Comparative Analysis of Machine Learning Algorithms versus Traditional Methods for Sustained Demand Forecasting under Varying Levels of Variability

Franco M. Novara[1,2]

[1] UTN Regional San Francisco, Av. de la Universidad 501, Argentina.
[2] Facultad de Ingeniería Química-UNL, Santiago del Estero 2829, Argentina.
fnovara@fiq.unl.edu.ar

Abstract. Demand forecasting is critical for manufacturing organizations to plan production and manage inventory efficiently. While traditional time-series methods, such as moving average and linear regression, offer simplicity and robustness, the rise of Machine Learning (ML) algorithms promises enhanced accuracy through complex data modeling. However, the superiority of ML is questionable in industrial settings where exogenous variables are lacking and there is a sole reliance on historical endogenous data. Furthermore, complex ML models can suffer from overfitting, reacting disproportionately to recent noise and inducing a "nervousness" in the system that severely impacts the stability of inventory coverage (Days of Supply). This paper presents a comparative analysis between ML algorithms and traditional forecasting methods for sustained demand. The study evaluates forecasting accuracy and projection robustness across different levels of demand variability, analyzing their consequent impact on inventory decision-making.

Keywords: Demand Forecasting, Machine Learning, Time Series, System Nervousness, Inventory Management.

Análisis comparativo de algoritmos de *Machine Learning* frente a métodos tradicionales para el pronóstico de demanda sostenida en el tiempo con distintos niveles de variabilidad.

Resumen. El pronóstico de demanda es crítico para que las organizaciones de manufactura planifiquen la producción y gestionen el inventario eficientemente. Mientras que los métodos tradicionales de series de tiempo, como el promedio móvil y la regresión lineal, ofrecen simplicidad y robustez, el auge de los algoritmos de *Machine Learning* (ML) promete una mayor precisión a través del modelado de datos complejos. Sin embargo, la superioridad de ML es cuestionable en entornos industriales en los que se carece variables exógenas y se depende

únicamente de datos endógenos históricos. Además, los modelos complejos de ML pueden sufrir de sobreajuste, reaccionando desproporcionadamente al ruido reciente e induciendo un "nerviosismo" en el sistema que impacta severamente la estabilidad de la cobertura del inventario (*Days of Supply*). Este artículo presenta un análisis comparativo entre algoritmos de ML y métodos de pronóstico tradicionales para demanda sostenida. El estudio evalúa la precisión del pronóstico y la robustez de las proyecciones a través de diferentes niveles de variabilidad en la demanda, analizando su consecuente impacto en la toma de decisiones de inventario.

Palabras clave: Pronóstico de demanda*, Machine Learning*, Series de tiempo, Nerviosismo del sistema, Gestión de inventario.

1 Introducción

El pronóstico de demanda constituye un pilar fundamental para la planificación y competitividad de diversas organizaciones, cobrando especial relevancia en aquellas dedicadas a la manufactura. Dependiendo del horizonte temporal considerado, las proyecciones cumplen roles distintos pero complementarios: a largo plazo, asisten en decisiones estratégicas como el dimensionamiento de instalaciones o el lanzamiento/discontinuación de productos; a corto y mediano plazo, se convierten en uno de los insumos principales para la generación de planes productivos, tales como el Programa Maestro de Producción (MPS, por sus siglas en inglés) (Vollmann et al., 2005).

La precisión de estos pronósticos adquiere un carácter crítico en sistemas productivos caracterizados por tiempos de abastecimiento extensos o por una baja flexibilidad operativa (por ejemplo, debido a tiempos de *changeover* prolongados). En estos entornos, que frecuentemente operan bajo estrategias de fabricación contra inventario (*Make-To-Stock*) para garantizar un alto nivel de servicio y una alta utilización de las instalaciones, un error en la estimación puede derivar en rupturas de *stock* o en costosos excesos de inventario (Silver, Pyke & Thomas, 2016).

En el ámbito de los pronósticos tácticos y operativos, cuando se dispone mayoritariamente de datos históricos de demanda, la literatura académica tradicional ha focalizado su atención en métodos basados en series de tiempo. Estos enfoques asumen que el comportamiento futuro guarda similitud con el pasado y emplean los datos disponibles para proyectar escenarios, considerando la presencia de aleatoriedad, tendencias y comportamientos cíclicos o estacionales (Nahmias, 2004; Chapman, 2000). Muchos de estos métodos clásicos, como la regresión lineal, el promedio móvil o la suavización exponencial, destacan por su simplicidad analítica y su probada robustez (Makridakis, Wheelwright & Hyndman, 1998).

Sin embargo, ante el abaratamiento de la capacidad de cómputo y la proliferación de grandes volúmenes de datos, ciertos métodos tradicionales de *forecasting* han comenzado a ser cuestionados. En este contexto, emergen poderosos algoritmos del campo del *Machine Learning* (ML), los cuales prometen una representación más detallada de la realidad al permitir la inclusión de múltiples variables exógenas y datos de contexto (tales como clima, campañas de *marketing*, indicadores económicos, etc.) en la

confección del pronóstico (Carbonneau, Laframboise & Vahidov, 2008; Bontempi, Ben Taieb & Le Borgne, 2012).

Si bien estos nuevos métodos basados en ML han demostrado un éxito rotundo en sectores como el *retail* de consumo masivo, el comercio electrónico y el energético, surge un interrogante válido respecto a su desempeño en entornos industriales más restrictivos, donde a menudo los métodos estadísticos simples siguen superando a los algoritmos complejos (Makridakis, Spiliotis & Assimakopoulos, 2018). El presente trabajo pretende evaluar si esta presunta superioridad de los modelos de *Machine Learning* se mantiene en escenarios donde se carece de abundantes datos de contexto —ya sea por tratarse de pronósticos de mediano plazo donde dichas variables aún son inciertas, o por limitaciones en la integración de datos de la compañía— y se cuenta únicamente con la demanda (variable endógena histórica, a veces reemplazada por las ventas) bajo distintos niveles de variabilidad.

La discusión sobre la superioridad de los modelos de ML frente a los métodos estadísticos clásicos ha cobrado gran relevancia a partir de los resultados empíricos de las competencias de pronóstico a gran escala, como la M4 y M5 de Makridakis. Estos certámenes evidenciaron que, en contextos con alta incertidumbre o datos limitados, los métodos tradicionales o las combinaciones híbridas simples suelen superar o igualar en consistencia operativa a los algoritmos de ML puros y complejos, alineándose con los hallazgos de este trabajo (Makridakis et al., 2018, 2022).

Asimismo, resulta imperativo analizar no solo la precisión (error medio) de los pronósticos, sino también la volatilidad de las proyecciones generadas. Modelos excesivamente complejos pueden sufrir de sobreajuste (*overfitting*) o reaccionar de manera desproporcionada ante cambios recientes (ruido) en la demanda. Esta sobrerreacción tiene un impacto directo y perjudicial en un ambiente productivo. Debe considerarse que el pronóstico no solo alimenta el plan de producción, sino que determina el dimensionamiento de la cobertura de inventario esperado, conocido como *Days of Supply* (DoS).

La métrica de cobertura suele calcularse proyectando el inventario actual frente a los pronósticos de demanda futura (ya sea un promedio estabilizado o proyecciones discretas mes a mes). Si se emplea un algoritmo de pronóstico que arroja valores erráticos, inconsistentes o hipersensibles a la variabilidad reciente, se corre el riesgo de inducir un profundo "nerviosismo" en el sistema de planificación (*System Nervousness*) (Ho, 1989; Heisig et al., 2010). Esto provocaría que las políticas de cobertura y las órdenes de producción cambien drásticamente de un período a otro, anulando la eficiencia operativa que se busca en la manufactura.

Por lo tanto, el objetivo de este artículo es realizar un análisis comparativo entre algoritmos de *Machine Learning* y métodos tradicionales (promedio móvil y regresión lineal) para el pronóstico de demanda sostenida, evaluando no solo el nivel de error tradicional, sino la robustez y estabilidad de las proyecciones frente a distintos niveles de variabilidad en la demanda, y su consecuente impacto en las decisiones de inventario.

2 Descripción de problema

El caso de estudio que motiva este análisis corresponde a una empresa dedicada a la manufactura y distribución de artículos de primera necesidad. Dada la naturaleza intrínseca de estos productos, su demanda global es inelástica y relativamente insensible a las fluctuaciones económicas, aunque sí existe aleatoriedad, tendencia y ciclicidad (asociado a las estaciones del año). La organización opera bajo un modelo de negocio *Business-to-Business* (B2B), por lo que no realiza ventas minoristas a consumidores finales, sino que sus puntos de entrega son grandes centros de distribución y clientes de envergadura.

Para el desarrollo de este análisis, se dispone de un conjunto de datos que abarca varios años de historia y se toman aquéllos más recientes, desde enero del 2024 (27 meses) a marzo de 2026, contemplando el desempeño de 37 productos diferentes. Resulta de especial interés para este estudio que dicho portafolio de productos exhibe una amplia gama de comportamientos, abarcando desde artículos con demanda altamente estable hasta aquellos con distintos (y en casos, muy elevados) niveles de variabilidad. La selección se corresponde con todos los productos que presentan demanda sostenida, no omitiendo ningún artículo de la cartera que presente dicho comportamiento.

Un aspecto crítico y diferenciador de este trabajo es el uso de datos de demanda, y no de ventas, en los modelos de pronóstico. En la práctica industrial, frecuentemente se utilizan los registros de "ventas" o "entregas" como una aproximación (*proxy*) de la demanda. Sin embargo, existe una brecha conceptual y operativa significativa entre la *lo solicitad por el cliente (demanda)* y la *cantidad entregada*. En contextos productivos reales, las restricciones de capacidad, los problemas logísticos o los quiebres de inventario (*stockouts*) provocan entregas parciales. Utilizar datos de ventas en estos períodos subestimaría la verdadera necesidad del mercado y entrenaría al modelo con ruido sesgado a la baja. Por consiguiente, este trabajo utiliza estrictamente registros de la demanda solicitada, asegurando que el modelo de pronóstico intente predecir la verdadera intención de consumo del mercado. No obstante, debe tenerse en cuenta que, ante situaciones de desabastecimiento, los actores de la cadena de suministro suelen reaccionar en los meses posteriores “inflando” sus pedidos. Esta distorsión, conocida como el "efecto látigo" (*bullwhip effect*) (Lee, Padmanabhan & Whang, 1997), introduce una complejidad adicional que, si bien se reconoce, excede el alcance analítico de este trabajo.

Finalmente, el esquema metodológico de evaluación de los modelos está planteado bajo un enfoque *one-step-ahead* (pronóstico a un paso hacia adelante). Si bien en la literatura de predicción de series temporales existen alternativas más complejas como el pronóstico multi-período (*multi-step-ahead*), el cual proyecta de manera simultánea varios períodos futuros mediante estrategias de modelado recursivas o directas (Bontempi, Ben Taieb & Le Borgne, 2012; Hyndman & Athanasopoulos, 2018), este último suele introducir una mayor acumulación de errores e incertidumbre a medida que se expande el horizonte de proyección. En contraposición, el enfoque *one-step-ahead* emula fidedignamente el ciclo de toma de decisiones en la realidad operativa mensual de la empresa: situado en el final del período actual t, el algoritmo de pronóstico debe ser capaz de utilizar exclusivamente la información histórica recolectada hasta ese

momento para predecir la demanda del período subsiguiente, *t+1*. Este enfoque garantiza que la evaluación del modelo refleje su desempeño en condiciones reales de ejecución operativa-táctica, minimizando el impacto de los errores acumulativos.

3 Metodología propuesta

Para llevar a cabo el análisis comparativo, se diseñó un experimento computacional estructurado basado en la construcción y evaluación de un catálogo exhaustivo compuesto por 75 configuraciones o variantes metodológicas. Estas variantes surgen de la combinación de algoritmos base, hiperparámetros (perfiles), granularidad temporal, alcance del modelo y el nivel de detalle respecto al destino de la demanda.

3.1 Tratamiento y granularidad de los datos

La disponibilidad de registros con resolución diaria permitió explorar diferentes estrategias para la agregación de la serie temporal antes del entrenamiento de los modelos:

- **Granularidad (Diario vs. Mensual):** en la modalidad *mensual*, los modelos se entrenan directamente sobre la serie de tiempo agregada por mes. En la modalidad *diario*, el algoritmo se entrena prediciendo el comportamiento día a día para luego agregar aritméticamente dichos pronósticos y obtener el valor total mensual.
- **Alcance (Individual vs. Conjunto):** en el enfoque *individual*, cada producto compite estrictamente con su propia historia de demanda. Por el contrario, el enfoque *conjunto* entrena una única serie temporal agregada que engloba a todos los productos activos; el pronóstico total resultante es luego distribuido (*top-down*) entre los productos en función de su participación porcentual reciente.
- **Uso del Destino (Sin destino vs. Con destino):** dado que se conoce a qué cliente o centro de distribución que genera la demanda, se evalúa el impacto de incorporar la dimensión de destino del pedido. La variante *con destino* incorpora esta dimensión como una variable categórica explicativa adicional, mientras que *sin destino* modela la demanda del producto ignorando a quién va dirigida.

3.2 Ingeniería de Características (*Feature Engineering*)

Para dotar a los algoritmos de *Machine Learning* de la capacidad para interpretar ciclos y relaciones, se definió un conjunto estandarizado de *features* (variables independientes). Estas incluyen un índice temporal continuo (*time_idx*) y codificaciones trigonométricas para capturar estacionalidad (*month_sin, month_cos*). En los modelos de granularidad diaria se añaden los componentes análogos para el día (*day_sin, day_cos*). Como variables categóricas se incluyen la identificación del producto (*id_prod*) y, de corresponder, el cliente (*destino*).

A nivel de preprocesamiento, se estandarizaron las variables numéricas mediante *StandardScaler* y las variables categóricas fueron transformadas utilizando *OneHotEncoder*. Esta transformación es fundamental para asegurar que los modelos de

aprendizaje no introduzcan sesgos debidos a las grandes diferencias de magnitud entre las variables numéricas continuas, y para prevenir que asuman relaciones de jerarquía u ordinalidad falsas entre las distintas categorías de clientes o productos.

3.3 Catálogo de Algoritmos Evaluados

El catálogo de 75 métodos se construye utilizando las siguientes familias algorítmicas, cada una con distintos perfiles o configuraciones de hiperparámetros:

- **Promedio Móvil Simple (SMA):** actúa como línea base o *baseline* de los métodos tradicionales. Se configuró exclusivamente en formato mensual, individual y sin distinción por destino, explorando ventanas de tiempo de 3, 6 y 12 períodos (3 variantes).
- **Regresión Lineal:** utilizado como método estadístico clásico (perfil estándar). Se combinó con todas las variantes de granularidad, alcance y destino (8 variantes).
- **Random Forest:** se evaluaron dos perfiles. Un perfil *Std* (80 estimadores, profundidad máxima 7) y un perfil *Deep* que busca mayor adaptación a patrones complejos (120 estimadores, profundidad 14). Su despliegue genera 16 variantes metodológicas.
- **Gradient Boosting:** se evaluaron mediante GradientBoostingRegressor. El perfil *Std* utiliza pérdida por error cuadrático medio (MSE), mientras que se incluyó un perfil *Robust* que utiliza pérdida de Huber, la cual es menos sensible a los valores atípicos (16 variantes).
- **XGBoost:** algoritmo avanzado de ensamble basado en árboles. Se contrastó un perfil *Std* (profundidad 5) contra un perfil *Conservative* (profundidad reducida a 4 y tasa de aprendizaje baja de 0.05), buscando reducir la propensión al *overfitting* frente a la variabilidad (16 variantes).
- **LightGBM:** similar al XGBoost pero optimizado computacionalmente. El perfil *Std* utiliza 31 hojas por árbol, mientras que el *Conservative* restringe su complejidad a 15 hojas, también con una tasa de aprendizaje conservadora de 0.05 (16 variantes).

3.4 Criterios de Selección y Filtros de Elegibilidad

A la hora de dirimir la competencia entre las 75 variantes metodológicas del catálogo para un producto y mes determinado, la métrica base empleada fue la minimización de la Desviación Absoluta Media (MAD) calculada sobre el período retrospectivo reciente. Sin embargo, seleccionar irrestrictamente el modelo con el menor MAD conlleva un riesgo operativo severo: los algoritmos complejos de Inteligencia Artificial pueden sufrir de "ilusión algorítmica", logrando un error promedio engañosamente competitivo mediante la compensación de errores extremos, para luego arrojar proyecciones tácticas completamente desfasadas de la inercia real del negocio. En la Figura 1 y 2, se muestran ejemplos, en los cuales el método menor MAD exhibe un valor muy bajo con un ajuste alto en los meses previos, pero proyecta un valor muy poco probable para el próximo mes.

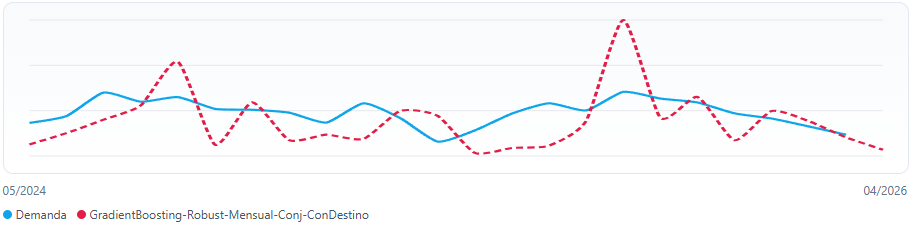

Fig. 1. Pronóstico de demanda con menor MAD, usando *Gradient Boosting Robust*, de 27.754 unidades mensuales para un producto de demanda mensual promedio de 180.746, siendo el último mes el de menor demanda en el último año con 95.000 unidades.

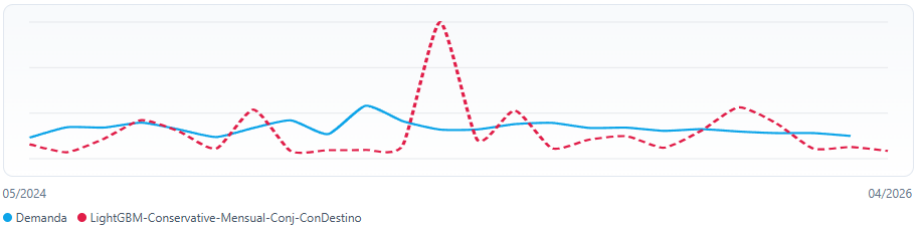

Fig. 2. Pronóstico de demanda con menor MAD, usando *LightGMB Conservative*, de 45.780 unidades mensuales para un producto de demanda mensual promedio de 171.738, siendo el último mes el de menor demanda en el último año con 134.400 unidades.

Para mitigar este riesgo, el algoritmo de selección fue confinado dentro de dos reglas de negocio o "protecciones" operacionales de cumplimiento obligatorio para que un método sea elegible como ganador predefinido:

- **Control del sesgo mediante una señal de rastreo (*Tracking Signal* o TS):** Se exigió que el método candidato presentara un valor absoluto de en la señal de rastreo estrictamente controlado ($|TS| < 3$). Esta métrica, que relaciona la suma acumulada de errores con el MAD, penaliza y descarta a los modelos que sistemáticamente se desvían de la realidad.
- **Filtro lógico de proyección atípica:** se calcula el promedio mensual de la demanda de los últimos 6 meses para el producto en cuestión; si el pronóstico arrojado por el modelo cae por debajo del 50% o supera el 150% de dicho promedio, la proyección se descarta. El filtro actúa como una heurística de control de riesgos que descarta soluciones matemáticamente óptimas pero inviables operativamente.
  Existen matices operativos vitales sobre este segundo filtro: el rechazo no implica que el registro sea un "error confirmado" (ya que la demanda real futura es aún desconocida), sino que la proyección se clasifica como "fuera de la banda histórica reciente". Este nivel de fluctuación resulta inaceptable para la estabilidad requerida en un entorno de producción *Make-To-Stock*.

4 Resultados y Discusión

4.1 Registro de Iteraciones Óptimas

Aplicando el estricto marco metodológico detallado en la Sección 3.4 (minimización de MAD, restricción estadística de $|TS| < 3$, y la exclusión de las proyecciones catalogadas como atípicas operativamente), el experimento evaluó el desempeño del catálogo entre enero de 2024 y marzo de 2026. Durante este horizonte de validación, tras descartar los modelos que incurrieron en sobreajuste (*overfitting*) o subestimaciones anómalas, se consolidaron 937 iteraciones exitosas de selección de "método ganador" para el portafolio de 37 productos.

4.2 Cuantificación del Riesgo: El Impacto de las Alucinaciones Algorítmicas

Al contrastar los resultados obtenidos bajo el marco metodológico estricto (exclusiones mencionadas) frente a un escenario de control donde se permitiera la libre selección de algoritmos omitiendo la restricción de proyección atípica, fue posible cuantificar matemáticamente la propensión al sobreajuste de los modelos avanzados en este entorno industrial.

En el escenario sin restricciones operativas, el sistema hubiera validado 996 iteraciones. Al aplicar los filtros, el total descendió a 937. El análisis detallado de las caídas revela que existieron 99 casos donde un modelo complejo logró engañar a las métricas retrospectivas (alcanzando un MAD competitivo y un TS aceptable), pero generó proyecciones operativas inestables arrojando un valor futuro absurdo.

La exclusión sistémica de estas 99 proyecciones peligrosas evidenció el rol de las heurísticas inerciales. El Promedio Móvil Simple (SMA) experimentó un salto de 365 a 405 victorias (40 casos). Esto arroja evidencia de que, frente a una proyección atípica de la Inteligencia Artificial, el SMA suele ser la alternativa subyacente más robusta para cubrir el vacío operativo de forma segura. En los 59 casos restantes, la volatilidad fue tal que ningún método del catálogo logró cumplir simultáneamente con los criterios de precisión histórica y coherencia estructural.

El impacto de este fenómeno fue especialmente crítico en la familia de Gradient Boosting. Sin restricciones, esta familia fue seleccionada en 376 oportunidades. Al aplicar el filtro, descendió a 310 (perdiendo 66 victorias). Esto demuestra empíricamente que el 17,5% de las veces que el Gradient Boosting lograba el mejor ajuste estadístico histórico puro, en la práctica estaba arrojando una proyección atípica e inoperable. Caídas similares en volumen relativo se observaron en Random Forest (18 casos) y Regresión Lineal (15 casos).

4.3 Desempeño General por Familia de Algoritmos y Variabilidad de Demanda

El análisis matricial de los resultados desafía frontalmente la asunción de que la Inteligencia Artificial, particularmente *Machine Learning*, domina indiscriminadamente

cualquier problema de series de tiempo. Cuando ML es sometido a controles estrictos de estabilidad operativa, las heurísticas tradicionales demuestran aún un valor vigente. La Tabla 1 muestra las victorias de cada familia algorítmica.

Table 1. Frecuencia de selección de métodos óptimos por Familia Algorítmica y Tipo de Variabilidad de Demanda (Sujeto a restricciones de TS y Proyección Atípica).

Familia de Métodos	Dem. Errática	Dem. Estable	Dem. Variable	Total	%
Promedio Movil Simple (SMA)	24	201	180	405	43,2%
Gradient Boosting (GB, XGB, LGBM)	53	112	145	310	33,1%
Random Forest	15	58	88	161	17,2%
Regresión Lineal	22	23	16	61	6,5%
Total General	114	394	429	937	100%

De la evidencia empírica se desprenden ciertas observaciones importantes.

- **Estabilidad del promedio móvil:** de manera individual, la familia más exitosa fue el promedio móvil simple, capturando el 43,2% de las victorias. Su dominio es abrumador en los escenarios de demanda estable y muy variable. La razón empírica es clara: al prohibirle a los algoritmos complejos "sobre ajustar" o compensar errores extremos, la inercia analítica del SMA se vuelve el ancla más segura. El SMA promedia el ruido sin sobre reaccionar y garantiza proyecciones que siempre recaen dentro de la banda histórica reciente. En particular, las variantes de largo plazo (SMA-12 y SMA-6) aportaron la mayor cuota de estabilidad (201 y 104 victorias, respectivamente).
- **El *Machine Learning* domina en escenarios erráticos:** cuando la demanda se vuelve esporádica e impredecible (demanda errática), las heurísticas inerciales como el SMA disminuyen su efectividad (apenas 24 victorias). En este segmento, los modelos de *Machine Learning* demostraron ser superiores. El Gradient Boosting lideró este cuadrante con 53 victorias. La capacidad del ML para encontrar patrones no lineales en el ruido resulta evidente, pero los datos confirman que su uso debe reservarse para aquellas series donde los métodos simples demuestran ser inherentemente incapaces.
- **Los perfiles conservadores aportan valor:** al analizar el perfil de los modelos cuyo pronóstico fue el mejor en un mes determinado, dentro de la familia de IA, los modelos configurados bajo un perfil "Conservative" (profundidad de árbol reducida y baja tasa de aprendizaje) lograron la mayoría de los aciertos válidos, reiterando que restringir matemáticamente el potencial de aprendizaje es vital para evitar el *overfitting* en contextos industriales de productos con perfiles de demanda maduro y un comportamiento sostenido en el tiempo.

4.4 Impacto de las Estrategias de Preprocesamiento y Agregación

Se evaluó el impacto de la granularidad temporal, el modelado cruzado y la información espacial. Los resultados se detallan en la Tabla 2.

Table 2. Impacto de las variantes de configuración e ingeniería de características sobre los 937 casos óptimos.

Variantes	Dem. Errática	Dem. Estable	Dem. Variable	Total	%
Granularidad Temporal					
Mensual	107	301	365	773	82,5%
Diario	7	93	64	164	17,5%
Alcance del entrena-miento					
Individual (Por producto)	85	304	319	708	75,6%
Conjunto (Múltiples series)	29	90	110	229	24,4%
Uso del destino					
Sin Destino (Agregado)	59	242	230	531	56,7%
Con Destino (Desagregado)	55	152	199	406	43,3%

De estos resultados se observan importantes situaciones:

- **Usar el detalle de transacciones diarias para una proyección mensual no genera un beneficio:** intentar predecir la demanda acumulando predicciones diarias probó ser altamente ineficiente y riesgoso. El 82,5% de las victorias correspondieron a entrenamientos puramente mensuales. Descender al nivel diario en series endógenas genera "ruido".
- **El destino promueve el sobreajuste:** los métodos que no recurrieron a la diferenciación por destino fueron superiores (56,7% de victorias). En un escenario de demanda errática, la segmentación por destino quedó prácticamente empatada (55 victorias) contra el agrupamiento general (59 victorias). Esto indica que, al menos en este caso, agregar granularidad espacial (cuántas categorías de clientes existen) suele inducir al algoritmo a memorizar ruido aleatorio particular de un cliente en lugar de abstraer el comportamiento real agregado del producto.
- **Las proyecciones por destino individual son mejores:** el enfoque de pronóstico por destino individual superó ampliamente (75,6%) al modelado conjunto (24,4%), indicando que la transferencia de aprendizaje entre clientes no siempre compensa la dilución de la identidad histórica propia de cada uno de ellos.

4.5 Impacto del Pronóstico en el Cálculo de Cobertura Avanzado

El fin último del pronóstico de demanda (**F**) y su respectivo error (**E**), equivalente en este caso a **1.25 * MAD** es determinar el tiempo de cobertura **R** (en meses) que el *stock* actual **Q** de la compañía puede soportar. Este cálculo nace de la ecuación de balance de inventario (Ecuación 1).

$$Q = F \cdot R + z \cdot E \cdot \sqrt{R} \quad (1)$$

Donde $z \cdot E \cdot \sqrt{R}$ representa el *stock* de seguridad dinámico. Despejando R, aplicando la resolvente con un reemplazo de variables ($X = \sqrt{R}$), se obtiene la expresión analítica para la cobertura temporal (Ecuación 2).

$$R = \left(-\frac{-z \cdot E + \sqrt{(z \cdot E)^2 + 4 \cdot F \cdot Q}}{2 \cdot F} \right)^2 \quad (2)$$

El análisis de esta ecuación teórica justifica plenamente la implementación del filtro de proyección atípica. Al ser el pronóstico el denominador implícito fundamental en el cálculo de la cobertura, permitir que un algoritmo proyecte matemáticamente una caída del 80% (frecuente en ML no supervisado que "sobre ajusta" picos) haría que la cobertura calculada se disparase significativamente de forma ficticia; inversamente, una proyección desmedidamente alta colapsaría la cobertura, disparando alertas de compras/cambios en los planes de producción masivos (Nerviosismo del Sistema). Al contener a los algoritmos dentro del rango [0.5, 1.5] del promedio histórico, se neutralizan las interacciones caóticas, logrando que el sistema capitalice la precisión del pronóstico sin exponer una la operación a una vulnerabilidad intrínseca de la herramienta de pronóstico.

5 Conclusiones

El presente trabajo abordó la problemática del pronóstico de demanda sostenida en entornos industriales tácticos, contrastando el desempeño de algoritmos avanzados de *Machine Learning* frente a métodos tradicionales, en un portafolio B2B de 37 productos. A través del análisis riguroso de más de 900 proyecciones sometidas a estrictos límites de control operativo, los resultados experimentales brindan evidencia en contra de una superioridad incuestionable de la IA, revalidando la vigencia de métodos más primitivos como el promedio móvil.

La conclusión central es que, para salvaguardar la estabilidad de un Programa Maestro de Producción, la simplicidad e inercia de métodos como el promedio móvil simple superan a los algoritmos complejos de *Machine Learning* en entornos de demanda estable y moderadamente variable. El SMA garantizó el 43% de los pronósticos seguros al minimizar la sobrerreacción al ruido, previniendo las proyecciones erráticas que caracterizan al sobreajuste de las redes no lineales.

En contraposición, se determinó que algoritmos avanzados como Gradient Boosting y Random Forest deben ser conceptualizados no como soluciones universales, sino como herramientas altamente especializadas para escenarios muy variables. Fue exclusivamente en el segmento de demanda errática donde estas técnicas lograron aislar patrones subyacentes superando a los modelos inerciales. Sin embargo, su éxito requiere indefectiblemente estar subordinado a lógicas de hiperparámetros conservadoras (perfiles restrictivos) y a restricciones de negocio duras, tales como la exclusión de

cualquier pronóstico que fluctúe fuera del 50 de la banda histórica reciente, ya sea por arriba o abajo del pronóstico (filtro de proyección atípica).

Adicionalmente, se encontró que inyectar mayor nivel de detalle a los modelos, ya sea aumentando la granularidad temporal (datos diarios) o segregando espacialmente la información (inclusión del destino), induce mayormente a la "ilusión algorítmica". En ausencia de variables causales exógenas contundentes, exigirle al modelo predecir el comportamiento individual de cada centro de distribución día a día empuja a la Inteligencia Artificial a memorizar ruido aleatorio, degradando la capacidad de generalización del pronóstico agregado mensual.

Resulta tema de investigaciones en curso el explotar el uso de datos de, por ejemplo, demanda no atendida que, sin dejar de ser endógenos, podrían ser buenos predictores de oscilaciones atípicas en la demanda futura como consecuencia de “pequeños” efectos látigo.

Finalmente, es importante señalar que, al basarse en un caso de estudio de 37 productos de una empresa particular enfocada en el sector B2B, los resultados cuantitativos de esta investigación no pueden generalizarse. Sectores con modelos de negocio diferentes (como el retail B2C con disponibilidad masiva de variables exógenas e interacciones dinámicas de mercado) podrían exhibir un comportamiento distinto frente a estas metodologías. No obstante, a partir de la ecuación analítica de cobertura de inventario (*Days of Supply*), se mostró que la minimización descontextualizada de métricas de error retrospectivas (MAD/RMSE) es operativamente peligrosa si no se acompaña de una restricción de volatilidad proyectiva. Las predicciones anómalas alteran drásticamente el discriminante de la función de cobertura, retroalimentando el Efecto Látigo en la cadena de abastecimiento y generando nerviosismo en la planificación. Por ende, la integración de la ciencia de datos en la manufactura resulta de gran valor únicamente cuando la capacidad algorítmica es encauzada mediante la prudencia de la investigación operativa tradicional.

Declaración sobre el uso de Modelos de Lenguaje de Gran Escala (LLM).
Durante la preparación de este trabajo, el autor utilizó Google Gemini con el fin de mejorar el lenguaje y la legibilidad del manuscrito. Tras utilizar esta herramienta/servicio, el autor revisó y editó el contenido según fuera necesario y asume plena responsabilidad por el contenido de la publicación.

6 Bibliografía

1. Bontempi, G., Ben Taieb, S., & Le Borgne, Y. A. (2012). Machine learning strategies for time series forecasting. *European Business Intelligence Summer School*, 62-77.
2. Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140-1154.
3. Chapman, S. N. (2000). *The Fundamentals of Production Planning and Control*. Pearson.
4. Heisig, G., Fleischmann, M., & Meyr, H. (2010). Planning stability in supply chain management. *Supply Chain Management and Advanced Planning*, 415-435.

5. Ho, C. J. (1989). Evaluating the impact of operating environments on MRP system nervousness. *International Journal of Production Research*, 27(7), 1115-1135.
6. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3), e0194889.
7. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 accuracy competition: Results, findings and conclusions. International Journal of Forecasting, 38(4), 1346-1374. https://doi.org/10.1016/j.ijforecast.2021.11.013
8. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.
9. Nahmias, S. (2004). *Production and Operations Analysis* (5th ed.). McGraw-Hill.
10. Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and Production Management in Supply Chains*. CRC Press.
11. Vollmann, T. E., Berry, W. L., Whybark, D. C., & Jacobs, F. R. (2005). *Manufacturing Planning and Control for Supply Chain Management* (5th ed.). McGraw-Hill.