Estados Unidos y Alemania, se está adoptando rápidamente la inteligencia artificial para optimizar procesos, reflejando el crecimiento acelerado de esta tecnología en las próximas décadas (Gonzalez & Robayo, 2020, p. 9). En el ámbito empresarial, especialmente en el sector retail, los modelos de predicción de la demanda están transformando los procesos de compra de productos para su venta al público. Los avances tecnológicos recientes han permitido almacenar grandes volúmenes de datos, mejorando las herramientas de visualización y el uso de modelos dinámicos y flexibles.

En el entorno actual, las empresas enfrentan una alta volatilidad en la demanda, ciclos de productos más cortos, competencia global y una menor fidelidad de los clientes, lo que exige tomar decisiones informadas en tiempo real para satisfacer a los clientes y mantener la competitividad. Los errores en las previsiones de demanda pueden resultar en inventarios excesivos o insatisfacción del cliente. Las empresas utilizan técnicas avanzadas de planificación para adaptar sus planes de producción a los cambios rápidos de la demanda, basándose en datos no solo internos, sino también de proveedores, distribuidores y clientes. Los sistemas ERP han integrado módulos que reciben información externa, pero la mayoría de los métodos estadísticos utilizados en la previsión de demanda dependen de datos históricos y reglas de decisión complejas, lo que dificulta su precisión. A menudo, las decisiones se toman en base a la experiencia e intuición, lo que condiciona los planes de demanda a factores del mercado. El pronóstico de demanda es clave en la gestión de la cadena de

suministro, ya que influye en el cálculo de costos, materias primas y precios, ayudando a optimizar inventarios. Sin embargo, los métodos tradicionales de previsión de demanda, basados en series temporales, son complicados debido a la relación incierta entre las variables históricas y las ventas futuras. Aunque los métodos estadísticos como los modelos autoregresivos de medias móviles han sido utilizados, el pronóstico sigue siendo incierto debido a factores internos y externos (Mena O'meara et al., 2006).

En este contexto y de acuerdo con Gonzalez & Robayo (2020), el Machine Learning se destaca por su capacidad de trabajar con variables complejas, como datos de redes sociales o fechas especiales, que afectan la demanda, y por su habilidad para adaptarse constantemente a medida que se incorporan nuevos datos. Este enfoque es especialmente útil en mercados cambiantes, donde la demanda puede fluctuar rápidamente. El objetivo de este trabajo fue contribuir en los procesos de producción y ventas con el uso de Machine Learning, buscando mejorar la eficiencia y precisión de las predicciones de demanda, y demostrar cómo esta tecnología puede aplicarse en diferentes sectores y empresas, beneficiando los procesos de compra a nivel local y nacional (p. 9). Según Mena O'meara (2006), se han comenzado a aplicar redes neuronales artificiales (ANNs) para mejorar los pronósticos, pero aún se requiere más precisión. Se propone el uso de redes neuronales difusas (FNN) con algoritmos genéticos para optimizar los pronósticos, aprendiendo reglas basadas en datos históricos y experiencia de expertos (p. 2).

## JUSTIFICACIÓN

En el sector manufacturero de PVC, la eficiencia en la planeación de la demanda es un factor estratégico para evitar desabastecimientos, reducir sobrecostos logísticos y asegurar la disponibilidad oportuna de productos en el mercado. Esta industria, altamente influenciada por la estacionalidad, la diversidad de referencias y la rotación diferenciada entre productos, exige un modelo de planificación que anticipe el comportamiento del mercado con base en datos confiables. Actualmente, la empresa cuenta con información completa sobre su facturación a mayoristas (*sell-in*), pero tiene visibilidad limitada de las ventas reales al consumidor final (*sell-out*). Esta situación, común en el sector, implica que muchas decisiones de producción y distribución se tomen a partir de una visión parcial de la demanda. El *sell-in*, al estar influenciado por políticas comerciales o decisiones tácticas del canal mayorista, no refleja necesariamente la dinámica real del mercado. Por el contrario, el *sell-out* representa directamente el consumo efectivo.

Este problema se acentúa en el contexto económico actual, donde las ferreterías, clientes clave en este canal, adoptan comportamientos más cautelosos, adquiriendo únicamente productos que efectivamente están rotando. Esto obliga a los mayoristas a operar con inventarios más ajustados, lo cual repercute directamente en la planta manufacturera, que debe producir con una mayor precisión y menor margen de error. Por ello, integrar ambas fuentes de información se convierte en una necesidad crítica para construir modelos de pronóstico robustos. Este proyecto busca resolver dicha brecha mediante la construcción de un modelo de analítica predictiva que combine datos de *sell-in* y *sell-out*, aplicando modelos avanzados como SARIMA y LSTM. Esta integración permitirá transitar de un enfoque reactivo a uno proactivo, mejorando la capacidad de anticipación de la demanda, optimizando la planificación de inventarios y fortaleciendo la toma de decisiones basadas en evidencia. Además, al aplicar una metodología estructurada como CRISP-DM, se garantiza un proceso sistemático, replicable y ético, con el potencial de ser escalado a otras empresas del sector manufacturero que enfrenten desafíos similares.

### 3. OBJETIVO GENERAL

Desarrollar un modelo analítico integral de planeación de la demanda que combine y aproveche la información de sell-in (ventas de la fábrica al mayorista) y sell-out (ventas del mayorista al consumidor final), con el fin de mejorar la precisión en la predicción de la demanda, optimizar la reposición de producto y la gestión de inventarios, contribuyendo así a una gestión más eficiente de la cadena de suministro y a la toma de decisiones comerciales más informadas. Esta integración permitirá reducir las brechas entre lo que se produce y lo que efectivamente se consume, ajustando dinámicamente los niveles de producción y distribución en función del comportamiento real del mercado.

### 4. OBJETIVOS ESPECÍFICOS

- Realizar un análisis descriptivo de los datos históricos de sell-in y sell-out con el propósito de identificar relaciones, correlaciones y patrones de comportamiento entre las ventas al mayorista y las ventas al consumidor final.
- Diseñar un modelo de predicción de la demanda basado en técnicas de series temporales y algoritmos de *machine learning*, que integre las variables de *sell-in* y *sell-out* para reflejar el comportamiento real del mercado.
- Evaluar el desempeño del modelo predictivo en distintos escenarios de producto, con el fin de validar su capacidad de anticipación y precisión frente a la demanda real, y formular recomendaciones operativas y tácticas que permitan a la organización alinear mejor su producción, inventario y distribución con base en los hallazgos del modelo.

### 5. ANTECEDENTES (máximo 2000 palabras)

De acuerdo con Lao León (2017), se considera que, para obtener pronósticos más certeros, no es recomendable utilizar un único método, sino una combinación de métodos cualitativos y cuantitativos, ya que sus resultados suelen complementarse. En 2012, como parte del perfeccionamiento del Modelo Económico-Social Cubano, se aprobó la creación de BioCubaFarma, que incluye a la Empresa Comercializadora y Distribuidora de Medicamentos de Holguín (EMCOMED), encargada de almacenar, distribuir y comercializar medicamentos a nivel mayorista en la región. El proceso de planificación en EMCOMED se basa en las solicitudes de sus clientes, pero estas, a menudo, se basan en percepciones subjetivas y no siempre se correlacionan de manera precisa con la demanda real de los productos, lo que genera faltas de stock o exceso de inventario de productos de lento movimiento. Teniendo en cuenta la necesidad de pronósticos acertados se plantea la utilización de herramientas capaces de establecer relaciones lineales y no lineales entre la demanda y las variables que influyen en ella. En este sentido, las Redes Neuronales Artificiales (RNAs) emergen como una solución efectiva, pues se consideran una técnica de la Inteligencia Artificial que emula el funcionamiento de la inteligencia humana y es capaz de aprender y adaptarse a patrones complejos, lo que permite optimizar los pronósticos de demanda (p. 45).

El procedimiento inicial se desarrolló exitosamente, brindando una visión integral de los procesos empresariales y su relación con la planificación. A través de un análisis de Pareto, se seleccionaron los medicamentos a pronosticar: Captopril 25 mg, Dipirona 300 mg y Heberprot-P 25 y 75. Usando datos del software Mistral® (2007-2014), se establecieron variables dependientes e independientes para entrenar un

modelo de redes neuronales con el algoritmo back-propagation. El pronóstico se centró en el segundo semestre de 2015, donde se utilizó el software Weka 3.6 y se diseñaron nueve redes neuronales, obteniendo el modelo P\_3 con los menores errores de aproximación. La validación mostró que los resultados fueron satisfactorios, con un error menor al de un pronóstico de suavizado exponencial simple. El pronóstico final indicó que el Captopril mantendría estabilidad, la Dipirona disminuiría y el Heberprot-P seguiría una tendencia irregular, confirmando que las redes neuronales fueron la técnica de predicción más efectiva (Lao León et al., 2017, p. 50).

Los alimentos perecederos requieren un manejo especial para prolongar su vida útil, siendo esencial un adecuado almacenamiento, especialmente debido al aumento del consumo de productos frescos y congelados. La demanda de estos productos es constante, pero su producción no lo es, por lo que el almacenamiento es crucial para asegurar su disponibilidad y, en ocasiones, para obtener mejores precios en el mercado. Además, la temperatura juega un papel clave en la conservación y comercialización de estos productos, y se estima que el 30% de la producción alimentaria mundial pasa por tratamiento frigorífico. La cadena de frío tiene un gran valor económico global, alcanzando los 80.660 millones de dólares anuales, y las pérdidas por mal manejo y almacenamiento son mucho mayores en países en desarrollo (hasta un 50%) que en los desarrollados (alrededor del 10%). Los almacenes son fundamentales en la cadena de suministro y deben garantizar la calidad del producto, adaptándose a los requerimientos térmicos y permitiendo una distribución eficiente. (Juárez et al., 2016, p. 388)

Para mejorar el desempeño de la cadena de suministro y la precisión de los pronósticos, es esencial el análisis de datos históricos. Los pronósticos ayudan a anticipar las necesidades de almacenamiento, lo que permite tomar decisiones más informadas sobre el tamaño y diseño de los almacenes. En el contexto de países en desarrollo como México, este artículo se enfoca en la aplicación de técnicas de pronóstico de series de tiempo para estimar el volumen de productos perecederos en cámaras frigoríficas, optimizando el uso de recursos y anticipando la necesidad de instalaciones adicionales. La empresa maneja 9 cámaras y no utiliza métodos cuantitativos para estimar los volúmenes de ingreso y egreso de los productos, basándose en la experiencia de los encargados. El análisis de datos históricos, especialmente mediante pronósticos de series de tiempo, se presenta como una solución clave para optimizar la gestión del inventario, mejorando la planificación y la toma de decisiones. Se realiza un análisis específico de dos productos, AP-1 y NE-2, que representan una porción significativa del inventario, y se destaca la necesidad de aplicar técnicas de pronóstico para prever los movimientos de estos productos en los próximos periodos y ajustar los recursos necesarios para su manejo. (Juárez et al., 2016, p. 391)

Según Juarez (2016), para el producto AP-1, se observa que la técnica de promedios móviles ponderados ofrece el pronóstico más ajustado para los ingresos, mientras que la técnica de suavización exponencial ajustada tiene el error más bajo para los egresos. En cuanto al producto NE-2, la técnica de promedios móviles ponderados también proporciona los resultados más precisos tanto para los ingresos como para los egresos, destacando una disminución del 25% en los ingresos durante el mismo período. Los análisis sugieren que ambos productos tendrán egresos superiores a los ingresos, lo que podría llevar a un desabastecimiento, especialmente del producto NE-2. Esta tendencia de egresos superiores a ingresos es consistente a lo largo del período de estudio, lo que indica la necesidad de ajustar la capacidad de almacenamiento y planificar estrategias de reposición. Se destaca la importancia de los pronósticos para optimizar la toma de decisiones y la planificación de recursos en el almacenamiento de productos perecederos, lo que permitirá una gestión más eficiente de inventarios y una mejor asignación de personal y recursos. Las recomendaciones incluyen la utilización de los pronósticos para diseñar nuevas estrategias y la optimización de la capacidad de las cámaras frigoríficas, así como la reestructuración del lay-out de las cámaras para mejorar la eficiencia en la recepción, acomodo y salida de los productos. (p. 393)

De acuerdo a Loaiza (2023), en el contexto empresarial colombiano, en la empresa Dyna & Cía S.A., se

implementaron modelos de machine learning como árboles de decisión y redes neuronales para predecir la demanda de productos. El objetivo principal fue evaluar cómo la integración de datos internos (ventas históricas) y externos (condiciones del mercado) impacta en la precisión de los pronósticos. La metodología empleada permitió validar la utilidad de estas técnicas en escenarios reales, resaltando su valor frente a modelos tradicionales (p. 12).

Dado lo anterior, los modelos de regresión no fueron efectivos para predecir las series de tiempo en este caso, ya que no existía una relación lineal clara entre la variable pronosticada (demanda) y la "Fecha". Aunque se intentó convertir el campo "Fecha" en variables como "Mes" y "Año" para evaluar su desempeño, este enfoque no resolvió el problema de predicción. De manera similar, el modelo Random Forest también enfrentó dificultades, ya que, aunque se asumió que las variables "Año" y "Mes" podrían explicar la demanda, no está claro que estas variables sean causantes de la misma, a diferencia de otras variables como la temperatura o el viento en pronósticos de lluvias. Por otro lado, el modelo ARIMA mostró buenos resultados en los primeros pronósticos, con menor error en los dos primeros periodos, pero presentó limitaciones al no identificar patrones más complejos como los cíclicos o estacionales, lo que dificultó las predicciones a largo plazo, dado que trataba todas las observaciones con la misma importancia. Finalmente, las redes neuronales resultaron ser útiles para el pronóstico de series de tiempo, ya que tienen la capacidad de aprender patrones complejos y consideran la relación entre el tiempo presente ( $t$ ) y el futuro ( $t+1$ ), así como su conexión con el pasado ( $t-1$ ). A diferencia de ARIMA y Random Forest, las redes neuronales no dependían tanto de las fechas, sino del histórico de la demanda (Loaiza, 2023, p. 40).

En términos generales los modelos de aprendizaje automático, como ARIMA y Redes Neuronales, demostraron un mejor ajuste, especialmente al trabajar con datos históricos de demanda, en comparación con los modelos de Regresión y Random Forest, que no son tan efectivos en problemas univariados. En general, los modelos de Machine Learning lograron reducir el error en un 23% en promedio respecto al método tradicional. Para el negocio, la métrica MAE resultó ser adecuada para comprender y comunicar los resultados, ya que las unidades de medida son relevantes en términos de inventario y ventas, facilitando la comunicación con personas no especializadas en Machine Learning. Este antecedente resulta relevante para el presente trabajo, ya que demuestra la viabilidad de aplicar analítica predictiva en empresas locales y refuerza la necesidad de combinar múltiples fuentes de información, como el sell-in y el sell-out, para construir modelos más robustos y alineados con la dinámica del consumidor final (Loaiza, 2023, p. 41).

Según Mazo (2011), las empresas en Colombia reconocen debilidades en su gestión logística, especialmente en áreas como el abastecimiento, la producción, la logística de salida y la formulación de estrategias para alcanzar mercados regionales, nacionales e internacionales. Esta conclusión se deriva de los resultados del Autodiagnóstico Logístico Exportador realizado por Proexport en colaboración con GS1-Logyca. El proyecto tenía como objetivo mejorar la cadena de abastecimiento dentro de la Política de Repotenciación de Exportaciones, buscando aumentar el aprovechamiento de nuevos acuerdos internacionales y fomentar las ventas de valor agregado al exterior. Además, de presentar un análisis cualitativo sobre la importancia de la Planeación de la Demanda, así como los criterios que deben considerarse al decidir qué modelo estadístico utilizar, según la disponibilidad de datos, conocimientos y presupuesto. (p. 3).

Tras el análisis del autodiagnóstico de Proexport y GS1-Logyca, se evidenció que el uso adecuado de información confiable en los pronósticos mejoraría la productividad y competitividad empresarial. Las técnicas informales, como regresiones y promedios ponderados, expusieron a las empresas a riesgos de destrucción de valor, altos inventarios o desabastecimiento, lo que generaba ventas perdidas y falta de eficiencia en marketing. No existió un "Predictor Óptimo", ya que la eficacia dependió del modelo y paquete estadístico utilizado. Se concluyó que la potencia de los modelos dependía del tipo de serie a analizar y el comportamiento de los datos, influenciados por estacionalidades, ciclos y tendencias. Además, los análisis de demanda debían complementarse con información de tendencias de consumo, análisis de mercado y monitoreos de



competidores (Mazo et al., 2011, p. 12).

De acuerdo con Campos & Zúñiga (2021), la empresa XYZ mantiene un acuerdo de nivel de servicio con sus clientes para abastecer semanalmente más de 300 productos de consumo masivo. El área Comercial toma las órdenes de compra entre lunes y sábado, pero la producción solo se realiza los martes, por lo que la demanda total se estima utilizando el promedio de las últimas ocho semanas. Esta incertidumbre genera dos problemas: si la cantidad ordenada es menor a la demanda, se produce insatisfacción del cliente; si es mayor, se incurre en costos de inventario por el exceso de producto almacenado. Con el objetivo de mejorar la predicción de la demanda y reducir estos costos, XYZ solicitó consultoría en analítica de datos. El desafío principal radicó en reducir el error en la estimación de la cantidad semanal demandada, enfocándose en los productos con mayor rotación de inventario. La empresa cuenta con un histórico de ventas de 156 semanas por producto, lo que servirá como base para desarrollar una metodología que optimice la gestión de inventarios y mejore la competitividad (p. 4).

En la empresa XYZ, se identificaron dos enfoques para abordar el problema: seleccionar los 12 productos de mayor rotación o escoger los productos más representativos de las tres categorías principales. Ambos enfoques llevaron a soluciones distintas en términos de algoritmos de pronóstico y reducción de costos. Los costos de inventario representaron el 4.7% de las ventas y el 15.8% de la utilidad neta para el modelo de 12 productos, mientras que para el modelo de tres categorías, los costos equivalieron al 2.4% de las ventas y el 7.9% de la utilidad neta. Los modelos ARIMA, Random Forest y SVM fueron los más efectivos en la reducción de costos, con un ahorro significativo de 1,305,320 MXN (67,516 USD) anuales en el modelo de 12 productos y 646,622 MXN (32,331 USD) anuales en el modelo de tres categorías. Se concluyó que pronosticar productos de manera individual genera mayores beneficios económicos que hacerlo por categorías, ya que las categorías presentan comportamientos heterogéneos. Además, se destacó la necesidad de implementar diferentes algoritmos de analítica y machine learning para mejorar la precisión de los pronósticos, y se recomendó un proceso continuo de actualización y calibración de los modelos (Campos & Zúñiga, 2021, p. 21).

Según Moreno Arenas (2016), para realizar una proyección efectiva, es crucial contar con una base de datos confiable. Sin embargo, en la empresa Servientrega S.A.S., falta dicha fuente confiable para la proyección de la demanda de materiales de empaque y embalaje, dado que la base de datos obtenida estaba alterada por manipulaciones previas y no reflejaban adecuadamente la demanda real en los Centros de Solución Puntos de Servicios (CDS). El objetivo del trabajo en mención es desarrollar un plan de pronóstico de demanda para estos materiales, utilizando un modelo de series de tiempo, con el fin de mejorar las compras, reducir las discrepancias en inventarios y cumplir con los pedidos de manera más eficiente. También, se buscaba mejorar la planificación de compras, ajustando las cantidades y fechas de entrega, y optimizar las relaciones con los proveedores para reducir tiempos de entrega y mejorar la cadena de abastecimiento (p. 2).

Para obtener un pronóstico de demanda más preciso, se utilizó el software Risk Simulator, que permitió realizar simulaciones y proyecciones de series de tiempo mediante el uso de Excel. Este software emplea técnicas como la simulación de Monte Carlo y optimización, y fue aplicado a los datos organizados según el periodo de tiempo establecido. En el análisis de los productos, se identificaron los siguientes métodos más adecuados para proyectar la demanda: método de Suavización Exponencial Doble fue el más preciso y el método Winter. La implementación de estos métodos permitirá una mejora significativa en la precisión de los pronósticos de demanda, reduciendo las variaciones y disminuyendo hasta un 58% los PQR (Pedidos quejas y reclamos), lo que impactaría positivamente la imagen de la empresa. Además, al establecer una planificación de compras estandarizada, se mejoraría la rotación de inventario, facilitando un control más acorde a la demanda real de cada producto (Moreno Arenas, 2016, p. 9).

## 6. FUENTE DE LOS DATOS (Cómo se obtendrán los datos)

Los datos utilizados en este proyecto provienen de los registros de facturación de la empresa (100% del sell-in) y de bases de datos de ventas finales reportadas por distribuidores seleccionados (20% del sell-out). Se implementarán técnicas de estadística y machine learning para estimar el comportamiento del mercado con base en la información disponible.

## 7. APLICACIÓN Y/O APOORTE ESPECÍFICO AL CAMPO

Este trabajo se enmarca principalmente en el área de la analítica de datos, más específicamente en el análisis predictivo aplicado a la planificación de la demanda. Aporta a la comprensión y modelado de comportamientos comerciales mediante el uso de técnicas de series temporales (SARIMA) y redes neuronales (LSTM), dos enfoques ampliamente utilizados en el campo de la ciencia de datos.

Adicionalmente, el proyecto establece conexiones directas con áreas como la logística, la ingeniería industrial y la gestión de la cadena de suministro, al proponer soluciones prácticas orientadas a mejorar la eficiencia operativa, reducir costos por sobrestock o desabastecimiento, y mejorar la alineación entre producción y consumo real.

Desde una perspectiva metodológica, también representa un aporte a la educación en ciencia de datos aplicada, al demostrar cómo una metodología como CRISP-DM puede ser implementada de forma estructurada en contextos empresariales reales, respetando criterios éticos y de confidencialidad. Finalmente, el proyecto puede ser útil como referencia para futuras investigaciones en sectores manufactureros que busquen integrar fuentes de datos dispares (*sell-in* y *sell-out*) para mejorar su inteligencia comercial y su toma de decisiones basadas en datos.

## 8. METODOLOGÍA O ACTIVIDADES ESPECÍFICAS

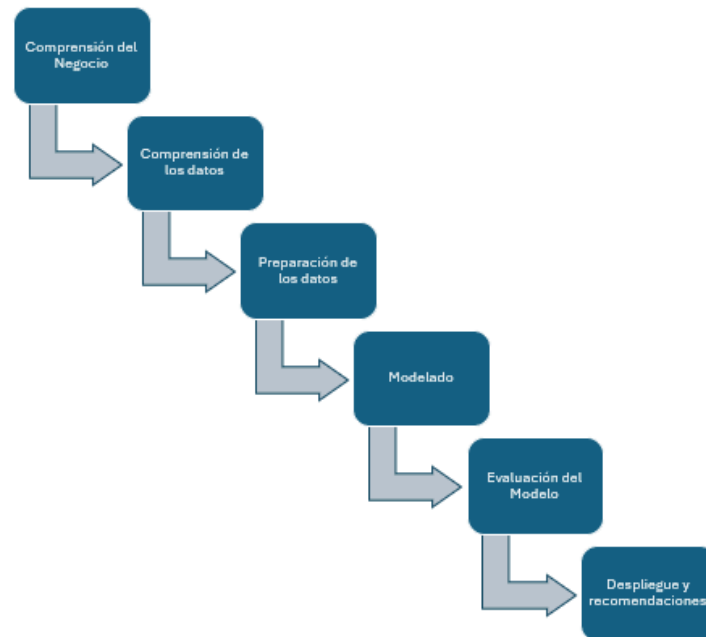
La metodología utilizada será CRISP-DM (*Cross Industry Standard Process for Data Mining*), ampliamente adoptada en proyectos de analítica de datos por su estructura flexible y centrada en el ciclo de vida del análisis. Este enfoque será clave para desarrollar un modelo de planeación de la demanda robusto, integrando datos de *sell-in* y *sell-out*.

CRISP-DM nos permite abordar el problema desde una mirada integral que parte de la comprensión del negocio hasta la entrega de recomendaciones accionables. Es una metodología iterativa y adaptable, lo que la hace especialmente útil en contextos donde los datos provienen de fuentes diversas como la venta a mayoristas y la venta al consumidor final, y es necesario armonizarlos para una visión completa. Además, facilita el diseño de procesos analíticos replicables, escalables y orientados a la toma de decisiones estratégicas.

En este trabajo, se aplicará el enfoque visual representado en el diagrama, el cual refleja las seis fases de CRISP-DM: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación del modelo y despliegue de resultados. Cada una de estas etapas se adaptará al contexto del sector manufacturero de PVC y a los objetivos definidos para este proyecto.



A continuación se presenta el diagrama de flujo que resume las fases del modelo CRISP-DM aplicado al presente trabajo:



### Fase 1: Comprensión del negocio

Se identificará la problemática del desalineamiento entre la producción y el consumo real en la industria del PVC. Se definirán los objetivos analíticos, los indicadores clave y el valor esperado del modelo para las áreas de planeación, comercial y logística. Se elaborará un plan de acción estructurado para el desarrollo del modelo.

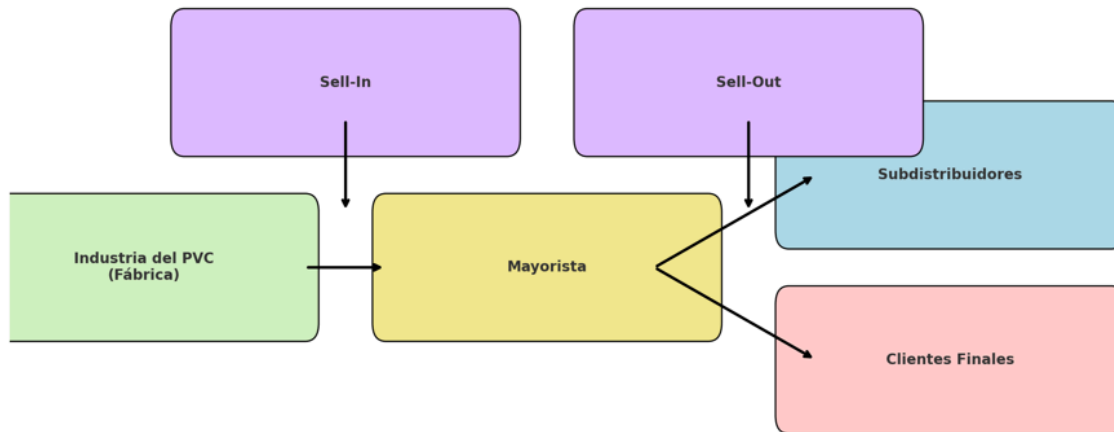
Actualmente, la planificación de la demanda en la empresa se basa exclusivamente en datos de *sell-in*, utilizando modelos estadísticos simples como los promedios móviles. Esta práctica, si bien ha sido útil en el pasado, presenta limitaciones críticas, ya que no considera el comportamiento real del consumidor final, sino únicamente el comportamiento de compra de los mayoristas, que puede estar influenciado por promociones, estacionalidades internas o acumulación de inventario por parte del canal. Como resultado, se generan desviaciones en la planeación que derivan en problemas de sobreproducción, falta de producto en el punto de venta y una gestión ineficiente del inventario.

El reto actual radica en la desconexión entre lo que se factura (*sell-in*) y lo que realmente se vende al cliente final (*sell-out*). La integración de ambas fuentes permitirá reducir esa brecha y comprender con mayor profundidad el comportamiento del mercado, al identificar patrones reales de consumo, reacciones del cliente ante ciertos productos, y comportamientos regionales. Esta visión más integral permitirá a las áreas de logística, comercial y planificación anticiparse mejor a la demanda, optimizando niveles de producción, puntos

de reabastecimiento y la rotación del portafolio, especialmente en productos de baja demanda.

Durante esta fase se priorizará la contextualización del problema en el entorno del sector PVC, se documentarán los procesos actuales, y se establecerá claramente la visión futura de un modelo de planeación más asertivo, fundamentado en la analítica avanzada y la realidad del mercado.

**Modelo de Relación Sell-In y Sell-Out en la Cadena de Abastecimiento**



## Fase 2: Comprensión de los datos

Durante esta fase se abordará la exploración detallada de dos fuentes de información fundamentales: por un lado, la base de datos de *sell-in*, que contiene el 100% de la facturación realizada por la fábrica a todos los mayoristas, y por otro, una muestra representativa cercana al 20% del *sell-out*, correspondiente a la información que algunos mayoristas reportan sobre sus ventas a subdistribuidores y clientes finales.

Aunque la cobertura del *sell-out* no es total, esta muestra representa un insumo valioso para comenzar a identificar patrones de consumo real, frecuencia de reposición y diferencias de comportamiento según zonas, categorías de producto o tipo de cliente. Se aprovechará esta información para caracterizar la demanda desde una perspectiva más cercana al consumidor, complementando el enfoque tradicional centrado en el canal mayorista.

En esta etapa se aplicarán técnicas de análisis exploratorio de datos (EDA), generación de gráficos descriptivos, cálculos de correlación entre variables clave y detección de estacionalidades. También se documentarán vacíos de información y se evaluará la consistencia temporal entre ambas fuentes. La comprensión detallada de los datos permitirá sentar las bases para la posterior preparación y modelado, alineando los objetivos del modelo con la realidad operativa y la calidad de los datos disponibles.

## Fase 3: Preparación de los datos

Durante esta fase se llevará a cabo la limpieza, transformación y estandarización de las bases de datos, con un tratamiento diferenciado para el *sell-in* y el *sell-out*. En el caso del *sell-in*, se cuenta con una base estructurada y de alta calidad, ya que proviene directamente del sistema de facturación de la fábrica, lo cual garantiza integridad, completitud y estandarización en variables como fecha de facturación, cliente, artículo,

volumen y valor.

Sin embargo, el *sell-out* representa un desafío mayor en esta etapa. Dado que cada mayorista maneja su propia lógica comercial y sus propios sistemas de información, los formatos, códigos y descripciones de productos varían considerablemente entre uno y otro. Para poder utilizar esta información de forma efectiva en el modelo, será necesario homologar los datos del *sell-out* al lenguaje operativo de la fábrica. Esto incluye procesos como la traducción de códigos de artículo, normalización de unidades de medida, y adaptación de nomenclaturas de zonas o canales, entre otros.

Este proceso de integración tiene como objetivo lograr una visión unificada del comportamiento de la demanda a lo largo de toda la cadena de abastecimiento, lo que permitirá a la fábrica tomar decisiones más asertivas sobre producción, distribución e inventarios, basadas en datos alineados a su propia lógica interna y a las dinámicas reales del mercado

#### **Fase 4: Modelado**

En esta fase se propone desarrollar el modelo de predicción de la demanda, integrando los datos de *sell-in* y *sell-out* preparados previamente. A partir de los objetivos del proyecto y de las características del problema, se plantea inicialmente el uso de dos enfoques complementarios: por un lado, el modelo SARIMA, por su solidez en la modelación de series temporales con estacionalidad, tendencias y autocorrelación; y por otro lado, el modelo LSTM (Long Short-Term Memory), por su capacidad para capturar patrones complejos y no lineales en secuencias largas de datos, especialmente útiles en contextos con dinámicas de mercado variables.

No obstante, al tratarse de un anteproyecto, se es consciente de que la elección final del modelo debe responder a un proceso comparativo riguroso. Por ello, se contempla la posibilidad de incluir y evaluar otros algoritmos como Random Forest Regressor o modelos híbridos, en función de su capacidad predictiva, facilidad de interpretación, robustez frente a outliers y adaptabilidad a los diferentes tipos de productos (alta y baja rotación).

Esta fase será crítica para garantizar que el modelo elegido no solo ofrezca buenos resultados teóricos, sino que sea viable para implementar operativamente dentro de la empresa. Por tanto, los modelos seleccionados inicialmente (SARIMA y LSTM) se tomarán como punto de partida para la experimentación, y estarán sujetos a validaciones cruzadas, análisis de errores y comparación bajo criterios objetivos. Esta estrategia no solo busca encontrar el modelo más preciso, sino también el más útil y escalable desde el punto de vista logístico y comercial.

Durante esta fase también se documentan las decisiones técnicas tomadas, los supuestos de modelado, y los hallazgos intermedios que contribuyan a entender mejor la dinámica entre lo que se factura (*sell-in*) y lo que realmente se vende (*sell-out*), y cómo dicha dinámica puede ser anticipada mediante modelos de analítica predictiva.

#### **Fase 5: Evaluación del modelo**

En esta fase se evaluará el rendimiento de los modelos seleccionados a partir de métricas cuantitativas como el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Porcentaje de Error Absoluto Medio (MAPE), que permiten medir con precisión la diferencia entre los valores predichos y los valores reales observados en el comportamiento del *sell-out*.

Además de evaluar el desempeño global del modelo, se realizarán pruebas específicas por tipo de producto (alta y baja rotación), zona geográfica y canal de distribución, con el fin de verificar la capacidad de

generalización del modelo y su utilidad en diferentes escenarios operativos. También se aplicará validación cruzada temporal (*time series cross-validation*) para asegurar la robustez del modelo ante variaciones estacionales y ciclos de mercado.

Se pondrá especial atención en analizar la capacidad del modelo para anticipar eventos críticos como picos de demanda, caídas abruptas o quiebres de inventario.

Finalmente, con base en los hallazgos, se definirán estrategias de mejora continua que garanticen la actualización periódica del modelo, asegurando su asertividad y vigencia en el tiempo conforme se recolecten nuevos datos de *sell-out* y cambien las dinámicas del mercado.

#### **Fase 6: Despliegue y recomendaciones**

Dado el carácter confidencial de la empresa y la naturaleza académica del proyecto, el despliegue del modelo no se realizará directamente en los sistemas operativos internos de la organización. En su lugar, se desarrollará un entorno de simulación controlado que emule los principales procesos de planeación, abastecimiento y distribución de la empresa, garantizando la protección de datos sensibles.

El modelo seleccionado será documentado en detalle y se implementará en un entorno técnico replicable utilizando Python y Power BI, lo que permitirá la visualización de sus resultados y su validación a través de escenarios históricos. Se generarán dashboards interactivos con visualizaciones como curvas de demanda proyectada, niveles óptimos de inventario, alertas de quiebre de stock y análisis por zonas o categorías, todo basado en datos anonimizados o simulados.

Además, se entregará un conjunto de lineamientos estratégicos y recomendaciones tácticas orientadas a mejorar los procesos de planificación y toma de decisiones, partiendo de las capacidades del modelo. Estos entregables permitirán a la organización evaluar la aplicabilidad del modelo de manera segura, y dejar abierta la posibilidad de futuras implementaciones reales.

Finalmente, se propondrá una hoja de ruta para la incorporación progresiva de nuevos datos *sell-out* y para la actualización periódica del modelo, con el fin de mantener su vigencia y asertividad a lo largo del tiempo.

Este enfoque metodológico permitirá no solo cumplir con los objetivos planteados, sino también establecer una hoja de ruta replicable para otros productos o industrias con problemáticas similares.

## **9. RECURSOS**

### **Recurso de personal**

El equipo humano que participa en la elaboración del proyecto está compuesto por :

- ❖ Javier Ignacio Acevedo Quiros- Ingeniero de Sistemas
- ❖ Willian Leonardo Lopez Garcia - Ingeniero Químico
- ❖ Jose Ignacio Arguello Torres - Ingeniero de Procesos Industriales

### **Recursos de hardware y software**

Hardware : Cada estudiante hará uso de su propio equipo de cómputo para participar en el proyecto, entre los

equipos un ASUS con procesador intel CORE i5 con 16 GB de RAM GPU 5200U.

Software : Se utiliza visual studio code para programar en Python, Docker como plataforma para crear, ejecutar y administrar la aplicación dentro de contenedores . El control de versiones de los avances se realizará a través del repositorio de GitHub, al cual tienen acceso todos los miembros del equipo de trabajo. En caso de ser requerido, se implementará MongoDB como motor de base de datos para la gestión de la información.

## 10. CRONOGRAMA DE ACTIVIDADES

ACTIVIDADES A REALIZAR	Semanas de ejecución de cada actividad																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
Actividad 1 (8 febrero - 21 de marzo ) - Exposición de propuestas de investigación por los tutores candidatos. - Formación del equipo de trabajo - Identificación y delimitación de la problemática a tratar en el anteproyecto - Seguimiento semanal con el tutor asignado.																		
Actividad 2 (24 marzo - 18 abril) - Encuentro con el equipo de trabajo para recibir orientación y dar seguimiento al proyecto - Diseño del plan de trabajo con sus respectivas actividades programadas. - Selección y formulación del título del trabajo de grado. - Elaboración de la introducción, justificación, objetivo general y objetivos específicos.																		
Actividad 3 (20 abril - 25 abril) - Exploración de estudios previos y literatura académica relacionada con la problemática de investigación. - Determinación del origen y tipo de datos necesarios para el desarrollo del estudio.																		
Actividad 4 (26 abril- 30 abril) - Evaluación de la aplicación y/o efecto específico en el ámbito relacionado con el proyecto de investigación.																		
Actividad 5 (03 mayo - 07 mayo) -Diseño de la metodología o actividades concretas. -Sesiones de seguimiento semanal con el tutor asignado.																		
Actividad 6 (09 mayo -15 mayo ) -Comprobación de los recursos tecnológicos, humanos y otros necesarios. -Desarrollo de posibles fuentes de financiamiento. -Redacción de los resultados que se esperan al implementar la propuesta. -Compilación de la bibliografía relevante.																		

## 11. FUENTES DE FINANCIACIÓN

Tipo	Categoria	Recurso	Descripcion	Fuente de Financiacion	Monto
Personal	Consultoria	Profesional	Experto Consultor Externo	Internos	0.00
Medios Disponibles	Instalaciones	Equipo	Portatil	Internos	0.00
		Equipo	Equipo de cómputo fijo	Internos	0.00
Medios Necesarios	Insumos	Impresora	Epson LS	Internos	0.00
		Otros	(Lapiceros , Folders )	Internos	30,000.00
Servicios	Servicio Publico	Datos , Internet	Internet , Comunicación	Internos	400,000.00
	Servicio Publico	Servicio Energia	Servicio Energia	Internos	450,000.00
	Transporte	Rodamiento , movilizacion	Rodamiento , Movilizacion	Internos	300,000.00
Total Fuentes de Financiacion					\$ 1,180,000.00

## 12. RESULTADOS ESPERADOS

- Obtención de un modelo analítico predictivo ajustado a la realidad operativa de la empresa manufacturera, capaz de anticipar la demanda con base en datos integrados de *sell-in* y *sell-out*.
- Aumento de la precisión en la planeación de la demanda, con una reducción significativa de los errores de predicción, especialmente en productos de baja rotación.
- Generación de visualizaciones dinámicas en dashboards con Power BI, que faciliten el seguimiento de patrones de consumo, alertas de inventario y curvas proyectadas de demanda.
- Mejora de la eficiencia logística mediante la reducción de quiebres de stock, sobreproducción o acumulación innecesaria de inventarios.
- Consolidación de una herramienta escalable que pueda ser replicada o adaptada por otras empresas del sector manufacturero interesadas en implementar modelos de planeación de la demanda basados en machine learning.
- Fomento de una cultura de toma de decisiones basada en datos, promoviendo el uso de herramientas de analítica avanzada en áreas como comercial, logística y producción.
- Validación académica y práctica de la metodología CRISP-DM como marco estructurado para abordar problemas reales de planificación en cadenas de suministro.

## 13. BIBLIOGRAFÍA

Juárez, A. C., Zuniga, C. A., Flores, J. L. M., & Partida, D. S. (2016). Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. *ESTUDIOS GERENCIALES*, 10. <https://doi.org/10.1016>

Loaiza, A. C. (2023). *Análisis de modelos basados en Machine Learning para la predicción de la demanda de productos en la empresa Dyna & Cía. S.A* (A. C. Loaiza, Ed.). Universidad de Antioquia.



Gonzalez, S. D., & Robayo, V. V. (2020). *PRONÓSTICO DE DEMANDA UTILIZANDO INTELIGENCIA ARTIFICIAL* (S. D. Gonzalez & V. V. Robayo, Eds.). UNIVERSIDAD ICESI.

Mazo, A. Z., Parra, P. A. M., & Giraldo, É. Y. G. (2011). LA PLANEACIÓN DE LA DEMANDA COMO REQUISITO PARA LA GESTIÓN DE LAS CADENAS DE SUMINISTRO EN LAS EMPRESAS EN COLOMBIA. *REVISTA POLITECNICA*, 11–22.

Lao León, Y. O., Rivas Méndez, A., Pérez Pravia, M. C., & Marrero Delgado, F. (2017). Procedimiento para el pronóstico de la demanda mediante redes neuronales artificiales. *Ciencias Holguín*, 23, 43–59. <https://www.redalyc.org/journal/1815/181549596004/html/>

Campos, D. L. J., & Zúñiga, E. L. (2021). MODELO DE FORECAST PARA PREDECIR LA DEMANDA SEMANAL DE ALIMENTOS Y BEBIDAS DE CONSUMO MASIVO. *UNIVERSIDAD DE LOS ANDES*, 26. <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/f93f41fd-7137-4169-8906-946815dea7b1/content>

DmitryIvanova, A. T., & Schönbergerb, J. (2019). Demand Forecasting: A Decision-Oriented Introduction to the Creation of Value. En *Global Supply Chain and Operations Management* (Ed.), *Global Supply Chain and Operations Management* (pp. 319–333). [https://doi.org/10.1007/978-3-319-94313-8\\_11](https://doi.org/10.1007/978-3-319-94313-8_11)

Haselbeck, F., Killinger, J., Menrad, K., Hannus, T., & Grimm, D. G. (2022). Machine Learning Outperforms Classical Forecasting on Horticultural Sales Predictions. *Machine Learning with Applications*, 19. <https://doi.org/10.1016/j.mlwa.2021.100239>

Martignon, M. (2016). Estimación y Pronóstico de Demanda: Una Diferenciación Práctica. *UNIVERSIDAD DE LOS ANDES*, 35. <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/e9e476b5-1cd2-47e1-bb6f-50b323184e8b/content>

tent

Lang, J. P. B. (2017). *Uso de modelos de forecast de demanda para el mercado de consumo masivo aplicado a empresa cervecera* [UNIVERSIDAD TORCUATO DI TELLA].  
<https://repositorio.utdt.edu/server/api/core/bitstreams/00c8422b-fae1-4370-8301-1c58c91507ab/content>

Moreno Arenas, R. (2016). Selección de un método de pronóstico de la demanda para la estandarización de compras de Empaque y Embalaje. *Universidad Militar Nueva Granada*, 14.  
<https://repository.umng.edu.co/items/62957789-627e-4ac0-9aba-158847912f08>

José, A. U., Enrique, G. S. L., Areli, B. D., & Elizabeth, N. G. G. (2018). Toma de Decisiones Basado en el Pronóstico de la Demanda: Caso de Estudio TAECOMA. *Revista de Ingeniería e Investigación Aplicada UPB/UPTap*, III, 15.  
[https://www.researchgate.net/publication/350671541\\_Toma\\_de\\_Decisiones\\_Basado\\_en\\_el\\_Pronostico\\_de\\_la\\_Demanda\\_Caso\\_de\\_Estudio\\_TAECOMA\\_Decision\\_Making\\_Based\\_on\\_the\\_Demand\\_Forecasting\\_TAECOMA\\_Case\\_Study](https://www.researchgate.net/publication/350671541_Toma_de_Decisiones_Basado_en_el_Pronostico_de_la_Demanda_Caso_de_Estudio_TAECOMA_Decision_Making_Based_on_the_Demand_Forecasting_TAECOMA_Case_Study)

Mena O'meara, N. A., Esteban, F. C. L., & Salort, E. V. (2006). Planificación de la Demanda en la Gestión de Cadena de Suministro con Redes Neuronales y Lógica Difusa. *Universitat Politècnica de València*, 9.  
[https://www.researchgate.net/publication/277733174\\_Planificacion\\_de\\_la\\_Demanda\\_en\\_la\\_Gestion\\_de\\_Cadena\\_de\\_Suministro\\_con\\_Red\\_Neuronales\\_y\\_Logica\\_Difusa](https://www.researchgate.net/publication/277733174_Planificacion_de_la_Demanda_en_la_Gestion_de_Cadena_de_Suministro_con_Red_Neuronales_y_Logica_Difusa)

#### 14. DATOS DE LA ORGANIZACIÓN

NOMBRE DE LA ORGANIZACIÓN:

RESPONSABLE DE LA ORGANIZACIÓN:

CORREO:

TELÉFONO:

DURACIÓN VINCULACIÓN (horas):

#### 15. FIRMAS

FIRMA DEL ESTUDIANTE:	FIRMA DEL DOCENTE DIRECTOR O TUTOR

**16. DATOS DE TRÁMITE COMITÉ DE INVESTIGACIÓN DEL PROGRAMA** (Espacio para diligenciar por el Comité del Programa)

No. CONSECUTIVO	
No. ACTA	
FECHA	