

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



FACULTAD DE INGENIERÍA EN ELECTRICIDAD Y COMPUTACIÓN

DEPARTAMENTO DE POSTGRADOS

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

MAGISTER EN CIENCIA DE DATOS

TEMA:

**“ANÁLISIS DE REDES NEURONALES RECURRENTES LSTM Y MODELOS DE VECTORES
AUTORREGRESIVOS VAR EN LA GENERACIÓN DE PRONOSTICOS DE SERIES DE TIEMPO
APLICADO A LAS VENTAS DE UNA EMPRESA DEDICADA A LA COMERCIALIZACIÓN DE
PRODUCTOS DE CONSUMO MASIVO”**

Presentado por:

ALEJANDRO DAVID OBANDO SANDOVAL

Guayaquil - Ecuador

Año: 2023

RESUMEN

Las ventas de productos de consumo masivo es una actividad que en años recientes ha mostrado un importante crecimiento dentro del entorno económico ecuatoriano, por tal motivo, las empresas que se dedican a esta actividad dedican especial atención a una eficiente y eficaz gestión de sus cadenas de suministros. Dentro de los macroprocesos que se incluyen en la gestión de cadenas de suministros, el primer lugar lo tiene la PLANIFICACION DE LA DEMANDA, este macroproceso consiste en poder anticipar de forma oportuna las necesidades de productos y/o servicios que el mercado requerirá a corto, mediano o largo plazo, a fin de poder desarrollar la planificación de demanda se requiere de una línea base de datos sobre la cual un equipo humano especializado pueda desarrollar tareas de análisis correspondientes y así lograr estructurar un conjunto de cifras como objetivos de ventas. Los pronósticos son la base para construir la planificación de la demanda.

En el presente proyecto de tesis se han utilizado técnicas estadísticas, específicamente los modelos de vectores autorregresivos (VAR) y técnicas de aprendizaje profundo, estas son las redes neuronales recurrentes con memoria a corto y largo plazo (LSTM) para generar proyecciones de ventas a corto plazo (14 días en el futuro), para llegar a la generación de dichos pronósticos, previamente se llevó a cabo el proceso de análisis exploratorio de datos y se utilizó el modelo K-MEANS para realizar una clasificación no supervisada de los artículos vendidos correspondientes a la muestra de datos utilizada en este proyecto, el resultado de la clasificación fue la identificación de 5 clusters de productos.

Los resultados obtenidos demuestran que para los 5 clústers de productos los pronósticos obtenidos con la aplicación de redes neuronales recurrentes con memoria a corto y largo plazo (LSTM) tienen un mejor desempeño frente a los pronósticos obtenidos con los modelos de vectores autorregresivos (VAR), esta afirmación se soporta en las métricas de error RMSE y MAPE calculadas para cada clúster y que serán presentadas en el capítulo #4 de presente documento.

ABSTRACT

Due to the fact that sales of mass consumer products are an activity that has shown substantial growth in recent years within the economic environment of Ecuador, companies that are engaged in this activity give special attention to the management of their supply chains in order to ensure that they are both efficient and effective. DEMAND PLANNING is at the top of the list when it comes to the macro-processes that are a part of supply chain management. In order to develop demand planning, a baseline of data is required, on which a specialized team can develop the corresponding analysis tasks, and thus structure a set of figures as sales targets. This macro-process consists of being able to anticipate in a timely manner the needs of products and/or services that the market will require in the short, medium, or long term; in order to develop demand planning, a baseline of data is required. The planning of supply and demand begins with forecasts.

In this thesis project, statistical techniques, specifically vector autoregressive models (VAR), and deep learning techniques, these are the recurrent neural networks with short and long term memory (LSTM) to generate short term sales projections (14 days in the future), have been used to arrive at the generation of such forecasts. Specifically, the generation of such forecasts has been accomplished through the application of these techniques. Previously, the exploratory data analysis procedure was carried out, and the K-MEANS model was used to do an unsupervised classification of the goods sold according to the data sample that was utilized in this project. The result of the classification was the identification of 5 product groups, and these clusters were based on the items that were sold.

This statement is supported by the RMSE and MAPE error metrics that were calculated for each cluster and that will be presented in chapter #4 of this document. The results obtained show that for the five product clusters, the forecasts obtained with the application of recurrent neural networks with short and long term memory (LSTM) have a better performance compared to the forecasts obtained with vector autoregressive models (VAR).

DEDICATORIA

Dedico este trabajo a Dios y a la Virgen, sin su bendición no me hubiese sido posible realizarlo.

A mi mamá, María Auxiliadora, por ser siempre un ejemplo de perseverancia, disciplina, rectitud y humildad, ella siempre será mi modelo a seguir.

A mi esposa María Fernanda, por cuidar de mí, por su compañía y apoyo incondicional.

A mis hijos, Angélica y Benjamín, con el deseo de que algún día lean este trabajo y les sirva de inspiración para sus propios proyectos.

A mi hermano Roberto, sé que llegarás muy lejos y cumplirás todas tus aspiraciones, nunca te detengas.

A todas las maravillosas personas que me brindaron su ayuda para que este proyecto haya sido concluido con éxito.

Alejandro Obando S.

AGRADECIMIENTO

A Dios y a la Virgen, por bendecir mi camino y permitirme finalizar este proyecto.

A mi mamá, María Auxiliadora, por su amor y por ser mi inspiración de crecimiento.

A mi esposa María Fernanda, por estar a mi lado demostrándome cada día su amor.

A mis hijos, Angélica y Benjamín, por ser mi motivación para seguir adelante cada día.

A todos los excelentes profesores con quienes tuve la gran experiencia de recibir clases durante este programa de maestría, pero en especial al profesor Phd. Sergio Baúz, de quien recibí valiosas aportaciones para mi desempeño profesional.

A Vicente, por su gran compañerismo.

Alejandro Obando S.

DECLARACION EXPRESA

Este Trabajo de Fin de Grado ha sido elaborado de forma respetuosa con los derechos de propiedad intelectual de terceros, como lo demuestran las citas que se incluyen en el documento y la inclusión de las fuentes de dichas citas en las referencias o bibliografías. Soy el único responsable de los hechos y doctrinas que se han expuesto en este Trabajo de Fin de Grado. Como consecuencia directa de ello, la totalidad de este trabajo se acredita a mi autoría. El legado intelectual del mismo es responsabilidad exclusiva de la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

Con la firma de esta declaración, reconozco que soy el único responsable del contenido, veracidad y amplitud del Trabajo de Grado al que se ha hecho referencia.



Firmado electrónicamente por:
ALEJANDRO DAVID
OBANDO SANDOVAL

Alejandro David Obando Sandoval

Autor

COMITÉ EVALUADOR

Phd. Sergio Bauz

TUTOR

MSc. Eduardo Cruz R.

REVISOR

ÍNDICE GENERAL

ÍNDICE GENERAL.....	VIII
LISTADO DE FIGURAS.....	XI
LISTADO DE TABLAS.....	XIV
CAPÍTULO 1.....	15
1. PLANTEAMIENTO DEL PROBLEMA.....	15
1.1 Descripción del Problema	15
1.2 Justificación	16
1.3 Objetivos.....	17
1.3.1 Objetivo General	17
1.3.2 Objetivos Específicos	17
1.4 Metodología:	17
1.5 Resultados Esperados:	18
1.6 Dataset:	19
CAPÍTULO 2.....	20
2. MARCO TEÓRICO Y REVISIÓN BIBLIOGRÁFICA.....	20
2.1 Análisis Exploratorio de Datos.....	20
2.2 K-MEANS para la generación de clústers de datos.....	21
2.2.1 Distancia de Manhattan:.....	22
2.2.2 Distancia Euclidiana:	22
2.2.3 Distancia de Minkowski:	22
2.2.4 Distancia de Mahalanobis:.....	22
2.3 Introducción a los pronósticos	23
2.4 Clasificación de métodos y tipos de pronósticos.	23
2.5 Las series temporales y sus componentes.....	25
2.5.1 Tendencia:	25
2.5.2 Estacionalidad:	25
2.5.3 Residuo:.....	26

2.6 Modelos de Vectores Autorregresivos VAR.....	26
2.7 Requerimientos generales de los modelos VAR.....	27
2.8 Transformación de la serie de tiempo	29
2.9 Evaluación de Raíz Unitaria	30
2.10 AIC versus BIC para la selección del orden del modelo	31
2.11 Causalidad de Granger	32
2.12 Fundamentos de Deep Learning	33
2.13 Redes Neuronales:	36
2.14 Redes Neuronales Recurrentes (RNN)	40
2.15 Redes Neuronales Recurrentes LSTM.....	41
2.16 Optimización de Hiperparámetros	47
2.17 Medidas de error del pronóstico	48
CAPÍTULO 3.....	50
3. DISEÑO E IMPLEMENTACIÓN	50
3.1 Introducción	50
3.2 El modelo de arquitectura tecnológica.....	50
3.3 Almacenamiento de los datos	51
3.4 Obtención de datos	52
3.5 Preprocesamiento y validación de datos	52
3.6 Generación de pronósticos y selección del mejor modelo	52
3.7 Diseño de Dashboard en Power BI para análisis de resultados	53
CAPÍTULO 4.....	55
4. ANÁLISIS DE RESULTADOS	55
4.1 Antecedentes	55
4.2 Análisis Exploratorio de datos	55
4.3 Resultados de los clústers de datos generados por el modelo K-Means.....	58
5. CONCLUSIONES Y RECOMENDACIONES	78
5.1 CONCLUSIONES	78

5.2 RECOMENDACIONES	80
6. REFERENCIAS.....	81
7. ANEXOS	84
Anexo 1.....	84

LISTADO DE FIGURAS

Figura 1	<i>Clasificación de los Métodos de Pronósticos</i>	24
Figura 2	<i>Componente de Tendencia en Series de Tiempo</i>	25
Figura 3	<i>Componente de Estacionalidad en Series de Tiempo</i>	26
Figura 4	<i>Componente de Residuos en Series de Tiempo</i>	26
Figura 5	<i>Proceso para Aplicación de Modelos VAR</i>	28
Figura 6	<i>Visualizaciones de la Función de Autocorrelación</i>	29
Figura 7	<i>Entrenamiento y Programación Explícita en Aprendizaje de Máquinas</i>	34
Figura 8	<i>Representación de capas en modelos de aprendizaje profundo</i>	35
Figura 9	<i>Inteligencia Artificial, Aprendizaje de Máquinas y Aprendizaje Profundo</i>	36
Figura 10	<i>Modelo de Redes Neuronales</i>	36
Figura 11	<i>Forma de la función de activación sigmoide</i>	38
Figura 12	<i>Forma de la función de activación tangente hiperbólica</i>	38
Figura 13	<i>Forma de la función de activación ReLU</i>	39
Figura 14	<i>Modelo de redes neuronales recurrentes</i>	40
Figura 15	<i>Compuerta de olvido</i>	41
Figura 16	<i>Compuerta de entrada</i>	42
Figura 17	<i>Compuerta de salida</i>	42
Figura 18	<i>Modelo de red neuronal recurrente</i>	43
Figura 19	<i>Representación del modelo de 1 paso</i>	45
Figura 20	<i>Representación del modelo de múltiples pasos</i>	46
Figura 21	<i>Representación del modelo de múltiples pasos y múltiples salidas</i>	46
Figura 22	<i>Componentes Principales de un Proyecto de Ciencia de Datos</i>	47
Figura 23	<i>Esquema del diseño del proyecto de implementación</i>	50
Figura 24	<i>Modelo de arquitectura tecnológica</i>	51
Figura 25	<i>Encuesta de la consultora Gartner</i>	54

Figura 26 <i>Participación en Ventas por Sucursal</i>	56
Figura 27 <i>Cobertura Comercial Directa de la Empresa</i>	56
Figura 28 <i>Participación en Ventas por Familia de Productos</i>	57
Figura 30 <i>Variables importe y frecuencia resumidas por producto</i>	59
Figura 31 <i>Determinación del Número de Clústers</i>	59
Figura 32 <i>Representatividad en Ventas por Clúster de Productos</i>	60
Figura 33 <i>Clúster 0 "Evolución Temporal y Forma de su Distribución de Probabilidades"</i>	62
Figura 34 <i>Clúster 1 "Evolución Temporal y Forma de su Distribución de Probabilidades"</i>	62
Figura 35 <i>Clúster 2 "Evolución Temporal y Forma de su Distribución de Probabilidades"</i>	63
Figura 36 <i>Clúster 3 "Evolución Temporal y Forma de su Distribución de Probabilidades"</i>	63
Figura 37 <i>Clúster 4 "Evolución Temporal y Forma de su Distribución de Probabilidades"</i>	64
Figura 38 <i>Clúster 0 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)</i>	65
Figura 39 <i>Clúster 1 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)</i>	66
Figura 40 <i>Clúster 2 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)</i>	66
Figura 41 <i>Clúster 3 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)</i>	67
Figura 42 <i>Clúster 4 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)</i>	67
Figura 43 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 0 y Costos Logísticos</i>	68
Figura 44 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 1 y Costos Logísticos</i>	69
Figura 45 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 2 y Costos Logísticos</i>	69
Figura 46 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 3 y Costos Logísticos</i>	70
Figura 47 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 4 y Costos Logísticos</i>	70
Figura 48 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 0 aplicando LSTM</i>	71
Figura 49 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 1 aplicando LSTM</i>	72
Figura 50 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 2 aplicando LSTM</i>	72
Figura 51 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 3 aplicando LSTM</i>	73
Figura 52 <i>Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 4 aplicando LSTM</i>	73

Figura 53 <i>Pantalla de Acceso al Sistema</i>	75
Figura 54 <i>Pantalla de Visualización General de Resultados</i>	76
Figura 55 <i>Pantalla de Visualización Detallada de Resultados</i>	77

LISTADO DE TABLAS

Tabla 1 <i>Resumen de Resultados de Clasificación ABC con base en \$ Ventas</i>	57
Tabla 2 <i>Variables y Registros del Dataset</i>	58
Tabla 3 <i>Composición de Clúster de Productos</i>	60
Tabla 4 <i>Concentración en \$ Ventas Clúster por Sucursales</i>	61
Tabla 5 <i>% Participación en \$ Ventas Clústers por Sucursales</i>	61
Tabla 6 <i>Principales Estadísticos Descriptivos por Clúster de Productos</i>	64
Tabla 7 <i>Principales Estadísticos Descriptivos por Clúster de Productos posterior a Imputaciones</i>	68
Tabla 8 <i>Resultados de Cálculos de Error del Pronóstico Método VAR</i>	71
Tabla 9 <i>Resultados de Cálculos de Error del Pronóstico Método LSTM</i>	74
Tabla 10 <i>Comparativo de Resultados de Métricas de Error</i>	74

CAPÍTULO 1

1. PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción del Problema

Desde hace más de una década las empresas en su mayoría pueden tener acceso a grandes volúmenes de datos de diversas fuentes y que crecen a un ritmo exponencial, esto se resume en las 3V's del BIG DATA (Marz, 2015), pero el tener acceso a estos datos no necesariamente significa contar con la capacidad de almacenarlos, procesarlos y extraer de ellos información valiosa de la forma adecuada.

Desde las empresas se ha detectado la importancia de poder explotar los datos a fin de evaluar lo que sucedió en el pasado, por qué sucedió, qué podría suceder en el futuro y cómo prepararse oportunamente para lo que podría suceder (Vaughan, 2020), el qué podría suceder y cómo prepararse son cuestionamientos que pueden responderse haciendo uso de la analítica predictiva, específicamente con el uso de pronósticos.

Para las empresas, sean estas manufactureras, comercializadoras, de servicios, híbridas, públicas o privadas, los pronósticos constituyen herramientas de mucha utilidad para brindar soporte en la generación de planes que les permitan alinear a los demás eslabones de sus cadenas de suministros: Compras, Producción, Almacenamiento y Distribución (APICS, 2022).

En el campo de la generación de pronósticos existen múltiples soluciones tecnológicas especializadas de gran potencia y que pueden generar pronósticos para grandes volúmenes de datos, sin embargo, estas soluciones generalmente tienen 2 características sumamente importantes: Requieren un elevado costo de inversión para ser implementadas, generalmente sólo utilizan métodos de pronósticos tradicionales para series de tiempo y finalmente los resultados de los pronósticos se presentan de forma rígida al usuario, usualmente como un reporte plano compuesto por filas y columnas, esto incrementa la dificultad para realizar el análisis de los resultados logrados.

Bajo estas circunstancias las empresas que aspiran a implementar soluciones tecnológicas para la generación de pronósticos requieren de una importante disponibilidad de recursos financieros, estar limitadas a técnicas de pronósticos convencionales y necesitar hacer uso de mucho tiempo para poder

llevar a cabo el análisis de los pronósticos obtenidos para finalmente poder utilizarlos en sus procesos de planeación.

1.2 Justificación

Un pronóstico consiste en calcular una estimación con respecto a lo que podría suceder en el futuro. En los negocios, poseer un dominio sobre esta actividad es de gran importancia al aplicarse en los procesos de planeación y definición de objetivos sobre variables de interés, por ejemplo: ventas, ingresos, costos, inventarios, etc., si bien es cierto que en casi el 100% de las ocasiones las decisiones se toman dentro de entornos con incertidumbre, es preferible asumir el riesgo y estimar lo que podría suceder, pero con un sustento sólido, a diferencia de no contar con algún elemento que brinde soporte a una decisión (Wilson, 2021).

Actualmente las organizaciones se encuentran en constante búsqueda de herramientas tecnológicas que faciliten estar preparados de forma oportuna ante las necesidades que a futuro demanden sus clientes, los pronósticos, como analítica predictiva generan información útil para el proceso de toma de decisiones anticipada.

A nivel de mercado existen herramientas tecnológicas que generan pronósticos, pero generalmente se enfocan en módulos especializados anexos a grandes sistemas informáticos como SAP, ORACLE, entre otros, y el adquirir este tipo de módulos requiere que las empresas posean una gran capacidad de inversión lo cual en la mayoría de los casos sólo es posible para las grandes compañías.

Con la implementación del presente proyecto se pretende generar una solución tecnológica para la generación de pronósticos haciendo uso de herramientas de código abierto que apliquen algoritmos tanto convencionales, así como de inteligencia artificial que sean aplicables a la generación de pronósticos para series de tiempo, con lo cual, las empresas no estarán limitadas por el requerimiento de un gran volumen de capital que usualmente es necesario para implementar este tipo de solución.

Finalmente, dentro del alcance del presente proyecto se considera la presentación de los pronósticos a través de un tablero interactivo construido sobre una plataforma de inteligencia de negocios, esto facilitará el análisis y uso de la información para los usuarios finales.

1.3 Objetivos

1.3.1 *Objetivo General*

Diseñar un modelo de pronóstico de ventas usando Redes Neuronales Recurrentes y Series Temporales para una empresa de consumo masivo.

1.3.2 *Objetivos Específicos*

- Estimar pronósticos de ventas generados a partir de técnicas de series temporales convencionales como Vectores Autorregresivos VAR.
- Establecer criterios de selección del mejor modelo de pronóstico de ventas usando series temporales convencionales y redes neuronales recurrentes LSTM con optimización de sus hiperparámetros.
- Desarrollar visualizaciones interactivas de los pronósticos de ventas mediante la implementación de un dashboard que ayude a la toma de decisiones oportunas.

1.4 Metodología:

El uso de técnicas de analítica predictiva representa una necesidad para las empresas que persiguen lograr una ventaja competitiva que les permita destacar en los mercados en los que operan (Vandeput, 2021). Para desarrollar el presente proyecto se llevará se considera el uso de la siguiente metodología:

Se utilizarán una base de datos estructurados que cuenta con aproximadamente 2 años de transacciones correspondientes a ventas para una empresa comercializadora de productos del sector de consumo masivo. Posteriormente se realizarán conexiones hacia la base de datos para poder tener acceso

a los mismos y realizar los procesamientos y análisis iniciales como: Transformaciones, limpieza, imputaciones, etc., estos pasos serán realizados haciendo uso de programas estadísticos y de inteligencia artificial de código abierto.

A continuación, haciendo uso de los mismos programas mencionados en el párrafo anterior se generarán los pronósticos de series de tiempo para cada uno de los productos que hayan sido vendidos por la empresa, para la generación de estos pronósticos se utilizarán las siguientes familias de técnicas convencionales: Suavizamiento Exponencial y Promedios Móviles Integrados Autorregresivos (ARIMA), así como Redes Neuronales Recurrentes.

Finalmente, los pronósticos generados que presenten la mayor precisión estadística serán presentados en un tablero interactivo que será desarrollado en una plataforma de inteligencia de negocios a fin de que los usuarios puedan desarrollar un análisis dinámico de los resultados obtenidos (Pronósticos) y con ellos puedan soportar la toma de decisiones.

1.5 Resultados Esperados:

Al finalizar el desarrollo del presente proyecto se obtendrá una herramienta que permita generar y visualizar de forma interactiva los pronósticos de Series de Tiempo para cada uno de los productos que hayan sido vendidos por la empresa.

Los pronósticos que se presenten a los usuarios serán aquellos que tengan el mayor nivel de precisión estadística que haya sido posible lograr mediante el uso de técnicas de optimización de parámetros e hiperparámetros.

Los pronósticos generados por la herramienta servirán como una base para la planeación comercial (Marketing y Ventas) que posteriormente pueda ser utilizada en los demás procesos de la cadena de suministro, por ejemplo: planeación de inventarios, planeación de abastecimiento, planeación de producción y planeación de distribución (APICS, 2018).

1.6 Dataset:

Para el desarrollo de este proyecto se puede acceder a una base de datos estructurados en la cual se encuentra registrada una muestra de las transacciones de ventas desde enero 2021 hasta octubre 2022, estos datos han sido anonimizados a través del uso de etiquetas descriptivas ficticias para los campos con datos categóricos y con el uso de factores de conversión para los campos con datos numéricos.

A lo largo del desarrollo de este proyecto se usaron datos correspondientes a ventas como fecha de venta, punto de venta, cantidad vendida, etc., y también datos de productos como código del producto, descripción de producto, etc., para más información sobre los campos utilizados se debe revisar el anexo 1 del presente documento.

Se considera que los datos utilizados en este proyecto puedan actualizarse de manera diaria, semanal, quincenal o mensual conforme se generan nuevas ventas y una vez finalizado un mes completo de ventas los modelos incluyan estos nuevos datos y excluyan los datos del mes más antiguo de la historia de ventas, de esta manera se puede capturar de mejor manera la evolución en el comportamiento de las ventas, así los modelos pueden actualizarse en su entrenamiento y generar nuevos pronósticos.

CAPÍTULO 2

2. MARCO TEÓRICO Y REVISIÓN BIBLIOGRÁFICA

2.1 Análisis Exploratorio de Datos

Los datos abarcan una colección o conjunto de: Palabras, fechas, números, enlaces a otros datos, incluso archivos multimedia como audio y videos, estos datos pueden estar asociados a diversos campos como: La vida cotidiana de los individuos, la economía, biología, ingeniería, negocios, etc.

Sabemos que el procesamiento de los datos nos ayuda a obtener información que al ser explotada correctamente se convierte en una ventaja competitiva para los individuos y/o empresas, sin embargo, la pregunta que debemos formularnos ahora es ¿Cómo podemos obtener información relevante de los datos?, la respuesta a esta pregunta es...aplicando “EDA”.

“EDA”, del inglés “Exploratory Data Analysis”, es un proceso que consiste en examinar el conjunto de datos disponibles para detectar: Los tipos de datos disponibles, Patrones, anomalías e incluso realizar comprobación de supuestos a través del uso de medidas estadísticas (Atwan, 2022).

Los componentes principales del EDA son:

- Resumir los datos.
- Aplicación de medidas estadísticas.
- Visualización de los datos.

Para desarrollar correctamente el EDA deben ejecutarse 4 pasos principales, los cuales son:

- Definir el Problema: Antes de intentar extraer información útil de los datos, primero se debe establecer el problema que se pretende resolver, esto permitirá enfocar el análisis de datos. En este paso es esencial definir las preguntas de negocio (Business Questions) a las que deberán darse respuestas.

- Preparación de Datos: Este paso se enfoca en extraer, transformar y cargar de datos una vez que estos se encuentran listos para ser utilizados en el análisis requerido.
- Análisis Estadístico: Este paso se enfoca en desarrollar análisis descriptivos, inferenciales, de correlación, de modelamiento, etc.
- Presentación de Resultados: Los resultados obtenidos deben ser presentados de forma comprensible y convincente a los interesados del negocio (Business Stakeholders), las herramientas frecuentemente utilizadas en este paso son tablas de datos y visualizaciones.

2.2 K-MEANS para la generación de clústers de datos

K-Means es uno de los algoritmos de aprendizaje no supervisado para clusterización más simples y populares que existen. Generalmente los algoritmos de aprendizaje no supervisado funcionan haciendo inferencias desde los datos que están siendo analizados a través de la aplicación de vectores de entrada sin apoyarse en el uso de etiquetas (Grus, 2019).

El objetivo de la técnica de K-MEANS es agrupar puntos de datos similares en un número “k” de clústers previamente definido, un clúster hace referencia a un conjunto de valores que comparten ciertas características comunes.

El número de clústers representará al número de centroides que serán considerados en el modelo, un centroide hace referencia al punto medio de cada subconjunto de datos que sea generado por el modelo de K-MEANS, cada valor estará ubicado con respecto al centroide a una distancia que minimice la suma cuadrática de las desviaciones (CHIANG, CHE, & HUNG, 2023).

En términos simples, la técnica de K-MEANS genera “k” centroides y ubica cada punto lo más cerca de cada centroide.

Es importante mencionar que la técnica de K-MEANS se basa en un proceso iterativo, pues al inicio genera de forma aleatoria la ubicación de los centroides y luego a través de cálculos repetitivos optimiza la posición de los centroides.

Finalmente, es relevante mencionar que al aplicar la técnica de K-MEANS se pueden utilizar diferentes formas de calcular las distancias entre cada valor y sus respectivos centroides (Rubiales, 2020), a continuación mencionaremos varias formas para calcular dichas distancias:

2.2.1 Distancia de Manhattan: Su cálculo corresponde a la suma de las desviaciones absolutas, en la ecuación 1 se muestra la expresión de cálculo correspondiente.

$$D = \sum_{i=1}^n |x_i - y_i| \quad (1)$$

2.2.2 Distancia Euclidiana: Su cálculo corresponde a la longitud que une a dos puntos, la ecuación 2 representa a la expresión de cálculo correspondiente.

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (2)$$

2.2.3 Distancia de Minkowski: Es una generalización de las distancias Euclidiana y de Manhattan, es utilizada para determinar la similitud entre distancias, en la ecuación 3 se muestra su expresión de cálculo.

$$d(p, q) = (\sum_{i=1}^n |p_i - q_i|^c)^{\frac{1}{c}} \quad (3)$$

2.2.4 Distancia de Mahalanobis: Es el cálculo de la matriz inversa de covarianza, mide como los valores de una característica cambian en función de otra característica, en la ecuación 4 se muestra su expresión de cálculo.

$$d_M(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)} \quad (4)$$

2.3 Introducción a los pronósticos

Un pronóstico se representa como el valor estimado hacia el futuro para una variable a través del uso de técnicas que aporten en la reducción de la magnitud del error, se considera además la aplicación del buen juicio y de la experiencia del quien genera los pronósticos. La finalidad de elaborar pronósticos es disminuir la incertidumbre de tal manera que se permita tomar decisiones con menos riesgo (Hyndman & Athanasopoulos, 2018).

Los pronósticos han tenido varias aplicaciones dentro del ámbito industrial y científico. Como por ejemplo en: logística, mantenimiento, salud, clima etc. (Bozarth, 2019).

Dentro del ámbito empresarial, se torna mucho más importante y debería ser un parte integral de todas las empresas. Cada vez que nosotros desarrollamos un plan, primero debemos generar pronósticos. Esto debe ser así para individuos, compañías con y sin fines de lucro, organizaciones gubernamentales; de hecho, esto debe ser así para cualquier entidad que realice un plan.

Producir buenos pronósticos es de gran impacto para la organización. La precisión de estos es el mayor determinante en los costos de inventario, nivel de servicio, estimación del personal, y otras medidas de rendimiento operacional (Jacobs, 2019).

En la misma dirección, es importante mencionar que un modelo de pronósticos es casi nunca 100% preciso. Sin embargo, un buen pronóstico será ligeramente superior o inferior al valor real, de manera general se espera una desviación de más menos el 10% (Nielsen, 2020).

2.4 Clasificación de métodos y tipos de pronósticos.

En la figura 1 se observa un esquema referente a la clasificación de métodos de pronósticos:

Figura 1

Clasificación de Métodos de Pronósticos



En el tratamiento de las series temporales para generar pronósticos existen dos métodos de disponibles: cualitativos y cuantitativos.

Por un lado, se utiliza los métodos cualitativos cuando no hay data disponible o relevante para realizar pronósticos. Por otra parte, se utilizan los cuantitativos cuando se satisfacen 2 condiciones: 1) existe suficiente información numérica sobre el pasado y 2) se asume que ciertos aspectos del pasado van a seguir existiendo en el futuro. (Rob J Hyndman et al). Dentro de los métodos cuantitativos podemos identificar los modelos: explicativos y de series temporales.

Los modelos explicativos, incorporan otras variables que explican a la variable a predecir y son conocidos como de regresión. Si bien es cierto, estos modelos son de gran utilidad ya que incorporan otra información que explica la serie, cuentan con problemas tales como: 1) no se tiene el conocimiento de todas las variables explicativa o regresoras 2) es necesario pronosticar las variables explicativas lo cual lo

vuelve muy complejo 3) el objetivo de predicción es conocer que va a pasar mas no conocer por qué pasa.
4) los pronósticos explicativos no son tan precisos como los de la serie temporales. (Rob J Hyndman et al).

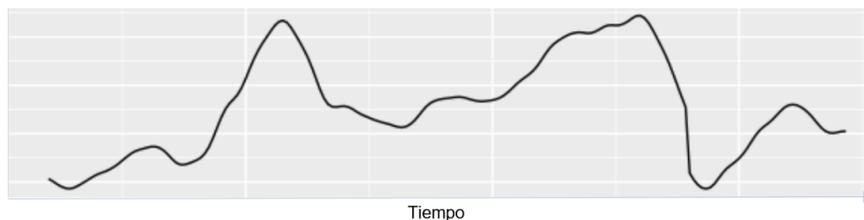
2.5 Las series temporales y sus componentes

Las series temporales son secuencias de observaciones ordenadas de manera cronológica y son los “input” para la generación de pronósticos. Una serie temporal está compuesta de tres componentes: tendencia, estacionalidad y residuo.

2.5.1 Tendencia: El componente denominado "tendencia" es el que ilustra si la serie temporal ha seguido en general un patrón ascendente o descendente en el pasado. Una subida en el valor de la serie puede deberse a diversos factores, entre otros, un aumento de la población, mejoras en los procesos productivos y/o de comercialización, la introducción de nuevos artículos o el establecimiento de sucursales adicionales, etc., en la figura 2 se presenta un ejemplo de cómo se visualiza la tendencia de la serie temporal.

Figura 2

Tendencia de la Serie Temporal

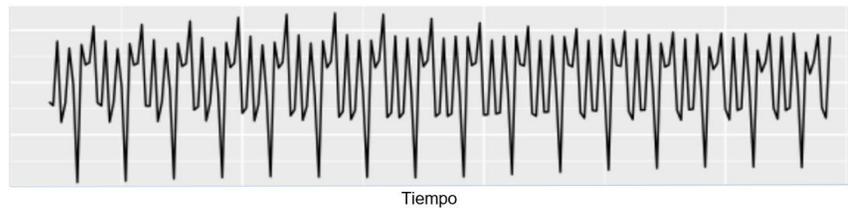


2.5.2 Estacionalidad: Este aspecto del comportamiento de la serie de tiempo se refiere al patrón de conducta que se repite en periodos iguales o inferiores a un año. Estos cambios en el comportamiento previsto podrían atribuirse a factores como el cambio entre estaciones (primavera, verano, otoño e invierno), ocasiones especiales o épocas de celebración, el clima u

otros factores., en la figura 3 se presenta un ejemplo de cómo se visualiza el componente de estacionalidad en series de tiempo.

Figura 3

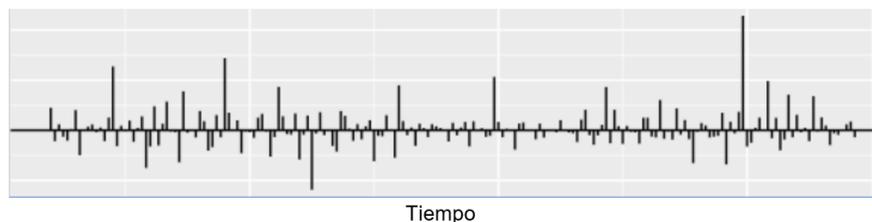
Componente de Estacionalidad en Series de Tiempo



2.5.3 Residuo: Es el componente que recoge todo lo adicional que se encuentra en una serie temporal y que no puede ser explicado ni por la tendencia ni por el componente estacional, en la figura 4 se presenta un ejemplo de cómo se visualiza el componente de residuos en series de tiempo.

Figura 4

Componente de Residuos en Series de Tiempo



2.6 Modelos de Vectores Autorregresivos VAR

Cuando se requieren generar pronósticos de series de tiempo, existe la posibilidad de trabajar con modelos univariantes o multivariantes, dentro de los modelos multivariantes se considera el supuesto de que existen variables predictoras (X_1, X_2, \dots, X_n) que son de utilidad para predecir el comportamiento de una variable dependiente (Y), este supuesto es fundamentado a través del estudio de correlación entre variables y además considera que dicha correlación es unidireccional, sin embargo, puede suceder que dos o más series de tiempo se encuentren correlacionadas en ambas direcciones, lo cual significa que si por ejemplo tenemos dos series de tiempo: La serie de tiempo 1 que representa a las ventas de una

empresa y la serie de tiempo 2 que representa los gastos operativos de la misma empresa, la serie de tiempo 1 puede ser predictora de la serie de tiempo 2 y viceversa, en estas circunstancias es muy útil disponer de un modelo que considere la bidireccionalidad de la correlación de las series y pueda pronosticarlas de forma simultánea.

Los modelos autorregresivos VAR consideran la bidireccionalidad en la correlación de las series de tiempo, las expresiones matemáticas para los modelos VAR se muestra desde las ecuaciones 5 hasta la 8.

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (5)$$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \quad (6)$$

$$y_{1,t} = C_1 + \phi_{1,1} y_{1,t-1} + \phi_{1,2} y_{2,t-1} + \epsilon_{1,t} \quad (7)$$

$$y_{2,t} = C_2 + \phi_{2,1} y_{1,t-1} + \phi_{2,2} y_{2,t-1} + \epsilon_{2,t} \quad (8)$$

2.7 Requerimientos generales de los modelos VAR

Para utilizar de forma correcta los modelos VAR para generar pronósticos de series de tiempo se deben ejecutar los siguientes pasos:

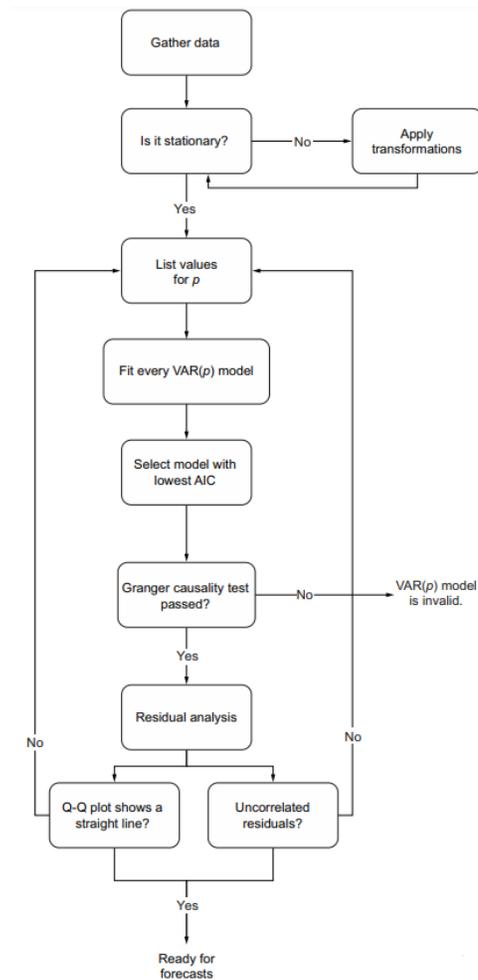
- Evaluar si la serie temporal es estacionaria, de no serlo deberán aplicarse procesos de transformaciones hasta cumplir esta condición.
- Evaluar una lista de posibles valores para el parámetro “p” del modelo.
- Utilizar técnicas como el cálculo de AKAIKE para la selección el mejor modelo con base a los distintos valores de “p”.

- Evaluar si todas las series de tiempo se encuentran correlacionadas entre sí, para esto se aplica evacuación de Causalidad de Granger, de no cumplirse la condición de correlación, entonces no será válido utilizar los modelos VAR para la generación del pronóstico.
- Finalmente se deberá aplicar el análisis de residuos y si estos no muestran estar correlacionados, entonces será válido aplicar los modelos VAR para la generación de pronósticos.

Todos los pasos detallados anteriormente se representan con la figura 5.

Figura 5

Aplicación de Modelos VAR



Nota. Tomada de Peixeiro, 2022

2.8 Transformación de la serie de tiempo

Aplicar transformaciones de series de tiempo tiene como objetivo lograr un patrón de comportamiento estacionario, a continuación, se comentará sobre el proceso de transformación a través de la diferenciación de la serie de tiempo (Manu, 2022).

La diferenciación es un procedimiento que nos ayuda a transformar las series temporales de NO estacionarias a estacionarias. Al diferenciar una serie de tiempo logramos estabilizar su varianza y su promedio, esto al reducir o eliminar la presencia de los componentes de tendencia y estacionalidad, la ecuación 9 representa a la técnica de diferenciación.

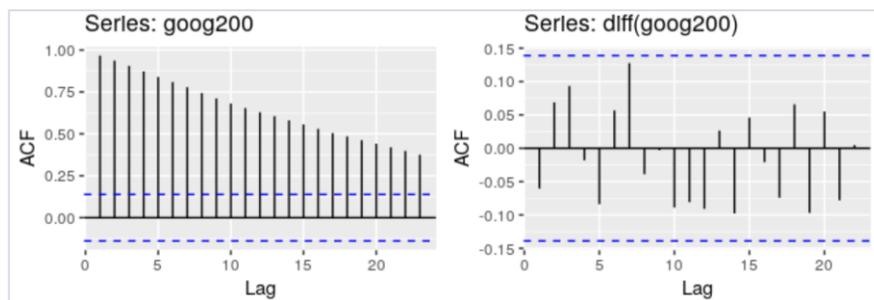
$$y'_t = y_t - y_{t-1} \quad (9)$$

La gráfica de la función de autocorrelación ACF es útil para identificar series temporales estacionarias y no estacionarias, en una serie temporal no estacionaria la gráfica de ACF decrece lentamente, en tanto que para series de tiempo estacionarias la gráfica de ACF se reduce de forma acelerada hacia cero.

En la gráfica 6 se presentan dos visualizaciones, hacia la izquierda se presenta a la función de autocorrelación para una serie temporal antes y hacia la derecha se tiene a la misma gráfica para la serie temporal diferenciada.

Figura 6

Visualizaciones de la Función de Autocorrelación



Nota. Tomado de Hyndman & Athanasopoulos, 2021

Al observar las gráficas anteriores se evidencia claramente un cambio en el comportamiento de las gráficas del ACF, podríamos concluir de forma empírica que al aplicar diferenciación a la serie de tiempo, esta se ha vuelto estacionaria, sin embargo podríamos plantearnos la siguiente pregunta: ¿Puede ser necesario aplicar más de una diferenciación a una serie de tiempo?, la respuesta es sí, aunque es poco frecuente, puede ser necesario aplicar más de una diferenciación a una series de tiempo para lograr un patrón de comportamiento estacionario. Cuando llega a ser necesario aplicar una segunda diferenciación a la serie de tiempo, su expresión de cálculo se muestra en la ecuación 10.

$$y_t'' = y_t - 2y_{t-1} + y_{t-2} \quad (10)$$

Una pregunta adicional que se puede plantear es: ¿Habría alguna diferencia en el procedimiento de cálculo de la diferenciación cuando la serie de tiempo es estacional?, la respuesta es sí, la expresión de cálculo correspondiente se muestra en la ecuación...

$$y_t' = y_t - y_{t-m} \quad (11)$$

2.9 Evaluación de Raíz Unitaria

La evaluación de raíz unitaria se utiliza en la determinación de si una serie temporal es estacionaria o no, y por lo tanto si necesita o no ser diferenciada. Cuando una serie de tiempo posee raíz unitaria se lo declara como no estacionario y por ende se esperaría que su comportamiento presente picos y valles de forma aleatoria, en tanto que si la serie de tiempo no posee raíz unitaria se la considera como estacionaria (Brockwell & David, 2016).

Existen múltiples pruebas estadísticas para evaluar si existe o no raíz unitaria en una serie de tiempo, a continuación se mencionan las más generalmente conocidas y utilizadas:

- Dickey Fuller
- Dickey Fuller Aumentada

- Elliott-Rothenberg-Stock
- Schmidt-Phillips
- Phillips-Perron
- Zivot-Andrews

2.10 AIC versus BIC para la selección del orden del modelo

Los criterios de información son herramientas estadísticas que permiten determinar de forma objetiva un modelo óptimo dentro de un conjunto de modelos, los criterios de información evalúan tanto la complejidad del modelo como su capacidad predictiva, es decir, la capacidad ser válido de forma general, más allá de los datos con los cuales fue entrenado el modelo (Shumway & Stoffer, 2017).

Tanto al cálculo del modelo de Akaike como el cálculo del modelo Bayesiano utilizan el algoritmo de máxima verosimilitud como criterio de bondad y un factor k que representa al número de parámetros del modelo como medida de complejidad.

En las ecuaciones 12 y 13 se muestran los modelos para los cálculos de Akaike y Bayesiano respectivamente.

$$AIC = 2k - 2\ln(L) \quad (12)$$

$$BIC = k\ln(n) - 2\ln(L) \quad (13)$$

Como se mencionó anteriormente, en el modelo del criterio de Akaike (AIC), “ k ” representan a la cantidad de parámetros y “ L ” representa al máximo de la función de verosimilitud. El valor de AIC disminuye en tanto se incremente la bondad del ajuste, se debe considerar si al incrementar parámetros al modelo se logra una mejor explicación de los datos o por el contrario, si al reducir parámetros al modelo podemos obtener una reducción significativa de AIC, es por este motivo que cuando seleccionamos un modelo se considera como ganador al que tenga el menor AIC.

En cuanto al criterio de información Bayesiano (BIC), nuevamente, “k” representa al número de parámetros, “L” representa al máximo de la función de verosimilitud y “n” representa al número de datos. En este modelo, la complejidad es determinada por una combinación de “k” y “ln(n)”, de esta manera, BIC penaliza con mayor fuerza a la complejidad del modelo en comparación con AIC, finalmente, se puede decir que BIC se enfoca en identificar el modelo más abstracto, más simple y que genere predicciones en un contexto más amplio.

Se puede concluir que tanto AIC como BIC realizan una medición simple de la complejidad del modelo, en los casos exista la presencia de modelos no lineales, o si existe interacciones entre parámetros, colinealidad o estructuras jerárquicas de factores aleatorio, entonces el medir la complejidad de los modelos a través del número de parámetros se queda corto (Juan G. Rubalcaba, 2023).

2.11 Causalidad de Granger

Para poder utilizar los modelos VAR en la generación de pronósticos uno de los requisitos es que exista correlación (en ambas direcciones) entre las series de tiempo, por tal motivo es fundamental poder demostrar la existencia de dicha correlación, es en este punto donde se debe utilizar la evaluación estadística de Granger (Rong, 2015).

La evaluación estadística de Granger aplicada por ejemplo a dos series temporales evalúa si los rezagos de la primera serie sirven para pronosticar los valores futuros de la segunda, de esta manera, la prueba estadística de causalidad de Granger debe ser aplicada a cada serie de tiempo de forma independiente, siguiendo este ejemplo, en las ecuaciones 14 y 15 se muestran las expresiones matemáticas aplicadas a cada serie de tiempo.

$$X_1(t) = \sum_{j=1}^{\rho} A_{11,j}X_1(t-j) + \sum_{j=1}^{\rho} A_{12,j}X_2(t-j) + E_1(t) \quad (14)$$

$$X_2(t) = \sum_{j=1}^{\rho} A_{21,j}X_1(t-j) + \sum_{j=1}^{\rho} A_{22,j}X_2(t-j) + E_2(t) \quad (15)$$

2.12 Fundamentos de Deep Learning

En las últimas décadas, el campo de la inteligencia artificial (IA) ha sido objeto de estudio para la aplicación en diversos campos, como la medicina, la manufactura, los servicios de entrega, entre otros.

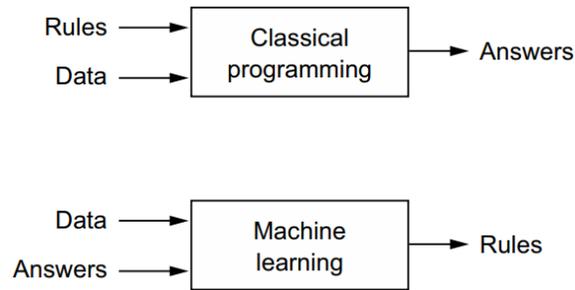
Es de suma importancia definir con claridad de qué estamos hablando cuando mencionamos términos como: Inteligencia artificial, aprendizaje de máquinas and aprendizaje profundo, para esto, a continuación, definiremos cada término.

Inteligencia Artificial: La inteligencia artificial “nació” en la década de 1950, cuando un grupo de expertos en el campo de la informática se plantearon la siguiente pregunta: ¿Las computadoras podrían llegar a pensar?, la respuesta a esta pregunta generó varias ramificaciones que continúan siendo estudiadas hasta el día de hoy. En concreto la inteligencia artificial puede definirse como el proceso de automatizar tareas intelectuales que regularmente son ejecutadas por los seres humanos (Géron, 2019).

Aprendizaje de máquinas: La forma en que habitualmente se logra que un computador lleve a cabo una tarea es a través de un programa informático que detalle de forma explícita las instrucciones (líneas de código) que un especialista en programación haya desarrollado, el aprendizaje de máquina tiene un enfoque totalmente diferente pues es la maquina la que examina datos de entrada con sus respectivas respuestas o incluso sin ellas para poder generar un resultado de tipo predictivo, en resumen, un sistema de aprendizaje de máquinas se entrena en lugar de programarse explícitamente como se puede observar en la figura 7.

Figura 7

Entrenamiento y Programación Explícita en Aprendizaje de Máquinas



Nota. Tomada de Peixeiro, 2022

A pesar de que el aprendizaje de máquinas se ha vuelto muy popular desde la década de 1990, rápidamente se ha posicionado como el subcampo más popular y exitoso de la inteligencia artificial.

Aprendizaje profundo (Deep Learning): Antes de llegar a una definición de aprendizaje profundo y lograr comprender su diferencia con respecto a otros enfoques de aprendizaje de máquinas, primero tener una idea clara de lo que hacen los algoritmos de aprendizaje de máquinas, estos algoritmos requieren 3 elementos:

Datos de entrada: Por ejemplo, si se desea lograr reconocimiento de animales, los datos de entradas serán imágenes de animales.

Ejemplos de datos de salida: Los resultados esperados podrían ser etiquetas con los nombres de animales, por ejemplo: perro, gato, conejo, etc.

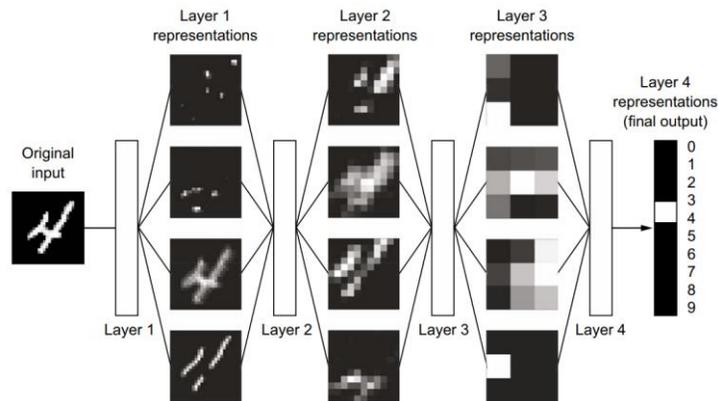
Métricas: Las métricas permiten demostrar si los algoritmos están o no realizando un buen trabajo, además, las métricas son las responsables del “aprendizaje” debido a que se utilizan para ajustar parámetros dentro de los algoritmos.

El aprendizaje profundo es un subcampo específico del aprendizaje de máquinas, la palabra “profundo” hace referencia a que los algoritmos que se utilizan están basados en arquitecturas de capas

denominadas redes neuronales artificiales (RNA), a mayor número de capas utilizadas en el modelo, mayor profundidad existirá, en la actualidad, el uso de modelos de aprendizaje profundo considera decenas o cientos de capas sucesivas, todas ellas aprenden ante las exposición a datos de entrenamiento, las arquitecturas de aprendizaje profundo tratan de imitar o recrear el aprendizaje del cerebro humano, la figura 8 representa las capas utilizadas en los modelos de aprendizaje profundo.

Figura 8

Representación de capas en modelos de aprendizaje profundo



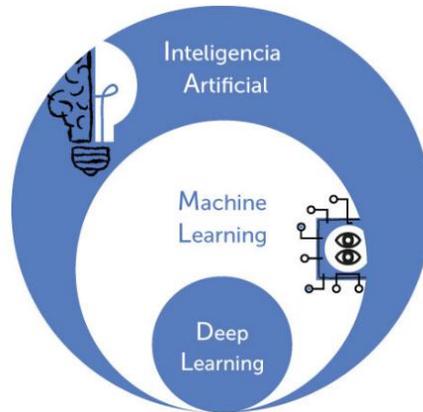
Nota. Tomada de Peixeiro, 2022

Complementando la definición, el aprendizaje profundo es una forma de aprender representaciones de datos en varias etapas (François Chollet, 2021).

Una vez revisadas las definiciones anteriores, se concluye que el aprendizaje profundo es un subcampo del aprendizaje de máquinas y a su vez el aprendizaje de máquinas es un subcampo de la inteligencia artificial tal como se observa en la figura 9.

Figura 9

Inteligencia Artificial, Aprendizaje de Máquinas y Aprendizaje Profundo



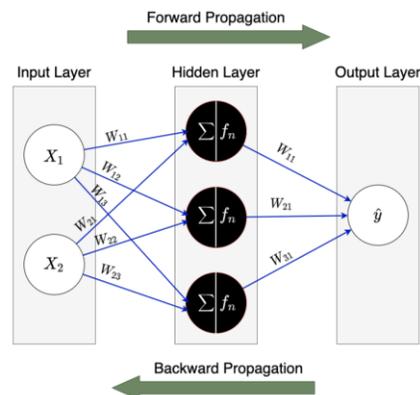
Tomado de <https://medium.com/>

2.13 Redes Neuronales:

La red neuronal artificial es conocida como perceptrón. El perceptrón es una estructura que está compuesta de capas. La red más simple, como mínimo se encuentra estructurada por tres capas. La primera corresponde a la capa de entrada, la segunda es la capa oculta y tercera es la capa de salida (Chollet, 2021), en la figura 10 se muestra un ejemplo de redes neuronales.

Figura 10

Modelo de Redes Neuronales



Tomado de Peixeiro, 2022

En la capa de entrada se encuentran las variables consideradas para el modelo, la capa oculta está compuesta por una función de activación y la de salida hace referencia al valor a predecir. El mecanismo por el cual este perceptron funciona se lo conoce como “forward propagation”.

Así mismo, Para obtener el valor predecir se aplica una función de activación no lineal a la multiplicación transpuesta de los valores de entradas por pesos aleatorios (w). La función también presenta un termino (w_0) que hace referencia al “bias” generalmente es igual a 1, con la ecuación 16 se expone la respectiva ecuación de cálculo.

$$\gamma = g(w_0 + x^t w) \quad (16)$$

El propósito de las funciones de activación es introducir no linealidad en la red neuronal. Entre las funciones de activación más comunes podemos encontrar: sigmoidea, tangente hiperbólica y unidad lineal rectificadora (ReLU). La aplicación de cada una de ellas dependerá de los valores que deseamos en nuestra salida, a continuación, se explican brevemente las 3 funciones de activación mencionadas.

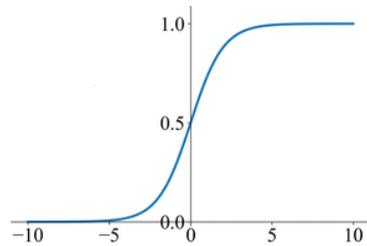
La función de activación sigmoide es generalmente utilizada en problemas de clasificación binaria, su forma es al de una “S” extendida tal como se muestra en la figura 11, de allí su nombre, esta función de activación recibe valores reales como entradas y genera salidas con valores entre cero y uno, es importante mencionar que esta función es débil con respecto al desvanecimiento del gradiente, las ecuaciones 17 y 18 muestran las expresiones matemáticas utilizadas en esta función:

$$g(z) = \frac{1}{1+e^{-z}} \quad (17)$$

$$g'(z) = g(z)(1 - g(z)) \quad (18)$$

Figura 11

Forma de la función de activación sigmoide



La función de activación tangente hiperbólica, en la figura 12 se observa la similitud en su trayectoria con respecto a la función de activación sigmoide, con la diferencia de que se encuentra centrada en cero, esta función devuelve como salidas valores entre -1 y 1 por lo cual, en la práctica suele ser considerada en lugar de la función de activación sigmoide.

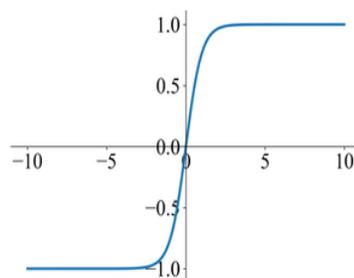
La función de activación tangente hiperbólica también presenta la debilidad del desvanecimiento del gradiente, las ecuaciones 19 y 20 muestran las expresiones matemáticas utilizadas por esta función:

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (19)$$

$$g'(z) = 1 - g(z)^2 \quad (20)$$

Figura 12

Forma de la función de activación tangente hiperbólica



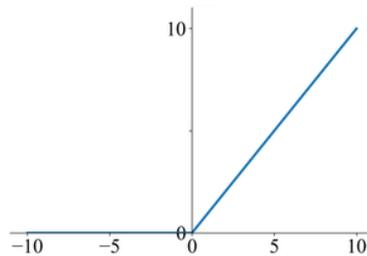
Dentro de las funciones de activación mayormente utilizadas se encuentra a la función de unidad lineal rectificada, uno de los motivos es que no presenta la debilidad del desvanecimiento del gradiente, esta función representa un bajo costo computacional en comparación a las dos funciones de activación antes mencionadas, esto debido a la simplicidad de su ecuación, en la figura 13 se puede observar la forma de la función de activación:

$$g(z) = \max(0, z) \quad (21)$$

$$g'(z) = \begin{cases} 1, & z > 0 \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

Figura 13

Función de activación ReLU



El proceso de aprendizaje de una red neural se la conoce como “entrenamiento” y el número de entrenamientos se lo conoce como “épocas”. En este proceso se busca modificar los pesos aleatorios (w) de tal forma que se minimice el error entre los datos de entrenamiento y los de test. El error se calcula aplicando una función de costo. Para problemas de regresión, como es el caso de la predicción de series temporales se utiliza el mean square error.

$$w^* = \underset{w}{\operatorname{argmin}} \sum_{i=1}^n \mathcal{L}(f(x^{(i)}; w), y^{(i)}) \quad (23)$$

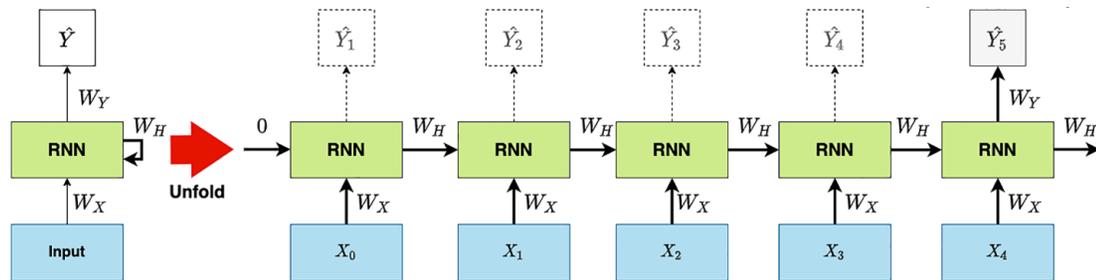
Al mecanismo de modificación de pesos se lo denomina “Backpropagation”. Donde estos pesos se recalculan optimizando la función de costo. Esta función, por su parte, se optimiza mediante una tasa de aprendizaje. Entre algunos de los algoritmos de optimización podemos encontrar: SGD (Descenso sobre la gradiente), Adam, Adadelta, Adagrad, RMSProp.

2.14 Redes Neuronales Recurrentes (RNN)

Los modelos de Redes Neuronales Recurrentes (RNN) se encuentran especialmente adaptados al procesamiento de datos secuenciales, por ejemplo: texto, audio y series de tiempo. La arquitectura de los de tipo RNN utiliza “capas ocultas” para retroalimentarse dentro de la red de forma tal que puede utilizar información pasada para generar predecir el próximo elemento en la secuencia, a esto se le denomina memoria, en la figura 14 se muestra el modelo de funcionamiento de este tipo de redes neuronales (MIKY, KALOOP, ELNABWY, BAKI, & ALSHOUNY, 2021).

Figura 14

Modelo de redes neuronales recurrentes



Tomado de Peixeiro, 2022

Las redes neuronales recurrentes poseen una importante limitación, únicamente poseen “memoria a corto plazo”, esto significa que conforme se avance en la secuencia de datos, las primeras observaciones dejarán de tener impacto en las predicciones del modelo, a este efecto se le conoce como “desvanecimiento del gradiente”, el gradiente es la función que se encarga de actualizar los pesos en cada

paso dentro de la red, en resumen, es la función que le permite aprender a la red, por lo tanto cuando el gradiente se desvanece hasta llegar muy cerca de cero se concluye que la red dejó de aprender.

Ante el efecto del desvanecimiento del gradiente, que se traduce en ocasionar un modelo con “memoria a corto plazo”, a través del uso un nuevo componente en la red se ha encontrado la manera de preservar información importante de la historia de los datos a largo plazo, las redes neuronales que consideran este componente son denominadas redes neuronales de tipo Long-Short-Term-Memory (LSTM) y serán explicadas a continuación.

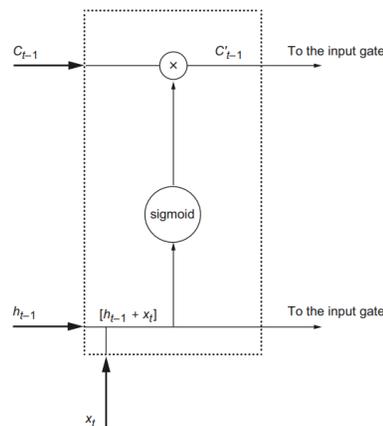
2.15 Redes Neuronales Recurrentes LSTM

Las redes neuronales recurrentes del tipo LSTM incorporan en su arquitectura una “celda de estado”, este componente evita que el gradiente desaparezca, lo cual permite que la red aprenda de un mayor número de datos históricos. Las redes neuronales recurrentes del tipo LSTM basan su arquitectura en tres compuertas:

La compuerta de olvido: Esta representada en la figura 15 y es la que discrimina que información de los datos históricos es relevante para las predicciones.

Figura 15

Compuerta de olvido

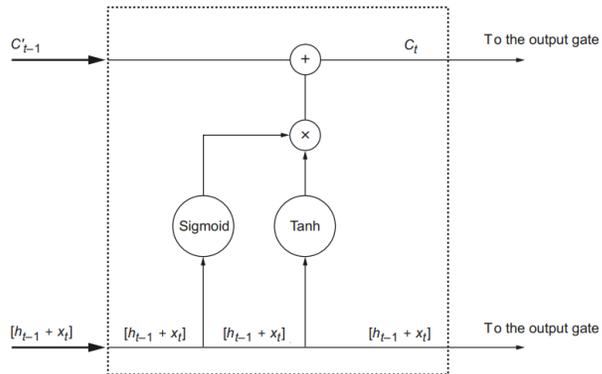


Tomada de Peixeiro, 2022

La compuerta de entrada: Esta representada en la figura 16 y es la que determina que información del paso actual es relevante para las predicciones.

Figura 16

Compuerta de entrada

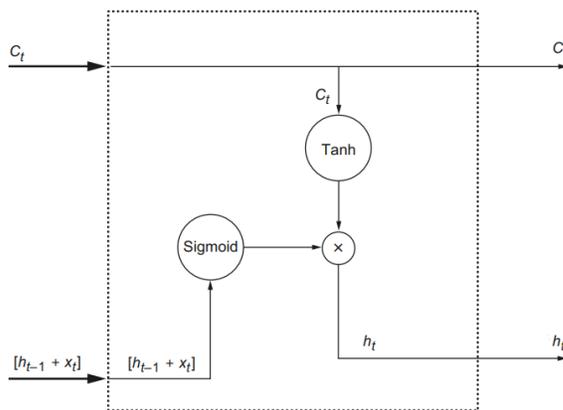


Tomada de Peixeiro, 2022

La compuerta de salida: Esta representada en la figura 17 y es la que determina que información será enviada a la siguiente célula de la red.

Figura 17

Compuerta de salida

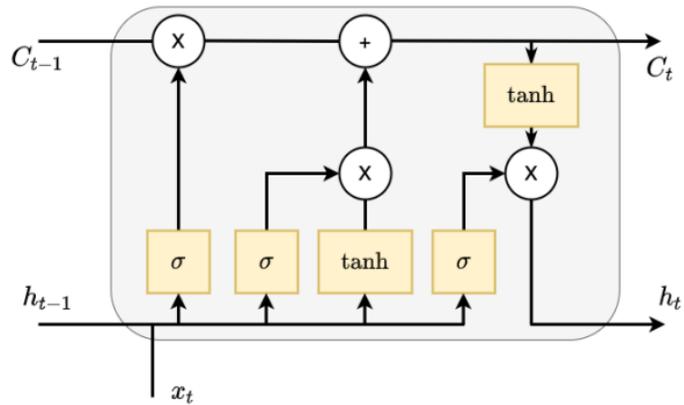


Tomada de Peixeiro, 2022

En la figura 18 se muestra la arquitectura de una red neuronal recurrente de tipo LSTM, se puede evidenciar que es una arquitectura de mayor complejidad en comparación a las redes neuronales recurrentes simple.

Figura 18

Modelo de red neuronal recurrente



Tomada de Peixeiro, 2022

A continuación, se muestran cada una de las ecuaciones consideradas en la arquitectura expuesta:

$$i_t = \sigma(w_{ii}x_t + b_{ii} + w_{hi}h_{t-1} + b_{hi}) \quad (24)$$

$$f_t = \sigma(w_{if}x_t + b_{if} + w_{hf}h_{t-1} + b_{hf}) \quad (25)$$

$$g_t = \tanh(w_{ig}x_t + b_{ig} + w_{hg}h_{t-1} + b_{hg}) \quad (26)$$

$$o_t = \sigma(w_{io}x_t + b_{io} + w_{ho}h_{t-1} + b_{ho}) \quad (27)$$

$$c_t = f_t \oplus c_{t-1} + i_t \oplus g_t \quad (28)$$

$$h_t = o_t \oplus \tanh(c_t) \quad (29)$$

Donde:

g_t : Es la modulación de la compuerta de entrada.

i_t : Es la compuerta de entrada.

f_t : Es la compuerta de olvido.

c_t : Es la celda de estado.

o_t, o_t : Son compuertas de salida.

σ : Es la función de activación, en este caso, es la función sigmoide.

Las técnicas de aprendizaje profundo destacan cuando son aplicadas a grandes y complejos conjuntos de datos, en estos casos es cuando los modelos pueden realizar inferencias sobre las relaciones entre características y así suelen dar lugar a pronósticos con un buen nivel de precisión.

Al hacer referencia a datos de tipo series de tiempo, un gran volumen de datos se podría dimensionar en más de 10,000 observaciones (Peixeiro, 2022), desde luego, esto es un valor de referencia. Ante situaciones donde se disponga de grandes volúmenes de datos es posible que las técnicas convencionales como ETS, ARIMA, SARIMA, VAR, etc., no logren tener un buen nivel de desempeño ya sea por el tiempo que consuman en generar pronósticos, así como en el nivel de precisión de dichos pronósticos, otro escenario donde las técnicas convencionales podrían estar limitadas es cuando existen varios niveles de estacionalidad en los datos, por ejemplo, si se analizara el consumo de energía por hora durante todo el año 2021 en el Ecuador, claramente existiría un mayor nivel de consumo energético durante las horas del día en comparación a las horas de la noche, además, se podría evidenciar fácilmente que el consumo de energía en fines de semana y festivos es diferente al consumo en días laborables ordinarios, finalmente también se podría detectar la variación en el consumo de energía durante los meses de verano en comparación a los meses de invierno, este es un escenario donde los datos contendrán varios niveles de estacionalidad y donde las técnicas de aprendizaje profundo pueden tener un mejor desempeño frente a las técnicas convencionales en cuanto a la generación de pronósticos.

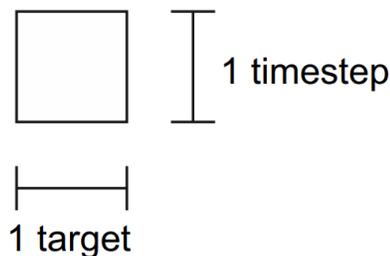
Adicionalmente es importante mencionar que las técnicas de aprendizaje profundo son recomendadas cuando los datos mantienen correlación en sus residuos, es decir, cuando no se logra obtener ruido blanco.

Existen diferentes tipos de modelos de aprendizaje profundo para pronosticar series de tiempo, estos son: Los modelos de un solo paso, los modelos de múltiples pasos y el modelo de múltiples pasos y múltiples salidas.

El modelo de un paso es el más simple de los tres modelos, su salida es el pronóstico de un paso en el futuro, sin importar la longitud de la cantidad de datos que se ingrese al modelo, este devolverá un valor escalar, al observar la figura 22 se aprecia una representación del modelo.

Figura 19

Representación del modelo de 1 paso

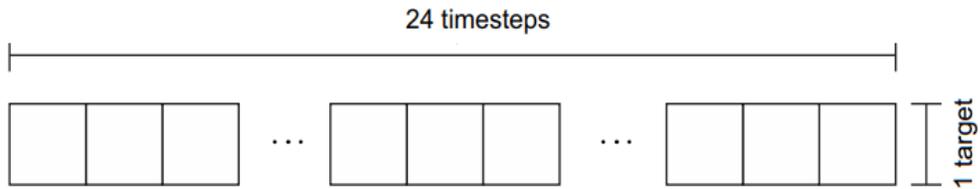


Tomada de <https://towardsdatascience.com/>

En el modelo de múltiples pasos, la salida es una secuencia de valores que representan a los pronósticos para varios periodos de tiempo en el futuro, al observar la figura 20 se observa una representación del modelo.

Figura 20

Representación del modelo de múltiples pasos

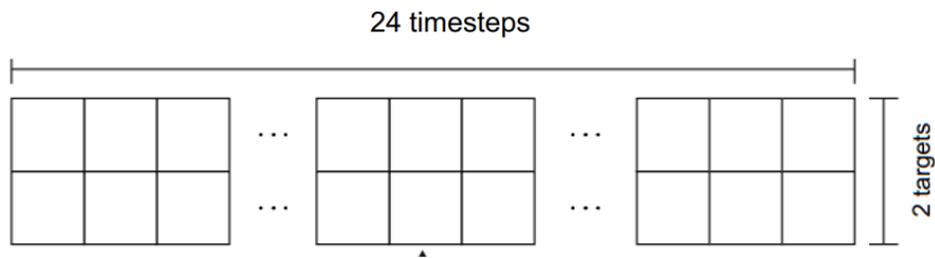


Tomada de <https://towardsdatascience.com/>

En el modelo de múltiples pasos y múltiples salidas, las salidas son secuencias de valores que representan los pronósticos para varios periodos de tiempo en el futuro para las diferentes variables u objetivos que se hayan considerado dentro del modelo, en la figura 21 se muestra una representación del modelo.

Figura 21

Representación del modelo de múltiples pasos y múltiples salidas



Tomada de <https://towardsdatascience.com/>

La ventaja de aplicar redes neuronales recurrentes de tipo LSTM es que el modelo conservará información histórica de los datos, lo cual es esencialmente útil en el análisis de datos secuenciales como lo son las series de tiempo.

2.16 Optimización de Hiperparámetros

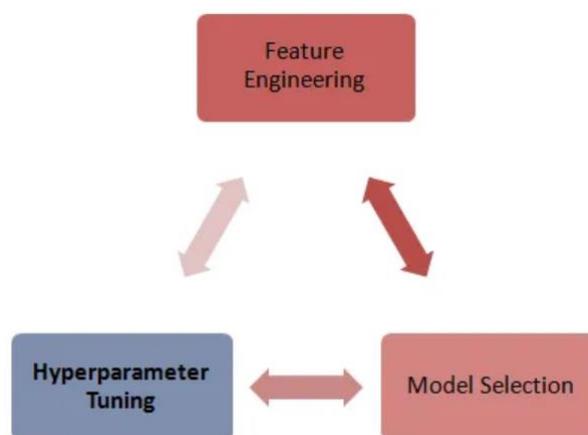
Una parte fundamental del proceso de creación de un modelo de machine Learning o Deep Learning es la optimización de sus hiperparámetros, los hiperparámetros son parámetros que se establecen al inicio de la construcción del modelo con la finalidad de controlar su comportamiento y aprendizaje (MICROSOFT, 2023).

Cada modelo de inteligencia artificial cuenta con sus propios hiperparámetros, de hecho, algunos de ellos son más importantes que otros, en el caso de las redes neuronales recurrentes por ejemplo podemos decidir el número de capas y neuronas de la red, el desempeño de un modelo de machine Learning o Deep Learning dependerá en gran medida de sus hiperparámetros.

De manera muy simple podemos dividir a un proyecto de ciencia de datos en 3 componentes principales: Obtención de datos e ingeniería de características, selección del modelo y finalmente optimización de hiperparámetros (HONG, LIANG, GONG, & YU, 2022), esto lo podemos apreciar en la figura 22.

Figura 22

Componentes Principales de un Proyecto de Ciencia de Datos



Tomada de <https://medium.com/>

La optimización de hiperparámetros puede ser un ardua y que demande mucho esfuerzo, es desde este punto que nos enfocaremos en la librería Grid Search, la cual es de mucha ayuda en este proceso.

Grid Search es un proceso de búsqueda exhaustiva dentro de un espacio dado para un conjunto de valores manuales de hiperparámetros, se prueban todas las combinaciones posibles de valores y evalúa los resultados haciendo uso de la técnica de validación cruzada (Sckit Learn, 2022).

2.17 Medidas de error del pronóstico

Los pronósticos poseen de manera implícita un margen de error y en función de qué tan grande o pequeño sea el error, así será el nivel de precisión o exactitud de la estimación; en tanto más pequeño sea el valor del error, el pronóstico será de mayor precisión y viceversa (Farrera, 2013).

Por tal motivo, es de suma importancia calcular diferentes mediciones del error relacionado a los pronósticos calculados mediante un método en particular y de esta forma lograr determinar si dichas estimaciones serán de utilidad o si será necesario utilizar otros métodos para lograr una mayor precisión en los resultados obtenidos.

Un error representa la desviación o la variación presente entre el valor real de la variable y su valor estimado, en la ecuación 30 se muestra su expresión de cálculo (Montemayor, 2013).

$$e_t = y_t - f_t \quad (30)$$

Sin embargo, podemos mencionar varias técnicas específicas para el cálculo del error del pronóstico, tales como:

El Error Medio (Mean Error): Su expresión de cálculo se detalla en la ecuación 31

$$ME = \frac{1}{N} \sum_{t=1}^N (y_t - f_t) \quad (31)$$

El Error Absoluto Medio (Mean Absolute Error): Su expresión de cálculo se detalla en la ecuación 32.

$$MAE = \frac{1}{N} \sum_{t=1}^N (|y_t - f_t|) \quad (32)$$

El Error Porcentual Medio (Mean Percentage Error): Su expresión de cálculo se detalla en la ecuación 33.

$$MPE = \frac{1}{N} \cdot \sum_{t=1}^N \frac{y_t - f_t}{y_t} * 100\% \quad (33)$$

El Error Absoluto Porcentual Medio (Mean Absolute Percentage Error): Su expresión de cálculo se detalla en la ecuación 34.

$$MAPE = \frac{1}{N} \cdot \sum_{t=1}^N \frac{|y_t - f_t|}{y_t} * 100\% \quad (34)$$

El Error Absoluto Porcentual Medio Ponderado (Weighted Mean Absolute Percentage Error): Su expresión de cálculo se detalla en la ecuación 35.

$$WMAPE = \frac{\sum_{t=1}^N w_t |y_t - f_t|}{\sum_{t=1}^N w_t |y_t|} * 100\% \quad (35)$$

La Raíz del Error Cuadrático Medio (Root Mean Squared Error): Su expresión de cálculo se detalla en la ecuación 36.

$$RMSE = \sqrt{\sum_{t=1}^N \frac{(y_t - f_t)^2}{N}} \quad (36)$$

CAPÍTULO 3

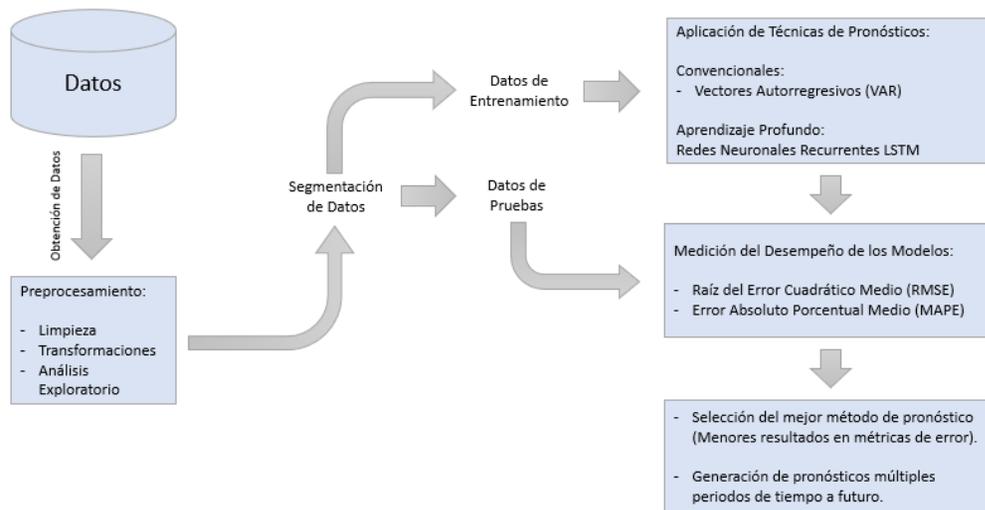
3. DISEÑO E IMPLEMENTACIÓN

3.1 Introducción

En este capítulo se explica al detalle el diseño de arquitectura tecnológica que es necesario para el almacenamiento, procesamiento y visualización de los resultados generados por los métodos aplicados. Además, se abordará el proceso y el criterio de selección de modelo utilizado para la generación de pronósticos, en la figura 23 se muestra el esquema del diseño a implementarse.

Figura 23

Esquema del diseño del proyecto de implementación



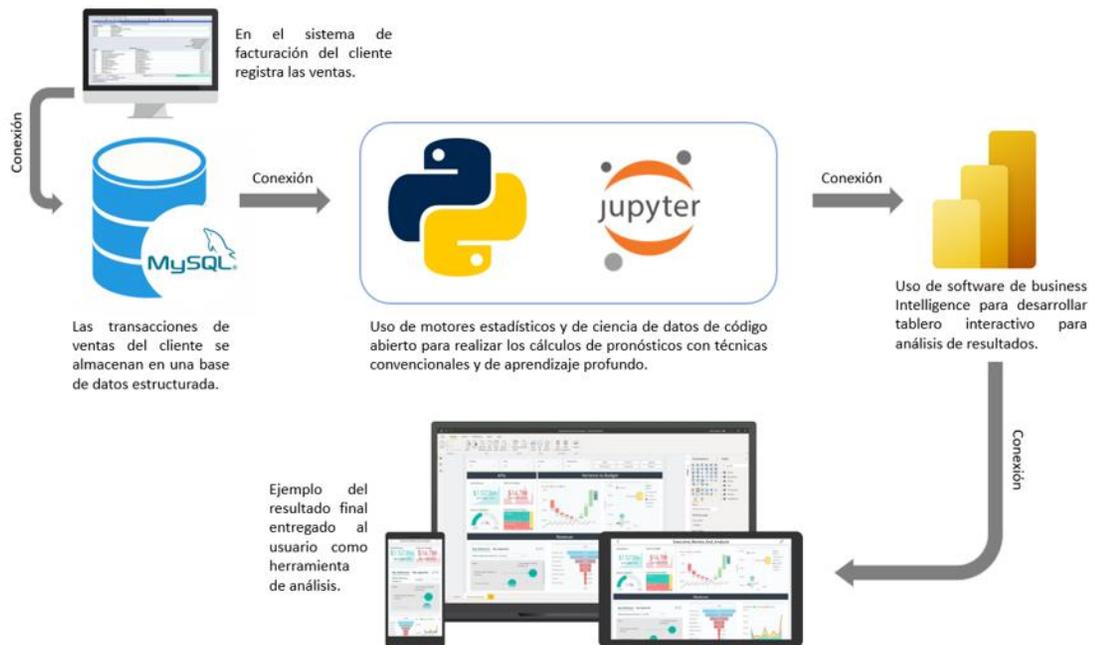
3.2 El modelo de arquitectura tecnológica

Para llegar a cumplir los objetivos planteados, es necesario que contar con herramientas para cada fase del proceso. Para el almacenamiento de los datos necesitaremos un gestor de base de datos que nos sirva como puente tanto para la recepción de la data transaccional, así como también para el almacenamiento de los resultados obtenidos. Además, es necesaria la utilización de ciertas plataformas que nos permita el procesamiento y análisis de información. Por último, se necesitará de una herramienta

de visualización por medio de la cual, el usuario podrá dar uso de los resultados generados por los algoritmos de pronósticos, en la figura 24 se muestra el modelo de arquitectura tecnológica para este proyecto.

Figura 24

Modelo de arquitectura tecnológica



3.3 Almacenamiento de los datos

Para el almacenamiento de los datos se utiliza MySQL, MySQL es un gestor de base de datos relacional que permite al usuario almacenar, ingresar, modificar y consultar información. El término “relacional” hace referencia a la estructura de la información puesto que, los datos están almacenados en tablas compuestas por registros (filas) y atributos (columnas) que se relacionan entre sí por medio de un campo clave identidad.

La función de esta herramienta es medular en la arquitectura tecnológica puesto que interviene en todas las fases del proceso, además, es la encargada de almacenar la información proveniente del punto de venta, así como también almacena los resultados que provienen de

los programas especializados en procesamiento y análisis de datos, finalmente, será el recurso para la herramienta de visualización.

3.4 Obtención de datos

La obtención de datos se realizará por medio de consulta al gestor de base de datos. La información para consultar proviene del sistema transaccional de ventas de la empresa. Esta información al ser estructurada y en esquema relacional nos permitirá extraer sencillamente los campos de nuestro interés como lo son: fecha, punto de venta, SKU y cantidad. Por medio del campo fecha, podremos extraer otros campos como año, mes y semana que será el input para la generación de los pronósticos.

3.5 Preprocesamiento y validación de datos

Para la fase de preprocesamiento y validación de datos se decidió utilizar el lenguaje de programación de código abierto Python a través del IDE Jupyter.

Python es un lenguaje de alto nivel de programación desarrollado a finales de los años ochenta por Guido Van Rosum. Este lenguaje es de gran popularidad gracias a la legibilidad de su código y su enfoque multiparadigma.

El enfoque de programación multiparadigma quiere decir que sirve para las tareas de programación orientada objetos, funcional, imperativo y reflexivo.

3.6 Generación de pronósticos y selección del mejor modelo

En este proyecto se realizarán pronósticos de series de tiempo considerando dos metodologías, la primera de ellas es de enfoque estadístico clásico, se trata de los modelos de Vectores Autorregresivos VAR, la segunda en una metodología de inteligencia artificial, específicamente Redes Neuronales Recurrentes LSTM, una vez se hayan generado pronósticos con cada metodología se considerarán dos métricas de error las cuales son RMSE y MAPE, únicamente aquellos pronósticos que generen las menores

métricas de error serán almacenados en la base de datos y posteriormente utilizados en la construcción del tablero (dashboard) de análisis de resultados para el usuario final.

3.7 Diseño de Dashboard en Power BI para análisis de resultados

Una vez escogido el mejor modelo para cada producto, es imperativo poder visualizar los pronósticos generados de forma intuitiva hacia el usuario. La herramienta escogida para esta fase es Power Bi. Power Bi es una herramienta de Microsoft orientado al análisis de datos que permite generar herramientas interactivas y que, es de mucha utilidad para la inteligencia empresarial.

Se ha escogido esta herramienta por los siguientes motivos:

- Como se muestra en la figura 25, la encuesta de la consultora Gartner muestra que hasta el año 2022 Power BI es líder en las plataformas de inteligencia empresarial (Gartner), lo cual hace probable que las empresas cuenten con esta herramienta.
- Al ser herramienta de Microsoft al igual que el gestor de base datos permite mayor flexibilidad a las consultas. Es posible generar consulta directamente en la base para luego ser consumida por Power BI, lo cual, disminuye el tiempo de procesamiento.
- Por medio del contexto de “filtro” es posible visualizar e interactuar solo con la información que el usuario considere relevante.

Figura 25

Encuesta de la consultora Gartner



Tomado de <https://www.gartner.com/>

CAPÍTULO 4

4. ANÁLISIS DE RESULTADOS

4.1 Antecedentes

En este capítulo, abordaremos al detalle la solución de la herramienta. Empezaremos por un análisis exploratorio de datos (EDA), esto con la finalidad de poder entender características propias de los datos y esto a su vez, nos permita realizar un preprocesamiento adecuadamente (Goldmeier & Gutman, 2021).

Una vez realizado el EDA, aplicaremos los modelos seleccionados y estudiados en la revisión de la literatura como los son los modelos tradicionales y de aprendizaje profundo LSTM.

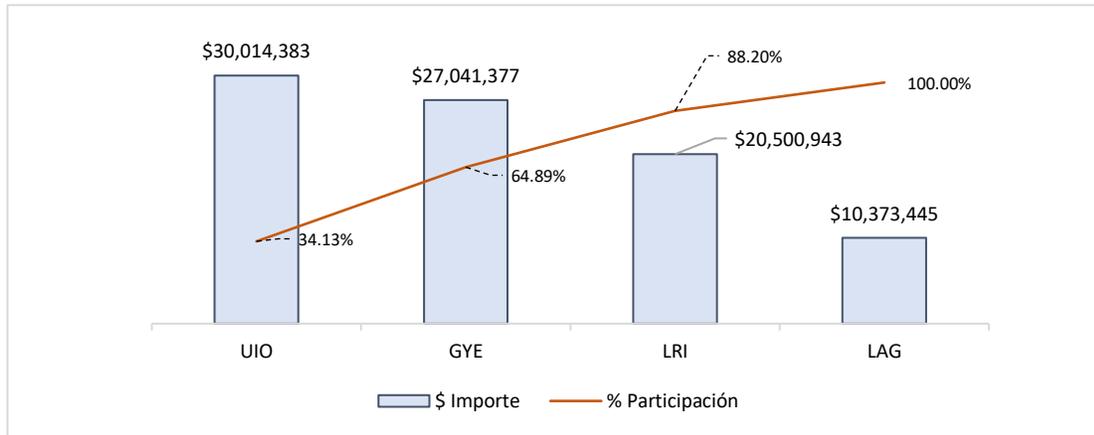
El objetivo principal de la aplicación de modelos es encontrar los parámetros e hiperparámetros óptimos para cada modelo, luego poder compararlos y finalmente seleccionar la técnica que genere los menores errores con base en las métricas RMSE y MAPE, posteriormente, explicaremos el análisis de resultados por medio de una plataforma de visualización que permitirá al usuario hacer uso de los pronósticos generados.

4.2 Análisis Exploratorio de datos

Para la empresa a la cual corresponde la muestra de datos que está siendo utilizada, es de suma importancia el desarrollo de su presencia comercial (ventas) en cada región, así como el crecimiento las diferentes familias de productos, en la figura 26 se muestra la participación en la venta en dólares por cada sucursal a nivel nacional, claramente se identifica que las sucursales UIO y GYE son las más representativas, acumulando entre ellas más del 64% del total de las ventas de la empresa, dejando así aproximadamente el 36% de las ventas en dólares en las sucursales LRI y LAG.

Figura 26

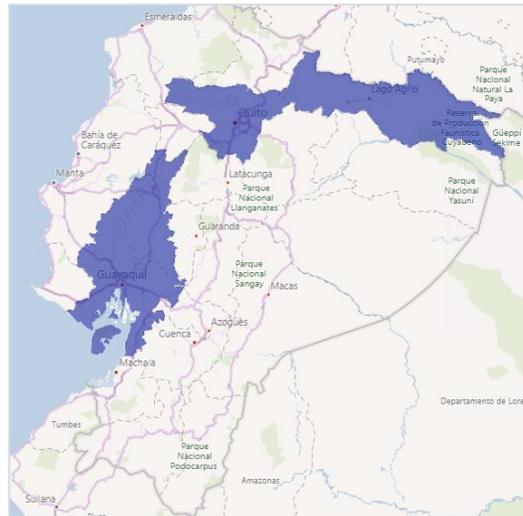
Participación en Ventas por Sucursal



Adicionalmente, en la figura 27 se muestra en color azul la cobertura comercial directa de la empresa a nivel nacional.

Figura 27

Cobertura Comercial Directa de la Empresa

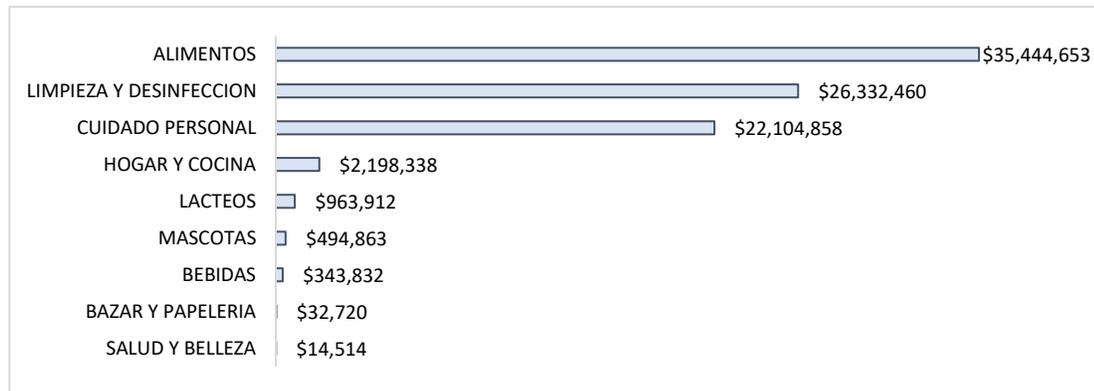


En la figura 28 se muestra la participación en ventas en dólares por cada familia de producto, se evidencia claramente que las 3 familias de mayor importancia para la compañía son: ALIMENTOS,

LIMPIEZA Y DESINFECCION y CUIDADO PERSONAL, estas 3 familias concentran aproximadamente el 95% del total de las ventas en dólares.

Figura 28

Participación en Ventas por Familia de Productos



Con respecto al portafolio de productos, se considera un total de 990 productos (skus) sobre los cuales se realizó la clasificación A-B-C con base en sus ventas en dólares, la tabla 2 presenta los resultados generados.

Tabla 1

Resumen de Resultados de Clasificación ABC con base en Ventas

Tipo	# Skus	% Skus	\$ Monto	% Monto
A	144	15%	\$ 70,312,504	80%
B	265	27%	\$ 13,213,981	15%
C	581	59%	\$ 4,403,664	5%
Totales	990	100%	\$ 87,930,149	100%

Con los resultados obtenidos se concluye lo siguiente:

Con el 15% del portafolio de productos (144 skus) se obtiene el 80% de las ventas (\$70 MM), con el 27% del portafolio de productos (265 skus) se obtiene el 15% de las ventas (13 MM) y con el 59% del portafolio de productos (581 skus) se obtiene el 5% de las ventas (4.4 MM).

Al continuar con el desarrollo del análisis exploratorio de datos, se explican las variables que componen el dataset, como se presenta en la tabla 2, el dataset cuenta con 822.295 transacciones y 13 columnas.

Tabla 2

Variables y Registros del Dataset

CAMPO	TIPO DE DATO
id	int64
fecha_venta	datetime64[ns]
cod_almacen	object
cod_articulo	object
articulo	object
marca	object
departamento	object
familia	object
proveedor	object
cantidad	float64
importe	float64
unidad	object
frecuencia	int64

Nota. 822.295 registros y 13 columnas

4.3 Resultados de los clústers de datos generados por el modelo K-Means

Para aplicar la técnica K-MEANS fueron consideradas las variables importe y frecuencia de venta a nivel producto, tal como se muestra en la figura 29.

Figura 29

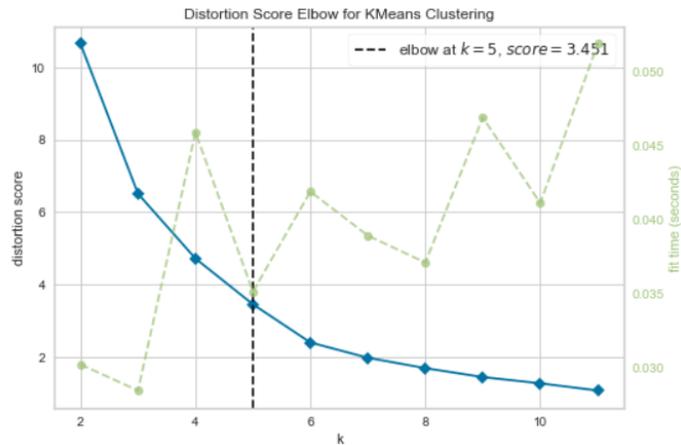
Variables importe y frecuencia resumidas por producto

	importe	frecuencia
0	9821.702	949
1	2555.454	242
2	5237.590	449
3	2608.134	225
4	3546.930	311

Para determinar el número óptimo de clústers se utilizó la técnica del codo (elbow), el resultado determinó 5 clústers de productos tal como se muestra en la imagen 30.

Figura 30

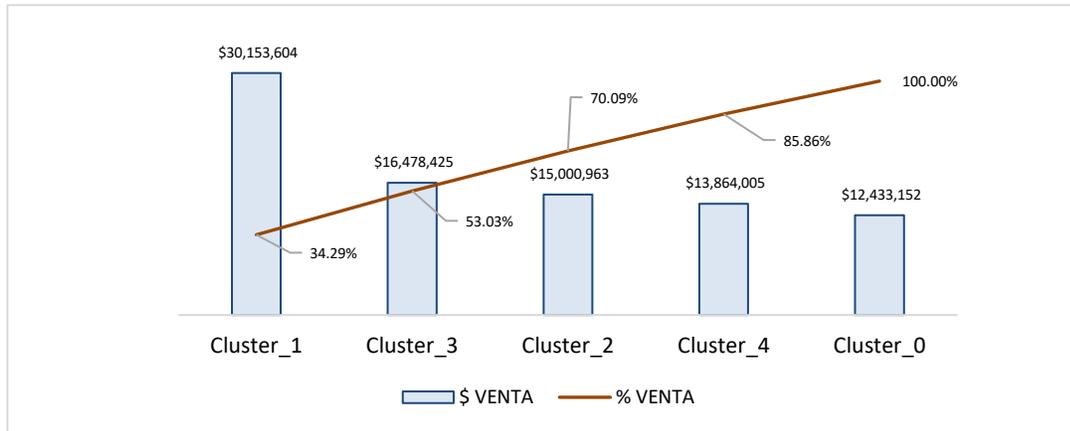
Determinación del Número de Clústers



Con los resultados obtenidos se generó el etiquetado de cada producto y posteriormente se realizó una agregación de los registros con base a las fechas de ventas y etiquetas de cada producto, de esta manera se obtuvieron 5 series de tiempo, en la figura 31 se muestra la representatividad en ventas en dólares de cada clúster de productos.

Figura 31

Representatividad en Ventas por Clúster de Productos



En la tabla 3 presenta un resumen de la composición de cada clúster de productos en cuanto al número de artículos, representatividad sobre el portafolio, concentración y representatividad de las ventas en dólares.

Tabla 3

Composición de Clúster de Productos

K-MEANS	# SKUS	% SKUS	\$ VENTA	% VENTA
Cluster_0	598	60%	\$ 12,433,152	14%
Cluster_3	239	24%	\$ 16,478,425	19%
Cluster_2	95	10%	\$ 15,000,963	17%
Cluster_1	54	5%	\$ 30,153,604	34%
Cluster_4	4	0%	\$ 13,864,005	16%
Total general	990	100%	\$ 87,930,149	100%

En la tabla 4 se expone un resumen de la concentración en ventas en dólares por cada clúster con respecto a las diferentes sucursales de la empresa.

Tabla 4*Concentración en Ventas Clúster por Sucursales*

CLUSTER	GYE	LAG	LRI	UIO	TOTAL
Cluster_1	\$11,030,596	\$2,209,491	\$5,937,273	\$10,976,244	\$30,153,604
Cluster_3	\$3,883,562	\$3,495,937	\$4,100,663	\$4,998,263	\$16,478,425
Cluster_2	\$4,080,799	\$1,401,953	\$3,768,011	\$5,750,200	\$15,000,963
Cluster_4	\$4,964,236	\$587,221	\$3,031,907	\$5,280,640	\$13,864,005
Cluster_0	\$3,082,185	\$2,678,843	\$3,663,089	\$3,009,035	\$12,433,152
TOTAL	\$27,041,377	\$10,373,445	\$20,500,943	\$30,014,383	\$87,930,149

En la tabla 5 se expone un resumen de la participación porcentual en ventas en dólares de cada clúster de productos con respecto a las diferentes sucursales de la empresa.

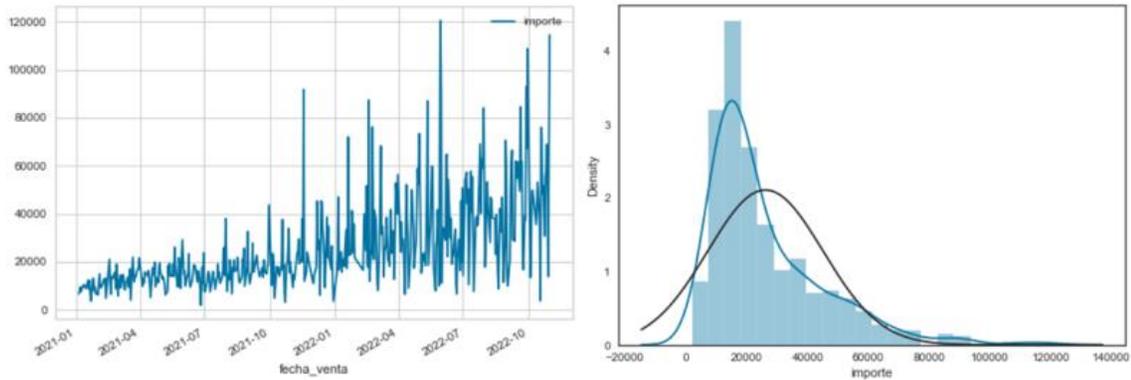
Tabla 5*% Participación en Ventas Clústers por Sucursales*

CLUSTER	GYE	LAG	LRI	UIO	TOTAL
Cluster_1	13%	3%	7%	12%	34%
Cluster_3	4%	4%	5%	6%	19%
Cluster_2	5%	2%	4%	7%	17%
Cluster_4	6%	1%	3%	6%	16%
Cluster_0	4%	3%	4%	3%	14%
TOTAL	31%	12%	23%	34%	100%

En la figura 32 se muestran 2 visualizaciones con respecto al primer clúster de productos (Clúster 0), hacia la izquierda se presenta una visualización de líneas donde se observa una tendencia creciente, así como una ampliación de tipo multiplicativa en el nivel de variabilidad a través del tiempo. En cuanto a la visualización de histograma se evidencia que la forma de la distribución de probabilidades no sigue un comportamiento normal.

Figura 32

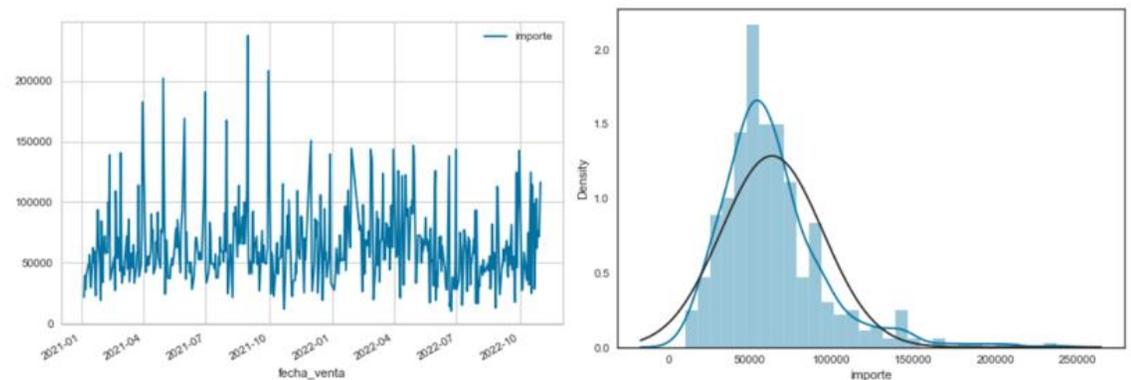
Clúster 0 "Evolución Temporal y Forma de su Distribución de Probabilidades"



En la figura 33 se muestran 2 visualizaciones con respecto al segundo clúster de productos (Clúster 1), hacia la izquierda se presenta una visualización de líneas donde no se observa la presencia de tendencia, también se denota mayor presencia de valores sumamente altos en la primera mitad de la serie con respecto a la segunda mitad. En cuanto a la visualización de histograma se evidencia que la forma de la distribución de probabilidades se asemeja a un comportamiento normal.

Figura 33

Clúster 1 "Evolución Temporal y Forma de su Distribución de Probabilidades"

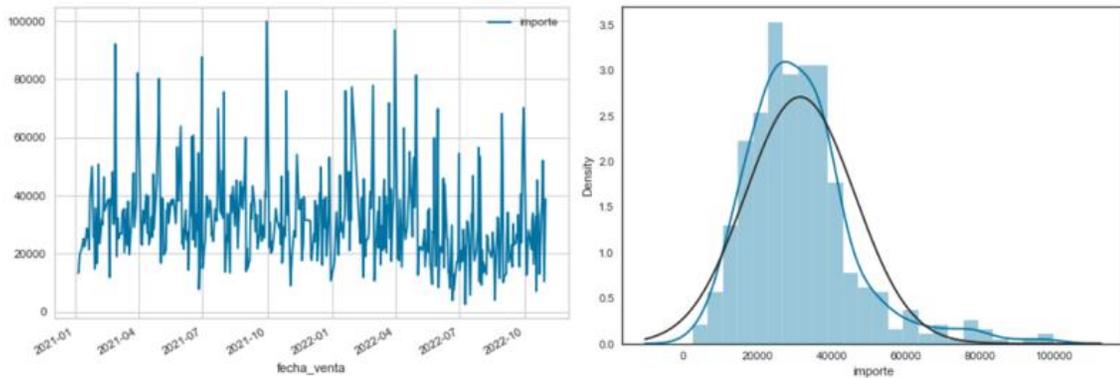


En la figura 34 se muestran 2 visualizaciones con respecto al tercer clúster de productos (Clúster 2), hacia la izquierda se presenta una visualización de líneas donde no se observa la presencia de

tendencia, también se denotan valores sumamente altos a lo largo de toda la serie de tiempo. En cuanto a la visualización de histograma se evidencia que la forma de la distribución de probabilidades se asemeja a un comportamiento normal con un leve sesgo hacia la derecha.

Figura 34

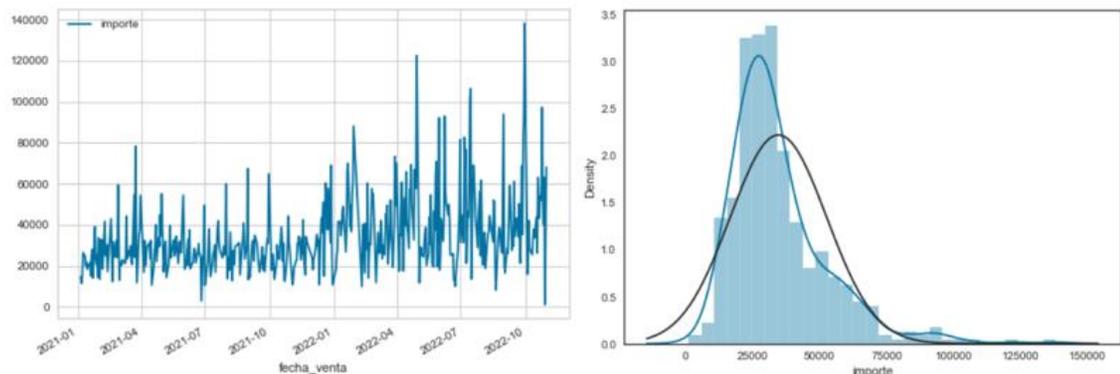
Clúster 2 "Evolución Temporal y Forma de su Distribución de Probabilidades"



En la figura 35 se muestran 2 visualizaciones con respecto al cuarto clúster de productos (Clúster 3), hacia la izquierda se presenta una visualización de líneas donde no se observa la presencia de tendencia, también se denotan valores sumamente altos a partir de la segunda mitad de la serie de tiempo con respecto a la primera mitad. En cuanto a la visualización de histograma se evidencia que la forma de la distribución de probabilidades no se asemeja a un comportamiento normal.

Figura 35

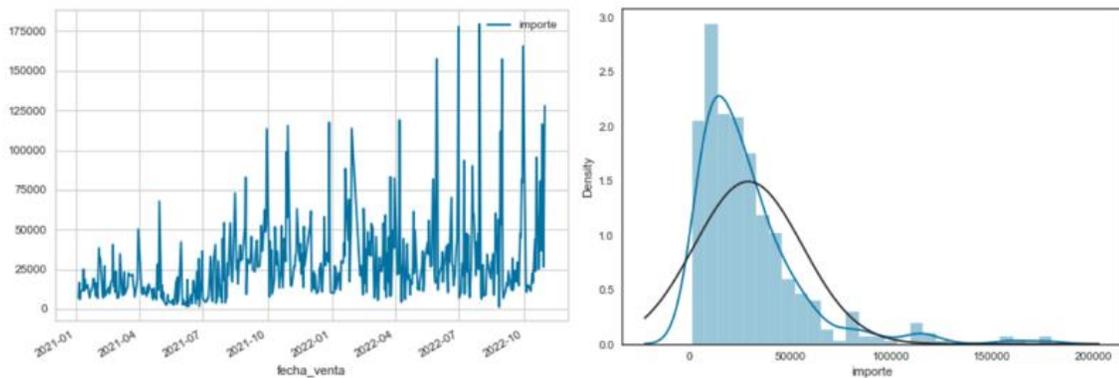
Clúster 3 "Evolución Temporal y Forma de su Distribución de Probabilidades"



En la figura 36 se muestran 2 visualizaciones con respecto al quinto clúster de productos (Clúster 4), hacia la izquierda se presenta una visualización de líneas donde no se observa la presencia de tendencia, también se denotan valores sumamente altos a partir de la segunda mitad de la serie de tiempo con respecto a la primera mitad. En cuanto a la visualización de histograma se evidencia que la forma de la distribución de probabilidades no se asemeja a un comportamiento normal.

Figura 36

Clúster 4 "Evolución Temporal y Forma de su Distribución de Probabilidades"



En la tabla 6 se muestran los principales estadísticos descriptivos para cada una de los clústers de productos.

Tabla 6

Principales Estadísticos Descriptivos por Clúster de Productos

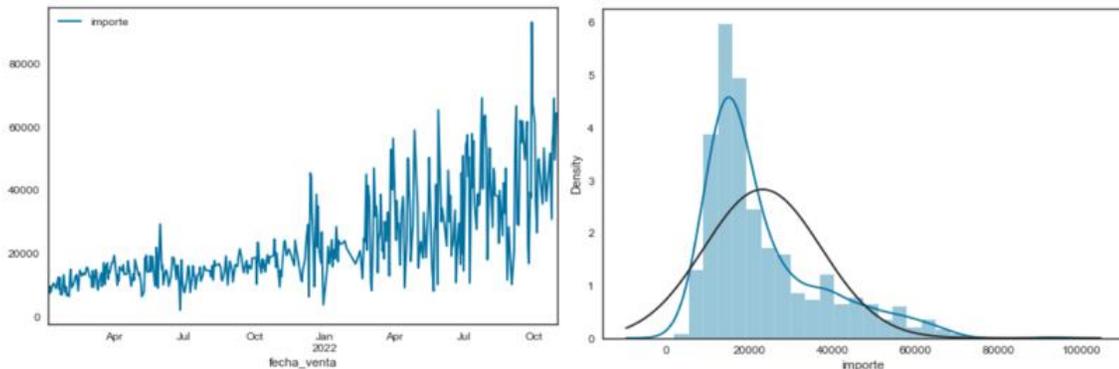
K-MEANS	MAXIMO	MINIMO	PROMEDIO	DESVIACION ESTANDAR	COEFICIENTE DE VARIACION
Cluster_0	\$ 120,386	\$ 1,912	\$ 26,120	\$ 18,950	0.725
Cluster_1	\$ 237,374	\$ 10,214	\$ 63,347	\$ 31,126	0.491
Cluster_2	\$ 99,860	\$ 2,573	\$ 31,514	\$ 14,761	0.468
Cluster_3	\$ 138,173	\$ 1,221	\$ 34,618	\$ 18,005	0.520
Cluster_4	\$ 179,432	\$ 1,195	\$ 29,126	\$ 26,783	0.920

Sin embargo, estas series presentaban datos faltantes y valores atípicos, por tal motivo fue necesario realizar imputación de datos en ambos casos, a continuación se analizarán nuevamente a cada una de las series de tiempo posterior a la aplicación de los procesos de imputación.

En la figura 37 se muestran 2 visualizaciones con respecto al primer clúster de productos (clúster 0), hacia la izquierda se observa una gráfica de líneas donde se aprecia que la magnitud de los valores máximos y mínimos se ha reducido, la serie de tiempo mantiene un comportamiento con tendencia creciente y la variabilidad en los datos se incrementa hacia el final de la serie, en cuanto al histograma, se observa que la distribución de probabilidad de los datos no se asemeja a una distribución normal.

Figura 37

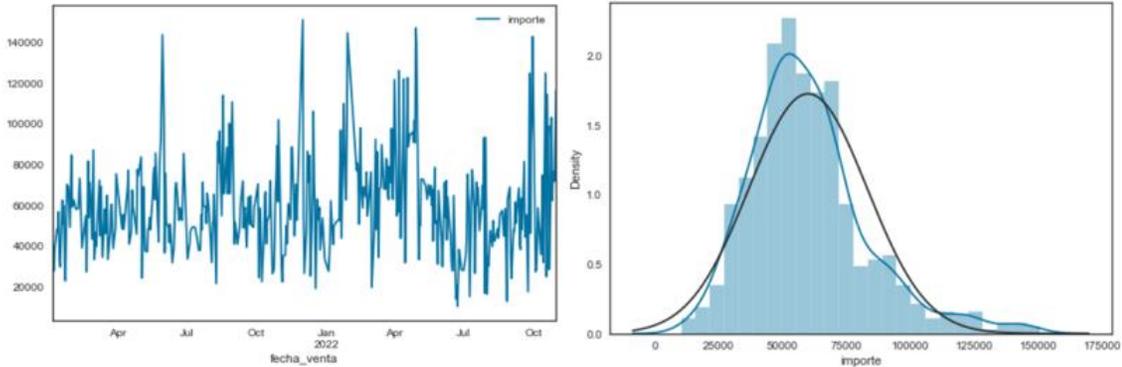
Clúster 0 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)



En la figura 38 se muestran 2 visualizaciones con respecto al segundo clúster de productos (clúster 1), hacia la izquierda se observa una gráfica de líneas donde se aprecia que la magnitud de los valores máximos y mínimos se ha reducido, la serie de tiempo no presenta un comportamiento con tendencia y la mayor variabilidad en los datos se refleja en segmentos específicos de la serie, en cuanto al histograma, se puede observar que la distribución de probabilidad de los datos se aproxima a la de una distribución normal.

Figura 38

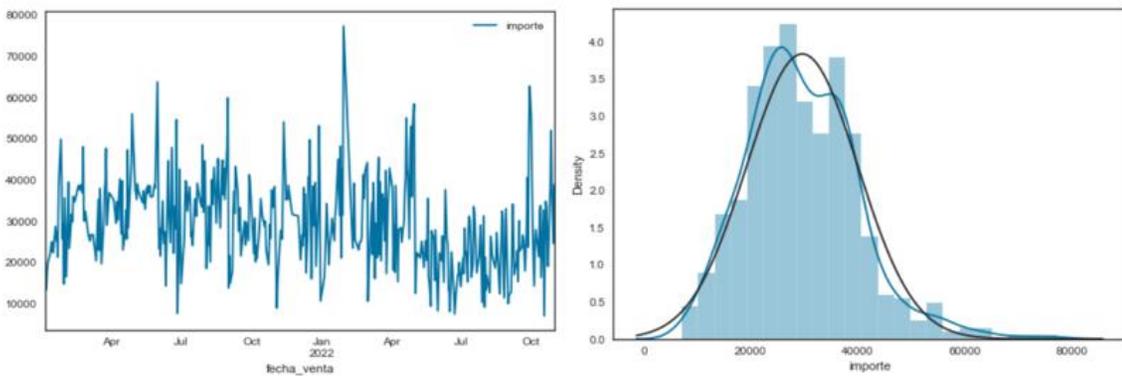
Clúster 1 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)



En la figura 39 se muestran 2 visualizaciones con respecto al tercer clúster de productos (clúster 2), hacia la izquierda se observa una gráfica de líneas donde se aprecia que la magnitud de los valores máximos y mínimos se ha reducido, la serie de tiempo no presenta un comportamiento con tendencia y la mayor variabilidad en los datos se refleja en segmentos específicos de la serie, en cuanto al histograma, se evidencia que la distribución de probabilidad de los datos tienen semejanza a una distribución normal.

Figura 39

Clúster 2 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)

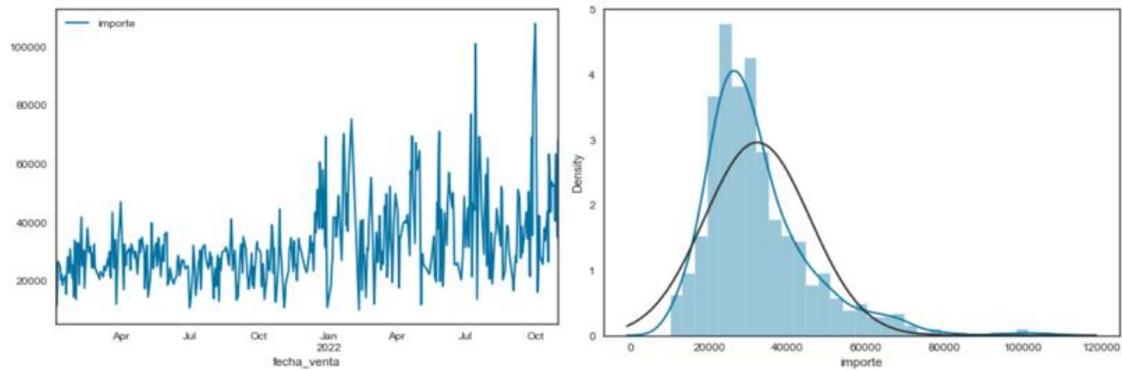


En la figura 40 se muestran 2 visualizaciones con respecto al cuarto clúster de productos (clúster 3), hacia la izquierda se observa una gráfica de líneas donde se aprecia que la magnitud de los valores máximos y mínimos se ha reducido, la serie de tiempo no refleja un comportamiento con tendencia y la

mayor variabilidad en los datos se refleja en segmentos específicos de la serie, en cuanto al histograma, es evidente que la distribución de probabilidad de los datos no sigue la forma de una distribución normal.

Figura 40

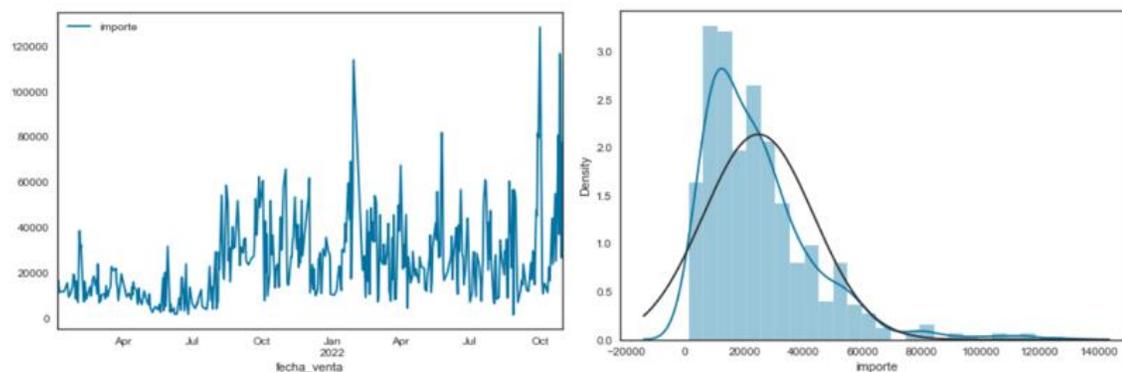
Clúster 3 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)



En la figura 41 se muestran 2 visualizaciones con respecto al cuarto clúster de productos (clúster 3), hacia la izquierda se observa una gráfica de líneas donde se aprecia que la magnitud de los valores máximos y mínimos se ha reducido, la serie de tiempo no refleja un comportamiento con tendencia y la mayor variabilidad en los datos se refleja en segmentos específicos de la serie, en cuanto al histograma, se observa que la distribución de probabilidad de los datos no sigue la forma de una distribución normal.

Figura 41

Clúster 4 Posterior a Imputación de Datos Faltantes y Puntos Singulares (Outliers)



En la tabla 7 se muestran los principales estadísticos descriptivos para cada clúster de productos una vez realizadas las imputaciones correspondientes.

Tabla 7

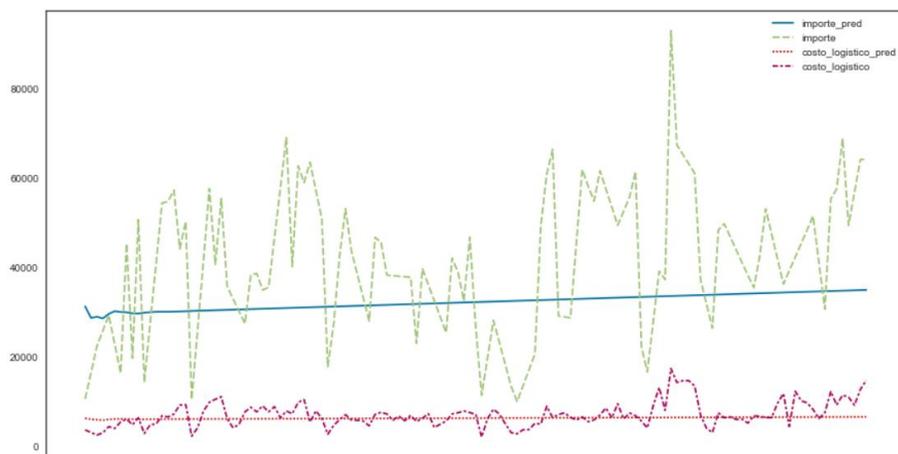
Principales Estadísticos Descriptivos por Clúster de Productos posterior a Imputaciones

K-MEANS	MAXIMO	MINIMO	PROMEDIO	DESVIACION ESTANDAR	COEFICIENTE DE VARIACION
Cluster_0	\$ 92,949	\$ 1,912	\$ 23,193	\$ 14,128	0.609
Cluster_1	\$ 150,783	\$ 10,214	\$ 59,913	\$ 23,118	0.386
Cluster_2	\$ 77,315	\$ 7,089	\$ 29,601	\$ 10,415	0.352
Cluster_3	\$ 107,633	\$ 9,978	\$ 32,193	\$ 13,491	0.419
Cluster_4	\$ 128,181	\$ 1,195	\$ 24,779	\$ 18,689	0.754

En la figura 42 se presentan los resultados de los pronósticos para las series de tiempo del clúster 0 y de los costos logísticos, en ambos casos se evidencia que los mejores pronósticos que pudo generar el modelo VAR son líneas rectas y específicamente para el caso del clúster 0 se incluye una leve tendencia creciente.

Figura 42

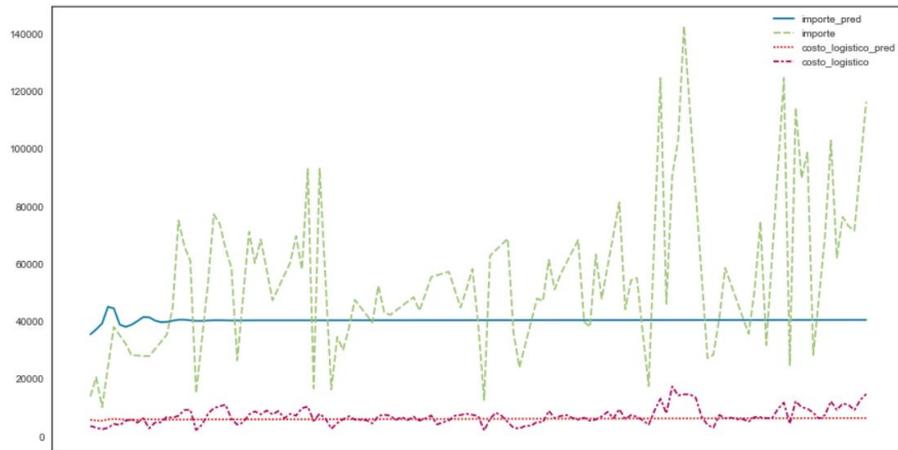
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 0 y Costos Logísticos



La figura 43 expone los resultados de los pronósticos con respecto a las series de tiempo del clúster 1 y de los costos logísticos, en ambos casos se evidencia que los mejores pronósticos que pudo generar el modelo VAR son líneas rectas horizontales.

Figura 43

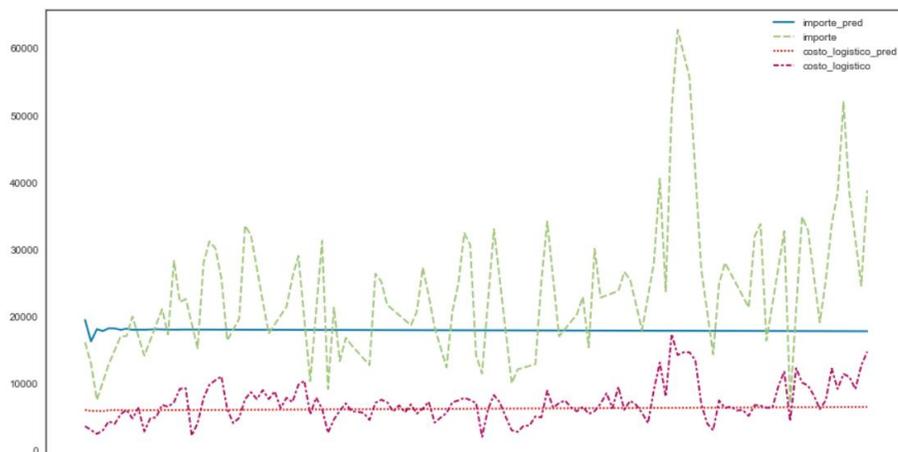
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 1 y Costos Logísticos



La figura 44 expone los resultados de los pronósticos con respecto a las series de tiempo del clúster 2 y de los costos logísticos, en ambos casos se evidencia que los mejores pronósticos que pudo generar el modelo VAR son líneas rectas horizontales.

Figura 44

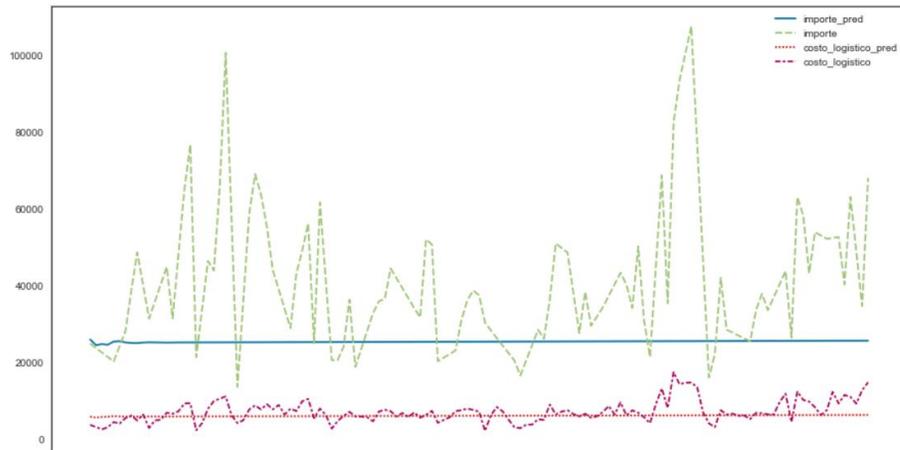
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 2 y Costos Logísticos



La figura 45 expone los resultados de los pronósticos correspondientes a las series de tiempo del clúster 3 y de los costos logísticos, en ambos casos se evidencia que los mejores pronósticos que pudo generar el modelo VAR son líneas rectas horizontales.

Figura 45

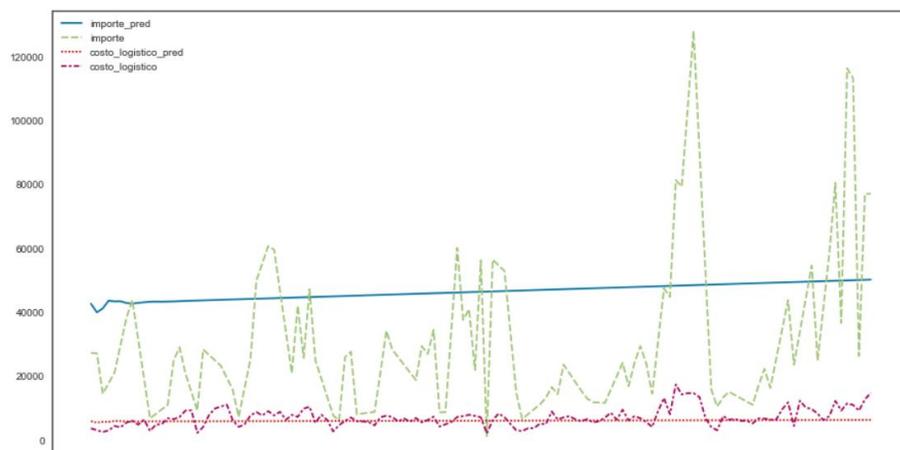
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 3 y Costos Logísticos



La figura 46 expone los resultados de los pronósticos correspondientes a las series de tiempo del clúster 4 y de los costos logísticos, en ambos casos se evidencia que los mejores pronósticos que pudo generar el modelo VAR son líneas rectas horizontales.

Figura 46

Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 4 y Costos Logísticos



En la tabla 8 se exponen los resultados obtenidos de las métricas correspondientes a los errores de pronósticos RMSE y MAPE en la aplicación del método VAR a cada clúster de producto.

Tabla 8

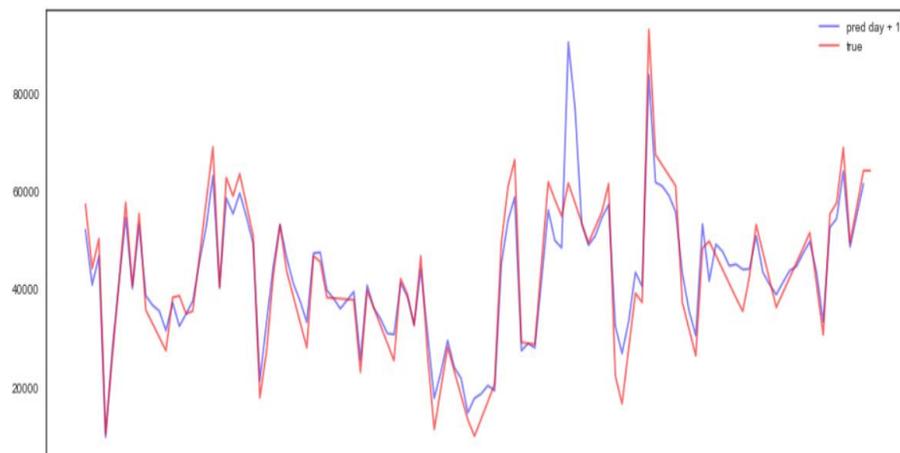
Resultados de Cálculos de Error del Pronóstico Método VAR

K-MEANS	RMSE	MAPE
Cluster_0	17458	43.99%
Cluster_1	28500	51.75%
Cluster_2	11155	43.36%
Cluster_3	23097	64.80%
Cluster_4	27985	52.85%

En la figura 47 se muestra el pronóstico generado para los datos de prueba correspondientes al primer clúster de productos (clúster 0) haciendo uso de redes neuronales LSTM, se evidencia que el pronóstico releja un buen nivel de aprendizaje en base a los datos con los cuales fue entrenado el modelo.

Figura 47

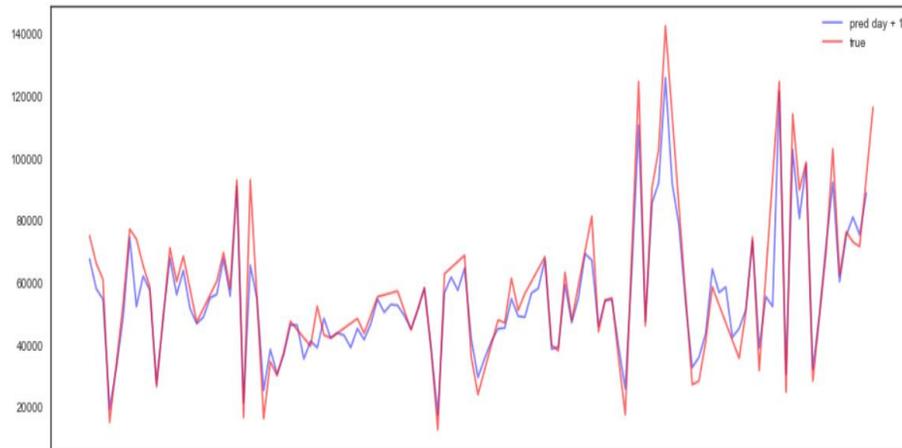
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 0 aplicando LSTM



En la figura 48 se muestra el pronóstico generado para los datos de prueba correspondientes al segundo clúster de productos (clúster 1) haciendo uso de redes neuronales LSTM, se evidencia que el pronóstico releja un buen nivel de aprendizaje en base a los datos con los cuales fue entrenado el modelo.

Figura 48

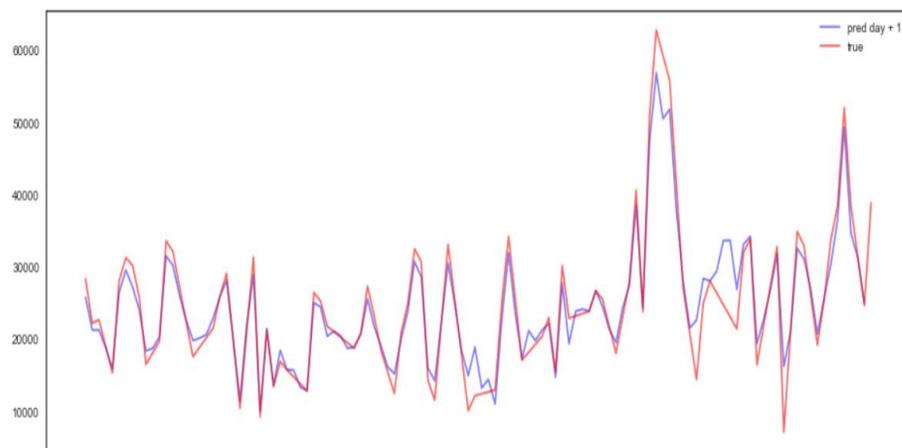
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 1 aplicando LSTM



En la figura 49 se muestra el pronóstico generado para los datos de prueba correspondientes al tercer clúster de productos (clúster 2) haciendo uso de redes neuronales LSTM, se evidencia que el pronóstico refleja un buen nivel de aprendizaje en base a los datos con los cuales fue entrenado el modelo.

Figura 49

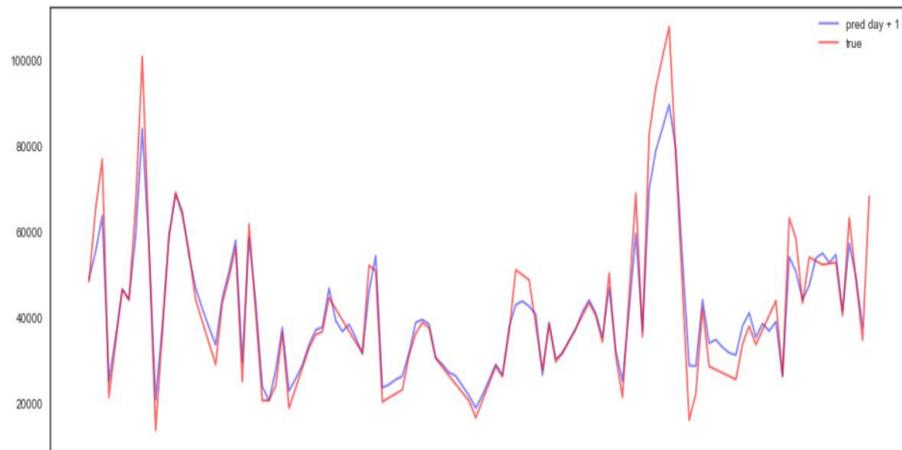
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 2 aplicando LSTM



En la figura 50 se muestra el pronóstico generado para los datos de prueba correspondientes al cuarto clúster de productos (clúster 3) haciendo uso de redes neuronales LSTM, se evidencia que el pronóstico refleja un buen nivel de aprendizaje en base a los datos con los cuales fue entrenado el modelo.

Figura 50

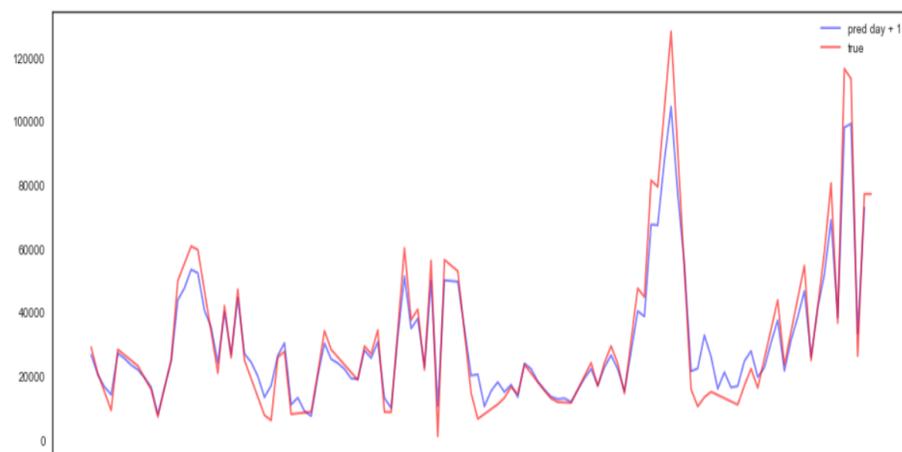
Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 3 aplicando LSTM



En la figura 51 se muestra el pronóstico generado para los datos de prueba correspondientes al quinto clúster de productos (clúster 4) haciendo uso de redes neuronales LSTM, se evidencia que el pronóstico refleja un buen nivel de aprendizaje en base a los datos con los cuales fue entrenado el modelo.

Figura 51

Pronóstico sobre Datos de Prueba para las Series de Tiempo Clúster 4 aplicando LSTM



En la tabla 9 se exponen los resultados de las métricas de los errores de pronósticos RMSE y MAPE correspondientes a la aplicación del método LSTM a cada clúster de producto.

Tabla 9

Resultados de Cálculos de Error del Pronóstico Método LSTM

K-MEANS	RMSE	MAPE
Cluster_0	5069	10.06%
Cluster_1	7931	9.51%
Cluster_2	2963	8.97%
Cluster_3	5712	8.97%
Cluster_4	6231	26.24%

Finalmente, en la tabla 10 se comparan los resultados obtenidos por ambos métodos de pronóstico aplicados a cada clúster de productos, se evidencia que en todos los casos, los pronósticos generados con LSTM presentan mejores resultados con respecto a los pronósticos generados con el método VAR.

Tabla 10

Comparativo de Resultados de Métricas de Error

K-MEANS	LSTM		VAR	
	RMSE	MAPE	RMSE	MAPE
Cluster_0	5069	10.06%	17458	43.99%
Cluster_1	7931	9.51%	28500	51.75%
Cluster_2	2963	8.97%	11155	43.36%
Cluster_3	5712	8.97%	23097	64.80%
Cluster_4	6231	26.24%	27985	52.85%

Con base en los resultados obtenidos, únicamente los pronósticos generados con LSTM será considerados en la construcción del tablero (dashboard) de presentación final de resultados en Power BI.

A continuación se presentará de forma resumida el tablero (Dashboard) final con el cual se busca generar una herramienta práctica y útil para los usuarios de la empresa.

En la figura 52 se presenta la pantalla de autenticación de usuarios, esta funcionalidad permitirá tener un acceso controlado de usuarios al aplicativo.

Figura 52

Pantalla de Acceso al Sistema



En la figura 53 se muestra una presentación general de los resultados de los análisis descriptivos y predictivos realizados dentro del presente proyecto, algunas de las preguntas de negocio para las cuales el usuario final podría obtener respuesta son: ¿Cuál es la representatividad en ventas en dólares por cada clúster de productos?, ¿Cuál es la representatividad en ventas en dólares por cada sucursal con respecto a cada clúster de productos?, ¿Existe un crecimiento o decrecimiento en las ventas en dólares en el año actual versus el año anterior a nivel de clústers de productos y/o sucursales?, ¿Cuál es el nivel de precisión de los pronósticos generados por el modelo?, etc.

Figura 53

Pantalla de Visualización General de Resultados

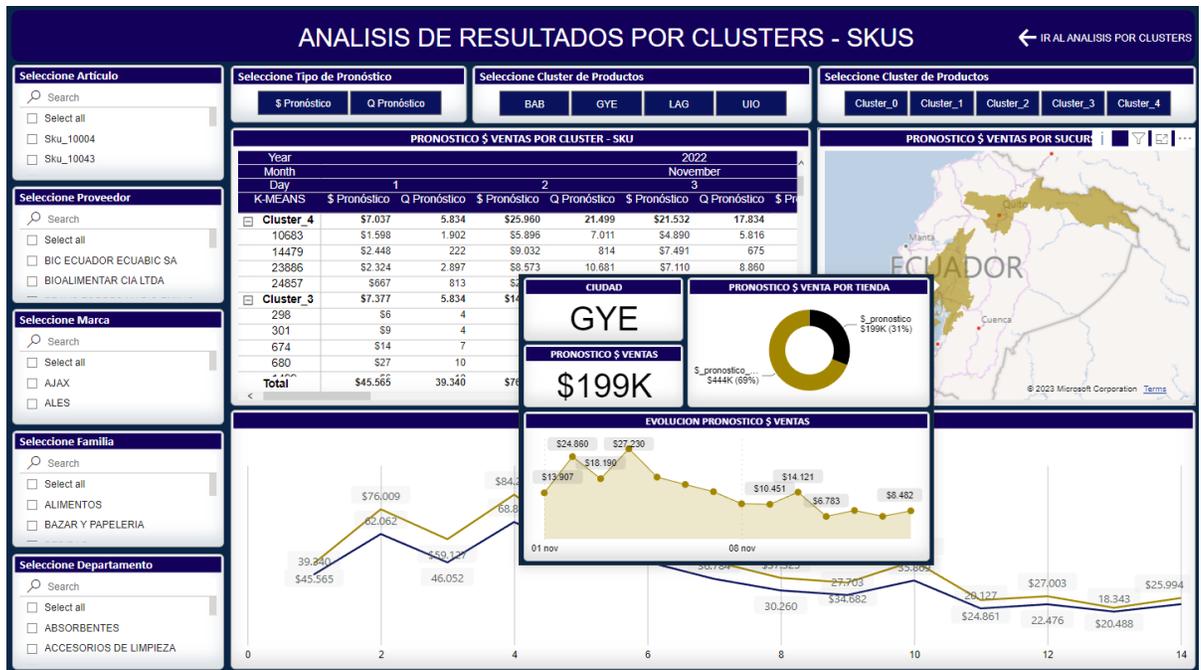


En la figura 54 se muestra una presentación detallada de los resultados de los pronósticos obtenidos, al mencionar que es a nivel de detalle se hace referencia a evaluar resultados a nivel producto ya sea en unidades o dinero, de esta manera el usuario puede analizar al de forma minuciosa los resultados a fin de obtener información de valor.

Los resultados que se muestran en esta pantalla son de utilidad en la planeación a corto plazo, es decir, son resultados con enfoque operativo, por ejemplo: Definir activaciones promocionales de corta duración o por otra parte servir para anticipar reposiciones de inventario a proveedores que manejen tiempos cortos de respuesta.

Figura 54

Pantalla de Visualización Detallada de Resultados



Seleccione Proveedor

- Select all
- BIC ECUADOR ECUABIC SA
- BIOALIMENTAR CIA LTDA

Seleccione Marca

- Select all
- AJAX
- ALES

Seleccione Familia

- Select all
- ALIMENTOS
- BAZAR Y PAPELERIA

Seleccione Departamento

- Select all
- ABSORBENTES
- ACCESORIOS DE LIMPIEZA

5. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- Inicialmente se consideró la posibilidad de generar pronósticos de series de tiempo para cada producto del portafolio, sin embargo, al realizar el análisis exploratorio de datos se detectó una elevada variabilidad en el comportamiento de venta de los productos, así como un importante volumen de datos faltantes, por ello, la mejor opción fue aplicar una técnica de aprendizaje no supervisado para conseguir la clusterización de productos, en este caso se obtuvieron 5 clústers.
- La cantidad de variables disponibles en el dataset es reducida, esto por motivos de confidencialidad de los datos, al aplicar la técnica de clusterización de K-MEANS únicamente fue posible considerar a las variables “importe” y “frecuencia”, se debe considerar que los moldeos multivariantes están diseñados para trabajar con un mayor número de variables.
- El uso de Redes Neuronales Recurrentes LSTM exige contar con un elevado número de datos históricos disponibles (> 10,000), este factor es sumamente relevante pues al disponer de un reducido número de datos fue necesario aplicar técnicas complementarias como el remuestreo a fin de permitirle al modelo “aprender” de la historia (datos de entrenamiento) y pueda generar pronósticos con un nivel de error aceptable, para este proyecto, 4 de los 5 clústers tuvieron pronósticos con un nivel de error (MAPE) máximo del 10%, solo 1 clúster (que representa el 15% de las \$ ventas) superó este umbral.
- Al aplicar los modelos de Vectores Autorregresivos VAR se evidencia que los pronósticos tienen niveles de error (MAPE) superiores al 10%, en este caso el nivel de variabilidad en los datos

representa un patrón de comportamiento que no pudo ser aprendido por estos modelos en comparación a los modelos LSTM.

- Finalmente, para este proyecto tomó consideración poder trabajar con acceso a nuevos datos de forma recurrente, es decir, a través de una frecuencia predefinida, por ejemplo: semanal, quincenal o mensual, de esta manera se les permite a los modelos de pronósticos aprender de nuevos datos para poder actualizar las proyecciones considerando cambios en el comportamiento comercial de la empresa.

5.2 RECOMENDACIONES

- Se recomienda incluir más campos de información en el dataset, esto con la finalidad de evaluar si los modelos de multivariantes pudiesen mejorar los resultados actualmente obtenidos.
- Para el clúster 4 de productos se obtuvo un pronóstico con un nivel de error (MAPE) superior al 10%, coincidentemente este clúster se encuentra conformado por 4 productos que poseen un comportamiento muy variable a través del tiempo, para próximos estudios se podría considerar la opción de pronosticar cada producto de forma independiente en lugar de agrupada para determinar si se logran obtener mejores pronósticos.
- Automatizar los procesos de actualización de data histórica y ejecución de los modelos de pronósticos con la finalidad de asegurar que los usuarios cuenten con la información más reciente para el desarrollo de sus análisis de negocios.
- Finalmente, considerando que se utilizan datos diarios de ventas, se recomienda que la actualización de estos datos, así como sus respectivos pronósticos se recalculen con frecuencia semanal, quincenal o mensual.

6. REFERENCIAS

APICS. (2018). *CPIM CERTIFIED IN PRODUCTION AND INVENTORY MANAGEMENT*. ASCM.

APICS. (2022). *CSCP CERTIFIED SUPPLY CHAIN PROFESSIONAL*. ACSM.

Atwan, T. A. (2022). *Time Series Analysis with Python Cookbook*. Packt.

Bozarth, C. (2019). *INTRODUCTION TO OPERATIONS AND SUPPLY CHAIN MANAGEMENT*. Pearson.

Brockwell, P., & David, R. (2016). *Introduction to Time Series and Forecasting*. Springer.

CHIANG, T.-A., CHE, Z.-H., & HUNG, C.-W. (2023, febrero). A K-Means Clustering and the Prim's Minimum Spanning Tree-Based Optimal Picking-List Consolidation and Assignment Methodology for Achieving the Sustainable Warehouse Operations. *Sustainability (Switzerland)*, 15(4), 6. doi:10.3390/su15043544

Chollet, F. (2021). *DEEP LEARNING WITH PYTHON*. MANNING.

Farrera, A. (2013). *Manual de Pronósticos para la toma de Decisiones*. EDITORIAL DIGITAL TECNOLÓGICO DE MONTERREY.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'REILLY.

Goldmeier, J., & Gutman, A. (2021). *BECOMING A DATA HEAD HOW TO THINK, SPEAK, AND UNDERSTAND DATA SCIENCE, STATISTICS, AND MACHINE LEARNING*. WILEY.

Grus, J. (2019). *Data Science from Scratch*. O'REILLY.

HONG, J., LIANG, F., GONG, X., & YU, Q. (2022, diciembre). Accurate State of Charge Estimation for Real-World Battery Systems Using a Novel Grid Search and Cross Validated Optimised LSTM Neural Network. *MDPI*, 15(24), 8. doi:10.3390/en15249654

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. Texts.

Jacobs, R. (2019). *OPERATIONS AND SUPPLY CHAIN MANAGEMENT THE CORE*. Mc Graw Hill Education.

Manu, J. (2022). *Modern Time Series Forecasting with Python*. Packt.

Marz, N. (2015). *Big Data: Principles and best practices of scalable realtime data systems*. MANNING PUBLICATIONS.

MICROSOFT. (2023, marzo 24). *Ajuste de hiperparámetros de un modelo*. Retrieved from MICROSOFT: <https://learn.microsoft.com/es-es/azure/machine-learning/how-to-tune-hyperparameters>

MIKY, Y., KALOOP, M., ELNABWY, M., BAKI, A., & ALSHOUNY, A. (2021). A Recurrent-Cascade-Neural network- nonlinear autoregressive networks with exogenous inputs (NARX) approach for long-term time-series prediction of wave height based on wave characteristics measurements. *Ocean Engineering*, 240(1), 12. doi:10.1016/j.oceaneng.2021.109958

Montemayor, J. (2013). *Métodos de Pronósticos para Negocios*. EDITORIAL DIGITAL TECNOLÓGICO DE MONTERREY.

Nielsen, A. (2020). *Practical Time Series Analysis*. O'REILLY.

Peixeiro, M. (2022). *Time Series Forecasting in Python*. MANNING.

Rong, J. (2015). *Time Series Econometrics: Granger Causality*. JRong.

Rubiales, A. (2020, julio 27). *Qué son las distancias, tipos y cómo calcularlas con Python*. Retrieved from Medium: <https://rubialesalberto.medium.com/qu%C3%A9-son-las-distancias-tipos-y-c%C3%B3mo-calcularlas-con-python-eb6a95d9e74>

Sckit Learn. (2022, diciembre 01). *Exhaustive Grid Search*. Retrieved from Sckit Learn: https://scikit-learn.org/stable/modules/grid_search.html

Shumway, R., & Stoffer, D. (2017). *Time Series Analysis and Its Applications*. Springer.

Vandeput, N. (2021). *Data Science for Supply Chain Forecasting*. De Gruyter.

Vaughan, D. (2020). *Analytical Skills for AI & Data Science*. O'REILLY.

Wilson, J. E. (2021). *PREDICTIVE ANALYTICS FOR BUSINESS FORECASTING & PLANNING*. GRACEWAY
PUBLISHING COMPANY.

7. ANEXOS

Anexo 1

Detalle de Campos del Dataset

VARIABLES	TIPO	DESCRIPCION
fecha_ventas	Date	Corresponde a la fecha en la que se realizó cada venta.
cod_almacen	Chr	Indica la identificación de la ubicación física del almacén donde se realizó la venta.
cod_articulo	Chr	Indica el código identificador de cada producto que fué vendido
articulo	Chr	Indica la descripción de cada producto que fue vendido
marca	Chr	Corresponde al nombre de la marca de cada producto que fué vendido
departamento	Chr	Corresponde al nombre del departamento comercial para cada producto que fué vendido
familia	Chr	Corresponde al nombre de la familia de cada producto que fué vendido
id_proveedor	Chr	Indica el código identificador del proveedor que suministra cada producto que fué vendido
frecuencia	Chr	Indica el número de veces que cada producto ha sido vendido
um	Chr	Indica la unidad de medida de cada producto que fué vendido
cantidad	Float	Corresponde al número de unidades vendidas
importe	Float	Corresponde al ingreso generado por las unidades vendidas