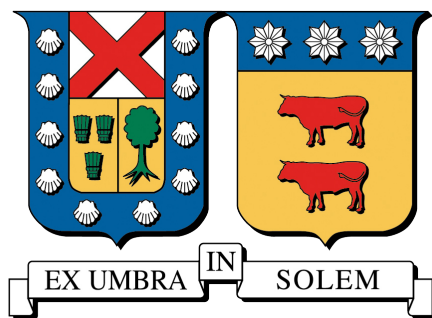




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

PRONÓSTICO DE DEMANDA MULTIVARIABLE USANDO REDES
NEURONALES EN INDUSTRIAS CON USO INTENSIVO DE ACTIVOS.

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

AUTOR

VÍCTOR DANIEL ÁLVAREZ BARRÍA

PROFESOR GUÍA : DAVID GODOY RAMOS

PROFESOR CORREFERENTE : GONZALO AMÉSTICA HERNÁNDEZ

SANTIAGO, 09 DE DICIEMBRE, 2022

Citas y Agradecimientos

*A mi familia,
por su paciencia y sabiduría inconmensurable que guió mi camino durante estos años
voraces.*

*A mi madre y hermana,
por ser los pilares fundamentales de mi vida. Sin su amor y apoyo jamás hubiera logrado
todo esto.*

*A todas esas personas que,
de mil formas distintas y únicas, aportaron en mi formación profesional y personal.*

“But remember this,... airplanes are not tools for war... airplanes are beautiful dreams.

Engineers turn dreams into reality.”

(H. Miyazaki, 2013)

“... Si puedes soñar sin que los sueños te dominen;

Si puedes pensar y no hacer de tus pensamientos tu único objetivo;

Si puedes encontrarte con el Triunfo y el Desastre, y tratar a esos dos impostores de la
misma manera...”

(R. Kipling, 1895)

Resumen

En las industrias con uso intensivo de activos la planificación de la demanda es clave. En la industria minera es importante dar continuidad operacional a través de pronósticos precisos de componentes críticos. Por otro lado, generar pronósticos de demanda precisos también es crucial en empresas e-commerce con grandes sistemas logísticos, pues es indispensable para la gestión de las operaciones y para la misma supervivencia del negocio.

Los pronósticos de demanda generalmente se basan en modelos univariantes usando solamente los datos históricos de ventas, lo cual los hace susceptibles a la incertidumbre propia de escenarios dinámicos en donde existen múltiples variables que interactúan con la demanda. Los modelos más robustos que se están desarrollando en la actualidad integran Data Science. Se estima que disminuyen los errores de pronóstico en un 30-50 %.

Dadas las posibles correlaciones entre los patrones de demanda de los productos o componentes, y también de las correlaciones entre estos mismos patrones de demanda y variables externas relacionadas, como festividades, feriados, datos de sensores, tasas de falla, entre otras variables relevantes, es que se genera una solución por medio del desarrollo de un modelo integrador usando redes neuronales Long Short-Term Memory (Deep Learning).

El objetivo principal es desarrollar modelos multivariantes de pronósticos de demanda para industrias con uso intensivo de activos a través del entrenamiento de redes neuronales con el fin de disminuir errores en escenarios de gran incertidumbre.

Se evaluaron 4 tipos de modelos en base a si entregaban el pronóstico de uno (uni-objetivo) o varios productos (multi-objetivo) y si eran desarrollados para pronosticar SKUs similares (por grupos) o no (por lotes), demostrando que los modelos UGR (uni-objetivos y por grupos) son los que alcanzan un menor error. Además, todos los modelos logran un menor error agregado por producto que los modelos ARIMA + GARCH.

A partir del uso de la metodología de Data Science de IBM y las estrategias propuestas de agrupamiento de variables e identificación de productos prioritarios se desarrollaron modelos multivariantes de pronósticos de demanda para estas industrias usando LSTM que

disminuyeron los errores en escenarios inciertos.

Se espera que en futuras investigaciones se ahonde en los modelos MGR (multi-objetivo y por grupos) en cuanto a una mejora en el agrupamiento de series de tiempo, como también contrastar diferentes tipos de variables externas (continuas o dummy) y su impacto en los modelos. Por último, el desarrollo de modelos usando LSTM y ARIMA+GARCH abriría varias áreas de futuras investigaciones.

Índice

Introducción	1
Problema De Investigación	2
Objetivos	4
Objetivo General	4
Objetivos Específicos	4
Marco Teórico	4
Pronósticos de Demanda en Gestión de Operaciones	4
Métodos de Pronóstico	6
Uso de Redes Neuronales en Pronósticos de Demanda	9
Principales Modelos	12
Variables, su Naturaleza y Agrupamiento	16
Pre-procesamiento de los datos	18
Validación de los Modelos de Redes Neuronales	19
Metodología	23
Metodología	23
Alcances y limitaciones	26
Herramientas Usadas	27
Data Requirements	27
Data Collection	27
Data Understanding	28
Data Preparation	30
Elección de SKUs prioritarios	30
Variables Externas	39
Estrategia de Agrupamiento	43

Base de Comparación	48
Modeling	51
Diseño de Estructuras de Redes Neuronales	51
Evaluation	55
Elección del Mejor Modelo	55
Resultados	57
Modelos ULO	57
Modelos UGR	57
Modelos MLO	60
Modelos MGR	62
Mejor Modelo	64
Conclusiones y Recomendaciones	67
Conclusiones Secundarias	69
Trabajos Futuros	70
Anexos	71

Introducción

Las empresas que usan activos intensivamente requieren de más capital y otros recursos financieros que otros factores de producción. Ejemplos clásicos de estas industrias son la minería, energía, telecomunicaciones, transporte, entre otros (Amendola, 2015). Sin embargo, en esta categoría también pueden incluirse empresas como Amazon o Walmart.com, los cuales poseen enormes centros de distribución y complejos sistemas logísticos destacados mundialmente (Bandara y col., 2019; Jeong y col., 2022; Voronova, 2022).

Pronosticar la demanda es el punto de partida para todas las actividades de planificación dentro de la cadena de suministros (Haberleitner y col., 2010), por lo que es de suma importancia en la gestión de operaciones. Esto se acentúa en industrias que usan activos intensivamente. Por ejemplo, en empresas e-commerce que cuentan con un inventario complejo de productos y un sistema logístico robusto con grandes centros de distribución y un alto grado de automatización, como Amazon.com o Walmart.com, la confiabilidad y precisión de los pronósticos de demanda a nivel de productos son factores habilitantes para la planificación de inventarios, tener precios competitivos, reducción de costos, cumplir con pedidos cada vez más específicos, alcanzar altos niveles de servicio y muchos otros temas importantes para la supervivencia del negocio dentro de este mercado dinámico y volátil (Bandara y col., 2019; Nguyen y col., 2021). Los pronósticos de demanda externa (venta de productos finales) también son importantes para los niveles estratégicos en gestión de operaciones, a partir de estos mismos se toman decisiones de capacidad y localización, por ejemplo con respecto a la ubicación, número, tamaño y expansión de capacidad de centros de distribución, estas decisiones deben tener en consideración factores como la proximidad a los clientes, sus hábitos de compra o variables externas que generen aumentos en la demanda que puedan sobrepasar a la capacidad planeada y requieran de una expansión futura, incluso se podrían tener en consideración los posibles envíos de productos provenientes de otros centros de distribución que sean gatillados por conexiones entre los diferentes canales de distribución o por requerimientos de los mismos clientes (C. Chen y col., 2021).

Los pronósticos son importantes pero no sólo para la predicción de la demanda externa cuando se habla de industrias con uso intensivo de activos. En la minería, por ejemplo, la gestión de componentes críticos es crucial para dar soporte a la continuidad operacional, no tener componentes en inventario para la reparación de maquinaria perjudica su disponibilidad afectando el rendimiento de la cadena productiva, lo cual tiene un fuerte impacto en los costos de inventario, mantenimiento y en los costos globales de operación (Chandriah & Naganahalli, 2021; Rosienkiewicz y col., 2017). Esta gestión de componentes críticos necesita políticas para la toma de decisiones de pedido con respecto a los mismos que se abastezcan de información proveniente de pronósticos robustos de la demanda interna (componentes para reparación de equipos, insumos, etc.), entre otras variables relevantes (Rosienkiewicz y col., 2017).

El resto de la presente memoria está organizada de la siguiente forma. Se expone el **Problema De Investigación** que surge como el resultado de la necesidad de negocio a resolver, luego se definen los **Objetivos** que se esperan alcanzar para dar solución a la problemática, después se realiza una revisión de la literatura existente relacionada al tema de la investigación en la sección **Marco Teórico**, en la sección siguiente se explica la **Metodología** seguida para alcanzar los objetivos propuestos, después se muestran los **Resultados** y las **conclusiones** obtenidas, para finalizar con una serie de **Trabajos Futuros** y líneas de investigación propuestas a seguir a partir del trabajo desarrollado en esta memoria.

Problema De Investigación

Como se mencionó anteriormente, los pronósticos de demanda, tanto interna como externa, son un factor clave dentro de las organizaciones del tipo capital-intensivo y su impacto se extiende a todos los niveles de la misma. Sin embargo, el desarrollo de modelos de pronóstico de carácter cuantitativo dentro de la industria en general se han enfocado en el uso y mejora de modelos derivados de suavizamientos exponenciales, de la familia ARIMA o de métodos de descomposición que comparten una similitud, son modelos univariantes o

de pocas variables, lo cual los hace muy susceptibles a errores producto de la incertidumbre propia de escenarios dinámicos y complejos en donde existen múltiples variables internas y externas que interactúan con la demanda (Abbasimehr y col., 2020; Bandara y col., 2019; Haberleitner y col., 2010; Moroff y col., 2021; Nguyen y col., 2021).

Los modelos más robustos que se están desarrollando en la actualidad integran Data Science para poder modelar una gran cantidad de variables usando una enorme cantidad de datos para el entrenamiento de redes neuronales, los tipos de neuronas que más sobresalen dentro de las investigaciones existentes son las del tipo LSTM (Long Short-Term Memory o Larga Memoria de Corto Plazo) por su capacidad de retener información a largo plazo y su flexibilidad estructural relacionada a los vínculos que permite establecer entre capas de neuronas no consecutivas (Abbasimehr y col., 2020; Bandara y col., 2019; Chandriah & Naraganahalli, 2021; Moroff y col., 2021), y han tenido muy buenos resultados. Sin embargo, se cree que estos modelos podrían mejorar aún más si se integraran variables exógenas como el clima o indicadores del desarrollo económico, además de explorar las relaciones de patrones de demanda entre productos (Bandara y col., 2019; Moroff y col., 2021). De acuerdo con un artículo de la consultora estratégica McKinsey con respecto a los pronósticos de demanda que usan Machine Learning (ML), “Los errores de pronóstico a menudo caen entre un 30 y un 50 por ciento” (2019), por lo que existe un costo de oportunidad considerable dada su capacidad para disminuir los errores de pronóstico. El poder computacional y la capacidad de almacenamiento cada vez son mucho menos costosos, y la cantidad de datos disponibles provenientes de diferentes fuentes internas y externas es cada vez mayor (Moroff y col., 2021). Por lo tanto, desarrollar y aplicar modelos multivariantes más robustos que usen técnicas de ML resulta competitivo y eficiente.

En consecuencia, surge la siguiente pregunta, ¿De qué forma se puede incorporar y desarrollar un modelo de pronóstico de demanda para este tipo de industrias que integre múltiples variables externas e internas de tal forma de robustecer los pronósticos disminuyendo el error?

Objetivos

Objetivo General

Desarrollar un modelo multivariable de pronósticos de demanda para industrias con uso intensivo de activos a través del entrenamiento de redes neuronales con el fin de disminuir errores e incertidumbre.

Objetivos Específicos

Diseñar una estrategia para el agrupamiento de las variables externas e internas a considerar en el modelo que sean relevantes para estas industrias.

Diseñar las estructuras de redes neuronales de los modelos de pronóstico a estudiar, en base a las estrategias de agrupamiento de las variables y al error alcanzado en la validación.

Distinguir el mejor modelo de pronóstico, dentro de los diseñados en el estudio para este tipo de industrias, en base al error alcanzado con nuevos datos.

Marco Teórico

Pronósticos de Demanda en Gestión de Operaciones

Tal y como se puede inferir de las secciones anteriores, la necesidad de pronosticar la demanda es de gran relevancia para la gestión de la cadena de suministros o Supply Chain Management (SCM). Aún más, se puede afirmar que es vital para la organización como un todo, ya que proporciona los datos de entrada para la planeación y control transversal de todas las áreas. Los niveles de demanda y su programación afectan en gran medida los niveles de capacidad, las necesidades financieras y la estructura general del negocio (Ballou, 2003, p.286-287).

Cada área dentro de una organización tiene sus problemas particulares de pronóstico. En logística, los pronósticos se relacionan con la naturaleza espacial y temporal de la demanda, su grado de variabilidad y aleatoriedad. La variación de la demanda en el tiempo,

su temporalidad, es resultado de las fluctuaciones ocasionadas por múltiples factores como, por ejemplo, la existencia de patrones estacionales. La mayoría de los métodos de pronóstico a corto plazo se relacionan con esta variación temporal conocida como serie de tiempo. La logística siempre deberá tener en consideración las dimensiones de tiempo y espacio, es decir, tendrá que ocuparse de conocer dónde va a tener lugar el volumen de demanda y cuándo. Los métodos de demanda entonces deberían ser capaces de reflejar diferencias geográficas que puedan afectar los patrones de demanda (Ballou, 2003, p.287-288).

Como se mencionó anteriormente, la variabilidad de la demanda también es un factor a considerar en los pronósticos, si los patrones de demanda tienen un comportamiento más bien regular entonces podrán descomponerse en componentes de tendencia, estacionalidad, estacionariedad, ciclicidad y aleatorias. Mientras que la componente aleatoria represente una pequeña variación en la serie de tiempo, usar métodos de pronóstico convencionales será suficiente. Cuando la demanda es intermitente producto de bajos volúmenes y mucha incertidumbre en cuanto a la cantidad y al instante en el que la demanda se presentará, la serie de tiempo es marcada como irregular. Estos patrones son muy difíciles de pronosticar usando los métodos convencionales y los artículos con este tipo de patrones pueden llegar a representar hasta el 50 % de la cartera de productos que las empresas más convencionales manejan (Ballou, 2003, p.288), ni pensar en negocios e-commerce con una amplia gamma de productos con volúmenes de ventas pequeños, en estos casos el porcentaje puede ser aún mayor. Por lo que pronosticar demanda irregular representa un gran desafío logístico.

El grado de aleatoriedad de la demanda tiene relación con si es independiente o dependiente, en la sección inicial se mencionó a la demanda externa e interna, esta clasificación es similar. La demanda independiente es representada por los clientes del negocio, mientras que la demanda dependiente es derivada a partir del plan maestro de producción que se encarga de planificar pedidos de partes y materiales necesarios para producir un producto final, un ejemplo puede ser el pedido de llantas al proveedor para la fabricación de automóviles, es fácil darse cuenta que el número de llantas puede ser un múltiplo del número de vehículos

a fabricar. Mientras que la demanda independiente conversa bien con métodos de pronósticos convencionales producto de su aleatoriedad, los patrones de demanda dependiente son altamente sesgados y no aleatorios, este sesgo alto es la causa de la poca necesidad de tener pronósticos para la demanda dependiente pues, si el pronóstico de la demanda independiente es certero, la demanda dependiente puede ser pronosticada perfectamente como un derivado de la demanda independiente (Ballou, 2003, p.288-289). Sin embargo, esto es cierto siempre y cuando exista esta relación directa entre producto y las partes o materiales y siempre que los pronósticos sean certeros. Para el pronóstico de la demanda interna de componentes críticos la historia es distinta, depende mucho de la política de mantenibilidad que tenga el negocio y no es un territorio de investigación muy explorado. Por lo que sigue siendo un área interesante para futuras investigaciones (Rosienkiewicz y col., 2017).

Métodos de Pronóstico

Se han mencionado hasta el momento las posibles variables que pueden afectar a los métodos de pronóstico pero no se ha profundizado en los tipos de pronósticos existentes. En esta sección se abordarán las principales ramificaciones de métodos de pronóstico, sus ventajas y desventajas, y sus principales características.

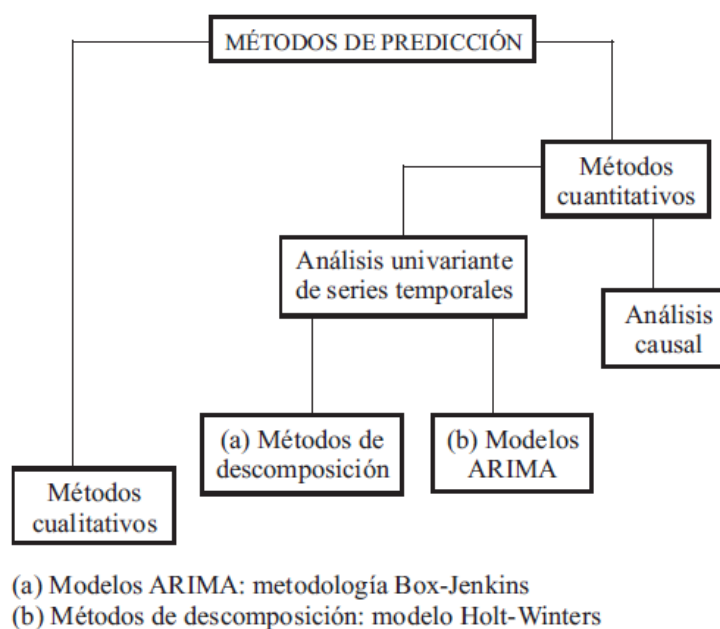
La primera clasificación entre familias comúnmente conocida es la distinción entre métodos cualitativos y cuantitativos. Los métodos cualitativos no usan información cuantitativa para la realización de pronósticos, se basan principalmente en la opinión de expertos, son útiles en situaciones en las cuales la información del pasado es nula o escasa (por ejemplo, lanzamiento de nuevos productos) o cuando los escenarios cambian generando discontinuidades estructurales (por ejemplo, nueva tecnología o cambios en las condiciones políticas) (Moroff y col., 2021), son usados para un alcance a largo plazo. Por otro lado, los métodos cuantitativos (usados para horizontes de corto a mediano plazo) tienen como meta extraer toda la información posible contenida en los datos y realizar estimaciones basándose en el patrón de conducta pasado (Guerrero y col., 2006), los mismos autores comentan que este

método puede dividirse en dos, como se observa en la figura 1, los métodos basados en el análisis univariante o univariable de series de tiempo y los basados en análisis causal. Ballou (2003, p.291-296) coincide con lo anterior describiendo a los métodos de análisis de series de tiempo univariantes como métodos de proyección histórica. La premisa básica, afirma, es que el patrón del tiempo futuro será al menos en su mayoría una réplica del pasado, sus principales herramientas son el uso de modelos matemáticos y estadísticos. Una de sus principales desventajas es que son modelos reactivos, haciendo una analogía, es como estar mirando con el espejo retrovisor, van actualizándose a medida que se obtiene nueva información, entonces si hay un cambio rápido tienen poca capacidad de adaptación rápida. Debido a esto, se dice que las proyecciones de estos modelos demoran los cambios fundamentales en las series de tiempo, y que son débiles para señalar los puntos críticos antes de que se presenten.

Los métodos causales por otra parte basan su construcción en la premisa de que el nivel de la variable pronosticada se deriva del nivel de otras variables relacionadas (Ballou, 2003, p.296). Estos métodos tienen una fuerte base econométrica, se intenta tomar en consideración variables explicativas que modelen factores externos que puedan explicar la variable dependiente (Guerrero y col., 2006), intentando replicar a los modelos de ecuaciones estructurales (SEM) estudiados en econometría. Ballou (2003, p.296) explica que estos métodos pueden dividirse en estadísticos y descriptivos. Los estadísticos engloban los modelos de regresión y econométricos, y los descriptivos cubren los modelos de entrada-salida, ciclo de vida y simulación.

Figura 1

Clasificación clásica de los métodos de pronóstico/predicción.



Nota. Este gráfico representa la clasificación más tradicional de los métodos de pronóstico, con ejemplos de los modelos más destacados dentro de la familia de los métodos de descomposición y de los modelos ARIMA. Tomado de *Guerrero y col., 2006*.

Según Ballou (2003, p.296), el principal problema de esta gran familia de métodos es que la mayoría de las veces resulta difícil encontrar verdaderas variables explicativas, y cuando se encuentran, su asociación con la variable que se pronosticará es, a menudo, preocupantemente baja. Las variables explicativas que guían a la variable de pronóstico en el tiempo son incluso más difíciles de encontrar. Con demasiada frecuencia, el tiempo para adquirir la información para las variables explicativas consume todo el tiempo o la mayor parte del periodo de meses o incluso semestres, en el que se encuentra que tales variables explican al pronóstico y son significativas dentro del modelo. Los modelos basados en técnicas de regresión y econométricas pueden experimentar un error de pronóstico importante debido a estos problemas.

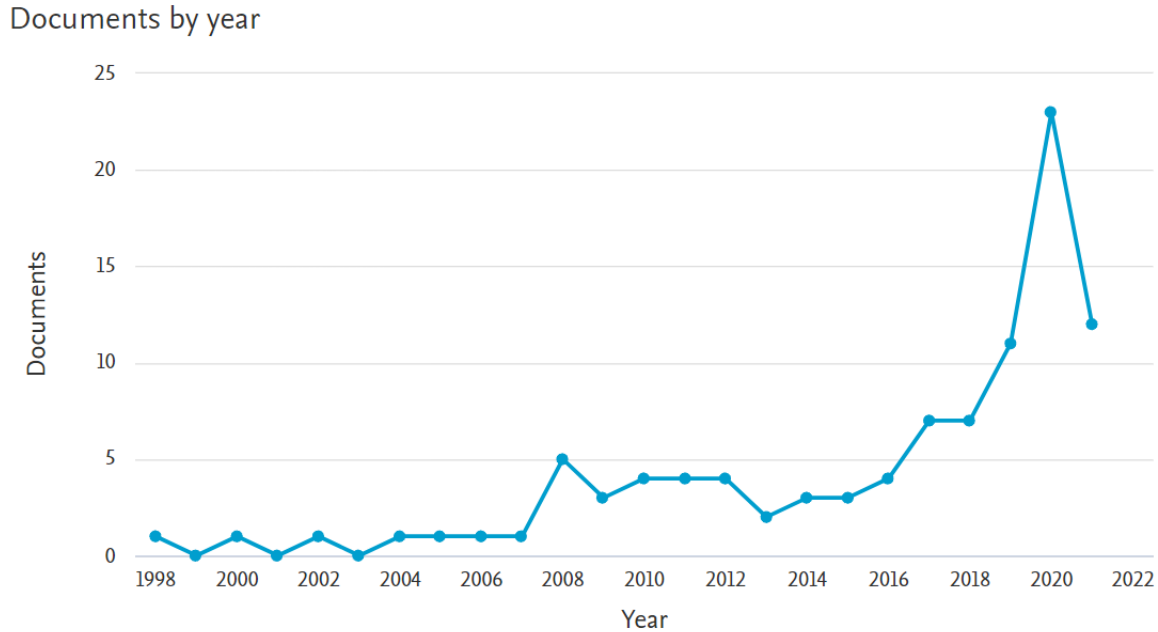
Uso de Redes Neuronales en Pronósticos de Demanda

En los últimos 20 años la tecnología ha tenido saltos gigantescos, desde la creación de smartphones, las redes sociales, el deep learning (DL), la inteligencia artificial (IA), hasta las computadoras cuánticas. Indudablemente estos avances tecnológicos han empapado a las industrias con herramientas y nuevas formas de hacer las cosas, los pronósticos de demanda no han quedado atrás. Actualmente se están usando y estudiando formas innovadoras y efectivas de realizar pronósticos de demanda, por ejemplo, el uso de métodos y modelos de Data Science como Random Forest, Árboles de decisión, Redes Neuronales Recurrentes (RNN) o Artificiales (ANN) (Moroff y col., 2021). Durante el último tiempo la publicación de documentos relacionados a los pronósticos de demanda mediante redes neuronales ha tenido un aumento exponencial como puede verse en la figura 2. Esto último respalda lo que se mencionó anteriormente, ha existido un aumento en el interés de estos modelos dentro de la industria. El futuro de la industria 4.0 tiene como uno de sus principales bastiones de desarrollo al uso de ML y DL dentro de los pronósticos de demanda (Alicke y col., 2019).

La forma de clasificar a los métodos que usan redes neuronales dentro de una familia varía, al igual que su alcance. De acuerdo con Ballou (2003, p.294), el uso de redes neuronales para el pronóstico de demanda tiene un alcance más a corto plazo y los describe como modelos matemáticos de pronóstico inspirados en el funcionamiento de las neuronas cerebrales, caracterizados por su habilidad para aprender a medida que se cuenta con mayor información, destaca además que su precisión de pronóstico parece ser mejor que con otros métodos de series de tiempo cuando la serie de tiempo no es lineal. Sin embargo, hay que destacar que esta apreciación sobre estos métodos tiene más de 20 años de antigüedad, durante todo este tiempo la tecnología ha evolucionado y se han desarrollado diversos tipos de neuronas con diferentes características y propósitos.

Figura 2

Documentos publicados relacionados con pronósticos de demanda mediante redes neuronales por año desde 1998 hasta 2021.



Nota. Este gráfico representa el número de documentos relacionados al uso de redes neuronales en pronósticos de demanda que fueron publicados por año desde 1998 hasta 2021. Tomado de *Analyzer of Search Results, 2021, Scopus*.

Sin avanzar mucho en el tiempo, autores como Zhang (2001) llegan a la conclusión de que un modelo simple de redes neuronales (Multilayer Perceptron o MLP) es bastante competente como para modelar y pronosticar series de tiempo lineales. Trabajos más recientes (Pacella & Papadia, 2021) destacan la efectividad y competitividad de los métodos que usan neuronas del tipo LSTM para modelar series de tiempo no lineales y su capacidad de lograr rendimientos superiores a los métodos de series de tiempo lineales más actuales, también destacan que el uso de neuronas del tipo Bidirectional Long Short-Term Memory (BLSTM), los cuales pueden procesar la data tanto hacia adelante como hacia atrás de la serie de tiempo, aumentan la precisión del enfoque LSTM en todos los casos considerados dentro del estudio.

Moroff y col. (2021) clasifican a las redes neuronales sólo como métodos cuantitativos. Además hacen una distinción entre modelos de ML como Random Forest y los de DL, donde ubican a las redes neuronales (en sus estudios usaron redes neuronales del tipo MLP y LSTM). Tsao y col. (2021) usan métodos de ML para el pronóstico de demanda B2B, integrando tendencias de búsqueda (variable cualitativa) dentro del modelo. Dividieron el problema en dos, primero ajustaron un modelo de serie de tiempo usando métodos de ML (específicamente Random Forest), y luego integraron dentro del mismo modelo términos de búsqueda del rastreador web (araña web) y tendencias de Google con el fin de recopilar señales de mercado relacionadas como índices de información externa para la industria estudiada, logrando con esto ajustar los resultados de pronóstico teniendo en cuenta las fluctuaciones de la demanda. Los autores además afirman que este enfoque puede aplicarse con otros métodos de ML y DL, queda entonces como una posible investigación futura la utilización de redes neuronales usando el mismo enfoque.

De esta forma se podría pensar que es necesaria una nueva rama dentro de los métodos cuantitativos para los algoritmos de ML y DL. Esto último debido a su versatilidad para modelar data cualitativa y cuantitativa, y su habilidad sobresaliente de capturar patrones y anomalías en la demanda, integrando las bondades de los métodos causales y de series de tiempo. Incluso pueden relacionar múltiples series de tiempo de diferentes tipos de productos como lo hicieron Bandara y col. (2019) en su investigación, usando neuronas LSTM, logrando capturar relaciones dentro del portfolio de productos de Walmart.com, alcanzando un rendimiento sobresaliente en comparación a las técnicas de pronóstico univariantes más modernas.

Queda claro entonces que, en los últimos 20 años, la capacidad, versatilidad y el potencial de las redes neuronales abarca mucho más que su uso en series de tiempo no lineales.

Principales Modelos

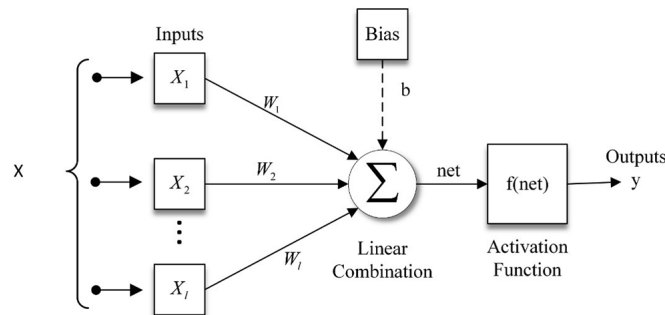
Se han mencionado tipos neuronas, herramientas y algoritmos de Data Science y Deep Learning pero no se ha ahondado en su funcionamiento. En esta sección se explicaran los principales modelos de redes neuronales, sus enfoques, ventajas, desafíos y oportunidades.

Primero se hará una sucinta explicación ilustrativa del funcionamiento de los principales tipos de neuronas usadas en la literatura para pronósticos de demanda (MLP, LSTM, RNN, ANN), esto por la enorme cantidad de información y teoría detrás de estas herramientas de DL, además de la veloz evolución del área, que podría dejar obsoleta la información muy detallada que pudiera plasmarse en esta sección.

La neurona más básica y antigua es el perceptrón simple (Du & Swamy, 2013) creada por Rosenblatt (1958), su estructura es muy buena para ejemplificar de forma general el funcionamiento de una neurona dentro de una red, como se puede observar en la figura 3, este tipo de neurona tiene entradas (inputs) x , una función de activación $f(\text{net})$ dentro de la neurona, y una salida (output) y . Cada input es ponderado por un peso w que determinará su relevancia para explicar al output mediante una combinación lineal, además existirá un bias b que representará al parámetro de ajuste, puede ser el símil del intercepto en una función lineal.

Figura 3

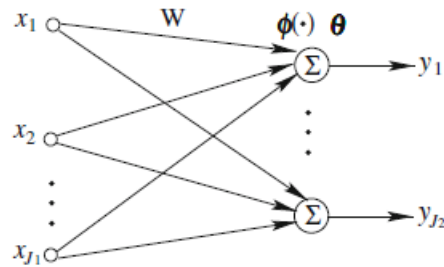
Estructura de un perceptrón.



Nota. Esta figura representa la estructura de un perceptrón. Tomado de Abbasimehr y col., 2020.

Figura 4

Arquitectura de red de una capa de perceptrones.

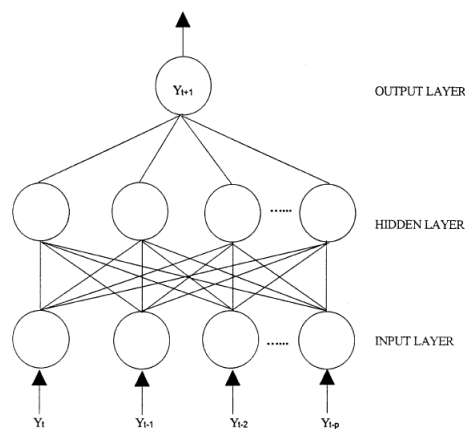


Nota. Esta figura representa una estructura de red para una capa de J perceptrones. Tomado de Du y Swamy, 2013.

La arquitectura que puede observarse en la figura 4 expone la posibilidad de modelar múltiples interacciones entre las variables de entrada, con el objetivo de explicar al output. Sin embargo, se debe precisar que un modelo de redes neuronales debe tener al menos una capa de neuronas de entrada, una oculta (entre capas), y una de salida (Pacella & Papadia, 2021), como puede verse en la figura 5.

Figura 5

Estructura de una red MLP.



Nota. Esta figura representa una estructura de red multicapa de perceptrones o MLP. Tomado de Zhang, 2001.

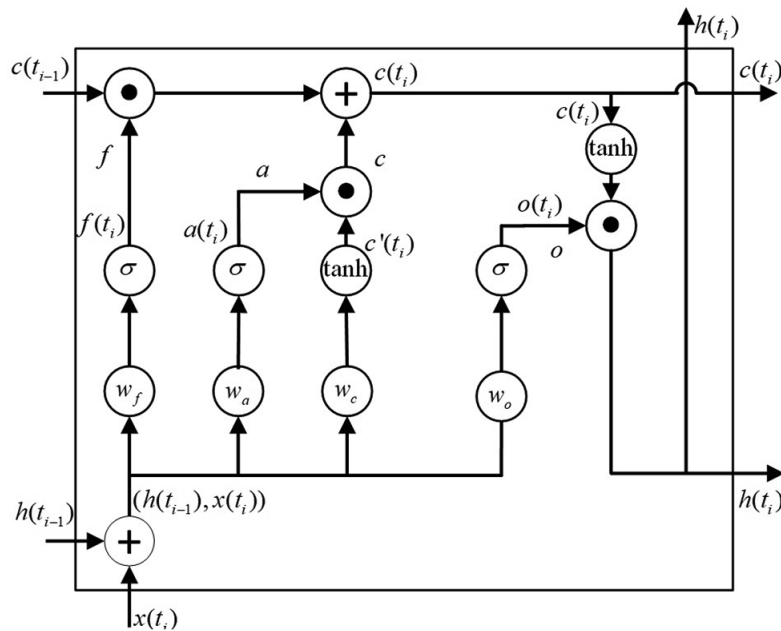
El tipo de red neuronal de la figura anterior se conoce como Multilayer perceptron o MLP y pertenece a la familia de las redes neuronales artificiales o ANN (Artificial Neural Networks), este tipo de redes se caracterizan por intentar representar a una red de neuronas cerebrales interconectadas por dendritas para procesar la información, son no paramétricas, es decir, que no son modelos que requieran de una distribución específica subyacente, y se basan en datos, por lo que la incorrecta especificación del modelo (selección de una forma funcional apropiada y la elección de variables a incluir) afecta menos al rendimiento de lo que lo haría con los modelos paramétricos (estimación puntual, estimación por intervalos donde se buscan intervalos de confianza, contraste de hipótesis). Los ANN con una capa oculta (hidden layer), como los MLP, son uno de los tipos de ANN más utilizados para el pronóstico de series de tiempo (Abbasimehr y col., 2020).

Durante el último tiempo se han desarrollado redes neuronales cada vez más complejas y con diferentes propósitos, como las redes neuronales convolucionales (Convolutional Neural Networks o CNN) para el procesamiento de imágenes (Y. Chen y col., 2021; Wang & Zhang, 2021) o las del tipo Transformer para el procesamiento de lenguaje natural (Natural Language Processing o NLP) (Vaswani y col., 2017). Algunas de estas redes neuronales más complejas se han vuelto bastante populares dentro de las investigaciones, como por ejemplo las redes neuronales recurrentes o RNN (Recurrent Neural Networks) y sus variantes como las LSTM o las BLSTM (Abbasimehr y col., 2020; Bandara y col., 2019; Chandriah & Naraganahalli, 2021; Nguyen y col., 2021; Pacella & Papadia, 2021). Las RNN se diferencian de las MLP en que las conexiones entre neuronas forman una estructura dirigida que incluye algunas conexiones autorreflexivas o recursivas consigo mismas (Abbasimehr y col., 2020), tiene como objetivo predecir el próximo paso de la secuencia de observaciones tomando como input los pasos previos de la secuencia, la idea de las RNN es hacer uso de la secuencia en las observaciones y aprender de las etapas anteriores para predecir tendencias futuras, se llaman ‘recurrentes’ ya que hacen lo mismo para cada etapa de la secuencia. La principal desventaja de las RNN es que recuerdan solo algunos pasos anteriores en la secuencia y, por lo tanto,

no son adecuadas para recordar secuencias de datos más largas. Esto se resuelve utilizando la celda de memoria introducida en las LSTM (Pacella & Papadia, 2021). Estas neuronas están diseñadas para resolver el desafío del largo plazo, lo cual logra que sean prácticas y excelentes en tareas que no pueden ser resueltas por otros RNN, como la predicción continua, el reconocimiento de voz, el modelado del lenguaje y traducción (Cheng y col., 2017).

Figura 6

Estructura de una neurona LSTM.



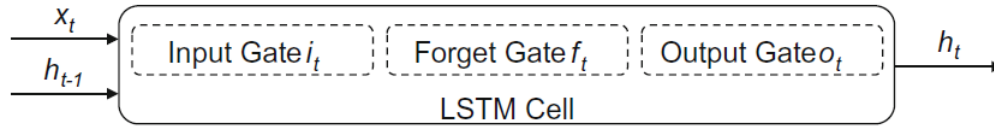
Nota. Esta figura representa un ejemplo de estructura de una neurona LSTM. Tomado de Abbasimehr y col., 2020.

La principal característica de las LSTM es que integran esta celda de memoria capaz de mantener información a lo largo del tiempo, la cual es controlada por varias unidades de compuerta o gate units (Cheng y col., 2017). La figura 6 representa la estructura completa de una LSTM. Sin embargo, resulta poco didáctico de explicar, es por esto que además se muestra en la figura 7 una estructura simplificada de una LSTM, en ella se observa que la neurona procesa la información para mantener un estado o status de la celda en función de un input común x_t y del output de la celda o neurona anterior h_{t-1} , luego decide que

información traspasar h_t y que información descartar mediante las gate units (input gate, forget gate y output gate) (Cheng y col., 2017).

Figura 7

Estructura de una neurona LSTM.



Nota. Esta figura representa la estructura simplificada de una neurona LSTM. Tomado de Cheng y col., 2017.

Las input gates (compuertas de entrada) eligen que data nueva será almacenada en la celda. Las forget gates entregan un valor entre 0 y 1, donde 1 muestra ‘mantener esto completamente’ y 0 implica ‘ignorar esto por completo’. Por último, las output gates deciden cuanto será lo que entregue de cada celda. El valor obtenido se basará en el estado de la celda junto con la data ‘filtrada’ y recién agregada (Pacella & Papadia, 2021).

Bandara y col. (2019) concluyen que la familia de las LSTM muestran resultados competitivos en todas las configuraciones que evaluaron en sus estudios con una gran cantidad de productos, demostrando una solidez en los pronósticos efectuados.

Variables, su Naturaleza y Agrupamiento

Los pronósticos de demanda con redes neuronales ha sido aplicado en diversas áreas y sobre diferentes variables, por ejemplo sobre componentes y partes en la industria minera (Rosienkiewicz y col., 2017), sobre muchos productos para encontrar jerarquías y relaciones entre sus series de tiempo para e-commerce (Bandara y col., 2019), sobre el clima y el consumo eléctrico para la industria de energía (Cheng y col., 2017), o de forma propuesta para variables cualitativas como tendencias de búsqueda para pronosticar demanda en la industria de servidores (soluciones cloud, data service, etc.) (Tsao y col., 2021).

Como se puede extraer del párrafo anterior, diversos estudios han tomado diversos tipos de variables para diferentes industrias obteniendo resultados destacables, pero dada la naturaleza de las variables y de los datos es que los diferentes tratamientos de la data y su agrupamiento van variando entre estudios, por ejemplo, Cheng y col. (2017) usaron información histórica del consumo eléctrico, información del clima e información temporal. Sin embargo, dentro de cada categoría habían muchas subcategorías de información que pudieron considerarse, es por esto que primero seleccionaron los atributos o características relevantes para el modelo usando modelos de Random Forest, de esta forma pudieron depurar la información para que la red neuronal fuera mucho más efectiva. Este es un ejemplo interesante pues ofrece un camino válido para la elección de variables (especificación del modelo).

Bandara y col. (2019) trabajaron con un portafolio de productos (de Walmart.com) bastante grande con diferentes comportamientos (series de tiempo dispares), para estos escenarios sostienen que aplicar estrategia de agrupación de series de tiempo pueden mejorar el rendimiento de las LSTM. Presentaron dos mecanismos de agrupación de productos para el preprocesamiento de la información. El primero consiste en lo que podría ser una clasificación ABC, bastante conocida para la gestión de inventarios, donde identificaron 3 grupos, el primero (G1) englobaba a todos los productos con un alto volumen de ventas y una baja cantidad de ventas cero (ventas con un valor cero en el dataset), el segundo (G2) representa a los productos con una densidad alta de ventas cero y bajas ventas, G3 abarca el resto de los productos, de esta forma se pueden identificar a los artículos que aportan más al negocio y, por lo tanto, cuáles son los pronósticos de ventas prioritarios a mejorar. El segundo mecanismo considera la aplicación de K-means (una herramienta de clasificación de Data Science) para identificar agrupaciones de productos, en este caso las características diferenciadoras vendrían dadas por las cualidades de cada serie de tiempo. Luego, para determinar el número de grupos o clústers óptimos para k-means aplicaron análisis de silueta, un método que indica si los elementos de un grupo están bien emparejados o no. En definitiva, hay formas de agrupar y tratar los datos bastante robustas, pero ciertas herramientas clásicas como la

clasificación ABC son totalmente aplicables y válidas.

Pre-procesamiento de los datos

Una vez elegidas las variables, los datos que representan esas variables deben ser pre-procesadas para que puedan ser usadas en el modelo, en algunos casos deben ser creadas a partir de meta datos (datos que provienen de otros datos) (Noah Silva, Curso Análisis de Negocios, 2021).

El **pre-procesamiento** comienza con la recolección y entendimiento de la data que representará a cada variable. Luego, esta data debe ser preparada para poder usarse dentro del modelo, esto principalmente se refiere a la depuración o limpieza de la data para su uso en los modelos, los principales puntos a considerar en esta preparación se muestran a continuación (Data Science Methodology Course, IBM Student Advocate Program, 2021).

- **Manipulación de los Datos Faltantes:** La existencia de datos faltantes (valores NaN) dentro del data set puede llegar a ser un verdadero desafío dependiendo del contexto. Si están relacionados a días donde no hubieron ventas es fácil reemplazar estos valores por ceros en el data set. Por otro lado, si representan fallas en los instrumentos encargados de la telemetría de activos físicos, como la medición de la temperatura en una caldera, no es recomendable el reemplazo por ceros, en estos casos es mejor reemplazar por el dato anterior, la mediana, el promedio, una interpolación por tendencia o eliminar estos datos si son muy pocos. Sin embargo, hay que tener en consideración que esta problemática puede venir solucionada con anterioridad de forma automática gracias a algún software empresarial o ERP del cual se extrajeron los datos (Bandara y col., 2019; Rosienkiewicz y col., 2017; Data Science Methodology Course, IBM Student Advocate Program, 2021; Clase Data Science, Curso Análisis de Negocios, 2021; María Elisa Farías, Curso Econometría, 2020).
- **Creación de variables categóricas:** Cuando las variables cualitativas representan una característica o cualidad, como por ejemplo indicar si ‘Es fin de semana’ o si

el mantenimiento realizado fue ‘Preventivo’ o ‘Correctivo’, es necesario transformar estas variables en números para que puedan ingresar a los modelos, en estos casos con un simple reemplazo por 0 y 1 para indicar pertenencia o existencia de la cualidad puede solucionarse este inconveniente (Data Science Methodology Course, IBM Student Advocate Program, 2021; David Godoy, Curso Modelos predictivos para la gestión de activos, 2022).

- **Escalamiento de los datos:** El escalamiento de los datos es importante sobre todo cuando las variables de entrada tienen órdenes de magnitud distintas, en estos casos es muy probable que sin el escalamiento los modelos tiendan a darle mayor peso a una variable por sobre las otras sólo por sus órdenes de magnitud, lo cual genera un sesgo. El escalamiento más común y conocido es el estándar, consiste en restar el promedio y dividir por la desviación estándar a cada dato (Data Science Methodology Course, IBM Student Advocate Program, 2021; Noah Silva, Curso Análisis de Negocios, 2021; Bandara y col., 2019). Por otro lado, la normalización utilizando la función logaritmo natural es muy importante en series de tiempo para poder relacionar o comparar las variaciones porcentuales de ventas (o retornos de acciones en un contexto financiero) en periodos de tiempo distintos, sobre todo al usar modelos de la familia ARIMA y ARCH, de esta forma se pueden hacer operaciones entre datos que no perderán normalidad y los modelos alcanzarán mejores rendimientos (Emilio Rojas, Cursos Finanzas II y Gestión de Inversiones, 2021).

Validación de los Modelos de Redes Neuronales

Teniendo la data pre-procesada, el siguiente paso es el desarrollo y entrenamiento del modelo propiamente tal. Existen diversos modelos de acuerdo con la naturaleza del problema, tal y como se vio en la sección de principales modelos en la página 12. Pero, ¿Cómo los podemos evaluar? Para poder determinar la precisión de los pronósticos de demanda usando redes neuronales la fórmula no es distinta a como se evaluaría con pronósticos de demanda

clásicos, es decir, usando a los errores de pronóstico, evaluar si lo que se predijo se acercó a lo que realmente sucedió.

Los tipos de medidas de error o rendimiento más usados para la evaluación de redes neuronales se basan en errores absolutos o errores al cuadrado, en particular al error absoluto medio (MAE) y el error cuadrático medio (RMSE), y sus derivados (Abbasimehr y col., 2020; Pacella & Papadia, 2021). Las expresiones de estas medidas se muestran a continuación.

$$MAE = \frac{1}{h} \sum_{t=1}^h |e(t)| \quad (1)$$

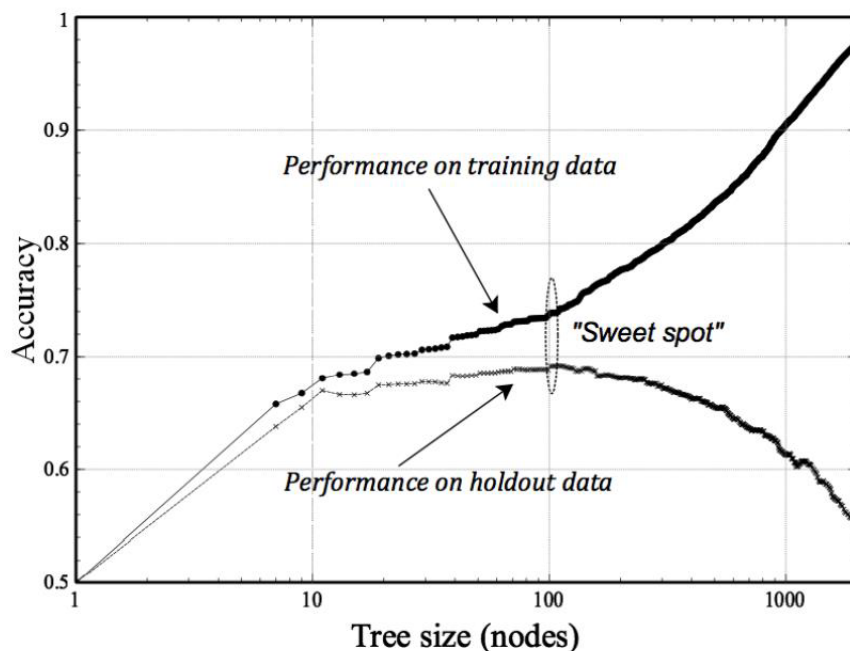
$$RMSE = \sqrt{\frac{1}{h} \sum_{t=1}^h |e^2(t)|} \quad (2)$$

Donde h es el número de datos para entrenamiento o testeo y $e(t)$ es la diferencia entre el valor pronosticado y el real. Estas medidas de error serán la base para las funciones de coste o pérdida que el modelo considerará para su mismo entrenamiento. Una red neuronal requiere que el data set sea dividido al menos en dos partes, una partición será para el entrenamiento de la red (training data), y la otra para evaluar o testear el modelo (holdout o testing data) (Noah Silva, Curso Análisis de Negocios, 2021; Data Science Tools and Data Science Methodology Courses, IBM Student Advocate Program, 2021). La red se irá corrigiendo (usando la función de pérdida) mediante el algoritmo de Back-propagation, este algoritmo corrige la red identificando los caminos mas probables por los cuales el error puede propagarse dentro de la red desde el final hacia el comienzo de la misma, basándose en los pesos que tiene cada segmento o dendrita que une dos neuronas (Rumelhart y col., 1986). Para poder determinar cuando parar de entrenar el modelo se usan las curvas de aprendizaje, un ejemplo puede verse en la figura 8 que ilustra curvas de aprendizaje para el entrenamiento de un árbol de decisión, al ser el mismo concepto, sirve para la explicación. El punto óptimo se dará cuando el modelo, usando la data de testeo o holdout, llegué a un máximo de precisión o a un mínimo error antes de volver a caer en su precisión o a subir su error. Además se observa que, si se traspasa este umbral, el modelo aumenta su precisión usando la training data, esto

no es bueno pues quiere decir que se acostumbró demasiado a estos datos y perderá precisión si se usa un data set distinto, en otras palabras, el modelo se sobre ajusta, perdiendo utilidad para usarse en otros escenarios.

Figura 8

Curva de aprendizaje.



Nota. Esta figura representa un gráfico de curvas de aprendizajes para el entrenamiento y testeo de un árbol de decisión. Tomado de *Curso Análisis de Datos*, 2021.

Para finalizar, cabe destacar que los modelos de redes neuronales para pronósticos de demanda han sido validados en un gran número de publicaciones como métodos que logran resultados por sobre los modelos más convencionales, como se mencionó durante el transcurso de las secciones anteriores. Por otro lado, han sido usados en diferentes contextos de pronóstico, teniendo casos de éxito destacados como el de Bandara y col. (2019) que lograron resultados destacables usando redes neuronales para el pronóstico de demanda de productos de walmart.com o el de Rosienkiewicz y col. (2017) sobre componentes y partes en la industria minera.

Tal y como se mencionó en las secciones iniciales de esta memoria, “Los errores de pronóstico a menudo caen entre un 30 y un 50 por ciento” (Alicke y col., 2019), este último dato se traduce, de acuerdo con Tsao y col. (2021), en la posibilidad de producir una reducción de hasta un 65 % en las ventas perdidas causadas por problemas de almacenamiento de inventario y una disminución del 10 % al 40 % en los costos de almacenamiento. Las oportunidades que ofrecen las redes neuronales y las herramientas de DS para la gestión de operaciones entonces traen consigo beneficios considerables.

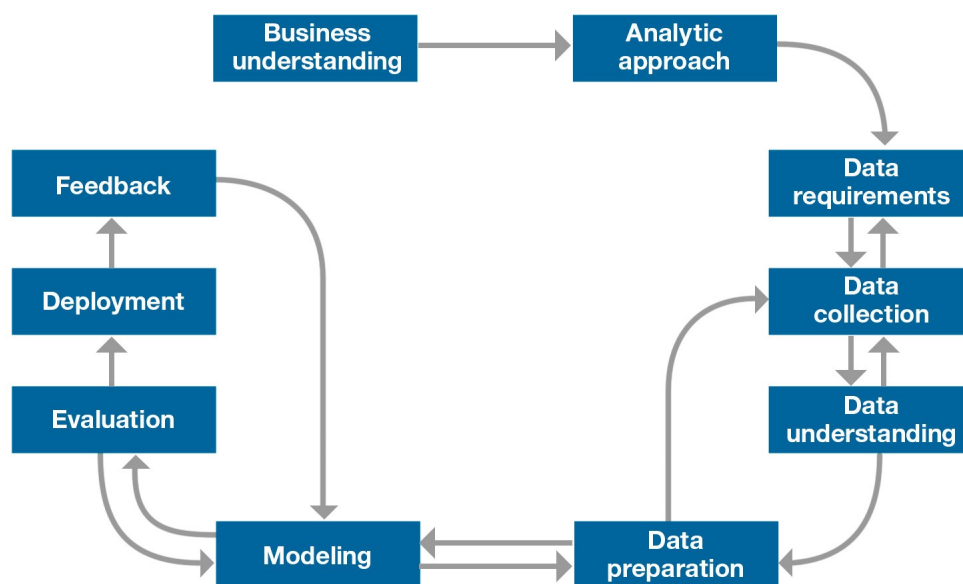
Metodología

Esta sección tiene como fin exponer y explicar el desarrollo de la serie de pasos recorridos para la obtención de los resultados de la investigación en base a los objetivos propuestos de esta memoria.

Metodología

Figura 9

Metodología de Data Science.



Nota. Esta figura representa la metodología de Data Science para la resolución, modelamiento y despliegue de soluciones. Tomado de *Data Science Methodology Course, IBM Student Advocate Program, 2021*.

La presente investigación se basó en la *metodología de Data Science desarrollada por IBM para la resolución, modelamiento y despliegue de soluciones*, expuesta anteriormente en la Figura 9 y de libre acceso. Esta metodología permitió cumplir con los objetivos específicos planteados de forma secuencial y ordenada. Y engloba las siguientes etapas:

- **Business Understanding:** Esta metodología comienza por un entendimiento de la

problemática de negocio a resolver. A partir de este entendimiento se podrá dar respuesta a interrogantes y problemáticas existentes teniendo claro el objetivo durante todo el camino a recorrer. Lo que se busca es tener claramente identificado el problema a solucionar con el fin de poder elegir correctamente los datos a usar para responder a la pregunta central. En esta memoria, el problema de negocio a resolver se desarrolló en la primera sección, la cual decantó en un objetivo general y objetivos específicos.

- **Analytic Approach:** Luego se define el **enfoque analítico** a utilizar para resolver la problemática. Esta decisión parte desde un nivel más general tomando alguno(s) de los siguientes 4 caminos: 1.- Enfoque descriptivo: Entrega de información sobre el estado actual. 2.- Enfoque diagnóstico: Se hace análisis estadístico para saber que esta pasando y sus causas. 3.- Enfoque predictivo: Pronosticar o entregar probabilidades de ocurrencia de un evento futuro. 4.- Enfoque prescriptivo: Entrega una solución al problema, al igual que prescribir una receta para un resfriado. Luego de elegir el enfoque se puede elegir el tipo de modelo a utilizar de acuerdo con el contexto del problema, por ejemplo el uso de árboles de decisión por su fácil comprensión pensando en usuarios con ningún dominio en estadística o data science. El enfoque de esta investigación se acerca más al prescriptivo, esto por que, además de poder pronosticar la demanda futura, se busca lograr una mejora en el rendimiento de los pronósticos, por medio del uso de redes neuronales recurrentes del tipo LSTM por su buena reputación en la literatura.
- **Data Requirements:** Dado el enfoque escogido en el paso anterior, se establecen los **datos requeridos** (elección de variables a considerar). El desarrollo de este paso y los posteriores se detallarán más adelante.
- **Data Collection:** Al **recolectar los datos**, estos pueden venir en cualquier formato. Por lo tanto, de acuerdo con el enfoque tomado y los resultados esperados, los datos deben pasar por un proceso de validación, a partir de esta validación se determina si es necesario recolectar más datos o descartarlos.

- **Data Understanding:** El **entendimiento de los datos** responde a la pregunta ‘¿Los datos recolectados son representativos del problema a resolver?’, en este sentido es que surge el análisis exploratorio de los datos (EDA, por sus siglas en inglés), este análisis busca determinar ciertas características a considerar producto de los datos mediante estadística. Gracias al análisis se podría retroceder a la etapa anterior para buscar corregir problemas relacionados con la recolección.
- **Data Preparation:** Cuando se **preparan los datos** es necesario tener en consideración los puntos descritos anteriormente en la sección de pre-procesamiento de los datos, es decir, Manipulación de datos faltantes, Creación de variables categóricas y Escalamiento de datos. En otras palabras, esta etapa tiene que ver con la limpieza, transformación y normalización de los datos. Esta etapa es necesaria para que los datos puedan servir de input para los modelos, sacando su máximo provecho.
- **Modeling:** El **modelamiento** tiene que ver con el desarrollo y entrenamiento de los modelos, pero también para determinar si los datos están completamente listos o si son necesarias modificaciones finales.
- **Evaluation:** La **evaluación** de los modelos desarrollados (elección del mejor modelo en base al error o precisión alcanzado) se asegura de la calidad de los mismos con el fin de que cumplan con los objetivos propuestos.
- **Deployment:** Cuando el modelo esta completamente evaluado, la metodología continúa con el **despliegue** de la solución a niveles productivos. Esta etapa se encarga de determinar cual es el rendimiento del modelo en un ambiente externo y como poder mejorarlo para que pueda alcanzar un nivel competitivo.
- **Feedback:** La **retroalimentación o Feedback** del usuario final es muy importante para afinar detalles y mejorar el modelo en cuanto a rendimiento e impacto. Los pasos involucrados en la retroalimentación engloban procesos de revisión, rastreo de resulta-

dos, medir la efectividad, entre otros.

Como puede observarse en la figura 9, esta serie de etapas tienen un carácter iterativo. Esta metodología busca que la solución mejore mientras se desarrolla y también después de haberse desplegado.

Alcances y limitaciones

La investigación concluyó en la etapa de evaluación de los modelos debido a que no estuvo dentro de su alcance el despliegue de estos modelos en entornos productivos o la generación de aplicaciones o de softwares empresariales.

Producto del tiempo y las herramientas de hardware y software, sólo se modelaron pocos productos. Las variables que se consideraron fueron pensadas para identificar patrones de venta relacionados a fechas relevantes, como los fines de semana, festivos, feriados, entre otros, esto se hizo para simplificar la búsqueda de variables o especificación de los modelos. Además, estas variables pueden ser representadas por variables binarias o Dummy, tomando el valor de cero o uno, lo cual no genera mayor sesgo relacionado a diferencias en los órdenes de magnitud entre las variables del modelo, siempre y cuando, claro está, se haya hecho el debido escalamiento a las variables a pronosticar previamente. Por otro lado, este tipo de variables binarias podrían perfectamente ser usadas para identificar la existencia de fallas de activos físicos en un momento determinado para un modelo de pronóstico de demanda de componentes mayores en minería, otra industria con uso intensivo de activos.

El principal objetivo de esta memoria es el desarrollo de un modelo multivariable con el uso de LSTM que logre disminuir los errores e incertidumbre, y en el camino también poder resaltar aspectos a considerar para el desarrollo de este tipo de modelos. Sin embargo, producto de las limitaciones técnicas y de tiempo, hay ciertos aspectos que sólo podrán teorizarse mas no comprobarse, este tipo de situaciones serán debidamente informadas en el transcurso de las secciones posteriores.

Herramientas Usadas

Los datos fueron manipulados usando el lenguaje de programación Python en la herramienta de Google llamada Colaboratory o ‘*Colab*’. Con un ambiente similar a un Jupyter Notebook, *Colab* permite correr archivos en la nube con una gran capacidad de procesamiento computacional gratuito, de hecho, todos los modelos y análisis de datos se corrieron en *Colab* usando esta capa gratuita. Las librerías usadas para la manipulación y análisis de los datos fueron principalmente Numpy y Pandas, los gráficos se generaron a partir de las librerías Matplotlib y Pandas. Las librerías usadas en el modelamiento se detallarán más adelante.

Data Requirements

Los datos requeridos para la investigación deben reflejar las ventas de múltiples productos a través del tiempo de una empresa con un sistema logístico robusto y, por tanto, presente en muchos países y con un gran inventario. También se deben buscar datos que sirvan como variables externas que expliquen los patrones de venta, la recolección y los pasos siguientes para este último requerimiento de datos será abordado más adelante.

Data Collection

El Data Set que se utilizará corresponde a dos años de registros de ventas e-commerce de una empresa de retail con presencia en varios países pero con punto de origen en Reino Unido y que vende principalmente artículos de regalo únicos para toda ocasión, sus clientes generalmente son mayoristas (venta B2B). El nombre de la empresa permanece en el anonimato pero, según su descripción, la empresa ocupa Amazon.co.uk para comercializar y vender sus productos (D. Chen y col., 2012; «Online retail dataset», 2019).

Los datos se obtuvieron del repositorio público **Kaggle**, una subsidiaria de Google LLC que sirve como comunidad online de Data Scientists y profesionales que usan ML y DL. *Kaggle.com* permite a los usuarios buscar y publicar conjuntos de datos (Data Sets), explorar y crear modelos en la misma página web y participar en concursos para resolver

desafíos de DS. Una muestra de los datos se ilustra a continuación.

Figura 10

Muestra de datos.

Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
...	...	...	...	...	...	...	...
581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	France
581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	France
581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	France
581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	France
581587	POST	POSTAGE	1	2011-12-09 12:50:00	18.00	12680.0	France

Nota. Esta figura ilustra una muestra de los datos a utilizar, específicamente los primeros y los últimos 5 registros del Data Set. Elaboración propia.

Como se puede observar en la Figura 10, el Data Set contiene 9 columnas. El número de factura (Invoice), un código identificador del producto o SKU (StockCode), la descripción del SKU (Description), la cantidad transada (Quantity), la fecha y hora de la factura (InvoiceDate), el precio unitario (Price), la identificación del cliente (Customer ID) y el país de destino (Country).

Data Understanding

Este apartado hará un breve análisis EDA con el fin de determinar posibles consideraciones a tener en cuenta para los siguientes pasos del pre-procesamiento de los datos como también para poder determinar si cumplen con los requerimientos anteriormente definidos. A partir del uso de la librería Pandas en Python se obtuvo el siguiente cuadro de descripción para cada columna.

Figura 11

Descripción de los datos.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
count	1067371	1067371	1062989	1.067371e+06	1067371	1.067371e+06	824364.000000	1067371
unique	53628	5305	5698	NaN	47635	NaN	NaN	43
top	537434	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN	2010-12-06 16:57:00	NaN	NaN	United Kingdom
freq	1350	5829	5918	NaN	1350	NaN	NaN	981330
mean	NaN	NaN	NaN	9.938898e+00	NaN	4.649388e+00	15324.638504	NaN
std	NaN	NaN	NaN	1.727058e+02	NaN	1.235531e+02	1697.464450	NaN
min	NaN	NaN	NaN	-8.099500e+04	NaN	-5.359436e+04	12346.000000	NaN
25%	NaN	NaN	NaN	1.000000e+00	NaN	1.250000e+00	13975.000000	NaN
50%	NaN	NaN	NaN	3.000000e+00	NaN	2.100000e+00	15255.000000	NaN
75%	NaN	NaN	NaN	1.000000e+01	NaN	4.150000e+00	16797.000000	NaN
max	NaN	NaN	NaN	8.099500e+04	NaN	3.897000e+04	18287.000000	NaN

Nota. Esta figura representa una tabla resumen de ciertas características de cada columna del Data Set. Elaboración propia.

Algunos puntos relevantes observados de la descripción inicial de los datos de la figura 11 se explican a continuación:

- Las transacciones principalmente tienen como **lugar de destino Reino Unido**, de más de un millón de transacciones, 981.330 no salen del país.
- Hay transacciones en **43 países distintos**.
- Hay más de 5000 productos distintos. El producto más vendido es el **"Portaluces Colgante Corazón Blanco"**.
- Observando el formato de fecha será necesario hacer un **agrupamiento previo** de los registros para que reflejen transacciones por producto a nivel de día y no de hora.
- Otro aspecto a considerar es que hay cantidades y precios negativos, lo que podría indicar devoluciones o cancelaciones, por lo que hay que descartar esos registros. Además, de la figura 10 se puede inferir que también hay códigos que no representan productos sino otro tipo de transacciones como costos de envío ('POSTAGE'), por ejemplo. Por lo que se debe hacer una **limpieza previa** considerando esos códigos.

- Se observa que la columna ‘Customer ID’ tiene formato flotante (número con decimales) cuando debería ser un string dado que es un código de identificación. También se debe **cambiar el formato** de fecha de la columna ‘InvoiceDate’ para poder usarse como índice en las tablas de los datos de input a nivel de día.

Dadas las características anteriormente descritas y el análisis exploratorio de los datos, se estimó que este Data Set cumple con los requerimientos para representar las ventas de una empresa e-commerce con un amplio catalogo de productos y con un uso importante de activos físicos como diferentes centros de distribución, un gran sistema logístico, entre otros. De igual forma podría extrapolarse y representar el registro de órdenes de compra de repuestos de activos físicos de una empresa minera, el cual también podría tener observaciones similares a los descritos en los puntos anteriores que requieren pre-procesamiento.

Data Preparation

Esta sección detallará los procesos que se llevarán a cabo para la preparación de los datos, previo al modelamiento.

Elección de SKUs prioritarios

Como se mencionó anteriormente, sólo se modelarán pocos productos, la elección de estos se detallará en este apartado.

Para poder determinar que SKUs deben ser modelados primero hay que determinar las características que van a identificar a los productos prioritarios de los demás. En este sentido, la característica o indicador que podría ser más simple y relevante de elegir es el beneficio por producto, un simple precio por cantidad vendida por cada SKU puede representar esta dimensión sin problemas. Una segunda dimensión interesante a considerar es el nivel de complejidad asociada al modelado de cada SKU, es más simple modelar un producto que tiene un comportamiento de ventas mas estable en el tiempo a modelar uno con ventas esporádicas. Para este fin, el indicador asociado a esta dimensión fue la cantidad de días

con ventas no nulas por producto, tomando como inspiración al trabajo hecho por Bandara y col., 2019.

Teniendo en consideración las observaciones del pre-procesamiento y los indicadores descritos en el párrafo anterior, se hizo una limpieza previa a los datos y se creó una tabla detallando estos indicadores por SKU, parte de la tabla se muestra a continuación.

Figura 12

Tabla de indicadores por SKU.

	StockCode	Description	cantidad	num_dias	precio	Revenue
0	10002	INFLATABLE POLITICAL GLOBE	5370	201	0.850	4564.500
1	10002R	ROBOT PENCIL SHARPNER	4	3	5.170	20.680
2	10080	GROOVY CACTUS INFLATABLE	315	26	0.390	122.850
3	10109	BENDY COLOUR PENCILS	4	1	0.420	1.680
4	10120	DOGGY RUBBER	608	62	0.210	127.680

Nota. Esta figura muestra las primeras 5 filas de la tabla de los beneficios acumulados y cantidad de días con ventas por SKU. Elaboración propia.

La tabla anterior se compone de 6 columnas, las primeras dos de izquierda a derecha son el código del producto o SKU y su descripción, respectivamente. La columna ‘cantidad’ tiene los valores de cantidad vendida por SKU durante los dos años de transacciones que contiene el Data Set, ‘num_dias’ es la cantidad de días con ventas no nulas por SKU, ‘precio’ es el precio por SKU y, por último, ‘Revenue’ es el beneficio acumulado (‘precio’ por ‘cantidad’) durante los dos años de transacciones del Data Set, esto por SKU.

Como pre-procesamiento, un punto relevante a mencionar es que sólo se consideraron las transacciones de Reino Unido para simplificar el problema de negocio, es cierto que la inclusión de diferentes lugares de destino ofrecía una oportunidad para ampliar la investigación, sin embargo, se prefirió limitar el alcance y sólo investigar considerando un lugar de destino para gestionar de mejor forma los tiempos de investigación y los recursos a disposición. A partir de la limpieza del Data Set considerando los puntos expuestos en la sección

de entendimiento de los datos, se redujo la cantidad de códigos a 4846, siendo este número la cantidad real de productos que maneja la empresa.

En base a la tabla de la cual se extrae la figura 12 se realizó una clasificación ABC de los 4846 productos de la empresa para determinar que productos son los prioritarios a modelar. La clasificación se hizo con ayuda del algoritmo de clustering K-means, este algoritmo, propuesto por Stuart Lloyd (1982) y Edward W. Forgy (1965), encuentra ‘k’ clústers o agrupaciones de datos minimizando la suma cuadrada de las distancias de cada punto al respectivo centroide de su clúster (Kanungo y col., 2002), Consta de dos fases, en la primera identifica los centroides de cada clúster y luego clasifica a cada observación de acuerdo con su distancia a cada centroide (Dhanachandra y col., 2015). Producto de su proceso de optimización iterativo, converge muy rápidamente, lo cual es uno de los factores que lo convierten en un algoritmo popular dentro del ML (L. Chen y col., 2021). Como este algoritmo fue usado para realizar una clasificación ABC, el hiper-parámetro K se definió como 3 (dado que son 3 grupos). Por otro lado, la clusterización se llevó a cabo tomando como base a las variables ‘num_dias’ y ‘Revenue’. Luego de haber generado los grupos y clasificar a cada SKU dentro de un clúster, se obtuvieron los siguientes resultados.

Figura 13

Resultados Clasificación ABC.

ABC	revenue	cantidad	%R	%C
A	1.361187e+07	798	73.925520	16.467189
B	2.895890e+06	916	15.727456	18.902187
C	1.905193e+06	3132	10.347024	64.630623

Nota. Esta figura muestra los resultados obtenidos de la clasificación ABC aplicada a los SKU en su totalidad. Elaboración propia.

La figura 13 se compone de 5 columnas, la primera de izquierda a derecha representa el grupo en la clasificación ABC, siendo A el grupo de los SKUs prioritarios o con mayor

relevancia considerando beneficio acumulado y cantidad de días con ventas. La columna ‘Revenue’ indica el beneficio acumulado que concentra cada grupo, la ‘cantidad’ muestra el número de productos que compone cada grupo. Las últimas dos columnas, ‘%R’ y ‘%C’, representan el porcentaje de beneficios que concentra cada grupo y la cantidad de productos en porcentaje con respecto al total, respectivamente. Estas columnas son las que contienen la información más relevante, el 16,48 % de los productos más importantes o relevantes (clasificados en el grupo A) concentran el 73,93 % de los beneficios totales. Estos resultados se acercan bastante a la proporción que plantea el principio de pareto o ley de los pocos vitales, el 80 % de las consecuencias proviene del 20 % de las causas (Edward Johns, Cursos Gestión de Calidad Total, 2021) o, en un contexto de control de inventarios, con el 20 % de los productos almacenados puede controlarse aproximadamente el 80 % del valor de los artículos del almacén (Raúl Stegmaier, Cursos Gestión de Operaciones I, 2021). Como se mencionó en la sección de ‘Alcances y limitaciones’, sólo se iban a usar pocos SKUs, específicamente, se eligieron los 5 SKUs con mayor cantidad de días con ventas y, luego, con el mayor beneficio acumulado, esta jerarquía en las dimensiones consideradas para el ordenamiento de los SKUs se estableció para tener los productos con la mayor cantidad de datos posible para ayudar al modelamiento y entrenamiento de las redes y, en segundo lugar, a los que también tuvieran el mayor beneficio acumulado. Los SKUs elegidos se exponen a continuación.

Figura 14

SKUs prioritarios por Clasificación ABC.

StockCode	Description	cantidad	num_dias	precio	Revenue	rev_dist	count_x	count_y	category	ABC
85099B	JUMBO BAG RED RETROSPOT	87941	611	1.95	171484.95	173000	1	1	0	A
21212	PACK OF 72 RETROSPOT CAKE CASES	71687	610	0.55	39427.85	41000	1	6	0	A
20725	LUNCH BAG RED RETROSPOT	31822	609	1.65	52506.30	53000	1	4	0	A
85123A	WHITE HANGING HEART T-LIGHT HOLDER	84909	604	2.95	250481.55	251000	1	1	0	A
20727	LUNCH BAG BLACK SKULL.	23034	580	1.65	38006.10	41000	1	6	0	A

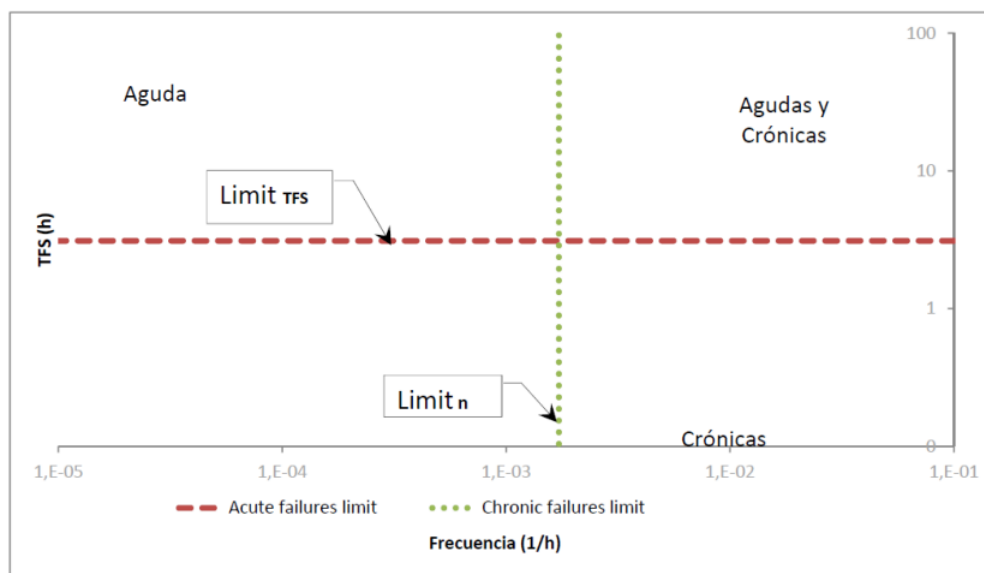
Nota. Esta figura muestra los 5 SKUs más prioritarios obtenidos de la clasificación ABC, junto con algunas características por SKU. Elaboración propia.

Con respecto a la figura 14 debe mencionarse que las columnas ‘rev_dist’, ‘count_x’, ‘count_y’ y ‘category’ son columnas auxiliares usadas para el proceso de clasificación. La columna ‘ABC’ indica el grupo al cual pertenece cada SKU. Se destaca que todos los SKUs elegidos son del grupo A, tal y como se esperaba.

Cabe señalar que la proporción 80/20 es un aproximado, si bien los resultados que se muestran en la figura 13 se acercan a esta proporción, lo obtenido en la clasificación ABC se contrastó con una variación del análisis Jack Knife, determinando productos prioritarios o críticos en base a las dos dimensiones usadas en la clasificación ABC, beneficio acumulado y cantidad de días con ventas por SKU.

Figura 15

Explicación Diagrama Jack Knife.



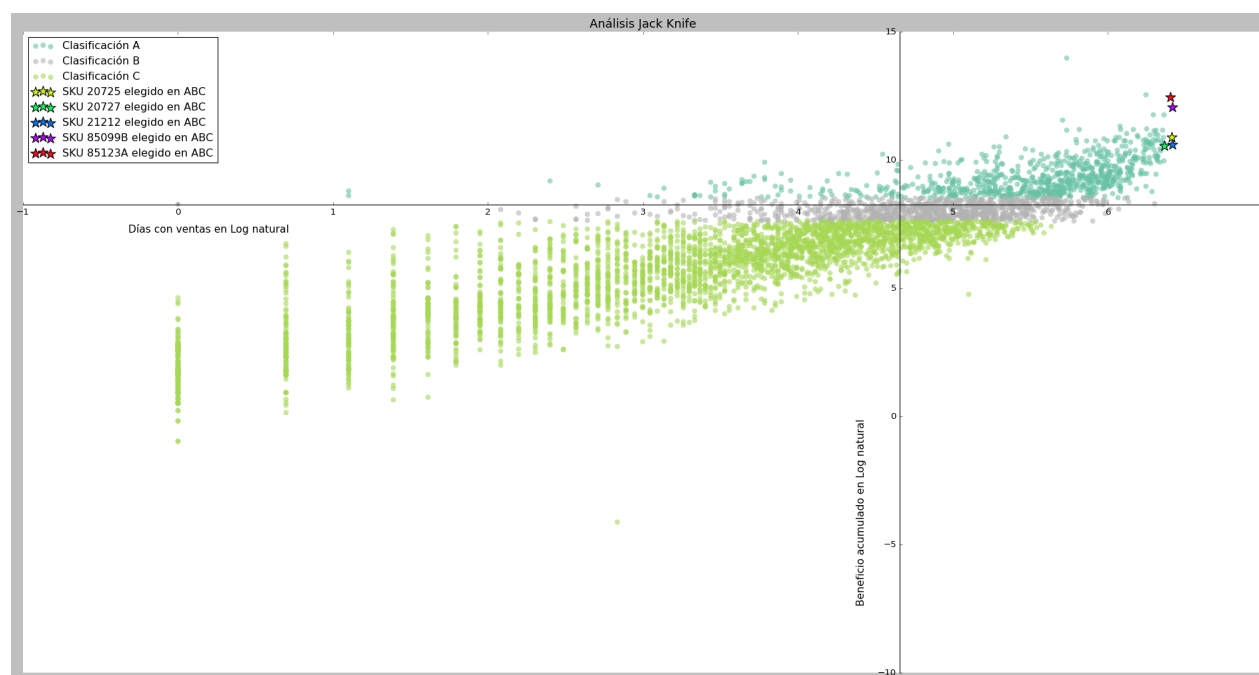
Nota. Esta figura ilustra los cuadrantes formados por los límites umbrales del diagrama Jack Knife. Tomado de *Knights, 2001*.

El análisis o diagrama Jack Knife o de dispersión logarítmicos se usa en la gestión de activos físicos para identificar de forma simple a las fallas que contribuyen más a los costos de los equipos, permitiendo también visualizar la relación entre el tiempo medio de reparaciones (MTTR) y la frecuencia de fallas, siendo estas dos dimensiones los ejes

del diagrama. Para el mejor trazado de iso-cuántas dentro del mismo gráfico para poder relacionar puntos similares, a los datos se les aplica la función logaritmo natural, de esta forma las iso-cuántas se transforman en líneas rectas que son más fáciles de trazar. Por último, los ejes están centrados en el promedio de los datos de cada dimensión. (Knights, 2001).

La figura 15 señala tres cuadrantes relevantes, el cuadrante de las fallas ‘Agudas’, las ‘Crónicas’ y las ‘Agudas y Crónicas’. Las fallas crónicas son las que ocurren con mayor frecuencia, solucionar su causa raíz aumenta el tiempo disponible de los equipos. Lo mismo ocurre con las fallas agudas, mientras más aguda la falla mayor es el tiempo necesario para repararlas y, en consecuencia, menos tiempo disponible. En ambos casos los costos de reparación y mantenimiento, y de lucro cesante o de oportunidad, tienen una relación directa con la existencia de estas fallas.

Teniendo en consideración lo explicado anteriormente, es evidente que las fallas críticas se encuentran en el cuadrante de las fallas ‘Agudas y Crónicas’, resolver estas fallas y darles seguimiento es prioritario. Este análisis se usó para determinar SKUs prioritarios, la única modificación que se hizo fue el cambio de las dimensiones a considerar, en este caso no se utilizó el MTTR y la frecuencia de fallas, se usaron las mismas dimensiones consideradas en la clasificación ABC, es decir, el beneficio acumulado y la cantidad de días con ventas por SKU. El diagrama resultante se muestra a continuación.

Figura 16*Diagrama Jack Knife modificado.*

Nota. Esta figura muestra el diagrama Jack Knife para elegir SKUs prioritarios junto con los SKUs elegidos en la clasificación ABC y los grupos formados en la misma clasificación. Elaboración propia.

En la figura 16 se observa con claridad que los SKUs escogidos por clasificación ABC (marcados con estrellas) son muy críticos o prioritarios de acuerdo con el diagrama Jack Knife, se encuentran en el cuadrante de los ‘Agudos y Crónicos’ y están en el extremo de la nube de puntos. Otra observación relevante es que la mayoría de los SKUs del grupo A también se encuentran en el cuadrante de los puntos críticos, por lo que se puede corroborar que la Clasificación ABC tuvo un buen rendimiento al reconocer a los SKUs prioritarios dentro del universo de productos.

Corroborando a los SKUs prioritarios a modelar, se le aplica una última transformación al data set original para poder generar una tabla que tenga a las ventas de cada producto en el tiempo separados en columnas distintas. Un extracto de esta tabla o Data Frame (DF) se observa a continuación (Los datos que contiene el DF de la figura 17 pueden

observarse gráficamente en la figura 41 en la sección de Anexos).

Figura 17

Ventas por día de los SKUs prioritarios.

	20725	20727	21212	85099B	85123A
2009-12-01	64.0	23.0	240.0	114.0	255.0
2009-12-02	12.0	77.0	599.0	246.0	270.0
2009-12-03	31.0	32.0	190.0	143.0	359.0
2009-12-04	73.0	72.0	180.0	587.0	167.0
2009-12-05	0.0	10.0	48.0	40.0	96.0
...	...	...	...	...	...
2011-12-05	50.0	24.0	18.0	66.0	234.0
2011-12-06	55.0	53.0	66.0	135.0	21.0
2011-12-07	35.0	40.0	50.0	89.0	175.0
2011-12-08	52.0	34.0	51.0	106.0	108.0
2011-12-09	42.0	11.0	9.0	15.0	4.0

Nota. Esta figura muestra los primeros y los últimos 5 días de ventas de cada SKU escogido.
Elaboración propia.

Las ventas de estos SKUs serán catalogadas de ahora en adelante como las variables internas consideradas. Con el objetivo de hacer un breve análisis EDA (**Data Understanding**) de las variables internas, se aplicó el test Dickey-Fuller aumentada (ADF) para determinar estacionariedad (media y varianza constantes a través del tiempo), usando la librería ‘Statsmodels’ en Python. Saber si las series son estacionarias es muy importante pues permite alcanzar una mejor predicción en modelos autorregresivos convencionales como ARIMA o GARCH, por lo que se relaciona con la facilidad de encontrar patrones y ser modelados. Los resultados de este test para cada SKU se detallan en la siguiente figura.

Figura 18*Test ADF por SKU.*

SKU	ADF Statistic	p-value
20725	-4,22	0,000600
20727	-26,66	0,000000
21212	-7,89	0,000000
85099B	-5,59	0,000001
85123A	-8,07	0,000000

Nota. Esta figura muestra los resultados obtenidos del test ADF para cada SKU prioritario escogido. Elaboración propia.

EL test ADF tiene como hipótesis nula que la serie es no estacionaria, mientras más negativo el estadístico alcanzado en el test, con mayor confianza se puede rechazar esta hipótesis nula (HO). En la figura 18, además de entregarse el estadístico alcanzado por SKU, se entrega el p-value, si es menor al 5 % (o 0,05) se rechaza la HO. Todos los SKUs presentan estacionariedad por lo que no deberían ser difíciles de modelar usando modelos convencionales.

En econometría es bastante común el uso del logaritmo natural principalmente para eliminar el efecto de las unidades de las variables sobre los modelos, otra ventaja es que reduce el orden de magnitud de las variables lo que produce una menor sensibilidad a los valores extremos o atípicos. Por último, si se calculan las rentabilidades o variaciones porcentuales a través del tiempo por medio del uso de logaritmo natural, estas rentabilidades o variaciones pueden sumarse sin mayor problema ya que la transformación permite que estén normalmente distribuidas y, por lo tanto, su suma no pierde normalidad (María Elisa Farías, Curso Econometría, 2020). Por los motivos explicados anteriormente, es recomendable aplicar esta transformación sobre las variables, por lo cual se llevó a cabo.

Un aspecto importante a considerar, como se mencionó en secciones anteriores, es el escalamiento de los datos. Los productos escogidos tienen diferentes órdenes de magnitud (incluso con la transformación logarítmica) y si se modelan juntos deben escalarse para no generar sesgos producto de estas diferencias en las órdenes de magnitud, el modelo podría

darle mayor peso a los SKUs con órdenes de magnitudes mayores, por ejemplo.

A modo de prueba se hizo un escalamiento del tipo Min-Max usando la librería ‘sklearn’ y se comprobó que este tipo de escalamiento no afecta la estacionariedad de las series. El escalamiento del tipo Min-Max aplica una transformación a los datos de tal forma que todos quedan dentro del rango $[0,1]$, otorgándole el valor 0 al dato menor y el 1 al dato mayor. La ecuación 3 describe la transformación usada en el escalamiento Min-Max, donde X_{Scaled} es el dato transformado o escalado, X es el dato original sin escalar, X_{max} es el dato más grande y X_{min} es el dato más pequeño.

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Este tipo de escalamiento fue el que se utilizó en el modelo, su elección se basó en que es el más común dentro de las investigaciones y a que no alteró la estacionariedad de las series. Cabe señalar que el escalamiento se debe hacer con respecto a los datos de entrenamiento, es decir, considerando el mínimo y máximo valor presentes en los datos de entrenamiento, esto se debe hacer siempre para no generar resultados sesgados. Si se considera todo el Data Set, el valor mínimo o máximo podría estar en los datos de testeo, lo cual no es realista, los datos de testeo deben poner a prueba todo el modelo. Por lo tanto, el escalamiento pasa a ser otro modelo a ‘entrenar’ dentro del modelamiento y que debe aplicarse a los datos de testeo, pero ajustado a partir de los datos de entrenamiento.

Como última observación del pre-procesamiento de las variables internas, las fechas sin registros o valores ‘NaN’ se reemplazaron con ceros. Al estar el DF a nivel de día, es más simple suponer que si no hay registros es porque no hubieron ventas o porque la empresa no operó durante esos días.

Variables Externas

“Las finanzas conductuales son una rama de las finanzas, que busca explicar las diferencias, desajustes o incoherencias que existen en los mercados financieros, al compararlos

con la teoría de la eficiencia de los mercados. La teoría de los mercados eficientes tiene un principio aceptado que es la racionalidad de sus agentes en la toma de decisión. Dado que el mercado está compuesto por personas, las finanzas conductuales toman como base los conceptos financieros y los factores psicológicos de las personas y de su comportamiento como conjunto, lo que viene a ser el mercado”(Emilio Rojas, Curso Gestión de Inversiones, 2021; Tversky y Kahneman, 1979).

Data Requirements. Uno de los tipos de incoherencias estudiadas son las anomalías de calendario, es decir, patrones de conducta en los precios de las acciones que son repetitivos y dependen de efectos producidos por los días de la semana, feriados, cambio de mes, inicio de año, etc. (Emilio Rojas, Curso Gestión de Inversiones, 2021). Este tipo de anomalías son las que se tomaron como base para la búsqueda de variables externas en el contexto de los pronósticos de demanda. Tal y como lo menciona Bandara y col., 2019, la volatilidad presente en la industria del E-commerce es ampliamente causada por efectos de festividades, ventas esporádicas, entre otras causas, por lo que incluir este tipo de variables relacionadas con el calendario resulta conveniente también por estos motivos.

Data Collection. El primer conjunto de variables externas que se incluyeron fueron las relacionadas al efecto feriados/festividades. Para estas variables se extrajo la información sobre días festivos, feriados o ‘Bank Holidays’ en Reino Unido entre el 2009 y 2011. Un extracto de estos datos se muestran a continuación.

Figura 19

Feriados y Festivos de Reino Unido.

Name	Bank Holiday	Date
New Year's Day	1	2009-01-01
Robert Burns Night (Burns Night)	0	2009-01-25
Holocaust Memorial Day	0	2009-01-27
Valentines Day	0	2009-02-14
Shrove Tuesday (Pancake Day)	0	2009-02-24

Nota. Esta figura muestra 5 días festivos de Reino Unido y si fue feriado (Bank Holiday) ese día. Elaboración propia.

Data Understanding. A partir de estos datos se generaron dos variables externas, la primera es una variable binaria que identifica las fechas que fueron feriados (Bank Holidays), y la segunda también es una variable binaria que identifica las fecha que fueron festivos pero no feriados. Otros efectos considerados fueron los cambios de mes, cambios de trimestre, cambios de año y fin de semana, los cuales también fueron representados como variables binarias o Dummy. Este conjunto de variables externas será nombrado como ‘Variables Externas I’. Un extracto de los datos preparados (con los valores de las series de tiempo escalados con Min-Max) considerando estas variables externas se observan a continuación.

Figura 20*Datos preparados variables externas I.*

Date	20725	20727	21212	85099B	85123A	Inicio de mes	Fin de mes	Inicio Trimestre	Fin Trimestre	Inicio Año	Fin Año	Es Fin de semana	Festivo	Feriado
2009-12-01	0.152381	0.008666	0.178439	0.074364	0.081888	1	0	0	0	0	0	0	0	0.0
2009-12-02	0.028571	0.029013	0.445353	0.160470	0.086705	0	0	0	0	0	0	0	0	0.0
2009-12-03	0.073810	0.012057	0.141264	0.093281	0.115286	0	0	0	0	0	0	0	0	0.0
2009-12-04	0.173810	0.027129	0.133829	0.382909	0.053629	0	0	0	0	0	0	0	0	0.0
2009-12-05	0.000000	0.003768	0.035688	0.026093	0.030829	0	0	0	0	0	0	1	0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2011-12-05	0.119048	0.009043	0.013383	0.043053	0.075145	0	0	0	0	0	0	0	0	0.0
2011-12-06	0.130952	0.019970	0.049071	0.088063	0.006744	0	0	0	0	0	0	0	0	0.0
2011-12-07	0.083333	0.015072	0.037175	0.058056	0.056198	0	0	0	0	0	0	0	0	0.0
2011-12-08	0.123810	0.012811	0.037918	0.069145	0.034682	0	0	0	0	0	0	0	0	0.0
2011-12-09	0.100000	0.004145	0.006691	0.009785	0.001285	0	0	0	0	0	0	0	0	0.0

Nota. Esta figura muestra un extracto de los datos preparados considerando las variables externas I. Elaboración propia.

El último efecto que se consideró dentro de las variables externas fue el de días de la semana, generando una variable binaria por cada día de la semana, excepto los días sábados y domingos ya que ya están contemplados en la variable ‘Es Fin de semana’. Los datos preparados considerando este segundo grupo de variables externas se expone a continuación.

Figura 21*Datos preparados variables externas II.*

20725	20727	21212	85099B	85123A	Date	Lunes	Martes	Miércoles	Jueves	Viernes
0.152381	0.008666	0.178439	0.074364	0.081888	2009-12-01	0.0	1.0	0.0	0.0	0.0
0.028571	0.029013	0.445353	0.160470	0.086705	2009-12-02	0.0	0.0	1.0	0.0	0.0
0.073810	0.012057	0.141264	0.093281	0.115286	2009-12-03	0.0	0.0	0.0	1.0	0.0
0.173810	0.027129	0.133829	0.382909	0.053629	2009-12-04	0.0	0.0	0.0	0.0	1.0
0.000000	0.003768	0.035688	0.026093	0.030829	2009-12-05	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
0.119048	0.009043	0.013383	0.043053	0.075145	2011-12-05	1.0	0.0	0.0	0.0	0.0
0.130952	0.019970	0.049071	0.088063	0.006744	2011-12-06	0.0	1.0	0.0	0.0	0.0
0.083333	0.015072	0.037175	0.058056	0.056198	2011-12-07	0.0	0.0	1.0	0.0	0.0
0.123810	0.012811	0.037918	0.069145	0.034682	2011-12-08	0.0	0.0	0.0	1.0	0.0
0.100000	0.004145	0.006691	0.009785	0.001285	2011-12-09	0.0	0.0	0.0	0.0	1.0

Nota. Esta figura muestra un extracto de los datos preparados considerando las variables externas II. Elaboración propia.

Todas las variables externas consideradas para los modelos son binarias. En un contexto de pronóstico de pedidos de componentes mayores, las fallas de los equipos que comparten un mismo componente a pronosticar podrían integrarse dentro de un modelo de pronóstico como variables binarias de una forma similar a como están consideradas las variables externas que se observan en la figura 20 y 21, por lo cual cumplen con los requerimientos de datos definidos anteriormente. Este ejemplo también intenta resaltar la versatilidad en la metodología de desarrollo que se siguió en esta memoria, pudiendo aplicarse a diferentes industrias como el retail o la minería.

Estrategia de Agrupamiento

“Las variables de entrada o input deben tener una relación entre sí, si hay variables que no tienen una relación directa, estas pueden agregarse al modelo pero como restricciones en una función de suma ponderada al final del modelo” (Ph.D. Pablo Salazar, Enterprise Architect and Artificial Intelligence Discovery Lead en IBM, comunicación personal, 13 de enero de 2022).

Como queda expuesto en el párrafo anterior, es importante el agrupamiento de las series de tiempo o variables internas para el entrenamiento y desarrollo de redes neuronales, como también lo es la especificación del modelo o elección de variables externas de tal forma que tengan relación o similitud con las variables internas, si bien las redes neuronales se caracterizan por no verse muy afectadas por una mala especificación, para tener mejores resultados es conveniente un buen agrupamiento de variables. Estos agrupamientos ayudarán a formar los Data Sets ya preparados o procesados para servir de input en las LSTM, estos Data Sets de input se conocen como ‘Bases de Conocimiento’ (KB, por sus siglas en inglés). Las KB son el resultado de la etapa **Data Preparation**.

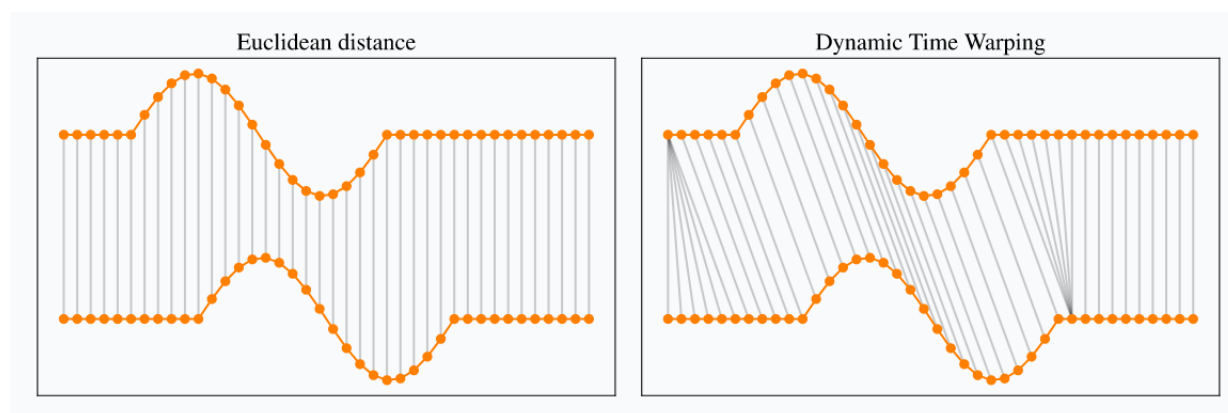
Agrupamiento de Variables Internas. Para el agrupamiento de las series de tiempo de los SKUs hay que tener en consideración que no se tienen disponibles las categorías y sub-categorías de los productos, esta dimensión podría considerarse como una restricción

de negocio para el agrupamiento, por ejemplo, definiendo grupos que exclusivamente sean de una misma categoría por decisión de la empresa. Sin embargo, la estrategia de agrupamiento debe tener como foco a las series de tiempo y sus patrones, lo cual es lo relevante para el entrenamiento de las redes neuronales.

Teniendo lo anterior en mente, se usó el algoritmo de clustering K-means pero modificado para series de tiempo por medio de la librería ‘tslearn’ en Python. La principal modificación en este algoritmo es el reemplazo de la utilización de la distancia euclidiana por la métrica de semejanza entre series de tiempo conocida como ‘Dynamic Time Warping’ o deformación dinámica del tiempo (DTW). Como puede verse en la figura 22, la diferencia entre ambas métricas es que DTW coincide con patrones distintivos de la serie temporal, lo que resulta en una evaluación de similitud más sólida que cuando se usa la distancia euclidiana que coincide con las marcas de tiempo, independientemente de los valores del eje y. Cabe señalar que, en ambos casos, la similitud devuelta es la suma de las distancias entre los valores de cada serie (Tavenard, 2021).

Figura 22

Distancia Euclidiana v/s DTW.

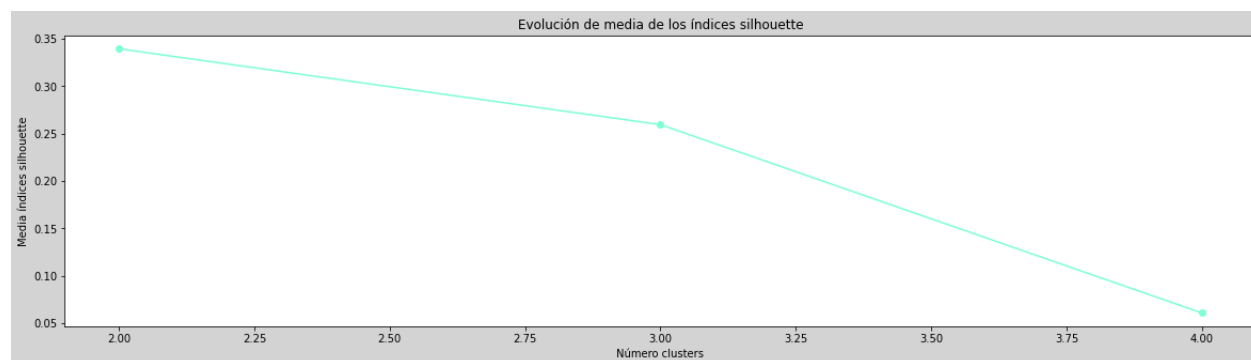


Nota. Esta figura muestra una comparación entre DTW y distancia euclidiana. Se debe tener en cuenta que, en aras de la visualización, las series de tiempo están desplazados verticalmente, pero los rangos de valores de características (valores del eje y) en realidad coinciden. Tomado de Tavenard, 2021.

Para determinar la cantidad de clústers óptimo se usó el índice de Silhouette, este índice es una medida de cuán similar es un objeto a su propio grupo (cohesión) en comparación con otros grupos (separación), de esta forma se obtiene una cantidad de grupos o clústers óptimo comparando los promedios de este índice (Rousseeuw, 1987). Los resultados de este índice se muestran a continuación.

Figura 23

Resultados Índice de Silhouette.



Nota. Esta figura muestra los valores promedio de los índices de Silhouette para cada número de clústers 'k' analizado. Elaboración propia.

Los resultados de la figura 23 indican que el 'k' óptimo es 2, esto es porque es el que alcanza un mayor valor promedio de Silhouette (0,34). Algo que debe mencionarse es que, si este índice está entre 0,26 y 0,50, las 'estructuras' generadas (grupos) son débiles y podrían ser artificiales. Al utilizar K-means en su forma tradicional, no se encuentran grupos sustanciales ($\text{Silhouette} = 0,07 < 0,25$), por lo que el uso de DTW en K-means si logra una mejora importante en el desarrollo de clústers de series de tiempo. Pese a lo débil de las estructuras, estos grupos se utilizaron de todas formas, entendiendo que clusterizar 5 series de tiempo es una tarea difícil de lograr y los resultados son aceptables. Los grupos resultantes se muestran a continuación.

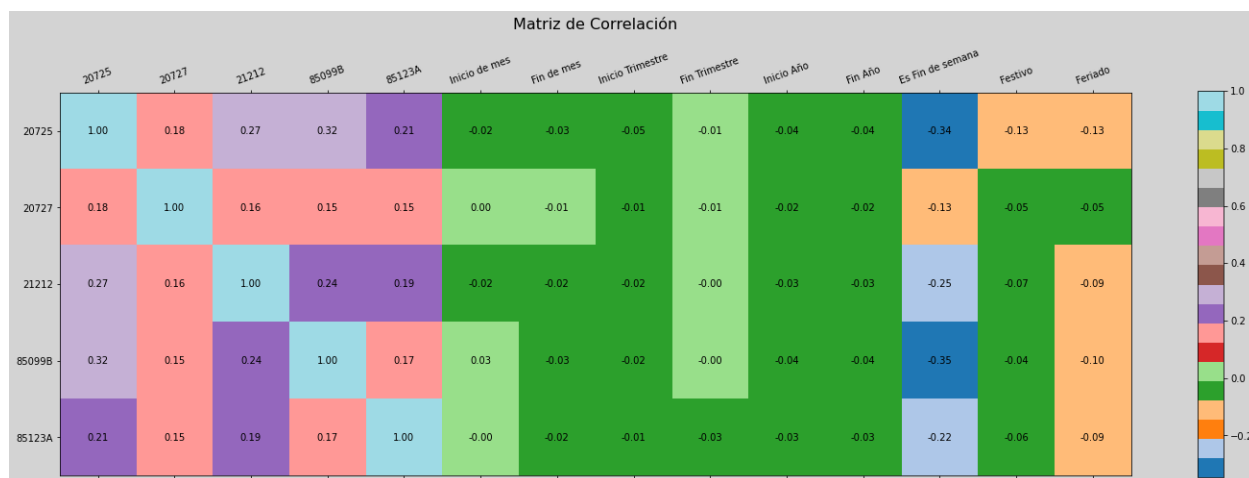
Figura 24*Resultados Grupos K-means.*

SKU	desc	Grupo_SKU
20725	Lonchera Roja Retrospot	1
20727	Lonchera Calavera Negra	0
21212	Pack Capsulas de papel Cupcake Retrospot 72 unid.	1
85099B	Bolson Rojo Retrospot	1
85123A	Candelabro Corazon Blanco Colgante T-light	0

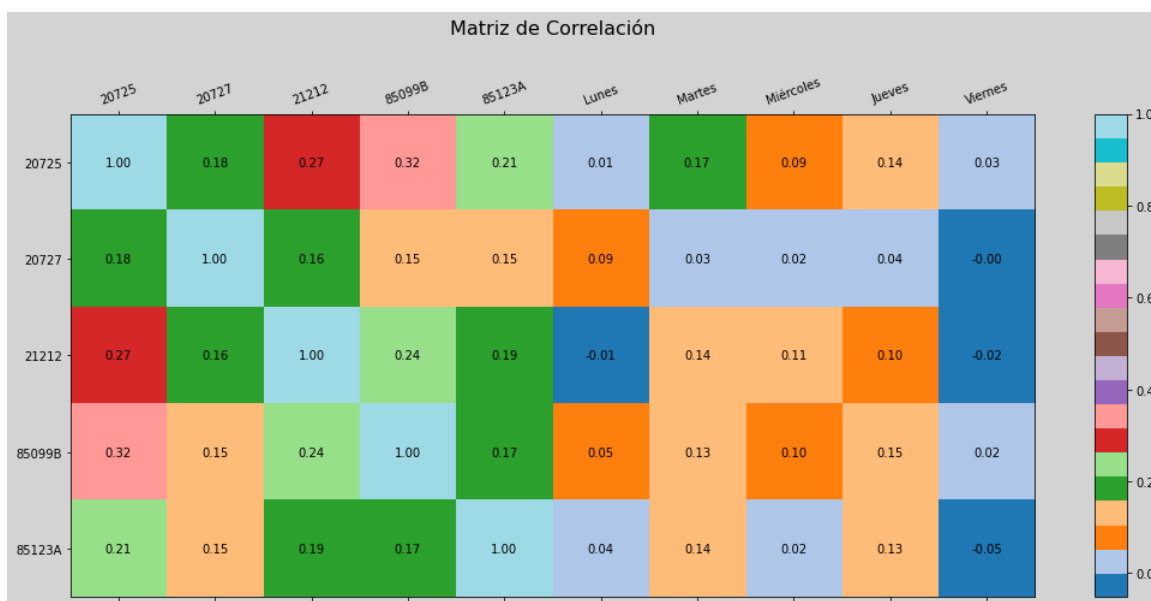
Nota. Esta figura ilustra los grupos resultantes del algoritmo de K-means modificado para series de tiempo. La columna ‘Grupo_SKU’ indica el grupo de cada SKU. Elaboración propia.

Como último comentario antes de pasar a las estrategias de agrupamiento de las variables externas, de la figura 24 se observa que todos los SKUs de la marca ‘Retrospot’ quedaron en el mismo grupo. Si se miran los gráficos de cada serie de tiempo de la figura 41, es posible inferir que el clustering generó un grupo con una menor volatilidad (Grupo 0), y otro con una alta volatilidad (Grupo 1).

Agrupamiento de Variables Externas. Dado que las variables externas no son variables objetivos a pronosticar, sino que sólo buscan mejorar el rendimiento del modelo, estas variables deben ser seleccionadas de acuerdo con las variables a pronosticar (internas). La elección de estas variables se hizo por medio del uso del coeficiente de correlación de Pearson, de esta forma se descartaron aquellas variables externas poco correlacionadas con las series de tiempo. Si bien este método es simple y no mide causalidad como, por ejemplo, el test de Granger, el probable error asociado a la elección de estas variables usando sólo un coeficiente de correlación ayuda a poner a prueba a las LSTM y su fama de ser resistentes a los problemas de especificación. Las matrices de correlación se exponen a continuación.

Figura 25*Matriz de correlación Variables Externas I.*

Nota. Esta figura muestra la matriz de correlaciones entre las series de tiempo y las variables externas I. Elaboración propia.

Figura 26*Matriz de correlación Variables Externas II.*

Nota. Esta figura muestra la matriz de correlaciones entre las series de tiempo y las variables externas II. Elaboración propia.

A partir de las matrices de correlación de las figuras 25 y 26 se hizo una primera depuración de las variables externas, descartando a aquellas variables con una correlación muy baja (cercana a cero) transversalmente con todas las series de tiempo de los SKUs. De esta forma, para el grupo de variables externas I sólo quedaron las variables ‘Es Fin de semana’, ‘Festivo’ y ‘Feriado’, en la matriz de la figura 25 puede observarse con mayor claridad la razón de esta elección, todos los cuadrantes en verde indican un bajo coeficiente de correlación. Con respecto al grupo de variables externas II, la única variable que se descartó fue el ‘Viernes’, en la figura 26 se puede observar que esta variable tiene correlaciones muy bajas transversalmente para todas las series de tiempo (con cuadrantes en color azul).

Después de realizar esta depuración de las variables externas, estas variables fueron agrupadas de acuerdo con su correlación a cada grupo de series de tiempo. Para el ‘Grupo 0’ de series de tiempo conformado por el SKU ‘20727’ y ‘85123A’ se consideró que las variables con mayor correlación eran el ‘Lunes’, ‘Martes’ y ‘Jueves’ (de las variables externas II), y ‘Es Fin de semana’ y ‘Feriado’ (de las variables externas I). Las variables externas con mejor correlación para el ‘Grupo 1’ fueron los ‘Martes’, ‘Miércoles’ y ‘Jueves’ (variables externas II), y ‘Es Fin de semana’, ‘Festivo’ y ‘Feriado’ (variables externas I). Para el ‘Grupo 1’ no se consideró la variable externa ‘Lunes’ por tener un coeficiente de correlación en una situación de borde pero más bajo respecto a las demás variables seleccionadas.

Base de Comparación

Como base de comparación o Benchmark, se utilizaron los modelos de la familia ARIMA y GARCH para el pronóstico univariable y autorregresivo de las series de tiempo y su varianza, respectivamente. Esto se hizo para poder evaluar a los modelos teniendo un criterio de base de comparación y rendimiento. El uso de esta familia de modelos para el benchmark es por su amplia utilización y buena reputación en diferentes industrias (por sobre otros modelos convencionales). Con el fin de poner a prueba de forma efectiva a los modelos, sólo el 70 % del dataset se utilizó para su entrenamiento y para determinar el rendimiento

efectivo de los mismos (testear con datos completamente nuevos) se usó el 30 % restante del dataset (data de testeo), es decir, las ventas desde mayo a diciembre del 2011. Esta división 70-30 se mantuvo para el entrenamiento y testeo de las redes neuronales.

Para la elección por SKU del modelo autorregresivo se utilizaron las librerías de Python ‘pmdarima’ y ‘pyflux’, los cuales contienen diferentes modelos de la familia ARIMA y GARCH, respectivamente. Estas librerías cuentan con funciones econométricas del software R que optimizan el modelamiento minimizando el criterio de información de Akaike (AIC). El AIC es una medida para la evaluación de modelos considerando la bondad de ajuste a partir de la máxima verosimilitud del modelo, y la complejidad a partir el número de parámetros (Emilio Rojas, Curso Gestión de Inversiones, 2021). Además, se obtuvo el RMSE, el MAE y el R2 ajustado, este último como indicador de bondad de ajuste del pronóstico, para cada SKU. Los resultados se exponen a continuación.

Figura 27

Modelos usados y benchmark por SKU.

SKU	MODELO	RMSE	MAE	R2 ajustado
20725	SARIMAX(0,1,1) x (2,0,[1,2],7) + EGARCH(1,1)	43,41	28,36	0,140
20727	SARIMAX(1,0,1) x (1,0,1,7) + SEGARCH(1,3)	34,73	22,97	0,0831
21212	SARIMAX(5,1,1) x (2,0,1,7) + EGARCH(1,1)	126,13	45,73	-0,0347
85099B	SARIMAX(1,0,1) x (2,0,[1,2],7) + EGARCH(1,1)	192,41	116,80	-0,0520
85123A	SARIMAX(5,1,2) + EGARCH(1,1)	264,22	164,22	-3,182

Nota. Esta figura muestra los modelos elegidos por las funciones de Python, el RMSE, MAE y R2 ajustado alcanzado por SKU. Elaboración propia.

La notación de los modelos se compone de una parte relacionada con la familia ARIMA (antes del ‘+’) y otra con GARCH (después del ‘+’). La primera parte puede escribirse como SARIMAX(p,d,q) x (P,D,Q,s), donde ‘p’ representa el número de rezagos o autorregresiones a considerar, ‘d’ son las diferencias necesarias para convertir la serie original en estacionaria y ‘q’ es el número de medias móviles o rezago de los errores a considerar. P, D y Q representan lo mismo que sus similares en minúsculas pero son con respecto a la compo-

nente estacional (en caso de considerarse, sino estos parámetros se descartan y no aparecen), ‘s’ es el periodo de la estacionalidad. La segunda parte asociada a la familia GARCH refleja el modelo elegido para explicar la varianza de la serie de tiempo, este tipo de modelos se ocupan generalmente para resolver los problemas asociados a la Heterocedasticidad (varianza no constante en el tiempo), EGARCH(p,q) intenta modelar la varianza dándole importancia a las variaciones positivas y negativas de forma asimétrica, mientras que SEGARCH(p,q) es una variación del EGARCH que tiene una mayor robustez frente a valores atípicos o outliers. Con respecto a los parámetros, ‘p’ es el número de rezagos o autorregresiones de la varianza y ‘q’ es el número de rezagos del cuadrado de los errores. Los gráficos de las predicciones generadas por estos modelos por SKU se pueden observar en Anexos (Fig. 42).

Un aspecto interesante a analizar es la predicción resultante del SKU ‘85123A’, que es posible observar en la figura 42, esta predicción generada por ARIMA + GARCH gráficamente deja de pronosticar correctamente al usar los datos de testeo de una forma estrepitosa (lo cual también se evidencia en el alto error alcanzado y su pésima bondad de ajuste observable en la figura 27), indicando un cambio en los patrones de comportamiento de venta de este SKU (cambio estructural), evidenciando una de las problemáticas que tienen este tipo de modelos, al existir un cambio estructural estos modelos dejan de ser efectivos completamente. Para determinar que efectivamente hay presencia en la serie de tiempo de un cambio estructural se aplicó el test de Chow, esta prueba econométrica intenta determinar si hay estabilidad permanente de los parámetros de un modelo de regresión si este se utilizara para modelar la serie de tiempo en cuestión, si no hay estabilidad esto implica que el patrón de la serie de tiempo es significativamente diferente en un punto (fecha) en particular y, por tanto, debería modelarse con dos regresiones diferentes (María Elisa Farías, Curso Econometría, 2021). El punto de ruptura es sugerido por el investigador, en este caso se tomó como punto propuesto al cambio de año (31 de diciembre del 2010 y 01 de enero del 2011). Los resultados de este test se muestran a continuación.

Figura 28

Resultados Test de Chow.

SKU	Chow Statistic	p-value
85123A	3,29	0,038

Nota. Esta figura muestra los resultados obtenidos del test de Chow tomando como punto de ruptura el cambio de año. Elaboración propia.

La hipótesis nula de este test es que no existe evidencia suficiente para decir que hay un punto de ruptura estructural en los datos. Como puede verse en la figura 28, el p-value es menor al 5 %, por lo cual se puede concluir con un nivel de confianza del 95 % que la serie de tiempo presenta un cambio estructural en el punto propuesto.

Modeling

Esta sección detallará los elementos importantes que fueron considerados en el modelamiento de las redes LSTM, junto con ciertas modificaciones hechas a las bases de conocimiento.

Diseño de Estructuras de Redes Neuronales

El diseño de las redes neuronales tuvo dos elementos importantes que fueron analizados. El primero de ellos fue la optimización de hiperparámetros de cada red construida, las redes neuronales tienen 4 hiperparámetros importantes a optimizar principalmente, la cantidad de capas de neuronas, la cantidad de neuronas por capa, la función de activación de las neuronas y la cantidad de veces que la data de input entrará a la red o que será recorrida por la red (Epochs). Un quinto hiperparámetro importante considerado en este estudio fue la decisión entre si la red debía entrenarse recorriendo la data en una sola dirección (de los registros más antiguos a los más actuales) o en ambas direcciones (Bidirectional LSTM o BLSTM).

Estos hiperparámetros se optimizaron por medio de las funciones de la librería ‘Keras

Tuner’, especializado en la optimización de redes LSTM de la librería ‘Keras’. Estas funciones buscan el mejor set de hiperparámetros, dentro de un espectro dado, minimizando el RMSE obtenido por iteración (tomando hiperparámetros de forma aleatoria) considerando los datos de validación, estos datos se obtienen a partir de la división de los datos de entrenamiento, una parte queda para el entrenamiento (75 % de los datos de entrenamiento) y la otra (25 % de los datos de entrenamiento) para la validación a partir de curvas de aprendizaje. El foco dado al uso de las funciones de ‘Keras Tuner’ dentro de esta investigación fue el poder vislumbrar aproximadamente el mejor set de hiperparámetros sin el propósito de llegar a un óptimo (se entrena a la red con un sólo epoch por iteración), esto para disminuir los tiempos en el diseño de la red, de esta forma se toma a este set resultante como un punto de partida para el diseño. De todas formas, los cambios en el diseño a partir de este set de hiperparámetros propuesto fueron siempre pocos. El número de epochs óptimo, como se explicó en secciones anteriores, fue optimizado a partir de las curvas de aprendizaje (usando también los datos de validación) durante el mismo entrenamiento de las redes.

El segundo elemento importante a considerar en el diseño fue el objetivo de la red, esto en cuanto su nivel de especialización o generalización. Las redes neuronales son un tipo de IA que puede resolver problemas específicos muy bien de acuerdo con la literatura, para entrenarse la red aprende de los datos con el fin de volverse muy bueno ejecutando una tarea, se especializa en ella, lo que se propone entonces es determinar hasta que nivel es conveniente especializar a la red considerando el rendimiento alcanzado y la dificultad para implementarse a niveles productivos para una cartera de productos grande, esto con el fin de entregar recomendaciones para el desarrollo de modelos multivariantes usando LSTM en entornos productivos para industrias con uso intensivo de activos.

El nivel de especialización se analizó en dos dimensiones, el primero considera la forma en como se pueden implementar este tipo de modelos con un amplio catálogo de productos, en ese sentido se consideraron sólo dos opciones, diseñar redes para cada grupo de SKUs con patrones de series de tiempo similares (por grupos, por ejemplo con K-means clustering para

series de tiempo) o diseñar las redes tomando una cantidad ‘n’ de SKUs prioritarios al azar sin importar su relación y repetir esto hasta tener toda la cartera de productos modelada (por lotes). En este estudio, se diseñaron redes considerando los 5 SKUs para ejemplificar el caso ‘por lotes’, y también se diseñaron redes para cada grupo generado del clustering para ejemplificar el caso ‘por grupos’, por limitaciones de tiempo y técnicas sólo hay un lote y el agrupamiento considera los mismos SKUs de este lote. Esta dimensión intenta capturar la especialización por medio de la consideración o no de los grupos de SKUs similares dentro del modelo, si se consideran hay un mayor nivel de especialización de los modelos.

La segunda dimensión se relaciona con la cantidad de valores pronosticados entregados por la red, si la red entrega el pronóstico de un SKU (uni-objetivo), sólo considerará este valor para minimizar el error alcanzado. Si la red entrega el pronóstico de todos los SKUs (multi-objetivo), la red considerará todos estos valores para minimizar el error alcanzado. Esta elección tiene relación con el nivel de especialización en el sentido de que, dada la cantidad de datos, el modelo debería tener un mejor rendimiento enfocándose en un sólo valor, especializándose en pronosticar una serie de tiempo.

Las KB que servirán como input para las redes entonces son tres. Dos para cada grupo de SKUs (0 y 1) junto con sus variables externas con mayor relación y la tercera considerando todos los SKUs (el lote) junto con todas las variables externas que no fueron descartadas. Por último, dado que el data set no es grande, se optó por aplicar una estrategia para el modelado incluyendo a las variaciones porcentuales diarias de las cantidades vendidas de los SKUs como variables extras en las KB, esto para replicar la componente ‘d’ de los modelos ARIMA y así poder ayudar a la red. Esta estrategia tuvo buenos resultados en trabajos anteriores de carácter confidencial pero con contextos similares por lo que se decidió incluirla. Las KBs utilizadas se ilustran a continuación.

Figura 29

KB del grupo 0.

	85123A	20727	Feriado	Es Fin de semana	Lunes	Martes	Jueves	% dif SKU 85123A	% dif SKU 20727
2009-12-02	5.598422	4.343805	0.0	0	0.0	0.0	0.0	0.057158	1.208311
2009-12-03	5.883322	3.465736	0.0	0	0.0	0.0	1.0	0.284900	-0.878070
2009-12-04	5.117994	4.276666	0.0	0	0.0	0.0	0.0	-0.765329	0.810930
2009-12-05	4.564348	2.302585	0.0	1	0.0	0.0	0.0	-0.553646	-1.974081
2009-12-06	5.327876	1.386294	0.0	1	0.0	0.0	0.0	0.763528	-0.916291

Nota. Esta figura muestra los primeros 5 registros de la KB para el grupo de SKUs ‘0’.
Elaboración propia.

Figura 30

KB del grupo 1.

	850998	20725	21212	Festivo	Feriado	Es Fin de semana	Martes	Miércoles	Jueves	% dif SKU 850998	% dif SKU 20725	% dif SKU 21212
2009-12-02	5.505332	2.484907	6.395262	0	0.0	0	0.0	1.0	0.0	0.769133	-1.673976	0.914623
2009-12-03	4.962845	3.433987	5.247024	0	0.0	0	0.0	0.0	1.0	-0.542487	0.949081	-1.148238
2009-12-04	6.375025	4.290459	5.192957	0	0.0	0	0.0	0.0	0.0	1.412180	0.856472	-0.054067
2009-12-05	3.688879	-1.000000	3.871201	0	0.0	1	0.0	0.0	0.0	-2.686145	-5.290459	-1.321756
2009-12-06	2.890372	3.258097	4.634729	0	0.0	1	0.0	0.0	0.0	-0.798508	4.258097	0.763528

Nota. Esta figura muestra los primeros 5 registros de la KB para el grupo de SKUs ‘1’.
Elaboración propia.

Figura 31*KB para el caso por lotes.*

	20725	20727	21212	85099B	85123A	Festivo	Feriado	Es Fin de semana	Martes	Miércoles	Jueves	% dif SKU 20725	% dif SKU 20727	% dif SKU 21212	% dif SKU 85099B	% dif SKU 85123A
2009-12-02	2.484907	4.343805	6.395262	5.505332	5.598422	0	0.0	0	0.0	1.0	0.0	-1.673976	1.208311	0.914623	0.769133	0.057158
2009-12-03	3.433987	3.465736	5.247024	4.962845	5.883322	0	0.0	0	0.0	0.0	1.0	0.949081	-0.878070	-1.148238	-0.542487	0.284900
2009-12-04	4.290459	4.276666	5.192957	6.375025	5.117994	0	0.0	0	0.0	0.0	0.0	0.856472	0.810930	-0.054067	1.412180	-0.765329
2009-12-05	-1.000000	2.302585	3.871201	3.688879	4.564348	0	0.0	1	0.0	0.0	0.0	-5.290459	-1.974081	-1.321756	-2.686145	-0.553646
2009-12-06	3.258097	1.386294	4.634729	2.890372	5.327876	0	0.0	1	0.0	0.0	0.0	4.258097	-0.916291	0.763528	-0.798508	0.763528

Nota. Esta figura muestra los primeros 5 registros de la KB para el caso por lotes. Elaboración propia.

Por último, para la KB del caso por lotes (Fig. 31) se descartó la variable externa ‘Lunes’ con el fin de reducir las variables del modelo y optimizar el uso de la capacidad computacional disponible. Esta variable es la que tenía la más baja correlación (en magnitud) con respecto las variables internas.

Evaluation

Esta sección detallará la forma en la cual se elegirá al mejor modelo posteriormente en *Resultados* con el fin de cumplir con los objetivos propuestos en esta memoria.

Elección del Mejor Modelo

Dado que el diseño de las redes en cuanto a la optimización de los hiperparámetros es un proceso mayormente automatizado, la elección del modelo esta sujeto al segundo elemento del diseño, el nivel de especialización de los modelos. Por lo tanto, los modelos candidatos se pueden resumir en la siguiente figura.

Figura 32

Cuadro resumen de los modelos candidatos.

	Por Lotes	Por Grupos
Uni-Objetivo	Modelo 1 (ULO)	Modelo 2 (UGR)
Multi-Objetivo	Modelo 3 (MLO)	Modelo 4 (MGR)

Nota. Esta figura muestra los modelos candidatos a evaluar, sus características y los nombres que los identificarán entre paréntesis. Elaboración propia.

El mejor modelo se escogerá considerando el RMSE, el MAE y el R2 ajustado alcanzado. De estas 3 métricas, la que tendrá mayor relevancia es el RMSE, ya que le da más peso a los errores grandes y a valores atípicos, y los datos están limpios por lo que su uso realmente capturará la presencia de errores significativamente grandes. En segundo lugar de relevancia, se considerará al MAE por su menor robustez. Por último, el R2 ajustado será tomado como indicador de bondad de ajuste del pronóstico con respecto al valor real.

Resultados

En esta sección se detallarán los resultados obtenidos de cada modelo, junto con los gráficos obtenidos y los errores alcanzados. Como recordatorio, para el entrenamiento de las redes se tomó el 70 % de los datos y el 30 % restante para su testeo (ventas desde mayo a diciembre del 2011). Las funciones de activación candidatas para las LSTM fueron ‘RELU’ y ‘sigmoide’, el algoritmo de optimización de todas las redes fue ‘adam’ y el error cuadrático medio (MSE) fue usado como función de pérdida o error a minimizar en todas las redes desarrolladas.

Modelos ULO

Los pronósticos usando modelos de redes uni-objetivos y por lotes (ULO) no dieron resultados aceptables, incluso desde la optimización de hiperparámetros. Al término del entrenamiento sólo entregaban como pronóstico un valor constante a través del tiempo, por lo cual se descartaron directamente y no se tomarán en consideración desde este punto.

Modelos UGR

Los pronósticos usando modelos de redes uni-objetivos y por grupos (UGR) si lograron entregar resultados aceptables. Los modelos de pronóstico en este caso usaron diferentes KBs de acuerdo con el grupo al cual pertenecía el SKU a pronosticar. Para cada SKU se optimizaron los hiperparámetros y se entrenó a la red, por lo tanto, se generaron 5 redes neuronales. Los resultados de estas redes se muestran a continuación.

Figura 33*Resultados modelos UGR.*

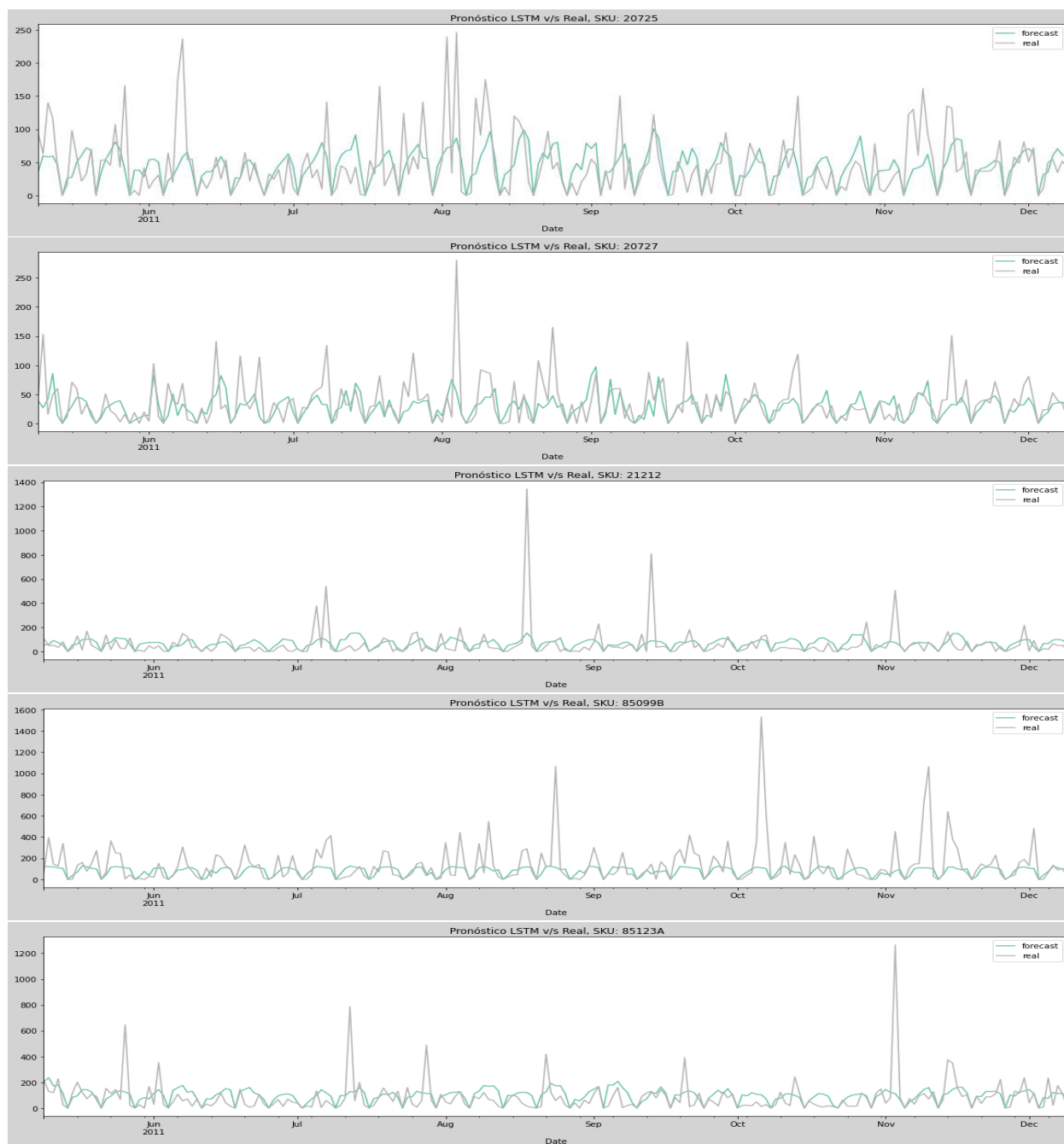
SKU	RMSE	MAE	R2 ajustado	Capas Ocultas	Neuronas por Capa	Función de activación	Epochs	Bidireccional	N° parámetros
20725	42,09	27,95	0,186	3	[352, 32, 96]	RELU	20	SI	1.600.321
20727	34,18	20,37	0,144	2	[128, 256]	RELU	60	NO	485.601
21212	118,86	55,25	0,0727	3	[352, 32, 96]	RELU	20	SI	1.600.321
85099B	184,19	92,56	0,0358	3	[352, 32, 96]	RELU	20	SI	1.600.321
85123A	124,72	66,17	0,0642	2	[64, 256]	RELU	37	NO	361.441

Nota. Esta figura muestra los hiperparámetros elegidos en la optimización, el RMSE, MAE y R2 ajustado alcanzado, y el número de parámetros de la red neuronal por SKU al usar modelos UGR. Elaboración propia.

Como se ve en la figura 33, los SKUs del grupo ‘1’ obtuvieron los mismos hiperparámetros en la optimización, a diferencia del grupo ‘0’. Sin embargo, los SKUs de este último grupo mencionado tienen redes con hiperparámetros muy similares. La columna ‘Neuronas por Capa’ indica la cantidad de neuronas de cada capa oculta de la red. Todos los pronósticos tienen una bondad de ajuste positiva y alcanzan un menor RMSE a los de la base de comparación (ver Fig. 27). Los gráficos con los pronósticos por SKU se exponen a continuación.

Figura 34

Pronóstico por SKU usando modelos UGR.



Nota. Esta figura muestra el pronóstico de ventas obtenido usando modelos UGR y las ventas reales por SKU desde mayo a diciembre del 2011. Elaboración propia.

Modelos MLO

Los pronósticos usando modelos de redes multi-objetivos y por lotes (MLO) utilizaron el mismo KB y sólo se diseñó una red neuronal para el pronóstico de todos los SKUs. Recordar que, al ser multi-objetivo, el modelo retorna un valor pronosticado para los 5 SKUs, a partir de estos valores la red se entrena para disminuir el MSE. Los resultados se muestran a continuación.

Figura 35

Resultados modelos MLO.

SKU	RMSE	MAE	R2 ajustado	Capas Ocultas	Neuronas por Capa	Función de activación	Epochs	Bidireccional	N° parámetros
20725	43,60	25,64	0,127	2	[480, 192]	RELU	20	SI	5.075.333
20727	39,09	23,25	-0,120	2	[480, 192]	RELU	20	SI	5.075.333
21212	126,25	66,62	-0,0462	2	[480, 192]	RELU	20	SI	5.075.333
85099B	175,13	89,25	0,128	2	[480, 192]	RELU	20	SI	5.075.333
85123A	135,21	87,32	-0,0998	2	[480, 192]	RELU	20	SI	5.075.333

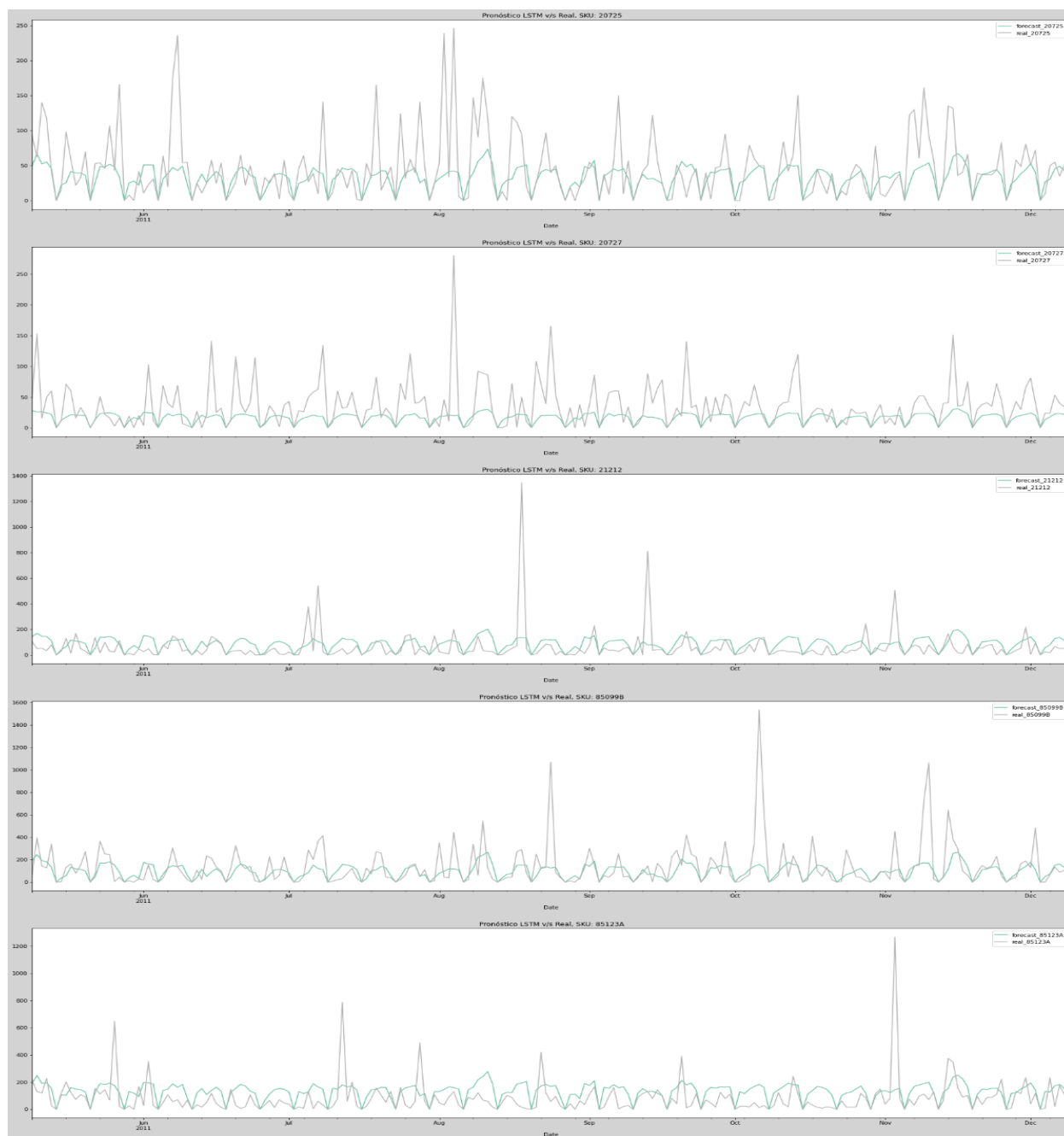
Nota. Esta figura muestra los hiperparámetros elegidos en la optimización, el RMSE, MAE y R2 ajustado alcanzado, y el número de parámetros de la red neuronal por SKU al usar modelos MLO. Elaboración propia.

La red desarrollada tiene una cantidad de parámetros elevada (ver Fig. 35), lo cual puede explicarse por la cantidad de variables de input consideradas y por el uso de BLSTM, lo cual incrementa la cantidad de parámetros ya que recorre en ambas direcciones las series de tiempo. Al mismo tiempo, es posible observar que se consideran pocas capas ocultas pero con una gran cantidad de neuronas (celdas, específicamente) LSTM por capa.

Sólo dos pronósticos obtienen bondades de ajuste positivas. En relación con el RMSE y el MAE, sólo 2 SKUs obtienen mejores pronósticos que el benchmark ('85123A' y '85099B'), los otros 3 obtuvieron resultados similares. De lo anterior se puede entender que la red no pudo evitar la especialización y se ajustó mejor a un par de SKUs, los cuales logran una bondad de ajuste positiva ('20725' y '85099B'). Los gráficos con los pronósticos por SKU se exponen a continuación.

Figura 36

Pronóstico por SKU usando modelos MLO.



Nota. Esta figura muestra el pronóstico de ventas obtenido usando modelos MLO y las ventas reales por SKU desde mayo a diciembre del 2011. Elaboración propia.

Modelos MGR

Los pronósticos usando modelos de redes multi-objetivos y por grupos (MGR) utilizaron diferentes KBs de acuerdo con el grupo al cual pertenecía el SKU a pronosticar, al igual que los modelos UGR. Como son redes multi-objetivo, entregan pronósticos de más de un SKU simultáneamente. Por otro lado, también se consideraron los grupos generados del clustering. En consecuencia, sólo se desarrollaron 2 redes neuronales. Los resultados de estas redes se muestran a continuación.

Figura 37

Resultados modelos MGR.

SKU	RMSE	MAE	R2 ajustado	Capas Ocultas	Neuronas por Capa	Función de activación	Epochs	Bidireccional	N° parámetros
20725	42,00	26,98	0,190	3	[352, 32, 96]	RELU	20	SI	1.600.707
20727	40,48	24,55	-0,201	2	[128, 256]	RELU	70	NO	485.858
21212	123,51	64,08	-0,00125	3	[352, 32, 96]	RELU	20	SI	1.600.707
85099B	176,27	89,66	0,117	3	[352, 32, 96]	RELU	20	SI	1.600.707
85123A	126,44	68,90	0,0382	2	[128, 256]	RELU	70	NO	485.858

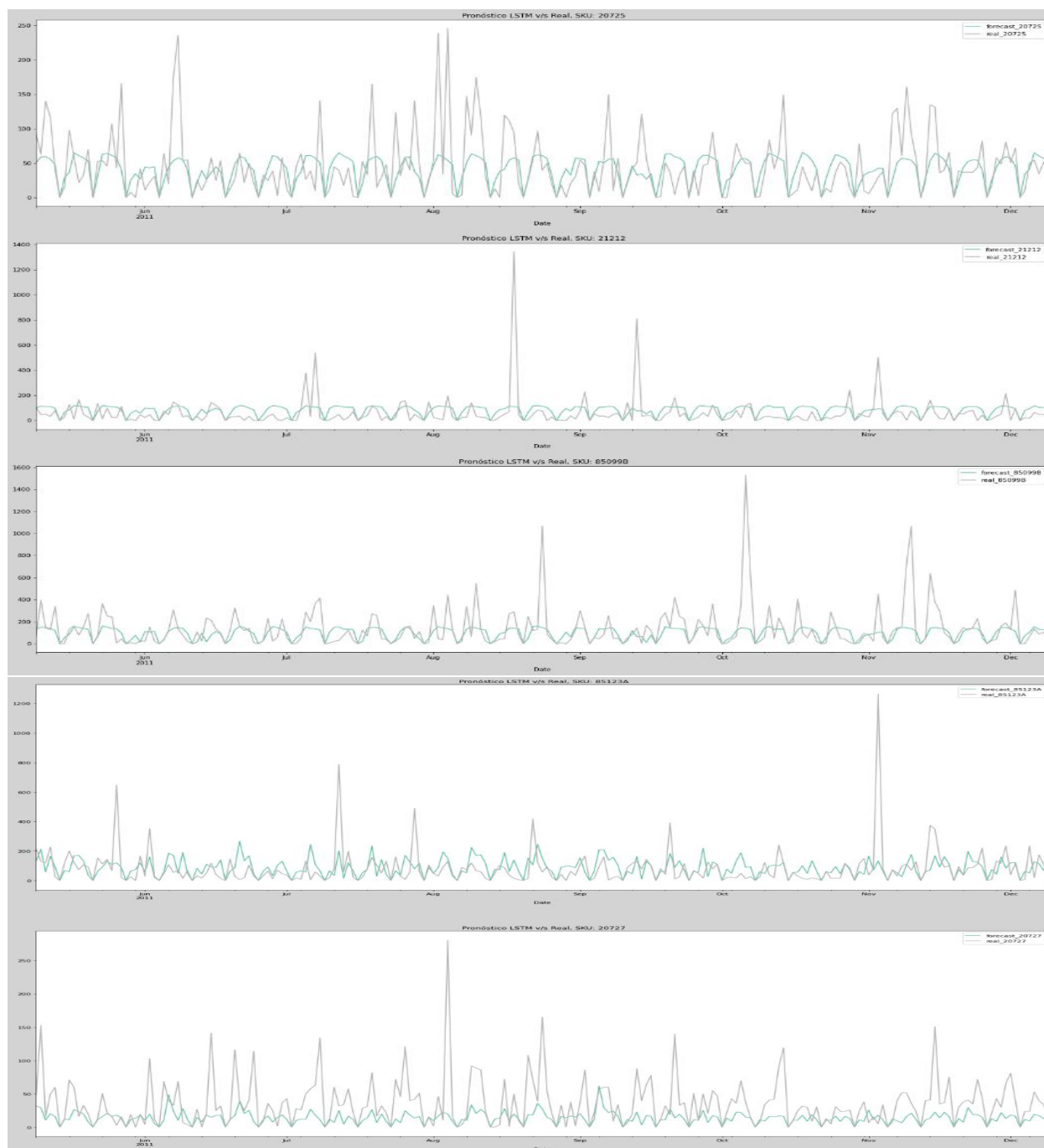
Nota. Esta figura muestra los hiperparámetros elegidos en la optimización, el RMSE, MAE y R2 ajustado alcanzado, y el número de parámetros de la red neuronal por SKU al usar modelos MGR. Elaboración propia.

La mayoría de los pronósticos logran tener un menor RMSE en comparación al benchmark y una bondad de ajuste positiva o muy cercana a cero (ver Fig. 37), lo cual es una mejora si se compara con los resultados obtenidos de los modelos MLO (ver Fig. 35).

Los resultados del grupo ‘0’ (‘20727’ y ‘85123A’) indican que la red se ajustó mejor al SKU ‘85123A’. Hubo una especialización de la red en pronosticar este SKU. Esto podría deberse a que los SKUs de este grupo en particular no son lo suficientemente similares entre sí, existiendo un sesgo producto del clustering, lo cual también explicaría las diferencias entre las redes desarrolladas para los SKUs de este grupo en los modelos UGR. Otro punto a destacar es la optimización de los hiperparámetros, para el grupo ‘1’ (‘20725’, ‘21212’ y ‘85099B’) es idéntica a la obtenida en los modelos UGR (ver Fig. 33), mientras que para el grupo ‘0’ es sólo similar. Los gráficos con los pronósticos por SKU se exponen a continuación.

Figura 38

Pronóstico por SKU usando modelos MGR.



Nota. Esta figura muestra el pronóstico de ventas obtenido usando modelos MGR y las ventas reales por SKU desde mayo a diciembre del 2011. Elaboración propia.

Mejor Modelo

A continuación, se presentará un consolidado de los resultados obtenidos por tipo de modelo con el fin de determinar el mejor modelo por SKU y de forma global.

Figura 39

Consolidado de los resultados obtenidos por modelo.

SKU	MODELOS UGR			MODELOS MLO			MODELOS MGR			BENCHMARK (ARIMA + GARCH)		
	RMSE	MAE	R2 ajustado	RMSE	MAE	R2 ajustado	RMSE	MAE	R2 ajustado	RMSE	MAE	R2 ajustado
20725	42,09	27,95	0,186	43,60	25,64	0,127	42,00	26,98	0,190	43,41	28,36	0,140
20727	34,18	20,37	0,144	39,09	23,25	-0,120	40,48	24,55	-0,201	34,73	22,97	0,0831
21212	118,86	55,25	0,0727	126,25	66,62	-0,0462	123,51	64,08	-0,00125	126,13	45,73	-0,0347
85099B	184,19	92,56	0,0358	175,13	89,25	0,128	176,27	89,66	0,117	192,41	116,80	-0,0520
85123A	124,72	66,17	0,0642	135,21	87,32	-0,0998	126,44	68,90	0,0382	264,22	164,22	-3,182

Nota. Esta figura muestra el RMSE, MAE y R2 ajustado alcanzado en cada modelo por SKU y su benchmark respectivo, destacado en verde se encuentra el modelo con menor RMSE. Elaboración propia.

Como puede verse en la figura 39, la mayoría de los SKUs obtienen el menor RMSE y la mayor bondad de ajuste al usar modelos UGR, el SKU ‘20725’ en particular tiene una diferencia muy mínima entre sus resultados al usar modelos UGR y MGR, por lo que podría considerarse como un empate, esto puede explicarse por lo idéntico de los hiperparámetros de las redes en ambos modelos (ver Fig. 33 y 37). Algo similar ocurre con el SKU ‘85099B’, sus resultados al usar modelos MLO y MGR presentan diferencias muy mínimas, este SKU fue el único que tuvo resultados sobresalientes al usar modelos MLO, lo cual podría explicarse por una especialización de la red en pronosticar este SKU, aun así, sus métricas son muy cercanas a las obtenidas por los modelos MGR.

De forma global, los modelos MLO no logran buenos resultados en comparación con el benchmark, superando el RMSE en sólo 2 SKUs. Dado lo anterior, se descarta a los modelos MLO de los candidatos a mejor modelo.

Con respecto a los modelos UGR, son los únicos que superan los resultados obtenidos en el benchmark para todos los SKUs, si bien no son los modelos con un menor RMSE para

todos los SKUs, son los que entregan los resultados más consistentes. Con estos modelos siempre se obtienen pronósticos con bondad de ajuste positivo y con un RMSE menor al obtenido en el benchmark.

Los modelos MGR tienen un gran potencial, para la mayoría de los SKUs se logra obtener un menor RMSE al benchmark y bondad de ajuste positivo. En algunos SKUs logra un menor RMSE que los modelos UGR, aunque esto se puede deber a una especialización de las redes, los SKUs involucrados ('20725' y '85099B') son del grupo '1', el cual obtuvo en todos sus SKUs un menor RMSE y una mejor bondad de ajuste que el benchmark, por lo tanto no hubo especialización de la red. El grupo '0' tuvo problemas con el SKU '20727', el cual obtuvo un peor pronóstico que el benchmark, la red evidentemente se especializó en el pronóstico del SKU '85123A', que obtuvo valores similares en las métricas alcanzadas en comparación con los modelos UGR. Esto último se puede deber a sesgos producto de un mal agrupamiento, lo que explicaría la especialización en la red del grupo '0', las series de tiempo no eran lo suficientemente similares como para agruparse.

A nivel agregado, es decir, sumando el error RMSE obtenido por SKU para cada modelo evaluado, todos los modelos usando redes LSTM lograron disminuir el error propuesto por el benchmark, el detalle de esto se muestra a continuación.

Figura 40

Variación del RMSE agregado por Modelo.

Modelo	RMSE Total	Disminución del RMSE con respecto a Benchmark (%)
UGR	504,04	23,73
MLO	519,28	21,43
MGR	508,70	23,03
Benchmark	660,90	-

Nota. Esta figura muestra el porcentaje de disminución del RMSE agregado de cada modelo con respecto al benchmark. Elaboración propia.

En conclusión, los pronósticos usando modelos de redes LSTM uni-objetivos y que consideran los grupos formados con el clustering K-means para series de tiempo (UGR) son

las mejores a nivel agregado, aunque son los que requieren mayor tiempo de desarrollo. Los pronósticos multi-objetivos y clustering (MGR) alcanzaron métricas similares a UGR, por lo cual tienen un buen potencial considerando el menor tiempo de desarrollo necesario, requiere de menos redes.

Otro aspecto relevante es el resultado obtenido en el SKU '85123A', el cual tiene un cambio estructural en su serie de tiempo que produjo un RMSE alto en el benchmark (observable también en su gráfico, ver Fig. 42). Al usar LSTM, el RMSE se redujo alrededor del 50 % en todos los modelos de redes considerados (UGR, MLO y MGR), esto es explicado por la capacidad de adaptación de las redes al momento de generarse el cambio estructural, lo cual se puede ver reflejado en los gráficos obtenidos y expuestos anteriormente (ver Fig. 34, 36 y 38), en ellos es posible observar que los modelos no están siguiendo un patrón de una serie de tiempo completamente diferente, como es el caso del benchmark. La bondad de ajuste también mejora al usar redes LSTM, lo cual conversa bien con las observaciones descritas anteriormente.

Conclusiones y Recomendaciones

Los pronósticos de demanda son un factor clave dentro de las organizaciones del tipo capital-intensivo y su impacto se extiende a todos los niveles de la misma. Sin embargo, el desarrollo de modelos de pronóstico dentro de la industria, en general, se han enfocado en el uso y mejora de modelos univariantes autorregresivos, como ARIMA, lo cual los hace muy susceptibles a errores producto de la incertidumbre, en donde existen múltiples variables que interactúan con la demanda. Por lo cual, se estudia el desarrollo de modelos multivariantes de pronósticos de demanda para estas industrias usando redes LSTM con el fin de disminuir el error alcanzado.

Primero, se logra diseñar una estrategia de agrupamiento de variables internas (series de tiempo de ventas) y externas (exógenas al negocio, como por ejemplo, influencia de festividades en los patrones de venta) que sean relevantes para robustecer los modelos. La estrategia se divide en dos partes, la primera se encarga de agrupar las variables internas por medio del uso del algoritmo de clustering K-means para series de tiempo. La segunda parte se encarga de agrupar las variables externas a través de una matriz de correlación, de esta forma se descartan aquellas variables externas con muy poca correlación con respecto a las variables a pronosticar (internas) y se buscan las variables externas más correlacionadas a cada grupo de variables internas generados en el clustering para el desarrollo de las KBs necesarias como input para las redes.

Luego, se diseñaron las redes neuronales de los modelos estudiados, considerando los agrupamientos de variables y el error de validación. Esto se hizo a partir del uso de funciones de optimización de hiperparámetros usando la librería ‘Keras Tuner’ y ‘Keras’ minimizando el error con respecto a los datos de validación. Por otro lado, también se consideraron los agrupamientos de variables formados considerando el nivel de especialización buscado para las redes neuronales. Esta especialización tiene dos dimensiones, la primera tiene que ver con la forma en como modelar un amplio catálogo de productos, se consideraron dos opciones, diseñar redes para productos con series de tiempo similares (por grupos) o

diseñar redes tomando ‘n’ productos prioritarios al azar y repetir hasta completar toda la cartera (por lotes). La segunda dimension tiene que ver con la cantidad de outputs que entrega la red, si la red entrega el pronóstico de un SKU es un modelo uni-objetivo y si el modelo entrega el pronóstico de múltiples SKUs es multi-objetivo, en este último caso la red se ajusta considerando el error de pronóstico de más de un SKU simultáneamente. A partir de estas dimensiones, se diseñan 4 tipos de modelos de redes LSTM (producto de las combinaciones de ambas dimensiones), los modelos ULO, UGR, MLO y MGR.

Por último, se distingue o escoge al mejor modelo, dentro de los 4 tipos propuestos, considerando el RMSE, MAE y R2 ajustado (bondad de ajuste) alcanzados al intentar pronosticar nuevos datos (ventas desde mayo a diciembre de 2011). La mejor alternativa son los modelos que usan redes LSTM uni-objetivos y que consideran en su KB sólo SKUs similares (UGR), tomando en consideración el RMSE y la bondad de ajuste alcanzados de forma agregada. Además, se toma como base de comparación o Benchmark a los pronósticos generados a partir de modelos de la familia ARIMA + GARCH desarrollados haciendo uso de las funciones de las librerías ‘pmdarima’ y ‘pyflux’, los cuales minimizan el AIC para encontrar al mejor modelo de la familia ARIMA y GARCH para cada SKU.

Todos los modelos de redes LSTM lograron tener un menor RMSE (de forma agregada) que el Benchmark, evidenciando su rendimiento superior a los modelos más convencionales (ARIMA + GARCH). Incluso, se logra evidenciar la buena adaptación de las redes LSTM frente a cambios estructurales en los patrones de ventas, lo cual habla de su buen rendimiento al enfrentarse a escenarios cambiantes e inciertos. Por lo tanto, se lograron desarrollar modelos multivariantes de pronósticos de demanda para industrias de capital intensivo por medio del entrenamiento de redes LSTM que disminuyeron el error de pronóstico en escenarios dinámicos y con gran incertidumbre.

Conclusiones Secundarias

Se propone una nueva estrategia para la identificación de productos prioritarios a modelar dentro de una gran cartera de productos utilizando una variación del análisis Jack Knife, usado originalmente en la gestión de activos físicos, en base al beneficio acumulado y la cantidad de días con ventas por SKU. Esta estrategia se contrasta con el método ‘Clasificación ABC’ usando K-means y obtiene muy similares resultados por lo que se comprueba su buen rendimiento para identificar productos prioritarios o críticos.

La investigación siguió la metodología fundamental de DS desarrollado por IBM capaz de proporcionar una estrategia de orientación, independiente de las tecnologías, los volúmenes de datos o los enfoques utilizados. A través de esta metodología se estableció la hoja de ruta para la elección, obtención, limpieza y pre-procesamiento de los datos utilizados, además de tener como foco a los objetivos a alcanzar durante toda la investigación, a partir de conocer la problemática y tener definido el enfoque analítico. La metodología sigue un camino con pasos iterativos que buscan la calidad de la solución desarrollada teniendo en el centro el problema de negocio a abordar por medio de objetivos y el enfoque analítico elegido. Gracias a esta metodología, la investigación pudo cumplir con los objetivos propuestos de forma ordenada y enfocada en el problema de negocio de inicio a fin. Bajo el mismo contexto, se destaca la necesidad de llevar a acabo una buena elección y manipulación de los datos para el correcto desarrollo de los modelos de ML y DL. Además, se comprueba que el uso de este tipo de herramientas de DS en el pronóstico de demanda logran una disminución en los errores considerable (más del 20 % en todos los modelos).

La investigación utilizó como ejemplo los datos de ventas de una empresa de retail e-commerce presente en múltiples países, con un robusto sistema logístico a disposición, para representar a una empresa que use activos intensivamente. Sin embargo, en diversos apartados de esta memoria se hace hincapié en que, tanto la metodología usada como las estrategias desarrolladas, pueden ser aplicadas en contextos propios de otras industrias con uso intensivo de activos como el pronóstico de la demanda de componentes críticos en minería.

Trabajos Futuros

Dentro de los modelos de redes LSTM evaluados se destacan los modelos multi-objetivo y que utilizan los agrupamientos de series de tiempo (MGR), estos modelos logran resultados bastante cercanos a los modelos UGR, incluso teniendo mejores resultados en algunos SKUs que los UGR. Las redes de ambos modelos (UGR y MGR) son bastante similares en cuanto a sus hiperparámetros y utilizan los mismos KBs. Por lo tanto, se presume que el menor rendimiento de los MGR con respecto a los UGR se debe al clustering, por lo cual sería muy relevante el estudio en profundidad de esta etapa con el fin de determinar si, efectivamente, los modelos MGR logran resultados iguales o mejores a los modelos UGR, en cuyo caso serían los mejores modelos por la eficiencia en su desarrollo.

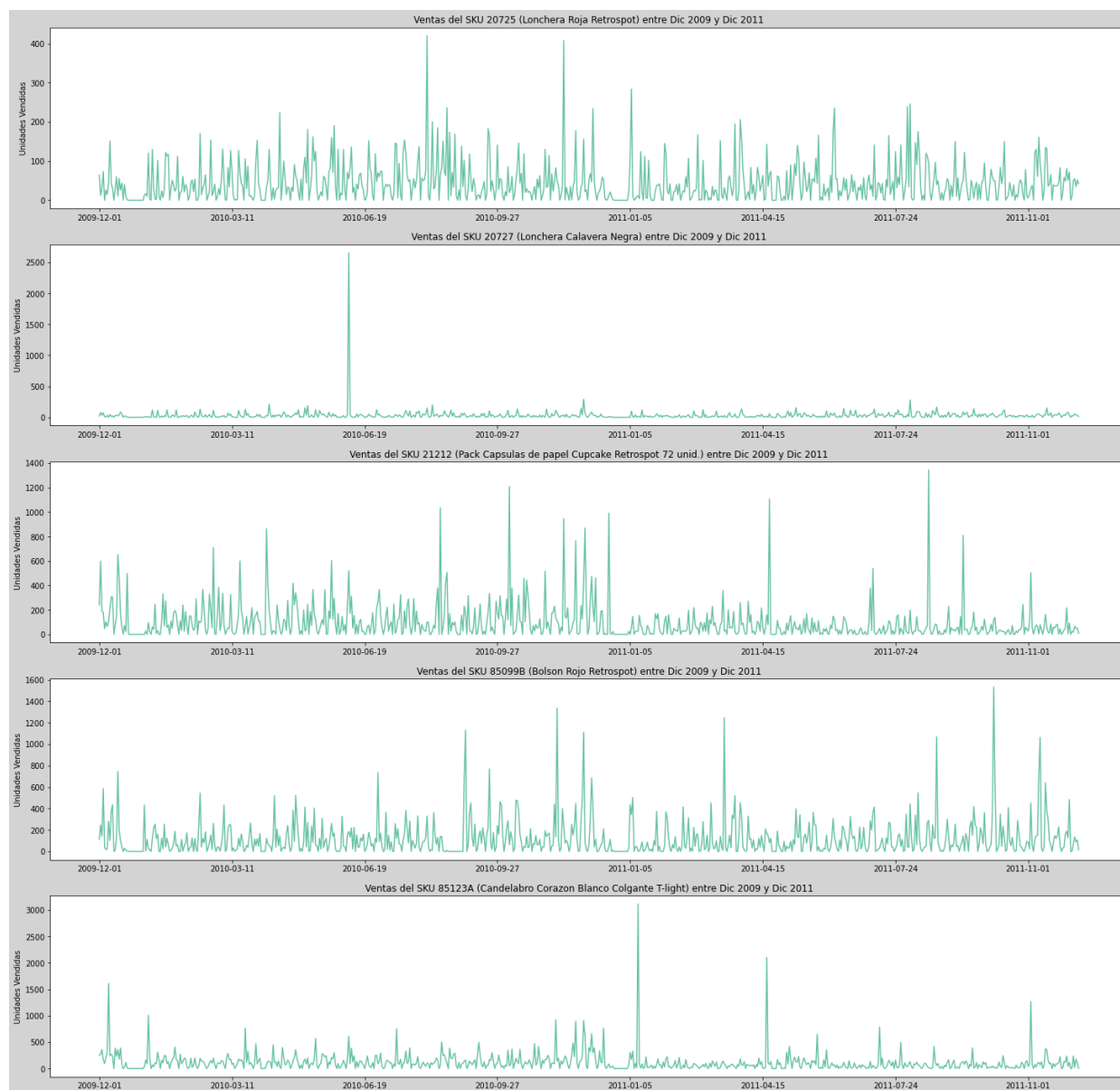
El ejemplo desarrollado en esta memoria sólo utilizó variables externas binarias o dummy, sería interesante una investigación con respecto al uso de variables continuas como variables externas. Por ejemplo, utilizando los datos de telemetría (temperatura, humedad, presión, rugosidad, etc.) para el pronóstico de demanda de componentes críticos en minería.

Otra investigación interesante a desarrollar sería el uso de redes LSTM en conjunto con modelos ARIMA con el fin de robustecer aun más los modelos de pronóstico. Por ejemplo, destinando los modelos ARIMA a establecer un patrón simple para usar de base, al cual se le agregarían los pronósticos de volatilidad o variación usando LSTM para armar el pronóstico de demanda completo.

Anexos

Figura 41

Anexo 1: Gráficos de ventas por día de los SKUs prioritarios.

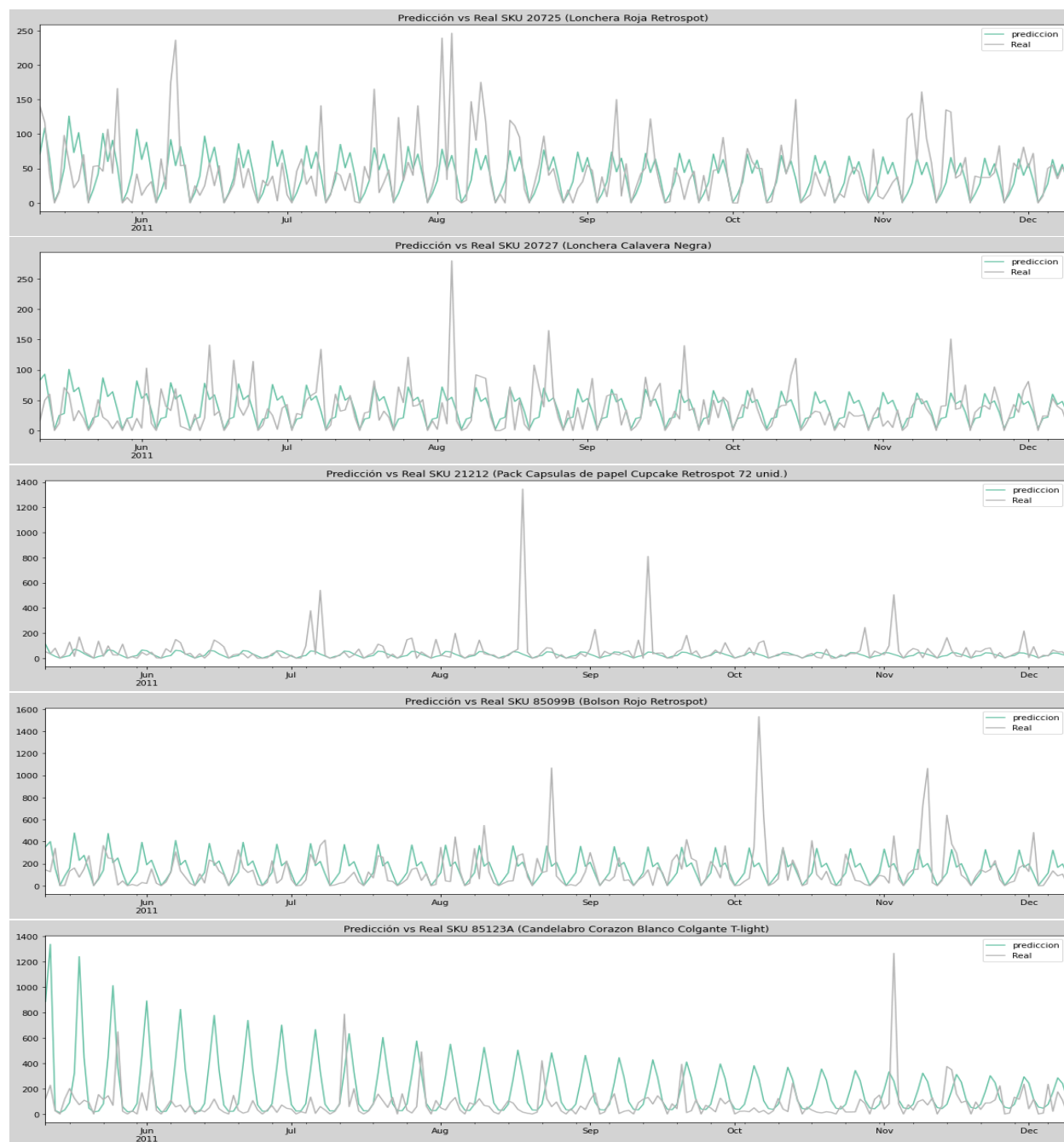


Nota. Esta figura muestra los gráficos de ventas de cada SKU escogido entre 2009 y 2011.

Elaboración propia.

Figura 42

Anexo 2: Pronóstico Benchmark por SKU.



Nota. Esta figura muestra el pronóstico de ventas obtenido como benchmark usando modelos ARIMA + GARCH y las ventas reales por SKU desde mayo a diciembre del 2011. Elaboración propia.

Referencias

- Abbasimehr, H. ., Shabani, M. . & Yousefi, M. . (2020). An optimized model using LSTM network for demand forecasting. *Computers and Industrial Engineering*, 143, 106435. <https://doi.org/10.1016/j.cie.2020.106435>
- Alicke, K., Rexhausen, D. & Seyfert, A. (2019, 14 de enero). *Supply Chain 4.0 in consumer goods*. <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/supply-chain-4-0-in-consumer-goods>
- Amendola, L. . J. . (2015, 20 de mayo). *Gestión Integral de Activos Físicos: Business Solutions Asset Management Maintenance Framework (Spanish Edition)* (3.^a ed.). PMM Institute for Learning.
- Ballou, R.H. (2003). *Business Logistics: Supply Chain Management* (5th). Prentice Hall College Div.
- Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. & Seaman, B. (2019, diciembre). Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology. En T. Gedeon, K. Wong & M. Lee (Eds.), *Neural Information Processing* (pp. 462-474). Springer Publishing. https://doi.org/10.1007/978-3-030-36718-3_39
- Chandriah, K. K. & Naraganahalli, R. V. (2021). RNN / LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-021-10913-0>
- Chen, C., Xu, X., Zou, B., Peng, H. & Li, Z. (2021). Optimal decision of multiobjective and multiperiod anticipatory shipping under uncertain demand: A data-driven framework. *Computers Industrial Engineering*, 159, 107445. <https://doi.org/10.1016/j.cie.2021.107445>
- Chen, D., Sain, S. L. & Guo, K. . (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal*

- of Database Marketing amp; Customer Strategy Management*, 19(3), 197-208. <https://doi.org/10.1057/dbm.2012.17>
- Chen, L., Shan, W. & Liu, P. (2021). Identification of concrete aggregates using K-means clustering and level set method. *Structures*, 34, 2069-2076. <https://doi.org/10.1016/j.istruc.2021.08.048>
- Chen, Y., Yi, H. ., Liao, C. ., Huang, P. . & Chen, Q. . (2021). Visual measurement of milling surface roughness based on Xception model with convolutional neural network. *Measurement*, 186, 110217. <https://doi.org/10.1016/j.measurement.2021.110217>
- Cheng, Y., Xu, C., Mashima, D., Thing, V. L. & Wu, Y. (2017). PowerLSTM: power demand forecasting using long short-term memory neural network. *International Conference on Advanced Data Mining and Applications*, 727-740.
- Dhanachandra, N., Manglem, K. & Chanu, Y. J. (2015). Image Segmentation Using K - means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, 54, 764-771. <https://doi.org/10.1016/j.procs.2015.06.090>
- Du, K. L. & Swamy, M. N. S. (2013). Perceptrons. *Neural Networks and Statistical Learning*, 67-81. https://doi.org/10.1007/978-1-4471-5571-3_3
- Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics*, 21, 768-769.
- Guerrero, J. F. J., Fernández, R. S. & Abad, J. C. G. (2006). La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turístico. *Revista Europea de Dirección y Economía de la Empresa*, 15(3), 185-198.
- Haberleitner, H., Meyr, H. & Taudes, A. (2010). Implementation of a demand planning system using advance order information. *International Journal of Production Economics*, 128(2), 518-526. <https://doi.org/10.1016/j.ijpe.2010.07.003>
- Jeong, H. Y., Song, B. D. & Lee, S. . (2022). Optimal scheduling and quantitative analysis for multi-flying warehouse scheduling problem: Amazon airborne fulfillment center.

- Transportation Research Part C: Emerging Technologies*, 143, 103831. <https://doi.org/10.1016/j.trc.2022.103831>
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R. & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, 24(7), 881-892.
- Knights, P. F. (2001). Rethinking Pareto analysis: maintenance applications of logarithmic scatterplots. *Journal of Quality in Maintenance Engineering*, 7(4), 252-263. <https://doi.org/10.1108/13552510110407041>
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129-137.
- Moroff, N. U., Kurt, E. & Kamphues, J. (2021). Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models. *Procedia Computer Science*, 180, 40-49. <https://doi.org/10.1016/j.procs.2021.01.127>
- Nguyen, H., Tran, K., Thomassey, S. & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57, 102282. <https://doi.org/10.1016/j.ijinfomgt.2020.102282>
- Online retail dataset*. (2019, 20 de noviembre). Consultado el 9 de octubre de 2022, desde <https://www.kaggle.com/datasets/lakshmi25npathi/online-retail-dataset>
- Pacella, M. . & Papadia, G. . (2021). Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management. *Procedia CIRP*, 99, 604-609. <https://doi.org/10.1016/j.procir.2021.03.081>
- Rosenblatt, F. . (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. <https://doi.org/10.1037/h0042519>

- Rosienkiewicz, M., Chlebus, E. & Detyna, J. (2017). A hybrid spares demand forecasting method dedicated to mining industry. *Applied Mathematical Modelling*, 49, 87-107. <https://doi.org/10.1016/j.apm.2017.04.027>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. <https://doi.org/10.1038/323533a0>
- Tavenard, R. . (2021, 8 de junio). *An introduction to Dynamic Time Warping*. Consultado el 23 de octubre de 2022, desde <https://rtavenar.github.io/blog/dtw.html>
- Tsao, Y. C., Chen, Y. K., Chiu, S. H., Lu, J. C. & Vu, T. L. (2021). An innovative demand forecasting approach for the server industry. *Technovation*, 102371. <https://doi.org/10.1016/j.technovation.2021.102371>
- Tversky, A. & Kahneman, D. (1979). An analysis of decision under risk. *Econometrica*, 47(2), 263-292.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 5998-6008.
- Voronova, O. . (2022). Improvement of warehouse logistics based on the introduction of lean manufacturing principles. *Transportation Research Procedia*, 63, 919-928. <https://doi.org/10.1016/j.trpro.2022.06.090>
- Wang, X. . & Zhang, W. . (2021). Anti-occlusion face recognition algorithm based on a deep convolutional neural network. *Computers Electrical Engineering*, 96, 107461. <https://doi.org/10.1016/j.compeleceng.2021.107461>
- Zhang, G. P. (2001). An investigation of neural networks for linear time-series forecasting. *Computers Operations Research*, 28(12), 1183-1202. [https://doi.org/10.1016/s0305-0548\(00\)00033-2](https://doi.org/10.1016/s0305-0548(00)00033-2)