



Análisis de modelos basados en Machine Learning para la predicción de la demanda de productos en la empresa Dyna & Cía. S.A.

Alejandro Correa Loaiza

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesora

Tatiana Mejía Lemus, Magíster (MSc) en Ingeniería

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia

2023

Cita	Correa Loaiza [1]
Referencia Estilo IEEE (2020)	[1] A. Correa Loaiza, “Análisis de modelo de pronóstico para la predicción de la demanda de productos en la empresa Dyna y Cía.”, Trabajo de grado especialización, Especialización en Analítica y Ciencia de Datos, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2023.



Especialización en Analítica y Ciencia de Datos, Cohorte IV.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Rector: John Jairo Arboleda Céspedes

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama.

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

TABLA DE CONTENIDO

I.	INTRODUCCIÓN	8
II.	RESUMEN	10
III.	ABSTRACT	11
IV.	PLANTEAMIENTO DEL PROBLEMA.....	12
	<i>A. Descripción del negocio</i>	12
	<i>B. Descripción del problema</i>	14
V.	JUSTIFICACIÓN	18
VI.	OBJETIVOS.....	19
	<i>A. Objetivo general</i>	19
	<i>B. Objetivos específicos</i>	19
VII.	MARCO TEORICO.....	20
	<i>A. Arima</i>	21
	<i>B. Redes Neuronales</i>	22
	<i>C. Regresión</i>	22
	<i>D. Random Forest</i>	23
VIII.	METODOLOGÍA	24
	<i>A. Adquisición de Datos</i>	25
	<i>B. Procesado de Datos</i>	25
	<i>C. Extracción y creación de variables significativas</i>	26
	<i>D. Algoritmo Machine Learning</i>	26
	<i>E. Evaluación y obtención de los modelos</i>	27
IX.	RESULTADOS.....	28
	<i>A. Ítem 43387</i>	32
	<i>B. Ítem 34631</i>	36

<i>C. Ítem 17001</i>	38
X. DISCUSIÓN	40
XI. CONCLUSIONES	41
XII. RECOMENDACIONES	42
REFERENCIAS	43
ANEXOS.....	44

LISTA DE TABLAS

TABLA I ATRIBUTOS	28
TABLA II PRODUCTOS SELECCIONADOS.....	30
TABLA III MODELOS ÍTEM 43387	34
TABLA IV MODELOS ÍTEM 34631	36
TABLA V MODELOS ÍTEM 17001	38

LISTA DE FIGURAS

Fig. 1 Cobertura Dyna.....	12
Fig. 2 Comportamiento del nivel de inventario.....	13
Fig. 3 Inventario vs Nivel de entrega en Dyna &Cía.....	15
Fig. 4 Representación de un modelo de redes neuronales.....	22
Fig. 5 Representación Random Forest	23
Fig. 6 Metodología Machine Learning.....	24
Fig. 7 Dataset Principal	29
Fig. 8 Top 10 en Ventas	31
Fig. 9 Gráfico 43387 vs Pronóstico Actual	32

SIGLAS, ACRÓNIMOS Y ABREVIATURAS

Core	El término core se utiliza para dar referencia a algo fundamental o esencial.
Csv	Es la extensión de un archivo de valores separado por comas.
Xls	Siglas de un archivo en formato de hoja de cálculo de Excel.
JIT	Traducción al español como justo a tiempo, refiriéndose a la metodología.
Nivel de entrega	Relación entre lo pedido por el cliente y lo que le fue entregado o facturado.
Lead time.	Tiempo que conlleva desde la solicitud hasta la recepción del pedido.
SKU	Código para identificar un producto.
MOQ	Mínima cantidad de pedido.
Dataset	Conjunto de datos.
Contact center	Departamento que facilita la comunicación con el cliente para brindar soporte, resolver consultas, ofrecer productos.
CEDI	Almacén que se encarga de recibir, almacenar y distribuir productos.
ERP	Sistema que permite integrar, gestionar diferentes procesos de una organización. Algunos ejemplos de ERP son SAP, Oracle, Siesa, NetSuite.
Retail	Es un tipo de comercio que se refiere a la venta de un inventario masivo a múltiples clientes.

I. INTRODUCCIÓN

Tradicionalmente, las empresas dedicadas a la comercialización de productos y servicios se han apoyado en la información de ventas para planear su operación a futuro, y aunque se han obtenido resultados útiles, no siempre lo planeado resulta muy acertado para ese viento futuro, generando consecuencias negativas para las organizaciones.

A partir de aplicaciones estadísticas y matemáticas se han obtenido mejoras en los pronósticos, sin embargo, en la era digital que nos encontramos, las organizaciones tienen acceso a una enorme cantidad de datos proveniente de diferentes fuentes gracias a la disponibilidad y velocidad [1] que se tiene en el presente, resultando esto en un insumo importante para tener en cuenta a la hora de generar predicciones.

Hoy en día, los datos en sí mismos no representan como tal un activo o un beneficio para las organizaciones, pero sí son el recurso para poder extraer información a partir de ellos. Se ha convertido entonces en un punto clave poder gestionar estos grandes volúmenes de datos y en otros esfuerzos más mayúsculos, el poder analizarlos de modo que se conviertan en información relevante para la toma de decisiones.

Sin embargo, con la llegada del Machine Learning o aprendizaje automático por su traducción al español, se ha abierto una nueva puerta que ayudaría a la planeación a través de la generación de predicciones a venetos futuros, permitiendo en muchas ocasiones una mayor precisión y flexibilidad ante los cambios del mercado. En resumen, el uso de técnicas de Machine Learning para pronosticar la demanda suponen una mayor eficiencia en la utilización de recursos y una ventaja para las organizaciones, al brindar la capacidad de adaptarse a nuevos datos y una gama de variables diferentes.

Así, este trabajo pretende presentar un análisis de la aplicación de modelos basados en el aprendizaje automático para predecir la demanda, que permita comparar y ofrecer nuevas herramientas que podrían ser útiles para tomar decisiones de forma más eficiente. De esta forma, en el primer apartado se encontrará el planteamiento del problema, donde se describe el contexto

en el que se desarrollará el modelo y la problemática actual. En el segundo acápite se presentan los objetivos, el marco teórico y la metodología, y por último se relacionan los resultados de la implementación de los modelos.

II. RESUMEN

Este estudio tuvo como objetivo analizar el uso de técnicas y modelos basados en Machine Learning para el pronóstico de la demanda en Dyna & Cía. S.A. La metodología utilizada se centró en modelos de Machine Learning que mejoran el rendimiento propio del modelo mediante el uso de datos.

Los resultados obtenidos indican que los modelos de aprendizaje automático superan al método utilizado actualmente por la empresa en términos de ajuste y precisión del pronóstico de la demanda. Esto significa que los modelos basados en Machine Learning ofrecen una mejora significativa en la capacidad de predecir la demanda de productos de Dyna & Cía. S.A., además de brindar más posibilidades para explorar y seleccionar mejores herramientas con el fin disminuir el error de la predicción.

Finalmente, el estudio demuestra la importancia en el uso de técnicas y modelos basados en Machine Learning para generar pronósticos de demanda más acertados para Dyna & Cía. S.A., lo que permite una toma de decisiones más informada y eficiente en la gestión de la demanda.

III. ABSTRACT

This study aimed to analyze the use of techniques and models based on Machine Learning for demand forecasting in Dyna & Cía. S.A. The methodology used focused on Machine Learning models that improve the model's own performance using data.

The results obtained indicate that Machine Learning models outperform the method currently used by the company in terms of demand forecast fit and accuracy. This means that the Machine Learning based models offer a significant improvement in the ability to predict the demand for Dyna & Cía. S.A. products, as well as providing more possibilities to explore and select better tools to decrease the prediction error.

Finally, the study demonstrates the importance of using techniques and models based on Machine Learning to generate more accurate demand forecasts for Dyna & Cía. S.A., which allows a more informed and efficient decision making in demand management.

IV. PLANTEAMIENTO DEL PROBLEMA

A. Descripción del negocio

La empresa Dyna & Cía. S.A. es una distribuidora de artículos de ferretería liviana con más de 60 años de existencia en el mercado colombiano. Contiene un amplio portafolio de productos, que suman alrededor de 6.000 referencias provenientes de más de 106 proveedores entre nacionales y extranjeros. Con estos se atiende varios canales como cadenas, exportaciones, y distribución (canal tradicional) que es donde se encuentra el core del negocio. Dyna atiende entonces alrededor de 7600 clientes repartidos en las diferentes regiones del país y más de 545 ciudades y poblaciones rurales.



Fig. 1 Cobertura Dyna

Fuente: Imagen tomada de presentación corporativa

En Dyna los pedidos pueden ser tomados a través de dos vías. La primera, por ventas en línea, que se dan mediante la tienda virtual (página web) [2], y el contact center donde el cliente puede hacer el pedido a través de una llamada y el asesor es quién diligencia el encargo. La segunda vía es la venta presencial que consiste en la visita de uno de los vendedores de la compañía y que está asignado a una zona, éste visita la ferretería con cierta frecuencia y el cliente hace el pedido a ese vendedor, quién monta el pedido en el sistema. Después de este proceso, los pedidos se gestionan y despachan desde el Centro de Distribución CEDI, siempre y cuando la referencia se encuentre con inventario disponible, de lo contrario esta queda para el histórico de demanda del producto y no como una venta.

Simultáneamente en el área de abastecimiento, a partir del histórico de pedidos y con la información del inventario, el área de compras realiza constantemente la revisión de los productos asociados a cada proveedor con una frecuencia establecida. La revisión comprende una

comparación entre el inventario que se tiene del producto al momento de la revisión, versus la demanda mensual del mismo que se calcula como un promedio de los últimos meses, todo esto con el fin de establecer cuántos meses de inventario se tiene. Cada producto tiene un máximo de inventario regido por la política de la compañía dependiendo de su clasificación, siendo la clasificación A para los productos con mayor participación en la venta y E la menor, así se define el tope de cuantos meses se puede tener de inventario de los diferentes productos. Finalmente, con estas variables mencionadas se toma la decisión de cuántas unidades comprar. Para mayor entendimiento, se tiene la siguiente imagen con el comportamiento del inventario:

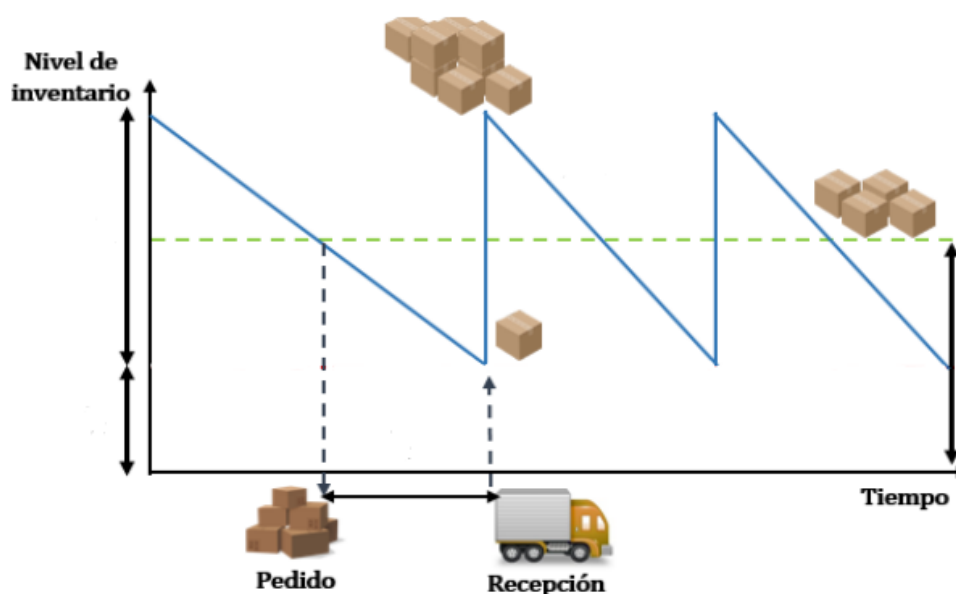


Fig. 2 Comportamiento del nivel de inventario.

Fuente:

<https://ciateq.repositorioinstitucional.mx/jspui/bitstream/1020/347/1/Innovacion%20de%20procesos%20para%20optimizar%20el%20punto%20de%20reorden.pdf>

Como se logra apreciar en la **Fig. 2**, el inventario va bajando a medida que las ventas se van dando a través del tiempo y en el transcurso de ese tiempo se hace la revisión. Después de la revisión se genera una orden de compra o pedido al proveedor quien iniciará la gestión y entregará el pedido después de un tiempo, para así nuevamente subir los niveles de inventario. De esta manera continua el ciclo, teniendo el propósito de no sobrepasar los niveles establecidos, ni quedarse sin unidades disponibles para la venta.

B. Descripción del problema

Es un hecho que la gestión del inventario es un tema tan relevante como complejo dentro de las compañías comercializadoras de productos. Y que en la práctica la venta de un producto no se comporta de forma lineal, puesto que se ve inmersa en muchos factores, como puede ser el aumento reiterado en el consumo del producto, variaciones debido a condiciones climatológicas, vacaciones o fechas especiales, comportamientos recurrentes como por ejemplo periodos de elecciones, etapas de bonanza o crisis económicas, además de picos o caídas irreconocibles.

Sin embargo, estar un paso o varios adelante, se convierte en aspecto clave para planear y estar preparados frente a las eventualidades. Esta predicción no deja de ser un riesgo, porque la inexactitud en el pronóstico de la demanda genera desaciertos a la hora de comprar, errores que no aportan a tener un abastecimiento adecuado, ocasionando excesos en algunos sku´s y agotados en otros, fenómeno que se presenta en Dyna hoy en día.

El área de compras se esfuerza por satisfacer de mejor manera los pedidos de los clientes, sin embargo, los vendedores de la compañía reportan de manera constante la pérdida de ventas por faltas de producto. Además, del otro lado se tiene una fuerte presión por parte de las directivas para disminuir el capital de trabajo invertido en el inventario.

En muchas ocasiones la empresa ha optado por subir sus niveles de inventario para no llegar a tener productos sin disponibilidad priorizando así la venta, sin embargo, aparte del esfuerzo en capital que esto requiere, no necesariamente un mayor inventario en general logra resolver el problema de agotados. Un ejemplo claro de lo expresado anteriormente se puede observar en la **Fig. 3**, donde se compara los días de inventario con el nivel de entrega del mismo periodo, refiriéndonos a nivel de entrega como lo que se logró facturar del valor total del pedido.

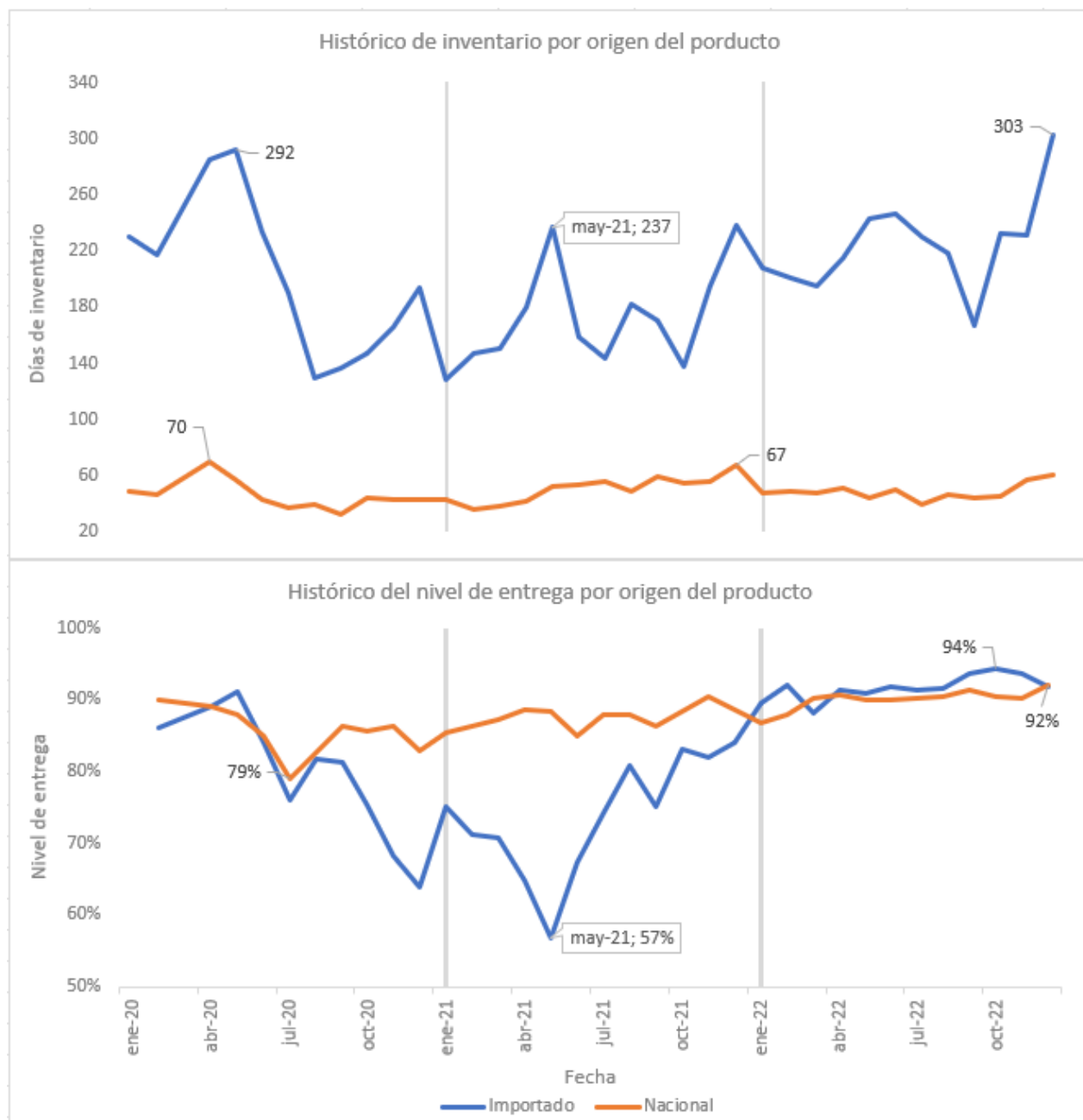


Fig. 3 Inventario vs Nivel de entrega en Dyna &Cía.

En el análisis se encuentra que no necesariamente un mayor nivel de inventario se traduce en un mejor nivel de entrega, dado que el inventario podría estar excedido en productos que no se están vendiendo, pero con faltantes en los que sí se requieren. Esto se observa claramente en el mes de mayo de 2021, donde se tenían 237 días de inventario en producto de origen importado, pero con un nivel de entrega con un 57%, significando una gran pérdida para las ventas de la compañía.

Es importante observar, que el nivel de entrega de los productos de origen nacional visto en la **Fig. 3** se ha mantuvo cercano al 90% al pasar de los años, sin tantas variaciones en el tiempo y sin caídas drásticas como si lo hizo el producto de origen importado, con un promedio del 82%. Lo anterior, logrado con un inventario mucho menor hablando de valores cercanos a la mitad de lo que cuesta el inventario de importado, sin omitir por supuesto que el inventario de productos importados está estrechamente ligado a los tiempos de espera de 6 meses promedio, pues en su mayor parte son traídos desde China. Consecuentemente, se tiene un inventario mucho mayor comparado con el de origen nacional, pero es ahí precisamente donde cobra importancia el tener un pronóstico y además de ello, un pronóstico preciso.

Los aspectos mencionados anteriormente, son los problemas comunes cuando de inventarios se habla, y aunque es claro que no se llega a la perfección y que siempre estarán presentes, la idea será propender por encontrar el balance e intentar porque ambos efectos en el inventario se den en porcentaje minoritario, ideal que busquen todas las empresas teniendo como referente el enfoque de justo a tiempo o por sus siglas en inglés JIT, "que puede resumirse en fábricas con productos estrictamente necesarios, en el momento preciso y en las cantidades debidas. Metodología para alcanzar la excelencia en una empresa" [3].

Bajo la premisa de ser más eficientes y sobre todo teniendo en cuenta la actual situación económica global, las empresas están buscando reducir sus niveles de endeudamiento, ya que dicha situación ha traído inflación y aumento en las tasas de interés, y que para Colombia los pagos de intereses pasaron a ser más del doble comparando mayo del 2022 contra mayo 2023, 6% y 13,25% respectivamente, según datos recogidos del Banco de la República de Colombia [4].

En cuanto a una empresa comercializadora y distribuidora de productos como lo es Dyna & Cía. S.A., la mayor parte de su capital de trabajo está en el inventario, en su gran mayoría obtenidos a través de créditos financieros. Entonces, una mejoría económica podría pensarse claramente a través de la optimización de sus inventarios, con el ánimo de afectar en menor medida el nivel de entrega frente a sus clientes.

En ese orden, se hace necesario realizar un análisis más profundo de la generación pronóstico de la demanda, pues las consecuencias de agotados y excesos reflejan un problema de inexactitud. Variables como la frecuencia de compra, el tiempo de espera, y la demanda; son pilares dentro del análisis del abastecimiento, sin embargo, como las dos primeras no dependen únicamente de la compañía que está realizando la compra, sino también de los procesos y la gestión del proveedor, será una primer medida y enfoque de este trabajo, el estudio de la demanda.

V. JUSTIFICACIÓN

En la actualidad, la inteligencia artificial es una de las herramientas más usadas para la resolución de todo tipo de problemas. El ingreso de las ciencias de la computación a todo lo cada vez más necesario comprender las posibilidades que ofrece esta nueva ciencia y cómo vincularla a las acciones cotidianas de las empresas e instituciones. En ese sentido, este trabajo analiza cómo a través de modelos de predicción de la demanda, es posible mejorar los procesos internos que usan las compañías dedicadas a las ventas, específicamente las del sector retail, en el momento de realizar las compras de productos que serán ofertados al público.

Los avances en la tecnología y disponibilidad de los datos en los últimos años han sido significativos. Lo anterior, ha llevado a tener más capacidad para almacenar datos, mejores herramientas para la visualización, y la posibilidad de utilizar modelos más dinámicos y flexibles. En el ámbito de los pronósticos ventas, Machine Learning ofrece la posibilidad de trabajar con variables que no son tan identificables o fáciles de manejar por los métodos tradicionales. Por ejemplo, la inclusión de datos de redes sociales, fechas especiales y demás, que podrán terminar influyendo en la demanda de productos o servicios. Además de trabajar con variables relevantes y algunas un poco más complejas, se resalta la capacidad de analizar grandes volúmenes de información, con la facilidad de adaptarse de forma continua a medida que se ingresan nuevos datos al modelo, resultando muy provechoso en entornos tan cambiantes y con demandas que fluctúan rápidamente.

De esta forma, este trabajo busca aportar al conocimiento generado en los últimos años que vincula el ingreso del Machine Learning a procesos de producción y ventas en el comercio, con el objetivo de promover la eficiencia y eficacia en las compras realizadas a través de modelos alternativos que puedan mejorar la predicción de la demanda. Además, este estudio también contribuye a comprender cómo funciona este sector de la economía y las diferentes formas en las que este campo del conocimiento puede aportar en él. Finalmente, también presenta una herramienta que puede ser replicada en diferentes sectores y empresas del departamento y del país, abriendo posibilidades a que sus procesos mejoren.

VI. OBJETIVOS

A. Objetivo general

Analizar el comportamiento del uso de técnicas y modelos basados en Machine Learning para el pronóstico de la demanda de productos en la empresa Dyna & Cía. S.A.

B. Objetivos específicos

- Examinar el modelo tradicional actualmente utilizado por la empresa para la predicción de la demanda.
- Identificar modelos de Machine Learning útiles para el trabajo con series de tiempo y generación de pronósticos.
- Evaluar el rendimiento de modelos Machine Learning en términos de reducción de la diferencia entre lo pronosticado y lo real.
- Comparar los resultados y sintetizar conclusiones obtenidas, proporcionando recomendaciones prácticas y consideraciones para la generación de pronósticos de la demanda.

VII. MARCO TEORICO

El término Machine Learning ha estado presente en el ámbito de la informática desde hace muchos años, sin embargo, se convirtió en un tema más común o explorado desde los años 90. Es uno de los campos de la inteligencia artificial que se enfoca en desarrollar algoritmos y modelos que se alimentan de los datos. En este se resalta una gran característica mencionada por Hinestroza [5], que se refiere al poder reconocer patrones dentro de los datos, brindado la posibilidad de adelantarse a una respuesta o predecir eventos futuros. Siendo esto un punto muy importante dentro de la planeación y la toma de decisiones en diferentes áreas, como menciona algunas Montgomery *et al.* [6] : administración de operaciones, mercadeo, finanzas y gestión del riesgo, economía, control de proceso industriales hasta en análisis demográficos.

Para el presente trabajo es importante definir el concepto de predicción o pronóstico que Contreras *et al.* [7] lo define como una herramienta capaz de entregar un valor estimado dada la probabilidad de eventos futuros. En cuanto al presente trabajo, el pronóstico será proporcionado a partir del histórico de pedidos o de demanda que han tenido los productos, es decir, de una secuencia de observaciones dadas en ciertos momentos se obtendrá un valor que podría darse en el futuro, a esta secuencia de datos ordenados cronológicamente la llamaremos series de tiempo.

Se entiende fundamental para el análisis de series de tiempo, el separar y comprender los diferentes componentes para obtener una visión más precisa y útil de los datos. En efecto, se tienen 4 componentes principales, es decir que podemos desagregar el comportamiento de una serie de tiempo en: tendencia, ciclo, estacionalidad e irregularidad como lo definen Heizer y Render [8].

Como menciona Carvajal [9], la tendencia en una serie de tiempo muestra la dirección que están presentando los datos a medida que pasa el tiempo, esta tendencia puede ser ascendente, descendente o plana, que significa que el comportamiento no tiene cambios. Por otro lado, se tiene el componente estacional y el cíclico que “se refiere a procesos que ocurren de manera más espaciada dentro del intervalo de observación, mientras que las componentes estacionales ocurren con menor separación en el tiempo” [10]. Finalmente, el componente irregular es importante

tenerlo en cuenta al analizar una serie de tiempo ya que representa la variabilidad no identificable y que no es sistemática, por lo tanto, no se debe utilizar para predecir patrones.

Si bien, no es sencillo desagregar y reconocer este tipo de patrones que se presentan en una serie de tiempo, existen hoy día varios modelos que han ayudado a tratar con los comportamientos antes mencionados para posteriormente generar predicciones, como lo son los modelos de Regresión Lineal, suavización exponencial y el promedio móvil que es el usado actualmente por la compañía Dyna y Cía. S.A., modelo que se basa en sacar una media de un conjunto de pedidos dados en un periodo de tiempo para estimar un valor futuro,

Seguido a los métodos tradicionales, se encuentran los modelos que han aparecido dentro las técnicas del Machine Learning, que vienen mostrando mejoría en los pronósticos debido a la capacidad de cómputo creciente de las máquinas y la posibilidad de realizar combinaciones e iteraciones de una forma mucho más veloz e ir aumentando la cantidad de opciones con el fin de poder realizar una comparativa que lleven a elegir un mejor modelo, como bien lo menciona en su artículo Fierro *et al.* [11].

Dentro los modelos basados en Machine Learning más comunes o usados, se encuentran los modelos ARIMA, modelos de Redes Neuronales, modelos de Regresión, y modelos Random Forest, modelos que a continuación se describen y menciona su utilidad:

A. Arima

El modelo ARIMA, es un modelo estadístico que combina las componentes o características de los modelos de autorregresión (AR), los modelos de media móvil (MA) [12] y la I que representa la necesidad de aplicar los componentes autorregresivos y los de media móvil. Los componentes anteriores se expresan como p, d, q, parámetros que tienen el objeto de ser iterados hasta tener el mejor rendimiento.

B. Redes Neuronales

Los modelos de Redes Neuronales son modelos que se retroalimenta de los errores para ir configurando su parametrización con el objetivo de dar un mejor resultado. Las Redes Neuronales son poderosas para el pronóstico de series de tiempo debido a su capacidad para aprender patrones complejos en los datos. Este modelo está inspirado en la estructura y funcionamiento del cerebro humano [13] como su nombre lo refiere, y que se puede ver representado en la **Fig. 4**.

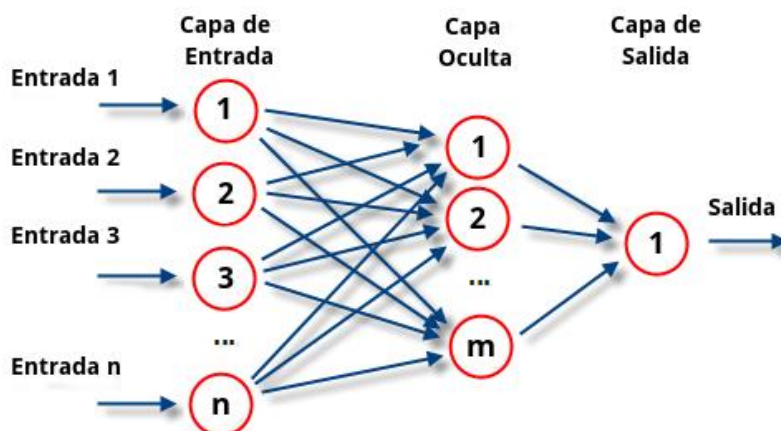


Fig. 4 Representación de un modelo de Redes Neuronales

Fuente:

<https://www.atriainnovation.com/que-son-las-redes-neuronales-y-sus-funciones/#:~:text=Las%20redes%20neuronales%20artificiales%20son,entrada%20hasta%20generar%20una%20salida.>

C. Regresión

Los modelos de Regresión son útiles cuando existe algún tipo de relación entre la variable o variables de entrada y la variable de salida, estas relaciones pueden ser lineales o no lineales lo cual ha generado muchas posibilidades y formas de ajuste para los datos. Algunos ejemplos de regresión son la Regresión Lineal, la Regresión Logística, la Regresión Polinomial, la Regresión Ridge, la Regresión Lasso, entre otras. “Estos modelos intentan explicar el comportamiento de una variable dependiente en relación con otras variables independientes” [14].

D. Random Forest

El Random Forest se compone de un conjunto de árboles de decisión donde cada árbol se entrena con un subconjunto diferente y variables de entrada diferentes, es decir, cada árbol tendrá una salida diferente y al final se combinan de forma ponderada para obtener la predicción final del modelo. En el caso específico de generación pronósticos, “el Random Forest tiene la capacidad de brindar valores mínimos y máximos de cada predicción cómo se muestra en la **Fig. 5**, con el objetivo de definir el intervalo más probable” [15].

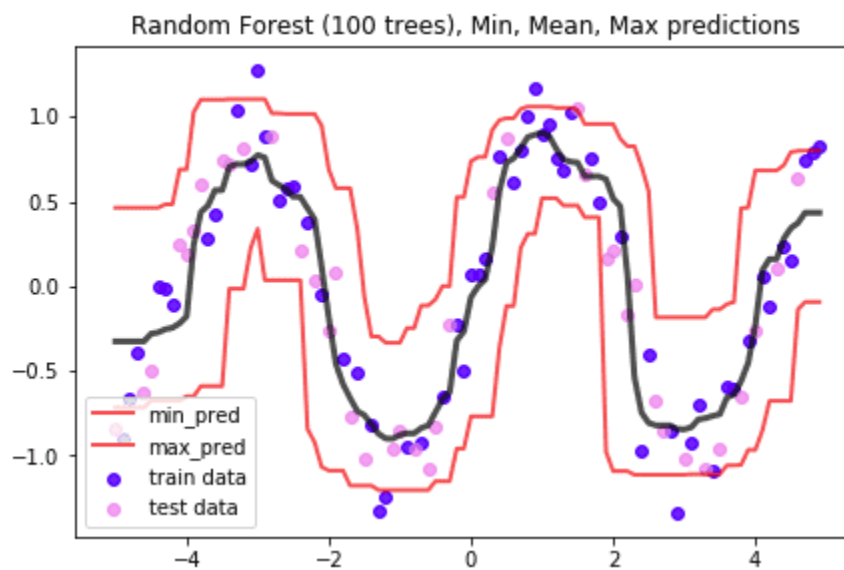


Fig. 5 Representación Random Forest

Fuente: <https://www.iartificial.net/random-forest-bosque-aleatorio/>

VIII. METODOLOGÍA

Dentro del campo de la Inteligencia Artificial IA, existen bien sea tipos de inteligencia o formas diferentes de abordar problemas, en las cuáles podría dividirse la IA según su aplicación. Por ejemplo, IA basada en reglas, IA de aprendizaje profundo conocida como Deep Learning o IA de aprendizaje automático o Machine Learning, que será el enfoque en el cual se desarrolla el presente trabajo debido a la capacidad de mejorar el rendimiento de tareas específicas como las predicciones, todo esto a través de los datos **Fuente especificada no válida.**

En este capítulo, se describe entonces la metodología utilizada para abordar la construcción de modelos basados en el aprendizaje automático, modelos que buscaran la predecir la demanda de productos en la empresa Dyna & Cía. S.A. En el mismo sentido, se detallan los pasos llevados a cabo como se muestra en la **Fig. 6**, desde la recopilación y preparación de datos hasta la selección y evaluación de los modelos. Además, se presenta la forma en que se realizaron las pruebas y se obtuvieron los resultados para la toma de decisiones.

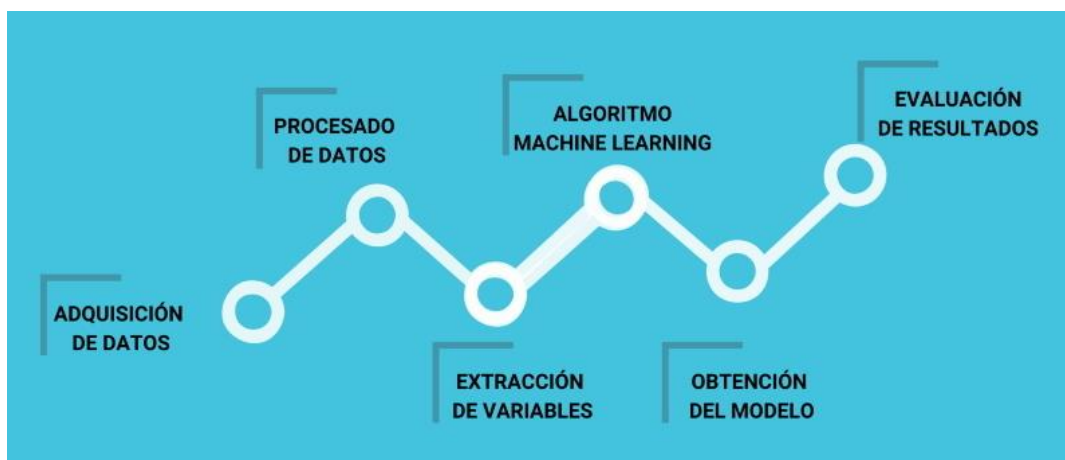


Fig. 6 Metodología Machine Learning

Fuente: <https://www.mastermarketing-valencia.com/marketing-digital/blog/machine-learning-ecommerce/>

A. Adquisición de Datos

En esa etapa, se llevará a cabo una identificación de las fuentes relevantes de información y lo más importante, que estuvieran disponibles para el estudio, con el fin de convertirse en el insumo de los modelos para la predicción de la demanda de productos en la empresa Dyna. Esto incluye el histórico de pedidos, maestro de productos, y unidades de empaque.

Se procederá luego a la extracción de información proveniente de las diferentes fuentes identificadas, realizando consultas a través del ERP de la compañía, prestando especial atención a garantizar la uniformidad de la información en todas ellas, ya que no se obtuvo una conexión directa a la base de datos de la empresa. Las consultas se descargarán en archivos con formatos csv, formato utilizado para almacenar y transmitir datos tabulares de manera sencilla y legible, donde los valores en están separados por comas.

La información de los pedidos se obtendrá a través de consultas en el ERP, ya que no se pudo obtener una conexión directa a la base de datos de la empresa. A través de esta forma se logra conseguir un histórico de pedidos desde el mes de enero de 2019 al mes de marzo del 2023, arrojando como resultado archivos en formato csv por mes.

B. Procesado de Datos

Utilizando el lenguaje de programación Python y con la ayuda de la librería Pandas, utilizada para el análisis y la manipulación de datos, se buscará consolidar un conjunto de datos que contenga el histórico de pedidos desde el año 2019. Y que además se unirán características relevantes como las unidades de empaque, el origen, estado y demás características relevantes en el estudio y que no se encontraban dentro la información suministrada en los pedidos.

Luego de la consolidación de la información, se realizará el preprocesamiento de los datos para garantizar que se encuentren en un formato adecuado para el análisis y entrenamiento de los modelos de Machine Learning. Esto contiene la transformación del tipo de variable, como la

diferenciación de los campos cuantitativos, cualitativos y de fecha. También incluye la eliminación de espacios en los datos y el tratamiento a datos nulos.

C. Extracción y creación de variables significativas

Una vez se tienen los datos en el formato adecuado para la ingesta de los modelos, se tiene como paso final para la construcción del conjunto de datos, la escogencia de las variables y generación de nuevos campos necesarios para el estudio, donde se justifica las decisiones tomadas, para luego sí, crear en un conjunto de datos que sea la base o insumo para los modelos. Como paso necesario para la evaluación de los modelos, se debe dividir el conjunto de datos en grupos de entrenamiento y prueba, dejando los meses de octubre de 2020 a marzo de 2023, como los 6 meses para realizar las pruebas, y todos los meses anteriores como insumo para el entrenamiento del modelo. El conjunto de entrenamiento se utilizará entonces para la construcción de los modelos de Machine Learning y el conjunto de prueba se usará para comparar el rendimiento final de los distintos modelos propuestos. Cabe resaltar, que la proporción de división se hace teniendo en cuenta que al trabajar con series de tiempo es de suma importancia conservar el orden y cronología en que se registraron los datos.

D. Algoritmo Machine Learning

En esta etapa, se trabajará con los algoritmos encontrados en medio de la investigación, como los más utilizados para abordar el tema de predicción de la demanda y que, además, estuvieran en función de las características de los datos. Estos modelos son ofrecidos dentro de las librerías de Scikit-Learn y Tensorflow permitiendo una gran facilidad en su implementación dentro del estudio presente.

Los modelos para aplicar serán ARIMA, Redes Neuronales, Regresión y Random Forest. Cada modelo tiene sus fortalezas específicas, por ejemplo, el modelo ARIMA es bueno para pronosticar series de tiempo cuando hay una relación entre los valores pasados y presentes de la serie ya que captura tendencias y los patrones repetitivos en los datos. Por otro lado, las Redes Neuronales son poderosas para el pronóstico debido a su capacidad para trabajar con interacciones

complicadas entre variables. Los modelos de Regresión, Puede ser útil cuando hay una fuerte relación entre la variable que se pronostica y las variables explicativas. Y, por último, está el modelo Random Forest que es efectivo en los pronósticos porque puede manejar variables de diferentes tipos y capturar interacciones no lineales.

Los modelos propuestos deberán mostrar su capacidad en la predicción de la demanda de tres productos estratégicos e importantes para la compañía Dyna & Cía. S.A., exhibiendo el poder para capturar patrones, relaciones complejas, dependencias de largo plazo y comportamiento no lineal en los datos, estos ajustes se podrán observar a través de la visualización generando gráficas con la ayuda de la librería Matplotlib.

E. Evaluación y obtención de los modelos

Para la evaluación de los modelos se empleará la métrica MAE (error medio absoluto) que puede implementarse con Scikit-Learn. Se destaca la importancia de utilizar una medida el error medio absoluto, pues esta métrica elimina la compensación que se da por diferencias positivas y negativas, proporcionando una medida parecida a la distancia que nos da muestra de que tan alejado está el pronóstico de lo real. Esta medida se analizará como las cantidades totales en que se desvía el pronóstico de la realidad.

Con el objetivo de asegurar una predicción más precisa de la demanda de productos, se realizará iteraciones de los diferentes hiperparámetros a través de ciclos y herramientas como GridSearch, que ayudan a optimizar cada modelo con miras a disminuir el error medio absoluto. Finalmente, se analiza y compara las métricas arrojadas por cada uno según su precisión frente a la predicción de la demanda.

IX. RESULTADOS

En primer lugar, se identificó que la fuente de datos más importante era la información de los pedidos realizados por los clientes a la empresa Dyna & Cía. S.A, los cuáles son tomados por los vendedores, personal del Contac Center o por la página web. Los pedidos contaban con más de 30 variables entre numéricas y categóricas, que contenían en su mayoría información relevante logísticamente referente a la ubicación y datos del cliente, pero no tan relevantes en términos de generación de pronósticos. Se procede entonces con la elección de las más importantes dentro del objetivo y las cuales se detallan en la

TABLA I:

TABLA I
ATRIBUTOS

Atributo	Descripción
Fecha	Hace referencia al día, mes y año en que se realizó el pedido.
Estado movto	Muestra el estado en que se encuentra el pedido como cumplido, aprobado, anulado
Item	Número de identificación interno de un producto.
Desc item	Descripción y características generales del producto.
Clasificacion venta pesos	Clasificación A, B, C, D, E según su participación en las ventas.
Clasificación rotacion unidades	Clasificación A, B, C, D, E según la cantidad vendida.
U.M.	Unidad de medida en que se vendió el producto.
Cant pedida	Hace referencia a las unidades solicitadas del cliente.
Valor bruto	Valor en pesos del pedido sin incluir impuestos.

Para el procesamiento y manejo de los más de 6 millones de registros que se encontraron del histórico de pedidos por ítem, fue útil la ayuda de Pandas, que en primera medida permitió dar una mirada inicial a la información contenida y quedarnos únicamente con las variables más relevantes mencionadas anteriormente.

Como tratamiento a las variables seleccionadas, se intentó convertir el campo “Fecha” en una variable de tiempo, sin embargo, se detectó la presencia de espacios lo que imposibilitaba su conversión. En otras herramientas se hubiera dificultado su detección y también su corrección, sin

embargo, con la función *strip* se logra el borrar los espacios en tan solo una línea de código (*ver Anexo A*), para posteriormente realizar el cambio a un formato de serie de tiempo necesario para la generación de pronósticos. Además de los espacios encontrados en las fechas, se evidencio la existencia de datos nulos dentro del conjunto de datos, que, al ser un porcentaje menor, se trataron eliminando los registros donde hacían presencia.

Paralelo a los problemas con el campo “Fecha”, se encontraron inconvenientes en los campos “Cant. pedida” y “Valor Bruto” pues el uso de comas como separador de miles era confuso para el programa, además la presencia del signo pesos no permitía el uso de los datos en “Valor Bruto” como un valor cuantitativo. Debido a lo anterior, se realiza el reemplazo de estos caracteres para disponer del dato crudo y posteriormente darle formato como número entero (*ver Anexo B*).

Con el fin de complementar información cualitativa y cuantitativa dentro del análisis del trabajo, se utilizó la función *Merge* (*ver Anexo C*) para relacionar el factor de las unidades de empaque en que se realizaron los pedidos, agregar también información del origen del producto; además se liga el estado del ítem; y por último se adiciona el indicador de compras. Conjuntamente a la fusión de nuevos campos, se crea una nueva variable llamada “Cant. Total” que representa la cantidad pedida expresada en unidades. Todo lo anteriormente mencionado además del factor, no se encontraba en los archivos de pedidos, pero tomaron importancia en la toma de decisiones y por tanto fueron incluidos para la construcción del conjunto de datos principal mostrado en la **Fig. 7**.

	Fecha	Estado movto.	Item	Desc. ítem	CLASIFICACION VENTAS PESOS	CLASIFICACION ROTACION UNIDADES	U.M.	Cant. pedida	Valor bruto	factor	ORIGEN PROVEEDOR	PROVEEDOR CCIAL	Estado	COMPRAS
0	2019-01-23	Cumplido	12032	VALVULA METAL CHEQUE VER 1/2 797183331	0005 - E (100%)	0005 - E (100%)	B3	2	60924	3	0001 - NACIONAL	0071 - GRIVAL	Activo	0001 - SI
1	2019-01-28	Cumplido	12032	VALVULA METAL CHEQUE VER 1/2 797183331	0005 - E (100%)	0005 - E (100%)	B3	1	30462	3	0001 - NACIONAL	0071 - GRIVAL	Activo	0001 - SI
2	2019-01-08	Cumplido	12032	VALVULA METAL CHEQUE VER 1/2 797183331	0005 - E (100%)	0005 - E (100%)	B3	2	58992	3	0001 - NACIONAL	0071 - GRIVAL	Activo	0001 - SI
3	2019-01-23	Cumplido	12032	VALVULA METAL CHEQUE VER 1/2 797183331	0005 - E (100%)	0005 - E (100%)	B3	2	60924	3	0001 - NACIONAL	0071 - GRIVAL	Activo	0001 - SI
4	2019-01-09	Cumplido	12032	VALVULA METAL CHEQUE VER 1/2 797183331	0005 - E (100%)	0005 - E (100%)	B3	1	29496	3	0001 - NACIONAL	0071 - GRIVAL	Activo	0001 - SI

Fig. 7 Dataset Principal

Teniendo en cuenta el proceso y conducto que siguen los pedidos en la plataforma, solo se tuvo en cuenta el histórico de pedidos que estén en estado aprobado, cumplido o comprometido, pues los demás significaron pedidos que por diferentes motivos no se gestionaron, y por lo tanto no deben formar parte de la demanda (*ver Anexo D*).

Además del primer filtro de información mencionado, para el objeto del negocio resultaría más costo-eficiente estudiar dichos pronósticos en los productos que son importados, debido a que en productos de origen nacional se tiene la posibilidad de reaccionar más rápido a un error, es decir, no representan como tal una amenaza grande en Dyna. Igualmente, al filtro en el origen, solo se tuvo en cuenta los productos que siguen activos en el portafolio de Dyna y que además fueran artículos de reposición regular, diferente a promocionales o productos de temporada (In & Out) para los cuáles no tendría mucho impacto generar una predicción.

Ahora bien, para efectos prácticos del estudio presente y con la intención de poder mostrar de manera clara el comportamiento de los modelos de Machine Learning propuestos, se tomaron tres ítems que cumplieran los criterios antes filtrados y que significaran un alto impacto para la empresa, así que se opta por los siguientes:

TABLA II
PRODUCTOS SELECCIONADOS

Ítem	Descripción
17001	Cerradura Cilíndrica Para Alcoba Con Pomo Acero Inox Nuovo
34631	Cerradura Cilíndrica Para Alcoba Con Manija Metálica Acero Inox Cádiz
43387	Soldador Inversor Para Electrodo 200 Amperios IE 8200/6/170 BVM

Los productos descritos en la **TABLA II**, representan productos estratégicos para Dyna ya que los dos primeros cuentan con los mejores niveles de rotación en la empresa Dyna & Cía. S.A., donde figuran en los dos primeros puestos; y el tercero, un producto de alto valor comercial aportando en gran participación a las ventas de la compañía y que por consiguiente también se encuentra dentro del top 10, como se puede observar en la **Fig. 8**:

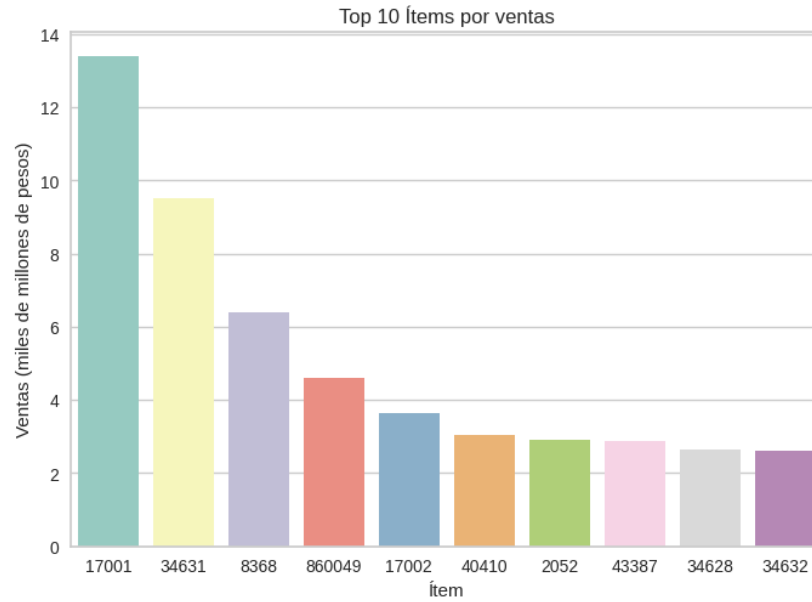


Fig. 8 Top 10 en Ventas

Se presenta a continuación los resultados obtenidos a partir de la implementación de los diferentes modelos en cada uno de los productos seleccionados anteriormente, con el objetivo de analizar el comportamiento del uso de técnicas basadas en Machine Learning enfocadas en el pronóstico de la demanda, donde se pronosticaron 6 meses a futuro en concordancia con el lead time que manejan los productos importados.

A. Ítem 43387

Para empezar, se presenta el gráfico en la **Fig. 9**, de los pedidos hechos por los clientes solicitando el soldador 43387, y aunque se tenga una base de datos desde enero del 2019, para este ítem específicamente hay registros desde el mes de septiembre de 2020 pues solo se lanzó hasta ese año. Hablando de temporalidad, es importante enunciar que los pedidos fueron agrupados por meses con el fin de establecer una frecuencia propia del análisis del negocio, teniendo entonces el mes/año asociado a la cantidad total demandada en dicho periodo. A continuación, cada punto se unió con el subsecuente para así poder evidenciar comportamientos a través del tiempo.

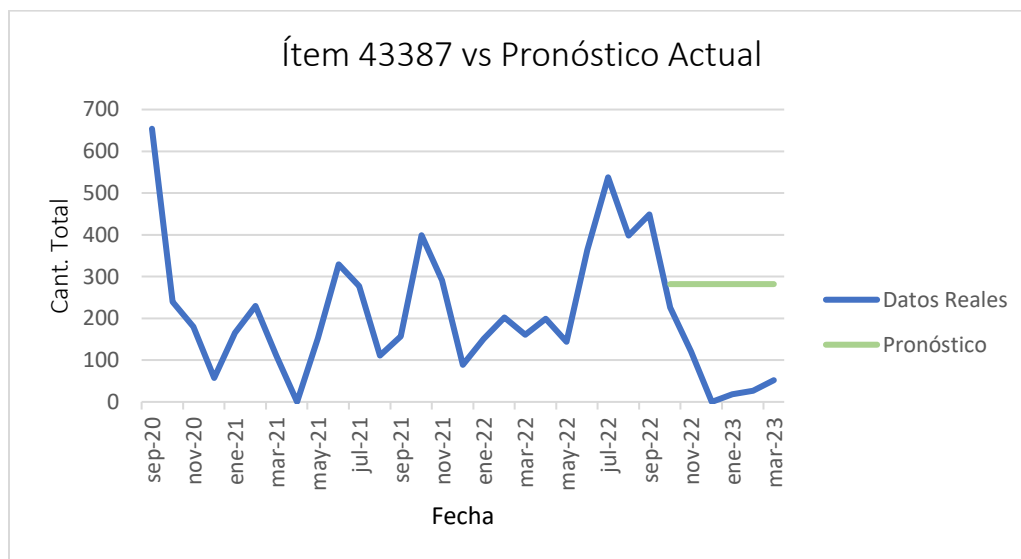


Fig. 9 Gráfico 43387 vs Pronóstico Actual

Para el ítem 43387 se observa un comportamiento volátil de la demanda a través de los meses, configurado con una línea de color azul y representando los datos reales de la demanda, percibiendo en primera instancia, una gran dificultad para observar una posible tendencia o algún comportamiento repetitivo en el tiempo, lo que implica que un promedio (línea en color verde) podría no ajustarse mucho al comportamiento. Se procede entonces a continuación a generar un pronóstico bajo el modelo ARIMA evidenciado en la **Fig. 10**.

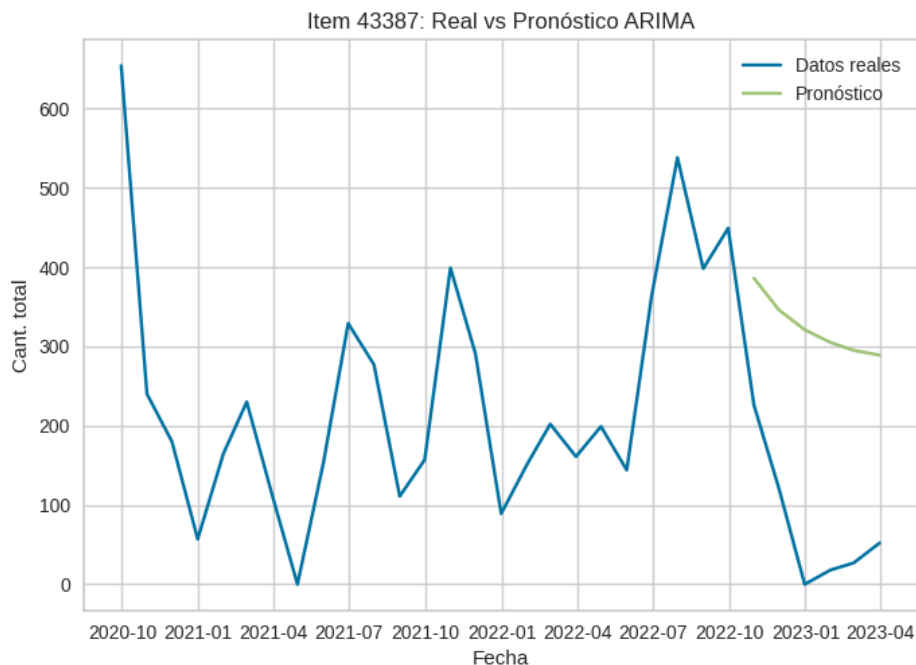
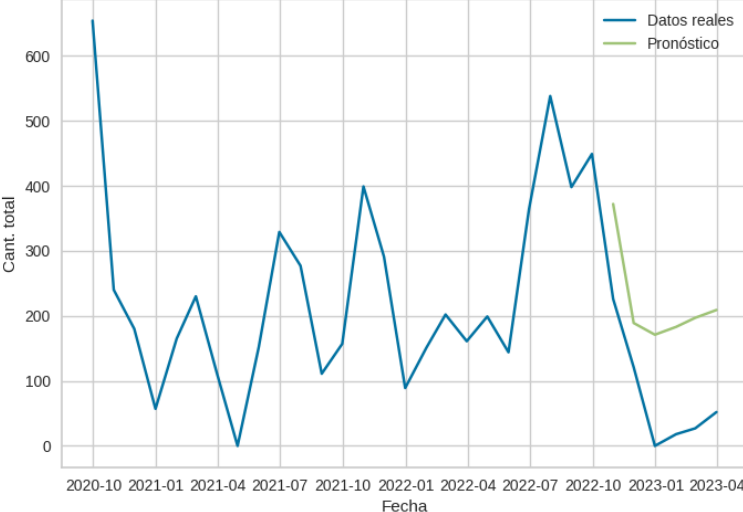
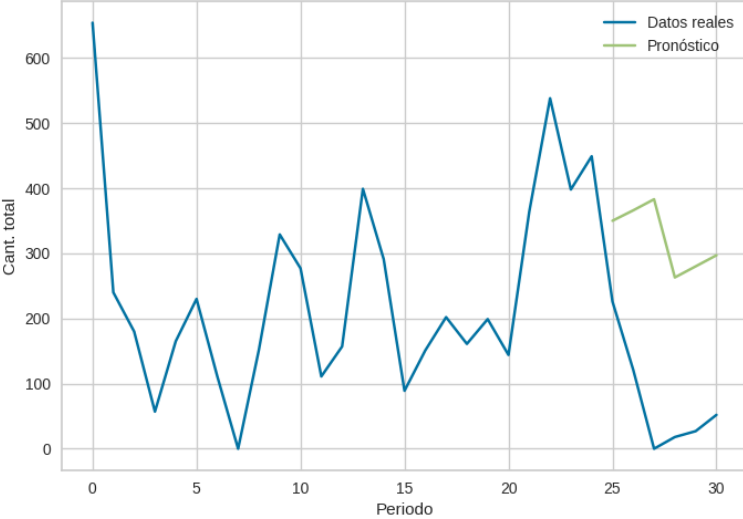


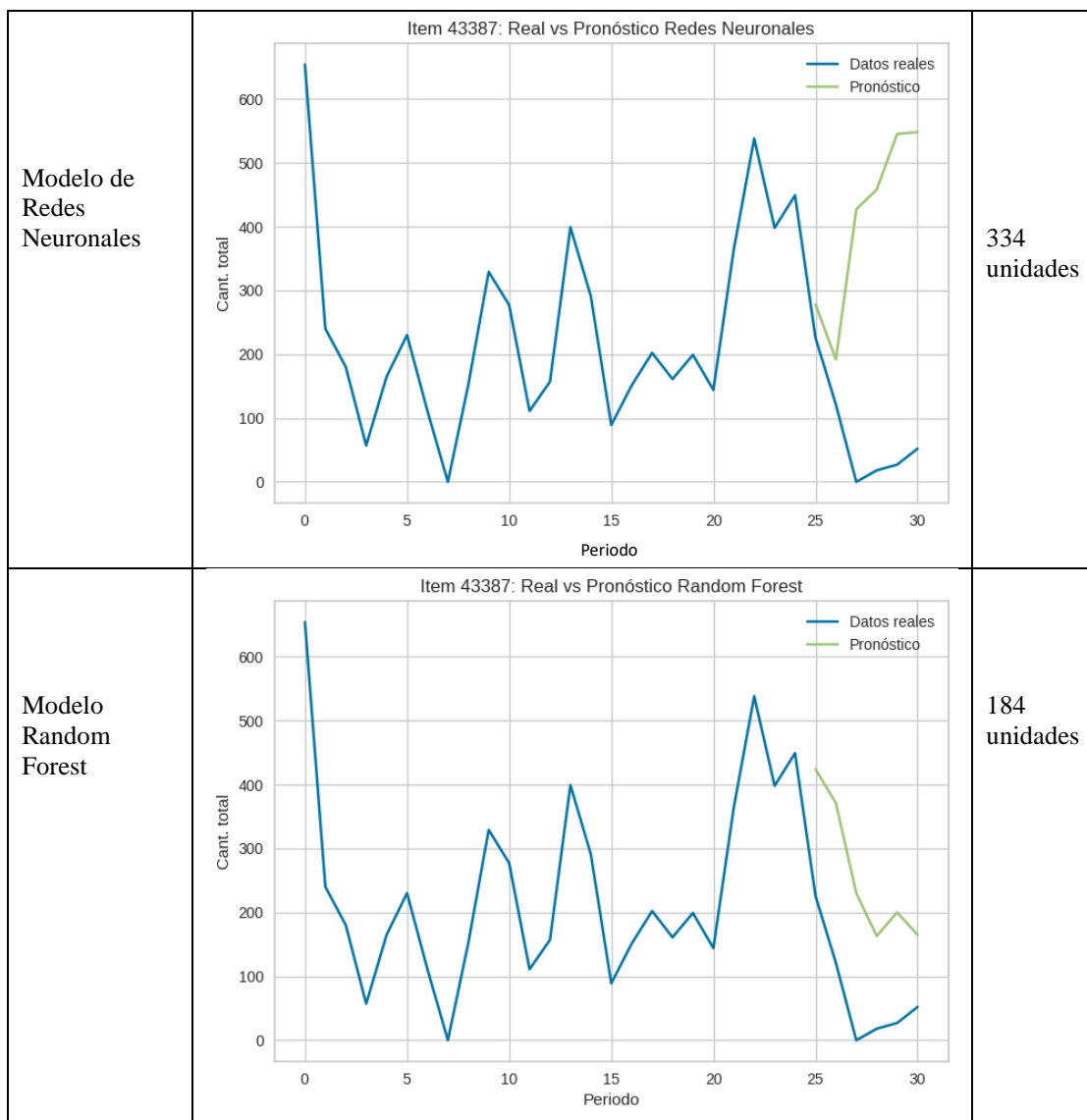
Fig. 10 Gráfico 43387 vs Pronóstico ARIMA

Aunque se nota una leve mejoría dentro de la predicción de la tendencia decreciente de los últimos meses, aún hay una diferencia muy grande entre los datos reales y el pronóstico que arroja el modelo. Con el fin de aminorar la distancia entre lo real y lo pronosticado, se realizaron iteraciones dentro de los modelos propuestos, a través de un ajuste de hiperparámetros (*ver Anexos E y F*) que permitieron disponer varias versiones de un mismo modelo y al final quedarnos con el óptimo. Seguido, se muestra los diferentes resultados obtenidos para el pronóstico de la demanda del soldador 43387 en la **TABLA III**:

Es pertinente recordar en este punto que los modelos son entrenados sin incluir los últimos 6 meses, ya que en dichos periodos es donde se requiere la predicción. es decir, que el modelo no tuvo conocimiento de estos datos, solo hasta el final, cuando los modelos fueron evaluados para su comparación contra estos datos reales.

TABLA III MODELOS ÍTEM 43387

MODELO	GRÁFICA	MAE
Modelo ARIMA	<p data-bbox="711 310 1027 331">Item 43387: Real vs Pronóstico ARIMA</p>  <p data-bbox="467 338 1206 846">The chart displays two data series: 'Datos reales' (blue line) and 'Pronóstico' (green line). The x-axis represents dates from 2020-10 to 2023-04, and the y-axis represents 'Cant. total' from 0 to 600. The real data shows significant volatility, with peaks near 600 and troughs near 0. The ARIMA forecast (green line) follows the general trend of the real data but shows a notable deviation starting around late 2022, where it predicts a sharp decline followed by a slight recovery, while the real data continues to fluctuate.</p>	146 unidades
Modelo de Regresión	<p data-bbox="711 867 1027 888">Item 43387: Real vs Pronóstico Regresion</p>  <p data-bbox="467 894 1206 1409">The chart displays two data series: 'Datos reales' (blue line) and 'Pronóstico' (green line). The x-axis represents 'Periodo' from 0 to 30, and the y-axis represents 'Cant. total' from 0 to 600. The real data shows a similar pattern of volatility as in the ARIMA model. The regression forecast (green line) follows the real data closely until period 25, after which it diverges significantly, predicting a much higher value (around 350-400) compared to the real data's low values (around 0-50).</p>	249 unidades

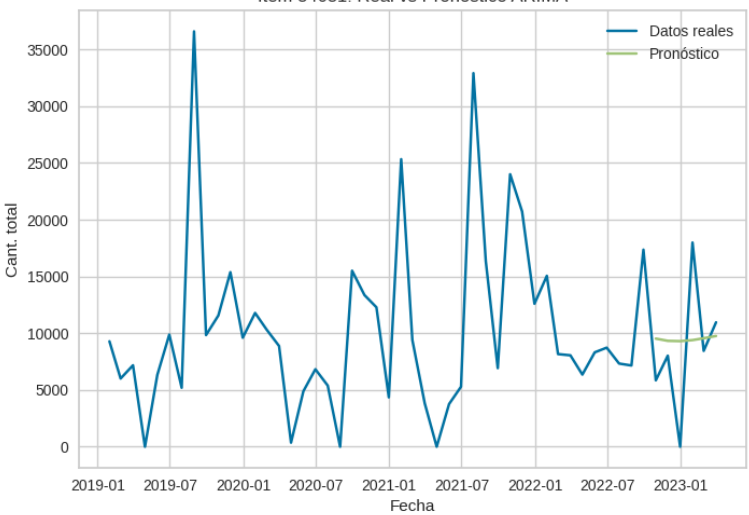
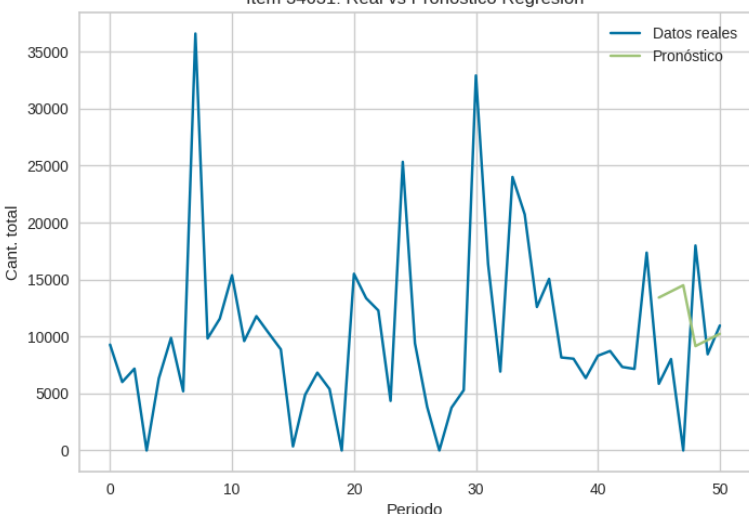


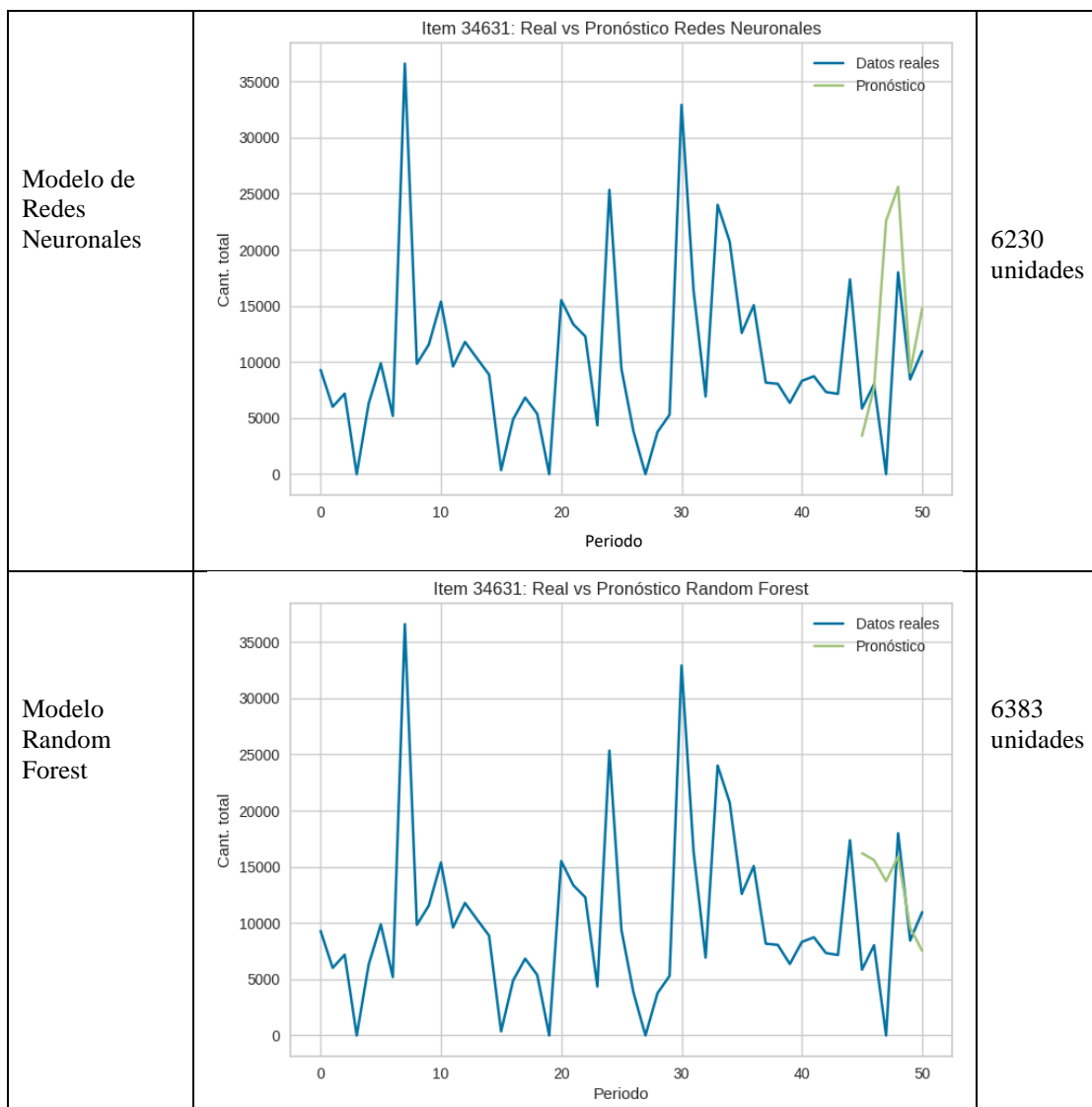
Gráficamente se observa que los modelos de Random Forest y ARIMA, tienen un buen rendimiento ya que se lograron ajustar de mejor manera al comportamiento de los datos de la demanda del ítem 43387. Inclusive, desde la mirada del error también hay una mejora, porque el MAE bajo el método actual que se utiliza en la empresa, es de 208, lo que quiere decir que, en los 6 meses pronosticados, dichas predicciones se alejaron en promedio 208 unidades comparados con la demanda real. Por el contrario, el modelo ARIMA obtuvo un MAE de 146 unidades que es el que obtuvo la mejor métrica.

B. Ítem 34631

A continuación, se presentan los resultados obtenidos para la cerradura 34631 que presenta varios datos atípicos evidenciados en los picos que tuvo la demanda y descubiertos gráficamente. Además, este producto tiene la curiosidad de tener una demanda de cero unidades en varios meses del rango en estudio (2019-2023) lo que se puede observar en la **TABLA IV**.

TABLA IV MODELOS ÍTEM 34631

MODELO	GRÁFICA	MAE
Modelo ARIMA		4200 unidades
Modelo de Regresión		6469 unidades



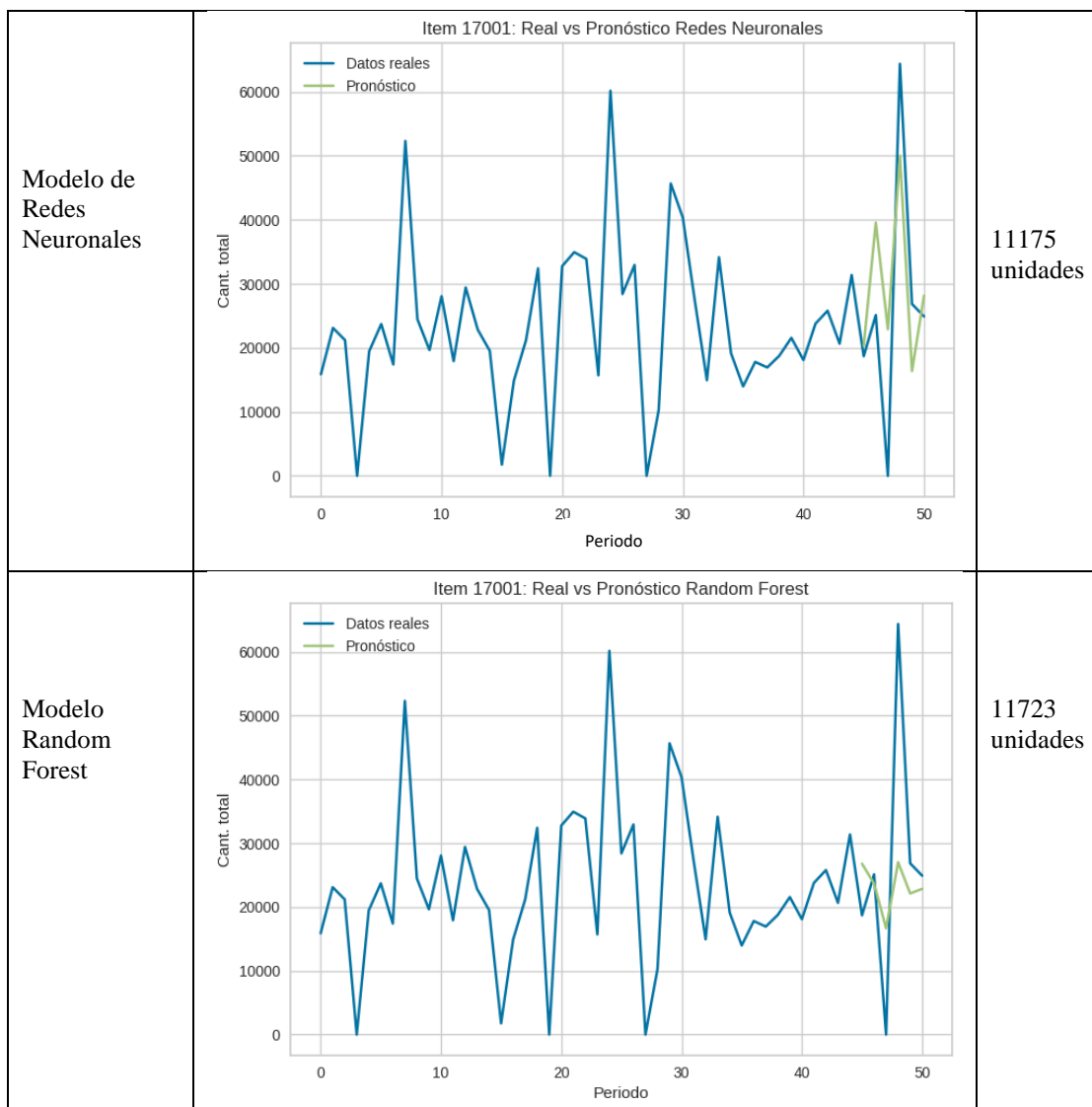
Aunque el ARIMA tiene un error mucho menor al de los otros modelos, gráficamente se puede observar que este modelo está suavizando demasiado la información, Si bien un sobreajuste es malo para un modelo, tampoco nos interesa un modelo muy general que no pueda adaptarse a cambios en la demanda. En ese sentido, el modelo de Redes Neuronales presenta un buen acercamiento a lo que muestra la demanda realmente con un MAE de 6230 unidades. Se debe hacer la salvedad que los MAE no son comparables con los de otros ítems, una muestra de ello son los dominios que tiene cada uno, donde la demanda promedio del soldador 43387 es de 209 unidades (*ver Anexo G*) y la media de la cerradura 34631 es de 10192 unidades.

C. Ítem 17001

Por último, se realiza el mismo ejercicio con la cerradura 17001, la cual representa el producto más vendido en la empresa (**TABLA V**).

TABLA V MODELOS ÍTEM 17001

MODELO	GRÁFICA	MAE
Modelo ARIMA		12167 unidades
Modelo de Regresión		12673 unidades



Las métricas comparadas en la tabla anterior no muestran diferencias muy grandes entre cada modelo propuesto, no obstante, el pronóstico arrojado por la red neuronal es muy eficiente pareciéndose al comportamiento real de la demanda en esos últimos 6 meses. De igual manera, es valido mencionar que todos los modelos aplicados a esta serie de tiempo tuvieron un menor MAE al del método actual con 13145 unidades.

X. DISCUSIÓN

Los modelos de Regresión no fueron muy acertados en la predicción de las series de tiempo presentadas, ya que como tal el ejercicio no comprendía una relación lineal entre la variable que se estaba pronosticando (demanda). Este modelo podría haber presentado dificultades porque el campo “Fecha” no era como tal una variable con relación en la demanda, sin embargo, se convirtieron en variables “Mes” y variable “Año” para evaluar su rendimiento.

Así como el modelo Regresión, el Random Forest también presenta un limitante parecido en problema de serie de tiempo tratado, y es que el modelo espera que a partir de unas variables de entrada se haga la predicción de una variable de salida, sin embargo, el ejercicio se pudo abordar haciendo el supuesto que el “Año” y el “Mes” eran variables que explicaban la demanda. No obstante, no es tan claro que “Año” y “Mes” fueran causantes de la demanda, como sí lo son, por ejemplo, la temperatura y el viento en un ejercicio de predicción de la lluvia.

El modelo ARIMA pudo predecir muy bien los primeros pronósticos, hablando más concretamente de la predicción de los dos primeros periodos, donde reflejó un error menor que en los demás modelos. Sin embargo, presentó el limitante de obviar patrones más complejos como los cíclicos o estacionales, y es ahí cuando se le dificulta realizar predicciones más alejadas ya que supone una misma importancia para todas las observaciones.

Por otro lado, Las Redes Neuronales fueron útiles para el pronóstico de series de tiempo debido a su capacidad para aprender patrones que no son sencillos. Además, cuenta con una gran característica muy aplicable al problema en estudio, pues supone una relación entre el t y el $t+1$, es decir, que el tiempo futuro tendrá relación a lo que pase en el tiempo presente, y éste a su vez será consecuencia del pasado $t-1$. A diferencia de los modelos ARIMA y Random Forest, éste no prestaba importancia mayor a las fechas sino al histórico de demanda.

XI. CONCLUSIONES

- Sí bien la forma actual usada por la empresa Dyna, ha sido de utilidad para poder tener un método que ayude a tener idea de la demanda futura, desde una mirada más exigente de lo que significa un pronóstico, puede quedarse muy corto pues un promedio supone un comportamiento lineal que no concuerda mucho con el comportamiento de la demanda de productos en Dyna.
- Es imprescindible entender que cada uno de los productos presentan un comportamiento de la demanda diferente, así mismo el método para pronosticar no puede ser el mismo para todos. Sin embargo, es importante mencionar que la utilización de un modelo Machine Learning no se convierte en una regla para realizar con todos los productos, pues su aplicación presenta una complejidad y un esfuerzo mayor al método usado en la actualidad.
- Según los resultados obtenidos, es evidente que los modelos de aprendizaje automático presentan un mejor ajuste vs el método usado por la empresa, traducido esto en mejores pronósticos para la demanda. Además, la utilización de Machine Learning abre un panorama más amplio de posibilidades, pues es relativamente sencillo aplicar distintos modelos cuando ya se cuenta con un conjunto de datos útil.
- Para los ejercicios en práctica, se tomaba como insumo principal la información del histórico de demanda, por ello resultaron mejor los modelos ARIMA y Redes Neuronales ya que los modelos de Regresión y Random Forest no son tan eficaces con problemas univariados. No obstante, todos presentaron un mejor ajuste y una disminución del error del 23% en promedio, comparándolos con los MAE del método actual.
- Para el negocio, la métrica del MAE fue muy acertada en términos de poder entender muy bien la problemática, pues las unidades son el lenguaje que se habla en términos de inventario o en la cantidad de ventas que se pueden perder por falta de disponibilidad. Además, esta simpleza permite comunicar y explicar de mejor manera los resultados obtenidos a personas no tan inmersas en el Machine Learning.

XII. RECOMENDACIONES

- Modelos como Random Forest y Redes Neuronales, tienen la capacidad de trabajar problemas más robustos y con muchas más variables, así que sería bueno explorar agregando otras variables al estudio, por ejemplo, periodos electorales en el país que por experiencia se sabe aumentan la demanda de varios productos. Además, sería interesante ver la relación de la demanda ante temporadas de lluvia o verano pues supondría cambios en la demanda de ciertos productos.
- Los modelos Machine Learning necesitan ser entrenados con muchos datos, así que el ingreso constante de nuevos datos al modelo podría mejorar mucho el rendimiento de éstos. En ese sentido, la implementación de una conexión directa a través de un proceso ETL (extraer, transformar, cargar), sería muy útil para la integración, preparación de datos y uso en los modelos.
- Las Redes Neuronales utilizan una inicialización aleatoria de los pesos, desactivación aleatoria de neuronas e introducción de ruido durante el entrenamiento. Estas técnicas pueden ingresar variabilidad en las predicciones, lo que lleva a diferentes resultados en cada ejecución. Es importante entonces tener en cuenta esta información para que el ajuste de los parámetros propenda además de reducir el error, buscar un modelo que no presente cambios drásticos en los resultados, pues no tendría sentido la elaboración de un pronóstico de demanda.

REFERENCIAS

- [1] A. E. P. Macedo, «Tecnología Integrada,» 24 Octubre 2016. [En línea]. Available: <https://tecnologiaintegrada.com.mx/2016/10/24/la-era-la-informacion/>. [Último acceso: 15 Mayo 2023].
- [2] Dyna & Cía S.A, «Dyna,» 2020. [En línea]. Available: <https://dyna.com.co/>. [Último acceso: 31 Mayo 2023].
- [3] P. ARNDT, «Just in Time: El sistema de producción Justo a Tiempo,» 2005. [En línea]. Available: <https://www.grin.com/document/43360>. [Último acceso: 01 05 2023].
- [4] Banco de la República de Colombia, «Banco de la República,» 28 Abril 2023. [En línea]. Available: <https://www.banrep.gov.co/es/noticias/jdbrc-decidio-mayoria-incrementar-25-pb>. [Último acceso: 02 Junio 2023].
- [5] D. H. RAMÍREZ, *EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD*, Pereira, 2018.
- [6] D. C. MONTGOMERY, C. L. JENNINGS y M. KULAHCI, «Introduction to Time Series,» de *Introduction to Time Series Analysis and Forecasting*, New Jersey, WILEY-INTERSCIENCE, 2008, pp. 2-3.
- [7] A. Contreras, C. Atziry, J. L. Marínez y D. Sánchez, «Análisis de series de tiempo en el pronóstico de la demanda,» *ESTUDIOS GERENCIALES*, p. 10, 2016.
- [8] J. Heizer y B. Render, «Principios de administración de operaciones,» de *Administración de operaciones*, México, 2009, p. 110.
- [9] P. Carvajal, «ESTUDIO DEL PRONÓSTICO DE LA DEMANDA DE ENERGIA ELECTRICA, UTILIZANDO MODELOS DE SERIES DE TIEMPO,» Universidad Tecnológica de Pereira, Pereira, 2003.
- [10] E. Montes, F. Calvete y M. Cralos, «APLICACIÓN DE SERIES DE TIEMPO EN LA REALIZACIÓN DE PRONÓSTICOS DE PRODUCCIÓN.,» *Revista Fuentes: El Reventón Energético*, vol. 14, nº 1, p. 10, 2016.
- [11] C. Fierro, V. Castillo y c. Torres, «Análisis comparativo de modelos tradicionales y modernos para pronóstico de la demanda: enfoques y características,» *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, vol. 12, nº 24, pp. 13-14, 2022.
- [12] S. de la Fuente, «SERIES TEMPORALES, MODELO ARIMA METODOLOGÍA DE BOX - JENKINS,» Universidad Autónoma de Madrid, Madrid.
- [13] ATRIA INNVATION, «Qué son las redes neuronales y sus funciones,» 22 Octubre 2019. [En línea]. Available: <https://www.atriainnovation.com/que-son-las-redes-neuronales-y-sus-funciones/#:~:text=Las%20redes%20neuronales%20artificiales%20son,entrada%20hasta%20generar%20una%20salida.> [Último acceso: 08 Junio 2023].
- [14] B. Buitrago, «Machine Learning — Modelos de Regresión I,» 14 Septiembre 2020. [En línea]. Available: <https://medium.com/iwannabedatadriven/machine-learning-modelos-de-regresi%C3%B3n-i-d293ae235e9a>. [Último acceso: 08 Junio 2023].
- [15] J. M. Heras, «IArtificial,» 18 Septiembre 2020. [En línea]. Available: <https://www.iartificial.net/random-forest-bosque-aleatorio/>. [Último acceso: 07 Junio 2023].

ANEXOS

Anexo A:

```
1 #Quitar espacios en blanco de la columna Fecha
2 datos1['Fecha'] = datos1['Fecha'].str.strip()

1 # Convertir la columna "Fecha" al formato de fecha '%m/%d/%Y'
2 datos1['Fecha'] = pd.to_datetime(datos1['Fecha'], errors='coerce')
```

Anexo B:

```
1 #Corregir problemas de formato con los valores
2 datos1["Cant. pedida"] = [float(str(i).replace(",","")) for i in datos1["Cant. pedida"]]
3 datos1['Valor bruto'] = [float(str(i).replace(",","").replace("$","")) for i in datos1['Valor bruto']]

1 #Corregir el tipo de variable
2 datos1['Cant. pedida'] = datos1['Cant. pedida'].astype(int)
3 datos1['Valor bruto'] = datos1['Valor bruto'].astype(int)
```

Anexo C:

```
datos2 = pd.merge(datos1,UM,left_on=['Item', 'U.M.'], right_on=['codigo_producto', 'unidad'])
datos3 = datos2.merge(maestro, on="Item", how='inner')
```

Anexo D:

```
1 # Para el histórico solo se tendrá en cuenta los pedidos que se hayan gestionado realmente
2 condicion1 = datos1['Estado movto.'] == 'Anulado'
3 condicion2 = datos1['Estado movto.'] == 'En elaboración'
4 condicion3 = datos1['Estado movto.'] == 'Retenido'
5 condicion4 = datos1['Estado movto.'] == 'En elaboración'
6 datos1 = datos1.drop(datos1[condicion1 & condicion2 & condicion3 & condicion4].index)
```

Anexo E:

```

1 # Crea el modelo de Random Forest
2 model = RandomForestRegressor(random_state=42)

1 # Define los hiperparámetros a probar
2 param_grid = {
3 |   'n_estimators': range(1, 101, 3)
4 | }

1 # Utiliza GridSearchCV para buscar los mejores hiperparámetros
2 grid_search = GridSearchCV(estimator=model, param_grid=param_grid, scoring='neg_mean_absolute_error', cv=5)
3 grid_search.fit(X_train, y_train)

```

```

> GridSearchCV
> estimator: RandomForestRegressor
  > RandomForestRegressor

```

Anexo F:

```

1 mejor_mae = np.inf
2 mejores_parametros = None
3
4 for combinacion in combinaciones:
5
6     modelo = ARIMA(train_data, order=combinacion)
7     modelo_fit = modelo.fit()
8     prediccion = modelo_fit.predict(start=len(train_data), end=len(train_data) + len(test_data) - 1)
9     prediccion = prediccion.round().astype(int)
10    mae = mean_absolute_error(test_data, prediccion)
11    mape = mean_absolute_percentage_error(prediccion, test_data)
12
13    if mae < mejor_mae:
14        mejor_prediccion = prediccion
15        mejor_mae = mae
16        mejor_mape = mape
17        mejores_parametros = combinacion

```

Anexo G:

	Cant. total
count	31.000000
mean	209.419355
std	158.323461
min	0.000000
25%	111.000000
50%	165.000000
75%	284.000000
max	654.000000