



UNIVERSIDAD RICARDO PALMA

ESCUELA DE POSGRADO

MAESTRÍA EN CIENCIA DE LOS DATOS

MODELO DE REDES NEURONALES ARTIFICIALES PARA EL
PRONÓSTICO DE LA DEMANDA DE STOCK EN UNA EMPRESA DE
CONSUMO MASIVO

TESIS

PARA OPTAR EL GRADO ACADÉMICO DE MAESTRO EN
CIENCIA DE LOS DATOS

AUTOR

RAFAEL TAIBE, OSMAN DENNYS

(ORCID: 0000-0001-7401-7428)

ASESOR

TRISTAN GOMEZ, LUDGARDO EDER

(ORCID: 0000-0002-7957-5717)

Lima, Perú

2024

Metadatos Complementarios

Datos de autor

Rafael Taipe, Osman Dennys

Tipo de documento de identidad del AUTOR: DNI

Número de documento de identidad del AUTOR: 45232490

Datos de asesor

Tristan Gomez, Ludgardo Eder

Tipo de documento de identidad del ASESOR: DNI

Número de documento de identidad del ASESOR: 45168254

Datos del jurado

JURADO 1: Roque Paredes, Ofelia, DNI N° 06243124, ORCID

0000-0001-8280-021X

JURADO 2: Cardenas Garro, Jose Antonio, DNI N° 47199993, ORCID

0000-0002-8343-2540

JURADO 3: Ruiz Orderique, German Junior Alexander, DNI N° 40717021,

ORCID 0000-0003-3195-6585

Datos de la investigación

Campo del conocimiento OCDE: 612027

Código del Programa: 1.02.00

DECLARACIÓN JURADA DE ORIGINALIDAD

Yo, Osman Dennys Rafael Taipe, con código de estudiante N° 202013280, con (DNI o Carné de Extranjería) N°45232490, con domicilio en Jirón Toribio Rodríguez de Mendoza # 32, distrito de Comas, provincia y departamento de Lima, en mi condición de egresado de la escuela de postgrado de La Maestría en Ciencia de los Datos, declaro bajo juramento que:

La presente tesis titulada; “Modelo de redes neuronales artificiales para el pronóstico de la demanda de stock en una empresa de consumo masivo”, es de mi única autoría, bajo el asesoramiento del docente Mg. Tristan Gomez Ludgardo Eder, y no existe plagio y/o copia de ninguna naturaleza, en especial de otro documento de investigación presentado por cualquier persona natural o jurídica ante cualquier institución académica o de investigación, universidad, etc; la cual ha sido sometida al antiplagio Turnitin y tiene el 14% de similitud final.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en la tesis, el contenido de estas corresponde a las opiniones de ellos, y por las cuales no asumo responsabilidad, ya sean de fuentes encontradas en medios escritos, digitales o de internet.

Asimismo, ratifico plenamente que el contenido íntegro de la tesis es de mi conocimiento y autoría. Por tal motivo, asumo toda la responsabilidad de cualquier error u omisión en la tesis y soy consciente de las connotaciones éticas y legales involucradas.

En caso de falsa declaración, me someto a lo dispuesto en las normas de la Universidad Ricardo Palma y a los dispositivos legales nacionales vigentes.

Surco, 09 de mayo de 2024



Osman Dennys Rafael Taipe

DNI:45232490

INFORME DE SIMILITUD DEL PROGRAMA ANTIPLAGIO TURNITIN

MODELO DE REDES NEURONALES ARTIFICIALES PARA EL PRONÓSTICO DE LA DEMANDA DE STOCK EN UNA EMPRESA DE CONSUMO MASIVO

INFORME DE ORIGINALIDAD

14%	13%	4%	5%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	1library.co Fuente de Internet	3%
2	hdl.handle.net Fuente de Internet	3%
3	opac.pucv.cl Fuente de Internet	2%
4	silو.tips Fuente de Internet	2%
5	Montes Ríos Edgar Alan. "Redes neuronales para la estimación de propiedades petrofísicas a escala sísmica usando atributos sísmicos y la prueba Gamma", TESIUNAM, 2019 Publicación	1%
6	alicia.concytec.gob.pe Fuente de Internet	1%
7	repositorio.unsaac.edu.pe Fuente de Internet	

		<1 %
8	www.slideshare.net Fuente de Internet	<1 %
9	rdu.iua.edu.ar Fuente de Internet	<1 %
10	cybertesis.unmsm.edu.pe Fuente de Internet	<1 %
11	vbook.pub Fuente de Internet	<1 %
12	docplayer.es Fuente de Internet	<1 %
13	janet-panic.com Fuente de Internet	<1 %
14	repository.unimilitar.edu.co Fuente de Internet	<1 %
15	Submitted to unsaac Trabajo del estudiante	<1 %
16	Submitted to Ministerio de Defensa Trabajo del estudiante	<1 %
17	Rosas Hernandez Javier. "Redes neuronales y algoritmos geneticos : análisis de equivalencia funcional aplicado a series de tiempo en inversiones", TESIUNAM, 2005 Publicación	<1 %

18	repositorio.utp.edu.pe Fuente de Internet	<1 %
19	studylib.es Fuente de Internet	<1 %
20	www.urp.edu.pe Fuente de Internet	<1 %
21	repositorio.uho.edu.cu Fuente de Internet	<1 %
22	es.wikipedia.org Fuente de Internet	<1 %
23	accedacris.ulpgc.es Fuente de Internet	<1 %
24	qdoc.tips Fuente de Internet	<1 %
25	Submitted to Pontificia Universidad Catolica del Peru Trabajo del estudiante	<1 %
26	logisticafiis.blogspot.com Fuente de Internet	<1 %
27	openaccess.uoc.edu Fuente de Internet	<1 %

DEDICATORIA

A mi esposa Vanessa, mi
abuela Alejandrina, mis padres
Carmen y Ursisino, hermanas
Lorena y Katy; quienes me
brindaron su apoyo
incondicional en este proceso.

AGRADECIMIENTOS

A los profesores de la maestría en Ciencia de Datos de la Universidad Ricardo Palma, por su dedicación y constante enseñanza, con un especial agradecimiento a la profesora Ofelia Roque Paredes.

A mi asesor de tesis por su paciencia y constante dedicación Eder Tristán Gomez.

A mis tíos Héctor Taipe Martínez, Fredy Taipe Martínez y Betty Taipe Martínez por su constante apoyo y consejos a lo largo de toda mi carrera, en especial a Héctor por ser como un padre y estar en los momentos difíciles.

Índice de contenidos

DEDICATORIA.....	vi
AGRADECIMIENTOS.....	vii
Índice de tablas.....	xii
Índice de figuras.....	xiii
RESUMEN.....	xvi
ABSTRACT.....	xvii
INTRODUCCIÓN.....	xviii
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA.....	1
1.1 Descripción del problema.....	1
1.2 Formulación del problema.....	4
1.2.1 Formulación general.....	4
1.2.2. Formulación específica.....	4
1.3 Importancia y justificación del estudio.....	4
1.3.1 Importancia.....	4
1.3.2 Justificación.....	5
1.4 Delimitación del estudio.....	6
1.5 Objetivos de la investigación.....	7
1.5.1 Objetivo General:.....	7
1.5.2 Objetivos Específicos:.....	7
CAPÍTULO II: MARCO TEÓRICO.....	7

2.1 Marco histórico	7
2.1.1 Redes neuronales artificiales recurrentes	7
2.1.2 Pronóstico de demanda de stocks	10
2.2 Investigaciones relacionadas con el tema	15
2.2.1 Investigaciones nacionales relacionadas con el tema	15
2.2.2 Investigaciones extranjeras relacionadas con el tema	17
2.3 Estructura teórica y científica que sustenta el estudio	26
2.3.1 Demanda de Productos	26
2.3.3 Métodos de pronóstico de demanda	27
2.3.4 Stock Productos	31
2.3.5 Planificación de la demanda	31
2.3.6 Gestión de demanda.....	32
2.3.7 Redes Neuronales Artificiales	33
2.3.8 Neurona Biológica.....	34
2.3.9 Neurona Artificial.....	36
2.3.10 Funciones de activación	37
2.3.11 Tipología de las RNA	40
2.3.12 Redes recurrentes Profundas	46
2.3.13 Problema de dependencia a largo plazo	46
2.3.14 Gradiente descendente	47
2.3.15 LSTM	47
2.3.16 Funcionamiento de la red LSTM.....	48

2.3.17 GRU.....	52
2.3.18 Series Temporales.....	53
2.3.19 Descomposición de una serie de tiempo	54
2.3.20 Evaluación de Modelos de Pronóstico.....	56
2.3.21 Mapa conceptual de las teorías que sustentan el estudio.....	57
2.4 Definición de términos básicos.....	58
2.5 Hipótesis	62
2.6.1 Hipótesis General	62
2.6.2 Hipótesis Específicas.....	62
2.6 Variables	62
CAPÍTULO III: MARCO METODOLÓGICO	65
3.1 Enfoque, Tipo, nivel, diseño y método de investigación.....	65
3.2 Población y Muestra	66
3.2.1 Población de estudio.....	66
3.2.2 Diseño muestral	66
3.3 Técnicas e instrumentos de recolección de datos	66
3.3.1 Procedimientos para la recolección de datos.....	67
3.4 Descripción de procedimiento de análisis de datos	68
3.5.1 Formulación del problema.....	68
3.5.2 Preparación de los datos	68
3.5.3 Feature engineering	69
3.5.4 Entrenamiento del Modelo	69

3.5.4 Evaluación del modelo	69
CAPITULO IV: RESULTADOS Y ANÁLISIS DE.....	70
4.1 Resultados	70
4.1.1 Análisis descriptivo	70
4.1.2 Tratamiento y construcción de las variables.....	79
4.1.3 Entrenamiento del modelo.....	79
4.2 Análisis de resultados o discusión de resultados	87
CONCLUSIONES Y RECOMENDACIONES	90
5.1 Conclusiones	90
5.2 Recomendaciones	92
REFERENCIAS BIBLIOGRÁFICAS	93
ANEXOS	100
Anexo 1: Declaración de Autenticidad	100
Anexo 2: Autorización del consentimiento para realizar la investigación ...	101
Anexo3: Matriz de consistencia de la investigación.....	102
Anexo4: Cuadro de base de datos.....	104
Anexo5: Tablas y gráficos que complementan la investigación.....	105

Índice de tablas

Tabla 1 Desviación de la demanda desde el año 2015 a 2019.....	3
Tabla 2 Métodos de pronóstico cualitativos	28
Tabla 3 Métodos de pronóstico cuantitativos – series de tiempo	29
Tabla 4 Métodos de pronóstico cuantitativos – causales.....	30
Tabla 5 Métodos de pronóstico cualitativos	36
Tabla 6 Matriz de operacionalización de variables	64
Tabla 7 Estadísticas descriptivas de las variables cuantitativas a nivel de día de venta	71
Tabla 8 Resultados del MAPE por centro de distribución	85
Tabla 9 Resultados del MAPE por centro de distribución	85
Tabla 10 Resumen de resultados del MAPE para los 3 experimentos por mes	86
Tabla 11 Resumen de resultados del MAPE para los 3 experimentos	87

Índice de figuras

Figura 1 Diagrama de Ichikawa para la identificación de quiebres de stock de la empresa de consumo masivo	2
Figura 2 Diagrama de Ichikawa para la identificación de quiebres de stock de la empresa de	32
Figura 3 Estructura de la Neurona Humana o Biológica.....	34
Figura 4 Modelo no lineal de una neurona.....	37
Figura 5 Función de activación escalón.	38
Figura 6 Función lineal y función mixta.	38
Figura 7 Función tangente hiperbólica.	39
Figura 8 Función sigmoideal.....	39
Figura 9 Función de activación de Gauss.....	40
Figura 10 Diagrama del esquema básico de una Red Neuronal Artificial (RNA).	41
Figura 11 Red feedforward con 1 capa de neuronas, 4 neuronas tanto en la capa de entrada como en la capa de salida.....	42
Figura 12 Red feedforward de múltiples capas totalmente acopladas, teniendo una capa de salida y una capa oculta.	43
Figura 13 Red recurrente sin bucles de auto-retroalimentación y sin ninguna neurona oculta.....	44
Figura 14 Red recurrente con bucles de auto-retroalimentación y con una capa de neuronas ocultas	45
Figura 15 Simple red recurrente.	46
Figura 16 Problema gradiente descendiente.....	48
Figura 17 Arquitectura de red LSTM.....	48

Figura 18 Ingreso de datos	49
Figura 19 Almacenamiento de datos	50
Figura 20 Actualización de memorias	50
Figura 21 Salida a siguiente neurona.....	51
Figura 22 Red recurrente con unidad completamente cerrada	52
Figura 23 Red recurrente con unidad cerrada mínima	53
Figura 24 Diagrama de cajas de la serie Airline.....	55
Figura 25 Diagrama de cajas de la serie Airline.....	56
Figura 26 Diagrama de los fundamentos teóricos que sustentan el estudio	58
Figura 27 Diagrama de proceso de recolección de datos	67
Figura 28 Diagrama de proceso de recolección de datos	68
Figura 29 Distribución de la demanda por centros de distribución.....	73
Figura 30 Distribución de ventas por departamentos del Perú.....	73
Figura 31 Distribución de ventas por categorías	74
Figura 32 Evolutivo de las ventas en toneladas de la empresa de consumo masivo	74
Figura 33 Distribución de las ventas en toneladas por año	75
Figura 34 Descomposición aditiva de la serie volumen de venta	76
Figura 35 Gráficos de dispersión de la venta versus las covariables	77
Figura 36 Mapa de calor de las correlaciones de las variables cuantitativas	78
Figura 37 Parámetros usados en arquitectura GRU	81
Figura 38 Funciones de perdida para entrenamiento y validación arquitectura GRU general.....	81
Figura 39 Volumen real vs estimado para arquitectura GRU general.....	82
Figura 40 Parámetros usados en arquitectura LSTM	83

Figura 41 Funciones de perdida para entrenamiento y validación arquitectura LSTM	83
Figura 42 Volumen real vs estimado para arquitectura LSTM general	84
Figura 43 Parámetros para el modelo general arquitectura LSTM	105
Figura 44 Parámetros para el modelo general arquitectura GRU.....	105
Figura 45 Parámetros para el modelo por centro de distribución arquitectura LSTM	106
Figura 46 Parámetros para el modelo por centro de distribución arquitectura GRU	106
Figura 47 Parámetros para el modelo por centro de distribución arquitectura LSTM	107
Figura 48 Parámetros para el modelo por centro de distribución arquitectura GRU	107
Figura 49 Volumen real vs estimado para arquitectura LSTM por centro.....	108
Figura 50 Volumen real vs estimado para arquitectura LSTM por categoría .	108
Figura 51 Volumen real vs estimado para arquitectura GRU por centro.....	108
Figura 52 Volumen real vs estimado para arquitectura GRU por categoría ...	109
Figura 53 Funciones de perdida para entrenamiento y validación arquitectura LSTM por centro	109
Figura 54 Funciones de perdida para entrenamiento y validación arquitectura LSTM por categoría.....	109
Figura 55 Funciones de perdida para entrenamiento y validación arquitectura GRU por centro	110
Figura 56 Funciones de perdida para entrenamiento y validación arquitectura GRU por categoría.....	110

RESUMEN

Uno de los principales problemas y retos que enfrentan las empresas de consumo masivo es la falta y excedente de stock en los almacenes, centros de distribución y finalmente en los puntos de venta; esto se dan por varias causas relacionadas a una planificación de demanda no adecuada.

El objetivo de la investigación es desarrollar un modelo de redes neuronales artificiales recurrentes, para que el pronóstico de demanda de stock sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo.

El método para usar será una técnica de redes neuronales artificiales recurrentes, en específico los métodos GRU y LSTM que ayudan a predecir mejor los casos con datos históricos. Existen investigaciones que han encontrado evidencia, donde el uso de estas técnicas tiene un mejor rendimiento en predicción.

Se realizaron 3 experimentos para esta investigación, el primero planteando de manera general, el segundo desarrollado para centros y el ultimo para categorías, en todos los casos nos muestra que el modelo LSTM tiene un MAPE menor que el modelo GRU, por lo cual podemos concluir que la arquitectura LSTM se ajusta mejor a la predicción de la demanda en toneladas para la empresa de consumo masivo.

El MAPE resultante del modelo actualmente implementado en la empresa de consumo masivo es de 20%, por lo cual podemos concluir que los modelos aplicados en esta investigación tienen un mejor desempeño obteniendo un MAPE promedio global de LSTM (5.8%) y GRU (7.5%).

Palabras clave: Redes neuronales, LST, GRU, MAPE, consumo masivo, demanda

ABSTRACT

One of the main problems and challenges faced by mass consumption companies is the lack and surplus of stock in warehouses, distribution centers and finally at points of sale; This occurs for several reasons related to inadequate demand planning.

The objective of the research is to develop a model of recurrent artificial neural networks, so that the forecast of stock demand is more accurate than the model currently implemented in the mass consumption company.

The method to use will be a recurrent artificial neural network technique, specifically the GRU and LSTM methods that help better predict cases with historical data. There are investigations that have found evidence where the use of these techniques has better performance in prediction.

3 experiments were carried out for this research, the first posing in a general way, the second developed for centers and the last for categories, in all cases it shows us that the LSTM model has a lower MAPE than the GRU model, so we can conclude that the LSTM architecture best suits the prediction of demand in tons for the mass consumption company.

The MAPE resulting from the model currently implemented in the mass consumption company is 20%, so we can conclude that the models applied in this research have a better performance obtaining a global average MAPE of LSTM (5.8%) and GRU (7.5%).

Keywords: Neural networks, LST, GRU, MAPE, mass consumption, demand

INTRODUCCIÓN

Una estimación no adecuada de la demanda trae como consecuencia una sobre producción de productos y peor aún un desabastecimiento. Las dos consecuencias mencionadas, generan pérdidas significativas a la empresa; una sobre producción genera costos relacionados a almacenamiento y desecho de productos por su vencimiento y un desabastecimiento también llamado quiebre de stock (out-of-stock) genera pérdidas en ventas ya que no se cuenta con el producto para vender y abastecer al cliente (bodegas, puestos de mercado y mini mayoristas) adicionalmente genera una insatisfacción en él, que en el largo plazo se traduce en la fuga del cliente por una mala experiencia.

Para el caso de quiebres de stock de productos (out-of-stock), las eficiencias se generan a través de la mayor venta, ya que el cliente podrá tener el producto cuando lo solicite. Por su contraparte no tener el producto en stock, genera pérdidas para la empresa, ya que el cliente va a buscar otras opciones en la competencia y en el largo plazo genera la fuga del cliente. Adicionalmente esto conlleva a generar una mala reputación de la empresa. Por otro lado un sobre stock en los almacenes y centros de distribución, generan una ineficiencia en los procesos de almacenamiento, por lo cual se traduce en costos más altos en alquileres de metros cuadrados y complejidad al momento de ejecutar los procesos logísticos.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción del problema

Las empresas líderes de consumo masivo en el mundo presentan problemas de *Out of Stock* o falta de stock en sus principales puntos de venta. Un estudio de la consultora Capgemini en el 2018 nos dice que la principal razón de los quiebres de productos está dada por las promociones, el bajo pronóstico de demanda y las malas prácticas en el almacenamiento.

En el caso de las empresas latinoamericanas, un estudio realizado por Corsten & Gruen (2016) que se titula *Retail Out of Stock: A Worldwide Examination of Extend, Causes and Consumer Responses*, el promedio de *Out of Stock* para las empresas de consumo masivo en América Latina es de 10% a 12% de casos en productos por punto de venta por lo cual, al reducir los quiebres significa una ganancia en ventas significativa para la empresa.

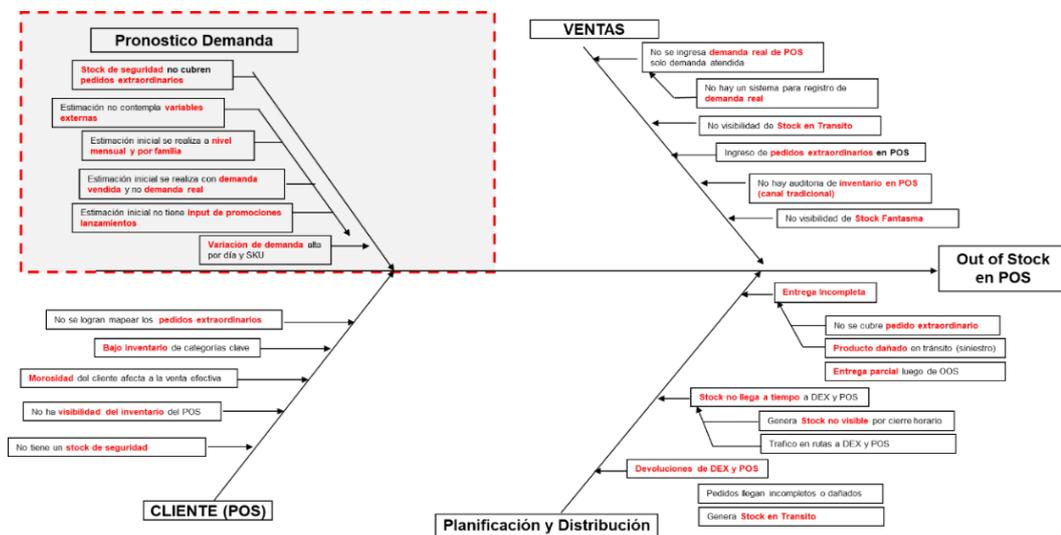
En el Perú las empresas de consumo masivo también son afectados por la falta de stock; un estudio en el año 2016 de la revista América Retail, menciona que en el Perú el 14.5% de los clientes que visitan un supermercado declaran encontrar quiebres de stock. Tal es el caso de la empresa líder de producción y comercialización de productos de alimentación de Perú. La empresa de consumo masivo tiene un portafolio de productos bastante variado, y una red de distribución participando en los mercados de consumo masivo Perú y B2B (mercado moderno y tradicional).

A continuación, mostramos las principales causas raíz del *Out of Stock* en la empresa. El diagrama de pescado muestra un diagnóstico integral de las causas raíz por cada área responsable del proceso. Y una de las causas principales de los quiebres de stock es una baja precisión en el pronóstico de demanda. La gerencia de demanda es la

responsable de generar los pronósticos de demanda de productos, su principal objetivo es velar por que la demanda pronosticada incluya todos los factores internos y externos para una buena estimación y tener una desviación mínima respecto a la demanda real, la cual es generada por los distintos clientes de la empresa.

Figura 1

Diagrama de Ichikawa para la identificación de quiebres de stock de la empresa de consumo masivo



Nota: Resultados de entrevistas en el 2020 a colaboradores. Elaboración: Propia

Una de las principales consecuencias de una estimación no adecuada de la demanda, es la sobre producción de productos y peor aún, un desabastecimiento de este. Las dos consecuencias mencionadas, generan pérdidas significativas a la empresa; una sobre producción genera costos relacionados a almacenamiento y desecho de productos por su vencimiento y un desabastecimiento también llamado quiebre de stock (*out-of-stock*) genera pérdidas en ventas ya que no se cuenta con el producto para vender y abastecer al cliente (bodegas, puestos de mercado y mini mayoristas) adicionalmente genera una insatisfacción en él, que en el largo plazo se traduce en la fuga del cliente por una mala experiencia.

La investigación aporta a la mejora de desviación del pronóstico de la demanda, en la empresa de consumo masivo. A continuación, mostramos el proceso que realiza la gerencia de demanda para calcular los pronósticos de demanda. La empresa utilizar un software ERP, para realizar su operación de ventas, logística, distribución y producción. Como parte de las soluciones analíticas que trae el software tiene un módulo destinado al cálculo de pronóstico de demanda. El proceso de pronóstico de demanda se realiza en base a la venta histórica a un nivel de cliente, categoría y mes, la frecuencia de este cálculo es mensual y todo el proceso se realiza dentro del módulo de pronóstico del ERP; otro aspecto importante es que las variables como promociones, lanzamientos de productos, quiebres en producción, campañas de marketing y el contexto externo, no están incluidos dentro del proceso de pronóstico de demanda por lo cual es un reto incluir dichas variables. La técnica usada por el software ERP para el cálculo de pronóstico de demanda es una serie de tiempo de tipo ARIMA, que corresponde a la familia de modelos autorregresivos, que básicamente funciona calculando medias móviles en base a rezagos de tiempo que podrían ser días, semanas, meses, etc. En promedio la desviación de la demanda para las categorías de consumo masivo está en 20% para el año 2019.

Tabla 1

Desviación de la demanda desde el año 2015 a 2019

Años	Desviación de la demanda
2015	18.5
2016	19.3
2017	21.2
2018	20.1
2019	20.4
Acumulado	20.2

Nota: Reportes anuales de pronóstico de demanda proporcionados por la gerencia de demanda con una adaptación propia.

1.2 Formulación del problema

1.2.1 Formulación general

En base a lo planteado en el diagnóstico del problema, se formuló la pregunta de investigación ¿Es posible desarrollar un modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo?

1.2.2. Formulación específica

- ¿Cuál es el algoritmo de redes neuronales artificiales recurrentes más eficiente para el pronóstico de demanda de stock en la empresa de consumo masivo?
- ¿Es posible definir la topología adecuada de las redes neuronales artificiales recurrentes para el pronóstico de demanda de stock y que sea más preciso que el modelo implementado en la empresa de consumo masivo?
- ¿Cuáles son las variables que tienen más influencia en el pronóstico de demanda de stock, usando los datos de la empresa de consumo masivo?

1.3 Importancia y justificación del estudio

1.3.1 Importancia

La presente tesis sería importante para las empresas de consumo masivo peruanas, específicamente para la gestión de la cadena de suministro en la Gerencia de Demanda, ya que al tener un mejor pronóstico de la demanda de stock de productos pueden tener una mejor planificación, distribución, almacenamiento y ventas de sus productos, generando eficiencia en su gestión, lo cual se traduce en ganancias para la empresa.

Corsten & Gruen (2016) en el estudio *Retail Out of Stock: A Worldwide Examination of Extend, Causes and Consumer Responses* realizado por *Grocery Manufacturers of América* en el 2016, el promedio de *Out of Stock* para las empresas de

consumo masivo en América es de 8% a 10% de casos en productos por punto de venta por lo cual, al reducir los quiebres hay una ganancia significativa para la empresa.

Según lo mencionado en el estudio anterior, la empresa de consumo masivo peruana podría llegar a tener en promedio un *Out of Stock* alrededor del 8% al 10%, y una de las principales soluciones que ayudaran a mejorar la ratio, es disminuir la desviación de pronóstico que actualmente tiene un valor de 20%.

1.3.2 Justificación

Teórico

El aprendizaje automático, en específico las técnicas de redes neuronales permiten identificar las variables internas y externas de la cadena de suministro que afectan directamente al pronóstico de la demanda de stock. El pronóstico de demanda se ha realizado desde el siglo XIX, pero de una forma limitada, con el avance computacional y científico de las redes neuronales artificiales ahora podemos amoldar su aplicación al pronóstico de la demanda de productos.

Práctica

El establecimiento y adopción de un modelo de redes neuronales artificiales podría predecir la demanda de stock de una forma más eficiente, lo cual lleva a un ahorro monetario para las empresas de consumo masivo logrando un menor quiebre de stock y un menor sobre almacenamiento de productos.

Metodológica

Mediante las etapas definidas de la implementación de las redes neuronales artificiales podemos lograr un modelo de pronóstico de demanda que considere variables

internas y externas de la cadena de suministro. Considerando el paso a paso metodológico que sigue la metodología de Crisp-DM.

1.4 Delimitación del estudio

Teórica

La presente tesis estará delimitada al sector de consumo masivo, con una aplicación en una empresa peruana y a nivel de técnica analítica está orientada al uso de redes neuronales artificiales en su aplicabilidad a predicción de demanda.

El procesamiento se realizará en el lenguaje *Python*, ya que hay varias librerías cargadas para desarrollar modelos de redes neuronales y series temporales por ejemplo TensorFlow y Keras. También usaremos SQL para extraer la información desde las fuentes origen.

Los algoritmos disponibles para usar en la investigación serían los siguientes:

- Redes neuronales *Long Short-Term Memory* (LSTM)
- Redes neuronales *Gated Recurrent Units* (GRU)

Existen más algoritmos que van por la rama de la estadística clásica, los cuales citaremos para conocimiento en el marco teórico.

Espacial

Respecto a la región geográfica se eligió al país de Perú dado que la ubicación de la casa matriz está ubicada en dicho país. Adicionalmente solo se realizará la predicción en las provincias del país donde exista un centro de distribución de la empresa.

Temporal

Usaremos datos históricos en el periodo enero 2016 a febrero 2020.

1.5 Objetivos de la investigación

1.5.1 Objetivo General:

Desarrollar un modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo.

1.5.2 Objetivos Específicos:

- Desarrollar el algoritmo de redes neuronales artificiales recurrentes más eficiente en predicción, para el pronóstico de la demanda de stock usando los datos de la empresa de consumo masivo.
- Definir la topología adecuada de las redes neuronales artificiales recurrentes para el pronóstico de demanda de stock usando los datos de la empresa de consumo masivo.
- Identificar las variables que tienen más influencia en el modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock, usando los datos de la empresa de consumo masivo.

CAPÍTULO II: MARCO TEÓRICO

2.1 Marco histórico

2.1.1 Redes neuronales artificiales recurrentes

El funcionamiento del cerebro y las redes neuronales ya eran un tema de debate y conversación a inicios del siglo XVIII, tal es el caso de Broca, P. (1824–1880), que demostró que existen áreas del cerebro que son responsables de funciones cognitivas. En ese tiempo, era de conocimiento para la comunidad científica que el cerebro se encontraba constituido por neuronas.

Posteriormente Olgi, C. (1843–1926), en el año de 1873, explicó un modo mediante el cual se podía observar a las células nerviosas de forma individual.

Adicionalmente, Ramón, S. & Cajal, D (1852–1934), volvieron a utilizar los conceptos ya antes mencionados, para realizar estudios donde se centraba en la comunicación de las neuronas y como se organizan en el cerebro.

Pitts, W. y McCulloch, W (1943), en su investigación afirman que el funcionamiento del cerebro fue tratado de explicar, mediante una red neuronal simple conectada y que inicialmente operaba conexiones lógicas básicas y en donde, cada una tiene el estado de "desactivada" o "activada", con un cambio a "activada" que sucede basándose en la estimulación de las neuronas que encuentran a su alrededor.

Donald, H. (1949) y su concepto denominado “*The Organization of Behavior*” (regla de actualización simple), la fortaleza de la conexión entre neuronas puede modificarse. Su gobierno, ahora llamado aprendizaje Hebbiano, sigue siendo un modelo influyente hasta hoy.

Minsky, M. y Edmonds, D. (1950), ensamblaron la “*Stochastic Neural Analog Reinforcement Calculator*” (SNARC), que es una red de 40 neuronas artificiales, conformada por 3000 tubos de vacío, hasta este momento ya se conocía el termino de Red Neuronal Artificial, en adelante se describirá mejor la terminología.

Rosenblatt, F. (1959), en la Universidad de Cornell construyó una simple red neuronal artificial que contenía 400 fotoceldas, las cuales se enlazaban al azar con otras 512 neuronas. El objetivo de esta investigación era enseñar acerca de las propiedades básicas en los sistemas de inteligencia, sin ser muy específicos en detalles relacionados seres vivos.

Bernie, W. (1960 a 1962), logró mejorar los algoritmos de aprendizaje de Hebb y llamó a sus redes *Adaline (Adaptive Linear Element)*.

Rosenblatt, F. (1962) logró mejorar el algoritmo de Bernie, utilizando perceptrones, su investigación también resalta que el algoritmo de aprendizaje puede adecuar las fortalezas de conexión de un determinado perceptrón a una que coincida con cualquier dato de entrada, siempre y cuando esta coincidencia exista.

Winograd y Cowan (1963) desarrollaron una forma de tratar grandes conjuntos de elementos, los cuales simbolizaban colectivamente bajo una concepción en particular, destaca en su investigación el uso de la robustez y al paralelismo.

Werbos, P. (1974), es pionero en su investigación con los algoritmos de tipo *back-propagation*, para redes de múltiples capas; su investigación fue la responsable del resurgimiento de las redes neuronales artificiales a mediados de los 70, y sus desarrollos cobran mayor fama en los años 80.

Kohonen, T. (1980) propuso un algoritmo compuesto por dos variantes (Universidad de Helsinki), cada una construye una representación gráfica basándose en las cualidades semejantes de los datos de ingreso proporcionados, la diferencia de las 2 variantes está en las dimensiones del gráfico. Estas redes el aprendizaje es de tipo *offline*, es decir establecen una primera etapa denominada de aprendizaje, el cual se establecen los pesos de las conexiones y la segunda etapa denominada de funcionamiento.

Hopfield, J. (1980), destaca en su investigación a la red recurrente, la cual está conformada por un conjunto de neuronas, en donde la capa de salida de cada una de ellas es aprovechada como retroalimentación, excepto para su misma entrada, lo cual le otorga estabilidad. La idea de Hopfield es similar a la propuesta por Jordan.

Rumelhart y McClelland (1986), en su investigación publicada “*Parallel Distributed Processing*”, dicho trabajo aplicaba el tipo de arquitectura de red neuronal *back-propagation*, lo que permitió ampliar las investigaciones de dicho algoritmo.

LeCun et al. (1988), desarrollaron *LeNet*, que es una red neural convolucional (CNN, por las siglas en inglés de *Convolutional Neural Networks*), la cual tiene su base en el "*neocognitron*" formulado en 1980 por Fukushima.

Ya con la llegada del siglo XXI surge el apogeo de las redes neuronales artificiales recurrentes, a consecuencia de una capacidad computacional y de cálculo nunca vista, tal es el caso de Cho, K. (2014), desarrolla las redes neuronales *Gated Recurrent Units* (GRU). Las redes GRU son como las memorias a corto plazo (LSTM) pero con una puerta de olvido y tienen menos parámetros que LSTM ya que carecen de una puerta de salida.

Hochreiter, S. y Schmidhuber, J. (2016) proponen las redes neuronales de tipo LSTM (*Long Short Term Memory*) que ya tienen una historia desde 1995, propuestas por y Jürgen, múltiples trabajos han desarrollado variaciones de las redes LSTM, por ejemplo, el 2016 GOOGLE usó una red LSTM, para sugerir mensajes en la app ALLO, en el 2017 FACEBOOK realizó unos 4500 millones de traducciones automáticas todos los días usando redes LSTM.

2.1.2 Pronóstico de demanda de stocks

En el siglo XVII, ya se hablaba de una forma cuantitativa del pronóstico de demanda, tal es el caso de Whitney, E. (1797) quien usó líneas de montaje para producir

mosquetes en masa para el gobierno de los Estados Unidos. Todas las partes del mosquete fueron producidas con las mismas tolerancias de ingeniería. Para que cada uno pueda insertarse en cualquier mosquete. Las previsiones sobre él se necesitaba una cantidad de ventas futuras de los artículos para determinar cuántos artículos producir y también de sus repuestos, por lo tanto, también se necesitaban pronósticos para cada pieza de reemplazo.

Posteriormente, en la década de 1800, se produjeron algunas calculadoras mecánicas que podían sumar y restar. La cara de las máquinas incluía filas y columnas de botones para presionar inserte los números. Estas máquinas fueron un poco útiles para la persona responsables de pronosticar las demandas futuras de cada artículo. Las calculadoras mecánicas se hicieron más comunes a principios de la década de 1900, y posteriormente pudieron sumar, restar, múltiples y dividir números.

Ya en el siglo XIX Henry H. (1944), ordenó que se desarrollará investigación para las Fuerzas Aéreas del Ejército de los Estados Unidos, acerca de las tecnologías que pudiese llegar a desarrollar el ejército en el futuro. Realizaron varios intentos, pero pronto se dieron cuenta que las técnicas cuantitativas no tendrían éxito, para superar estas deficiencias, se organizó el proyecto RAND que consistía en consultar a varios expertos su opinión y dar retroalimentación hasta llegar a un consenso, cuya metodología se sigue usando hasta la fecha, junto con modificaciones y reformulaciones se logró llamar el Método Delphi.

En los años 50, Brown et al.(1950) fueron los pioneros en los modelos de pronóstico de demanda que todavía son de uso común. Los tres utilizaron un método llamado suavizado exponencial que solo requiere demanda del período de tiempo más

actual. En la década de 1950, no era común almacenar y guardar múltiples períodos de tiempo de demanda.

El procesamiento electrónico creció en la década de 1950 cuando se introdujeron las cintas magnéticas para almacenar los números. Las cintas y las tarjetas se utilizaron juntas para procesar el número para todo tipo de empresas comerciales. El uso fue para nómina, facturación, inventario, cuentas por cobrar, etc. Pero como antes, la matemática primaria, las funciones de los sistemas eran básicas, tales como sumar, restar, multiplicar y dividir. Las personas responsables de desarrollar métodos de previsión de demanda estaban limitadas a las capacidades de las máquinas disponibles para ellos en esa época.

Robert, G. (1956), fue el primero en aplicar el método de pronóstico llamado suavizado exponencial. El método requería la demanda del tiempo más reciente. período (mes, semana) y un coeficiente del período de tiempo anterior. El método es se llama suavizado exponencial simple y requiere un coeficiente (llamado nivel) del período de tiempo anterior. El método de suavizado exponencial único generados pronósticos de patrones de demanda horizontal. El método combina la demanda actual. entrada con el coeficiente pasado y es un paso matemático fácil de aplicar, ya sea por procesamiento de datos o por una calculadora de mano.

Holt, C. (1957), agregó un componente de tendencia al método d suavizado exponencial y esto permitió pronósticos que incluían tendencias hacia arriba o hacia abajo. El modelo requiere dos coeficientes (nivel y pendiente). Para aplicar el método, la demanda necesita el período de tiempo más actual, junto con dos coeficientes del pronóstico modelo.

Dalkey (1969), manifiesta que las técnicas cuantitativas y cualitativas de demanda siempre han estado presentes a lo largo de la historia, tal es el caso del método Delphi que es el método cualitativo de predicción de demanda más famoso.

Winters, P. y Holt, C. (1967), agregaron componentes estacionales a la exponencial suavizando los pronósticos. El modelo de pronóstico requiere tres coeficientes, (nivel, pendiente, y estacional). Se introdujeron dos modelos, el modelo multiplicativo y el modelo aditivo. Como en los modelos de suavizado exponencial anteriores, solo la demanda del período de tiempo más actual y tres coeficientes del período de tiempo anterior son necesarios para generar el pronóstico.

Barnett (1988), menciona que el pronóstico de demanda cuantitativa es usado desde 1974, un caso de aplicación se dio en las empresas de servicios eléctricos de EE. UU. Hicieron planes para duplicar la capacidad de generación de electricidad en función de las previsiones de un crecimiento anual de la demanda del 7 %. Los pronósticos servirían para construir más plantas generadoras de cinco a diez años antes de que entren en línea. Pero la demanda solo creció a una tasa de solo 2 %. Los cual llevo perdidas de millones de dólares.

Barnett (1988), en la misma línea menciona que en los años 80, las empresas empezaron a ser más exhaustivos con las técnicas de pronóstico de demanda, pero para productos nuevos que tenían poca historia de ventas, el *forecast accuracy* obtenido aún era bajo. Un ejemplo de aplicación es el mercado de computadoras, un modelo pronóstico un potencial de 27 millones de unidades para 1988; otro predijo 28 millones de unidades para 1987. De hecho, solo se habían enviado 15 millones de unidades en 1986. Por lo cual hay evidencia que con un buen pronóstico de demande se puede ser más eficiente en la producción de productos nuevos.

Para dar solución a tal complejidad se hace uso intensivo de aprendizaje automático junto con los modelos tradicionales de pronóstico de series temporales para generar pronósticos semanales y diarios mejorados mediante el uso de datos de señal de demanda histórica y datos de punto de venta.

Berry, C. y Levinsohn, A. (1995), desarrollaron funciones de demanda de bienes heterogéneos. Este enfoque y sus generalizaciones se han aplicado desde entonces a numerosos tipos de bienes para comprender la estructura de mercados subyacentes. Algunas aplicaciones incluyen Nevo (2001), que utilizó métodos BLP para medir el poder del mercado en cereales, y Villas-Boas (2009), quienes investigaron los efectos de una política que prohíbe la discriminación de precios.

Ramesh, V. (2018) en los últimos años, el crecimiento de las señales de información que dejan los clientes o usuarios a las organizaciones ha crecido exponencialmente desde el inicio del año 2000, llamada la era de la información, adicionalmente la inteligencia artificial y el aprendizaje automático están preparados para revolucionar el pronóstico de la demanda basadas en AI, ofreciendo capacidades innovadoras para extraer conocimiento de conjuntos de datos masivos reunidos de cualquier cantidad de fuentes internas y externas. La aplicación de algoritmos de aprendizaje automático descubre ideas e identifica tendencias perdidas por los pronósticos tradicionales configurados por humanos. Utilizando grandes datos y una potencia de procesamiento masiva en la nube, AI puede probar y refinar simultáneamente cientos de modelos avanzados, mucho más allá de lo que es posible con el pronóstico de demanda tradicional.

Finalmente Prabhudesai et al. (2020) comentan el concepto de Demand Sensing, el cual ha tomado una gran relevancia debido a la necesidad de las empresas de bienes de

consumo envasados (GPC) que representan algunas de las industrias más grandes del mundo. La gestión de la cadena de suministro en las GPC es compleja porque se suministran varios productos a través de múltiples canales y distribución métodos. Los productos siguen patrones de orden complejos caracterizados por eventos promocionales, influencias estacionales, desastres naturales, etc. Dada esta complejidad, es crucial generar pronósticos precisos a corto plazo de cantidades de pedidos que reflejen la demanda realista para productos Dichas previsiones permiten a las empresas impulsar una respuesta eficiente de la cadena de suministro a Mejorar el servicio al cliente.

2.2 Investigaciones relacionadas con el tema

2.2.1 Investigaciones nacionales relacionadas con el tema

- Guzmán Valle (2018) en su tesis de maestría titulada “Aplicación de métodos determinísticos y dinámicos para calcular los pronósticos de la demanda en la cadena de suministro de la Empresa Chiclayana Sipán Distribuciones S.A.C. utilizando Minitab y Oracle Crystal Ball” la cual tuvo como objetivo principal: “proponer diversos métodos de pronóstico de ventas tanto determinístico como dinámicos, basado en el análisis de series de datos históricos de los productos que forman parte de la demanda en la cadena de suministro de la empresa” y usó como Metodología: Se usaron métodos de pronóstico de demanda determinísticos para productos que representan el 80% de la venta de la empresa Sipan Distribuciones SAC.

La investigación finaliza con los siguientes resultados:

Al utilizar los diferentes métodos de análisis, tomando como referencia el predictor de la herramienta *crystall ball* dentro de su modelo de riesgos, se aprecia quede los 08 SKU seleccionados como SKU importantes según modelo de inventarios ABC, 07 de ellos puede ser predecibles pues a pesar de que todos

tienen un patrón de comportamiento identificable y con un error o MAPE, menor al 30%, en uno de ellos el SKU CD Kolynos Super White 100cc no se pudo comprobar el cumplimiento de los supuestos de aleatoriedad y normalidad de los residuos en los 03 métodos seleccionados para su análisis.

- Sandivar Rosas (2018) en su tesis de maestría titulada “Pronóstico de la demanda de postulantes de un centro superior universitario, basado en redes neuronales” que tiene como objetivo: “lograr es estimar la demanda de postulantes para un centro superior universitario basado en redes neuronales” y a seguido la siguiente metodología: En la investigación tomó datos históricos de los últimos 22 años relacionados a la cantidad de postulantes, número de ingresantes e ingresantes matriculados por proceso de admisión de cada carrera. La investigación donde se aplicará es la Facultad de Química e Ingeniería Química de la Universidad Nacional Mayor de San Marcos. Para el desarrollo de la investigación se van a utilizar técnicas de inferencia estadística y redes neuronales y se propone contrastar la efectividad de resultados que brinden cada una de estas, luego seleccionar quella que se aproxime a los resultados reales. Con los resultados obtenidos se podrán tomar posibles decisiones que beneficien a los futuros profesionales de estas carreras haciéndolas más atractivas y que respondan con profesionalismo al mercado laboral competitivo.

La investigación finaliza con los siguientes resultados: Se aprecia una disminución de los postulantes a la carrera de Química, esto nos muestra preliminarmente los resultados del modelo. Los modelos aplicados en esta investigación como las redes neuronales son más precisos que los modelos clásicos, por lo cual tiene un menor error respecto a la realidad. Algunas de las conclusiones de la investigación, es realizar actividades en los colegios con los

alumnos de los últimos grados para explicar los beneficios y alcances, generar ferias con los egresados destacados para los alumnos más jóvenes de la carrera para evitar la deserción.

- Chau de la Cruz & Montero (2018) en su tesis de pregrado titulada “Desarrollo de algoritmos de pronóstico de demanda y planificación de compras para reducir excedentes y faltantes en la gestión de Stocks de un retail farmacéutico” la cual tuvo como objetivo principal: “proponer diversos métodos de pronóstico de ventas tanto determinístico como dinámicos, basado en el análisis de series de datos históricos de los productos que forman parte de la demanda en la cadena de suministro de la empresa” y usó como Metodología: Se usaron métodos de pronóstico de demanda determinísticos para productos que representan el 80%.

La investigación finaliza con los siguientes resultados:

Se puede apreciar que, al aplicar los diferentes métodos de análisis de los datos, y usar las herramientas de *crystall ball* para los modelos de riesgos del mismo, uno de los primeros resultados es que los 8 SKU seleccionados como SKU importantes, 7 de ellos puede ser predecibles a pesar de que todos tienen un patrón de comportamiento identificable y un MAPE menor al 30%, por ejemplo el SKU CD Kolynos Super White 100cc no cumplió todos los supuestos estadísticos de aleatoriedad y normalidad.

2.2.2 Investigaciones extranjeras relacionadas con el tema

- Siwerz y Dahlén (2017) en su investigación titulada: “*Predicting sales in a food store department using machine learning*” tuvo como objetivo: Demostrar que los modelos de aprendizaje automático como los algoritmos de MLP(algoritmo de perceptrón multicapa, es un algoritmo de redes neuronales), RBFN (red de funciones de base radial es un algoritmo de redes de neuronales) y SVM

(máquinas de soporte vectorial, es un algoritmo de *machine learning*) logran una mejor precisión cuando predicen el número de productos vendidos en un departamento de tienda de alimentos.

Metodología: La investigación aplicó los métodos para series de tiempo basados en SVM, MLP y RBFN. Estos métodos generalmente se comparan con los tradicionales métodos estadísticos, los cuales han sido los populares por varios años, adicionalmente se comparan con los trabajos desarrollados por (Di Pillo, 2013).

Muestra: Los datos que usaron son proporcionados por un por una empresa de alimentos sueca y consiste en las ventas diarias de un departamento en una tienda desde el año 2012 hasta el año 2016. El instrumento para obtener los datos, fueron las bases de datos de la empresa sueca.

La investigación finaliza con los siguientes resultados:

- Se eligió usar MAPE y RMSE como medidas de error. A medida que el SVM obtuvo un promedio más bajo MAPE y RMSE que MLP y RBFN. Una razón para dar este resultado podría ser que este estudio utilizó datos de un departamento de tienda de alimentos en lugar de un solo producto.
- Además, los modelos en este estudio fueron entrenados en un conjunto de datos más grande que también podría ser un factor que contribuye a las mediciones de error generalmente más bajas. El desempeño de la SVM no fue inesperado ya que similar a (Levis & Papageorgiou, 2005) utilizó un SVM para predecir las ventas mensuales de un solo producto y mostró un MAPE similar en comparación con este estudio.

- Nuestro resultado está en línea con (Krause-Traudes, 2008) quien concluyó que los SVM pueden ser usado para predecir ventas agregadas en alimentos en una tienda.
- No esperábamos el bajo rendimiento del MLP, ya que múltiples estudios han demostrado su uso en la predicción de ventas. El MLP muestra desviaciones estándar sorprendentemente altas en ambos MAPE y RMSE en comparación con los otros modelos. Sin embargo, nuestro resultado podría ser una consecuencia de la configuración de parámetros utilizada ya que hay múltiples formas de elegir los pesos iniciales y la cantidad de neuronas en la capa oculta.
- Moskalev (2019) en su investigación titulada: “*Demand forecasting for fast moving products in grocery retail Advanced time series regression modelling*” tuvo como objetivo: Desarrollar el mejor modelo de previsión de la demanda es para la industria de comercio minorista de comestibles.

Metodología: Se aplicó las técnicas de regresión armónica dinámica, Poisson GLM con red elástica, MLP y LSTM de dos capas. Las empresas minoristas de comestibles tratan de reducir el deterioro de sus productos, así como maximizar sus beneficios. Productos de rápido movimiento, o productos con muchas ventas y la rotación rápida es particularmente importante para pronosticar con precisión debido a su alta volúmenes de ventas.

Muestra: Los datos que usaron son proporcionados por un por una empresa de venta minorista de productos comestibles, los datos usados corresponden a los años 2014 a 2018. El instrumento para obtener los datos, fueron las bases de datos de la empresa.

La investigación finaliza con los siguientes resultados:

- Los modelos que usan regresión armónica dinámica no son muy inferiores en desempeño a los modelos usando redes neuronales como MLP y LSTM con múltiples capas.
- Los modelos basados en redes neuronales MLP y LSTM son muy complejos que requieren un ajuste cuidadoso de los hiper parámetros para funcionar correctamente. Estos modelos no son buenos para responder hipótesis planteadas.
- Los resultados con respecto a redes neuronales y regresión armónica dinámica están en línea con los resultados presentados por (Arunraj & Ahrens, 2015, págs. 321-335), que utilizaron SARI- MAX y SARIMA- QR pronostican una única demanda de productos de rápido movimiento. Los autores informaron que las series temporales lineales los modelos con regresores externos muestran mejores resultados en MAPE, mientras que las redes neuronales son mejor en RMSE.
- Pavlyshenko (2018) en su investigación titulada “*Machine Learning Models for Sales Time Series Forecasting*” tuvo como objetivo: “Considerar los principales enfoques y estudios de caso sobre el uso del aprendizaje automático para previsión de ventas. Se ha considerado el efecto de la generalización del aprendizaje automático. Este efecto se puede usar para hacer predicciones de ventas cuando hay una pequeña cantidad de datos históricos para series temporales de ventas en el caso de que se lance un nuevo producto o tienda.”

Metodología: Se aplico las técnicas de redes neuronales, regresión lasso, random forest, Arima multivariado y luego un *stacking models* para ensamblar los modelos desarrollados.

Muestra: Los datos que usaron corresponden a las ventas por día y producto de la empresa grupo Bimbo en el periodo de 2014 a 2016. El instrumento para obtener la base de datos fue la web de competencia *Kaggle Competition Grupo Bimbo Inventory Demand*.

La investigación finaliza con los siguientes resultados:

- Los modelos enfocados en regresión tuvieron mejores resultados que los modelos de series de tiempo, por lo cual los datos de ventas se pueden explicar mejor con modelos donde hay una dependencia lineal.
- El efecto de la generalización del aprendizaje automático consiste en capturar los patrones en todo el conjunto de datos. Este efecto se puede usar para hacer predicción de ventas cuando hay una pequeña cantidad de datos históricos para series de tiempo de ventas específicas en el caso cuando se lanza un nuevo producto o tienda.
- El uso del *Stacking* permite tener en cuenta la diferencias en los resultados para múltiples modelos con diferentes conjuntos de parámetros y mejorar la precisión en la muestra de validación.
- Fildes et al.(2018) en su investigación titulada “*Retail forecasting: research and practice*” tuvo como objetivo: Presentar los problemas de pronóstico que enfrentan los grandes minoristas, desde lo estratégico hasta lo más importante, operacional, desde la tienda hasta los canales de distribución de la competencia a medida que las ventas se agregan productos a marcas a categorías y a la empresa en general.

Metodología: Usaron técnicas univariadas de predicción tales como ARIMA y TBATS. También aplicaron modelos multivariantes de predicción tales como SVM y RNN, los indicadores usados para evaluar fueron MAPE y RMSE.

Muestra: Los datos que usaron corresponde a ventas por mes y producto de las empresas Ada, Sainsbury, Tesco, Waitrose en el periodo de 1995 a 2015. El instrumento para obtener los datos, fueron las bases de datos de las empresas.

La investigación finaliza con los siguientes resultados:

- Los pronósticos agregados admiten las decisiones estratégicas y se discuten en tres niveles: las ventas minoristas agregadas en un mercado, en una cadena y en una tienda.
- Los pronósticos de nivel de producto generalmente se relacionan con decisiones operativas donde la jerarquía de datos de ventas a lo largo del tiempo, se examina el producto y la cadena de suministro.
- Muchos de los métodos complejos propuestos han proporcionado muy poca evidencia para convencer de su valor, lo que plantea más preguntas de investigación. En contraste, algunos métodos econométricos ambiciosos Se ha demostrado que superan todas las alternativas más simples, incluidas las utilizadas en la práctica.

(Kasun , Peibei , Christoph , & Hansika , 2018) en su investigación titulada *“Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology”* tuvo como objetivo proponer un nuevo modelo de pronóstico de demanda, usando variables adicionales del negocio mediante un algoritmo LSTM y no solo una variable, la cual es la forma tradicional de trabajar. Adicionalmente proponen un marco sistemático de procesamiento de datos para el negocio de Walmart.com, adicionalmente el indicador usado para medir la eficiencia del modelo es el MAPE. Metodología: Usaron técnicas de redes neuronales de tipo LSTM para realizar la predicción de ventas. Muestra: Se evaluó el dataset de Walmart.com, primero un conjunto de 1724 SKUS que pertenecen a

la categoría de productos domésticos, que consta de 15 subcategorías diferentes. Y otro data set que corresponde a 18, 254 skus, que corresponden a un super departamento que consta 16 categorías diferentes. El instrumento para obtener los datos, fueron las bases de datos de las empresas.

La investigación finaliza con los siguientes resultados:

- Desarrollaron un enfoque de redes LSTM para pronosticar la demanda de ventas, es un modelo global a través de los artículos disponibles en una jerarquía, el cual tiene mejores resultados que el modelo ARIMA usado actualmente por Walmart.com
- Cuando las ventas de los SKUS son escasas, una estrategia es la agrupación de productos con lo cual se obtiene menor MAPE en el pronóstico.
- El modelo propuesto es más eficiente que el modelo univariante ARIMA tanto a nivel de SKUS como por categoría, comparando el indicador MAPE resultante.

- Lanham et al.(2017) en su investigación titulada “*A Comparative Study of Machine Learning Frameworks for Demand Forecasting*”

tuvo como objetivo proponer varios modelos de pronóstico usando librerías de *Deep Learning*, específicamente keras y scklearn. Metodología: Al desarrollara los diferentes modelos se midió el rendimiento predictivo, tiempo de ejecución, escalabilidad y facilidad de uso. Los modelos desarrollados fueron *Gradient Boosting*, *Factorización Machines* y *Deep Neuronal Network (DNN)*; para medir la eficiencia de los modelos se usó el indicador NWMSLE.

Muestra: Los datos utilizados en esta investigación provienen de la competencia Kaggle, que tiene como objetivo pronosticar la demanda de millones

de artículos a nivel de tienda y día para una cadena de supermercados de América del Sur. Instrumento: Los datos se proporcionan en diferentes tablas llamadas entrenamiento y prueba.

La investigación finaliza con los siguientes resultados:

- La previsión de la demanda a nivel granular es un problema analítico complicado con múltiples series de tiempo que se propagan en conjunto y se ven afectadas por factores externos como el precio del petróleo, los días festivos, etc.
- El excedente de inventario y los "desabastecimientos" son problemas clave en la gestión de tiendas minoristas empresa que afecta a los márgenes finales.
- Una práctica herramienta digital impulsada por este modelo podría ayudar al administrador de categorías a decidir las existencias de inventario diario de millones de artículos en varias ubicaciones de tiendas.
- El modelo también asume la ausencia de eventos catastróficos (como terremotos), en presencia de los cuales la varianza se dispararía drásticamente aquí.
- Ashwini y Vijaykumar (2019) "*Retail Giant Sales Forecasting using Machine Learning*" tuvo como objetivo pronosticar las ventas de Global Mart, que es una empresa de venta minorista en línea, que toma los pedidos y realiza entrega a todas partes del mundo, la granularidad del pronóstico tiene que ser a nivel de depósito, segmento y categoría. Un objetivo adicional es encontrar los dos segmentos más rentables en oferta y demanda. Metodología: Para el cálculo del pronóstico de ventas y demanda usaron inicialmente un modelo autorregresivo (ARIMA), iniciando con una descomposición espectral de la serie, tales como tendencia,

estacionalidad, ciclicidad y ruido, para luego realizar un suavizado, al final se aplicó un modelo ARMA y el indicador para evaluar el rendimiento del modelo fue el MAPE. Muestra: Los datos corresponden a la empresa Global Mart, una empresa de venta minorista a nivel mundial, las variables de la base de datos son, ordenes, ventas, mercado, cantidad, segmento y fecha.

La investigación finaliza con los siguientes resultados: Los segmentos de CAPAC y UE son los más rentables, así que el stock asignado a estos segmentos será más alto que los otros. Lo más relevante es que el modelo pudo explicar adecuadamente el comportamiento para el segmento de EU, ya que tienen el MAPE más bajo, manteniendo el inventario alrededor de 400 toneladas.

- Prabhakar y Sayiner (2018) en su investigación titulada “*Demand forecasting for a large grocery chain in Ecuador*” tuvo como objetivo: Proponer un modelo de pronóstico de demanda para Corporación Favorita que es una firma ecuatoriana que cuenta con más de 50 supermercados en todo el país. La mayoría de sus tiendas se concentran en Quito, la capital del país. Cadena de supermercados almacenan más de 4000 artículos y en promedio hay 1695 transacciones por día por tienda (casi 100,000 transacciones por día en todas las tiendas).

Metodología: Usaron varios modelos *de machine learning* y estadísticos clásicos para realizar el pronóstico de venta, tales como: Regresión lineal, *Gradient Boosting* y redes neuronales, calibrando los parámetros adecuados para cada modelo, adicionalmente para la validación se dividió en una muestra de entrenamiento y validación tomando como indicador de eficiencia de pronóstico el RMSE, MAE y Rsquared.

Muestra: Los datos minoristas acumulados para las ventas unitarias estaban disponibles a nivel de estado, ciudad, tienda y artículo para cada día entre

2013 y 2017. Los artículos se clasifican además por familia, clase y si es perecible. Cada tienda fue clasificada por su grupo, tipo y ubicación definidos por el cliente. Además, las vacaciones, también se incluyeron los datos relevantes por ubicación y los precios del petróleo durante el período de tiempo de observación.

La investigación finaliza con los siguientes resultados:

- Predecir las ventas, especialmente en las tiendas de comestibles, es de suma importancia para el propietario de una tienda, teniendo en cuenta la granularidad de la predicción.
- En la situación actual, debido al tamaño del conjunto de datos y las dificultades computacionales asociadas se eligió al modelo de *Gradient Boosting* como el modelo ganador. Pero el autor menciona además que, si le dedicara más tiempo a seguir modelando, es posible que el *modelo* de redes neuronales pueda superarlo.
- Además, este problema de predicción puede considerarse como una predicción de series de tiempo para utilizar las relaciones entre los distintos parámetros de series de tiempo para mejorar nuestra predicción. Podemos usar modelos como ARIMA, HoltWinters. nuestro modelo óptimo.

2.3 Estructura teórica y científica que sustenta el estudio

Teorías y modelos que sustentan el estudio

2.3.1 Demanda de Productos

Chase et al. (2009), afirma que según el padre del marketing, Philip Kotler, “la demanda es el deseo que se tiene de un determinado producto pero que está respaldado por una capacidad de pago. Se refiere a las cantidades de un producto que los

consumidores están dispuestos a comprar a los posibles precios del mercado. Aclarando, la demanda es la cantidad de bienes o servicios que el comprador o consumidor está dispuesto a adquirir a un precio dado y en un lugar establecido, con cuyo uso pueda satisfacer parcial o totalmente sus necesidades particulares o pueda tener acceso a su utilidad intrínseca”.

La administración de la demanda implica reconocer el mercado a manera de conocer la demanda real de los productos, adicionalmente existe un equilibrio en oferta y demanda con lo cual mediante análisis cuantitativos podemos encontrar ese punto el cual será el óptimo de la demanda.

2.3.2 Pronóstico de demanda de productos

Chase et al. (2009) afirman que “es un proceso que consiste en utilizar datos pasados para determinar acontecimientos futuros. Los pronósticos a menudo son utilizados para poder predecir la demanda del consumidor de productos o servicios, aunque se pueden predecir una amplia gama de sucesos futuros que pudieran de manera potencial influir en el éxito. Pronosticar es el arte y la ciencia de predecir los eventos futuros, puede involucrar el manejo de datos históricos para proyectarlos al futuro, mediante algún tipo de modelo matemático, puede ser una predicción del futuro subjetiva o intuitiva o bien una combinación de ambas, es decir, un modelo matemático ajustado por el buen juicio de un administrador”.

2.3.3 Métodos de pronóstico de demanda

Chase et al.(2009), afirman que los pronósticos se pueden clasificar en dos tipos básicos: cuantitativos y cualitativos.

Los pronósticos cualitativos o subjetivos

Involucran algunos factores importantes tales como el juicio, la intuición, emociones, experiencias personales del que toma la decisión, y sistemas de valores para alcanzar un pronóstico. Hace poco uso de la información histórica y casi siempre consideran un horizonte de mediano y largo plazo.

Tabla 2

Métodos de pronóstico cualitativos

Método	Descripción	Usos	Horizonte de tiempo
Investigación de mercados	Es un procedimiento donde se aplican técnicas para validar hipótesis previamente planteadas.	Pronóstico de las ventas totales de la compañía o validación de venta de productos nuevos.	Medio Largo
Agregados de la fuerza de ventas	Estos pronósticos se basan más en el juicio experto, se realiza preguntas a expertos en el área de ventas y que conocen el mercado para luego sacar un promedio de los resultados de cada opinión.	Buena fuente para identificar cambios en las preferencias de los clientes. El agregado de fuerza de ventas puede ser inexacto cuando la compensación al personal de ventas se basa en cubrir una cuota.	Corto Mediano

Método Delphi	Un panel de expertos que tienen acceso a un cuestionario previamente validado, este cuestionario trata de responder el problema a través de preguntas específicas.	Pronósticos de ventas a largo plazo para planeación de capacidad o instalaciones.	Medio Largo
---------------	--	---	----------------

Nota: Adaptado de Chase et al.(2009)

Los pronósticos cuantitativos

Nahmias y Olsen (2017) afirman que “este tipo de métodos como métodos objetivos y los define como aquellos en los que el pronóstico se deriva de un análisis de datos. Un método de series de tiempo es aquel que usa sólo valores pasados en cuanto al fenómeno que se desea predecir. Los modelos causales son aquellos que usan datos provenientes de fuentes distintas a las series que están pronosticando, es decir, pueden existir otras variables con valores que están vinculadas de alguna forma a lo que se está pronosticando. Al igual que con los métodos Cualitativos, a continuación, se presenta una breve descripción de algunos métodos Cuantitativos que existen para el cálculo de pronósticos”.

Tabla 3

Métodos de pronóstico cuantitativos – series de tiempo

Método	Descripción	Usos	Horizonte de tiempo
Promedio móvil	Cada observación es un valor a ingresar en el promedio, mientras van avanzando la movilidad de la observación el promedio va cambiando.	Útiles para pronosticar ventas e inventarios en los almacenes	Corto

Suavización exponencial	Esta técnica es similar al promedio móvil, excepto que los puntos que son más recientes reciben mayor ponderación. El nuevo pronóstico será igual al anterior más cierta parte del error de pronóstico pasados. La nivelación exponencial doble o triple son versiones complejas del modelo básico que explican la variación de tendencia de la serie.	Aplica a los mismos casos que el promedio móvil	Corto
Técnicas Box Jenkins	Complejo procedimiento iterativo basado en computadora que produce un modelo de promedio móviles integrado y auto regresivo, que se ajusta para los factores de tendencia y estacional, estima los parámetros apropiados de ponderación, valida el modelo y repite el ciclo según sea el apropiado.	Limitado debido al costo de los productos que requieren de pronósticos muy exactos a corto plazo	Corto Mediano

Nota: Adaptado de Chase et al.(2009)

Tabla 4

Métodos de pronóstico cuantitativos – causales

Método	Descripción	Usos	Horizonte de tiempo
Modelo Matemáticos	Un modelo lineal o no lineal ajustado con los datos de series de tiempo, normalmente mediante regresión. Incluye las líneas de tendencia, polinomios, logaritmos, lineales, series de Fourier, etc.	Aplica lo mismo que el promedio móvil, pero con limitaciones debido al costo y usos con pocos productos	Corto Mediano
Descomposición de series de tiempo	Método para descomponer una serie de tiempo en componentes estacionales, de tendencia y regularidad.	Es adecuado para identificar puntos críticos y es una excelente herramienta de pronóstico para el periodo de tiempo	Corto Mediano

mediano-largo, es decir de 3 a 12 meses.

Análisis de regresión	Relaciona la demanda con otras variables que "causan" o explican su nivel. Las variables se seleccionan sobre la base de significancia estadística. La disponibilidad de programas de regresión por computadora hace de esta técnica una de las populares.	Limitado debido al costo de los productos que requieren de pronósticos muy exactos a corto plazo	Corto Mediano
Modelos econométricos	Un modelo econométrico es un sistema de ecuaciones de regresión interdependientes que describe las ventas de cierto sector económico. Los parámetros de la ecuación.	Útil para pronósticos de ventas por clases de productos para planeación corto y mediano plazo	Corto Mediano

Nota: Adaptado de Chase et al.(2009)

2.3.4 Stock Productos

Chase, Jacobs, y Aquilano (2009) afirman que el “stock es un término anglosajón el cual indica la cantidad de productos o materias primas que posee un comercio en su almacén a la espera de su venta o comercialización. El stock es una inversión que debe de realizar el comercio para asegurarse de atender el flujo de demanda de manera adecuada para el buen funcionamiento del mismo, y así poder llevar a cabo sus actividades de manera normal sin necesidad de paralizar sus ventas por falta del producto. Se puede utilizar la palabra stock como sinónimo de inventario y existencias. Con referencia al inventario, es la relación ordenada y detallada del conjunto de bienes que posee el patrimonio de una empresa y por otro lado, las existencias son los productos o materias primas que posee la empresa para responder a la demanda”.

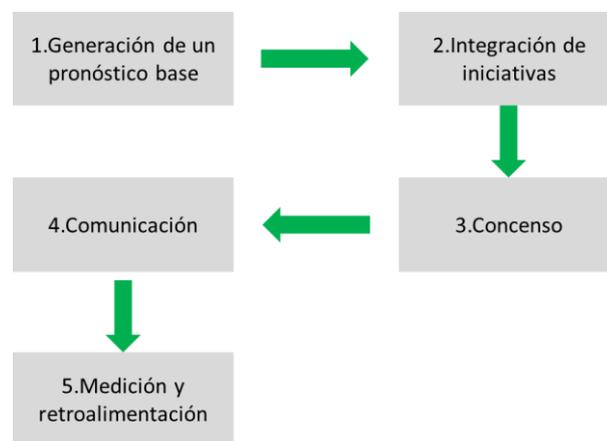
2.3.5 Planificación de la demanda

Chase, Jacobs, y Aquilano Chase et al.(2009) (2009), afirman que “una cadena de suministro es una red entre una empresa y sus proveedores para producir y distribuir un

producto específico al comprador final. Esta red incluye diferentes actividades, personas, entidades, información y recursos. La cadena de suministro también representa los pasos necesarios para llevar el producto o servicio de su estado original al cliente. Las empresas desarrollan cadenas de suministro para que puedan reducir sus costos y seguir siendo competitivas en el panorama empresarial”.

Figura 2

Diagrama de Ichikawa para la identificación de quiebres de stock de la empresa de



Nota: Adaptado de Chase et al.(2009)

2.3.6 Gestión de demanda

James, A. (2014) afirma que, “hace muy poco, un fabricante tenía que intentar predecir qué tendría que producir en base al pasado. Tendría que desarrollarse un pronóstico de ventas de lo que se vendió el año o el mes pasados”.

Ese pronóstico de ventas, a su vez, fue la base de un plan de producción para la fábrica. Y la producción de la fábrica fue la base de lo que consiguió enviado al cliente. Pero las ventas históricas no son el predictor más confiable del futuro. Con demasiada

frecuencia, lo inesperado ocurre en el mercado. UNA el producto se pone de moda por alguna razón inexplicable o un evento imprevisto aumenta las ventas.

Hasta la década de 1990, la predicción era un ejercicio en solitario por empresas por su cuenta y, a menudo, sin consultar el suministro y socios de la cadena. Entonces se les ocurrió a algunos ejecutivos de la cadena de suministro que dos cabezas podrían ser mejores que una para tratar de predecir el futuro. En 1995, el gigante minorista Wal-Mart Stores Inc. y la industria farmacéutica la compañía Warner-Lambert (ahora parte de Pfizer) hizo una prueba de colaboración previsión de reposición de inventario junto con la asistencia de la firma consultora Benchmarking Partners y dos softwares.

2.3.7 Redes Neuronales Artificiales

Matich (2001) afirma que “las redes neuronales artificiales están inspiradas en las redes neuronales biológicas, aunque poseen otras funcionalidades y estructuras de conexión desde la perspectiva biológica”. Las características principales son las siguientes:

- Autoorganización y adaptabilidad: Esta característica de los algoritmos corresponde al aprendizaje adaptativo, por lo cual ofrecen mejores posibilidades de procesamiento robusto.
- Procesado no Lineal: Aumenta su inmunidad frente al ruido de los datos, mejorar su proceso.
- Procesado Paralelo: Normalmente se usa un gran número de nodos de procesamiento, con alto nivel de interconectividad. El elemento básico de computación (modelo de neurona) se le llama habitualmente nodo o unidad. Recibe un input desde otras unidades o de una fuente externa de datos. Cada input tiene un peso asociado w ,

que se va modificando en el llamado proceso de aprendizaje. Cada unidad aplica una función dada f de la suma de los inputs ponderadas mediante los pesos.

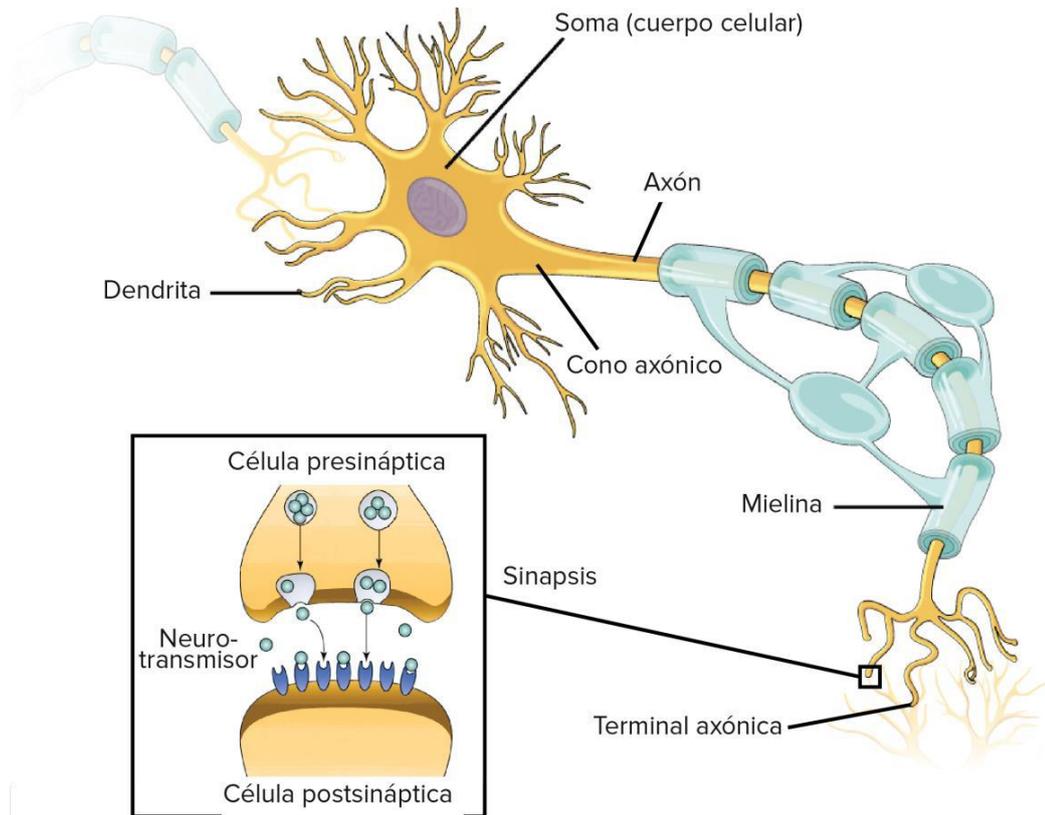
2.3.8 Neurona Biológica

Haykin (2009) Afirma que “poder comprender el cerebro se ha facilitado gracias al trabajo pionero de Ramón y Cajal (1911), al conceptualizar cada neurona como la unidad básica de la estructura del cerebro, una similitud sería un grano de arena en la viga de concreto de un edificio. Simbólicamente, la puerta lógica del chip de silicio es más rápida que la neurona en un orden comprendido entre 5 y 6 veces; los eventos lógicos se producen en nanosegundos, en tanto que los eventos neuronales suceden en milisegundos. No obstante, el cerebro equilibra la velocidad de operación relativamente lenta de una neurona con un número verdaderamente asombroso de neuronas (células nerviosas) y sus interconexiones masivas entre ellas, dando como resultado una estructura enormemente eficiente denominada cerebro”.

Ponce (2010) En la misma línea, menciona que “la célula llamada a transformar la información es la neurona biológica que acoge, por medio de sus dendritas, los impulsos o señales provenientes de neuronas que están a su alrededor y envía a través del axón, señales emitidas en su propio cuerpo”.

Figura 3

Estructura de la Neurona Humana o Biológica.



Nota: Adaptado de Ponce (2010)

Cada neurona se encuentra conformada por los siguientes elementos primordiales:

- Dendritas
- Soma o Cuerpo de la célula
- Axón o Cilindro
- Sinapsis

La capacidad de comunicarse es una cualidad fundamental de la neurona. Las señales de entrada son percibidas por las dendritas y el soma; el soma las armoniza, las une y envía señales de salida; dichas señales son: eléctricas y químicas. La señal eléctrica, es un impulso eléctrico generado por la neurona y transita por todo el largo del Cilindro Eje, mientras la señal química es la señal que, viaja en la sinapsis entre una neurona y la siguiente.

Para construir una analogía de la actividad sináptica con las RNA debemos suponer que, las señales percibidas por la sinapsis de cada neurona son los ingresos de datos a la RNA; las cuales son simplificadas (compensadas o atenuadas) por medio del parámetro denominado peso, el cual está relacionado con la sinapsis. Cada señal de ingreso podría ocasionar que la neurona sea estimulada (sinapsis excitatoria o sinapsis con peso positivo) o sea inhibida (sinapsis inhibitoria o sinapsis con peso negativo).

Tabla 5

Métodos de pronóstico cualitativos

Neurona Biológica	Neurona Artificial
Dendrita	Entrada
Soma	Neurona
Sinapsis	Peso/Conexión
Axón	Salida

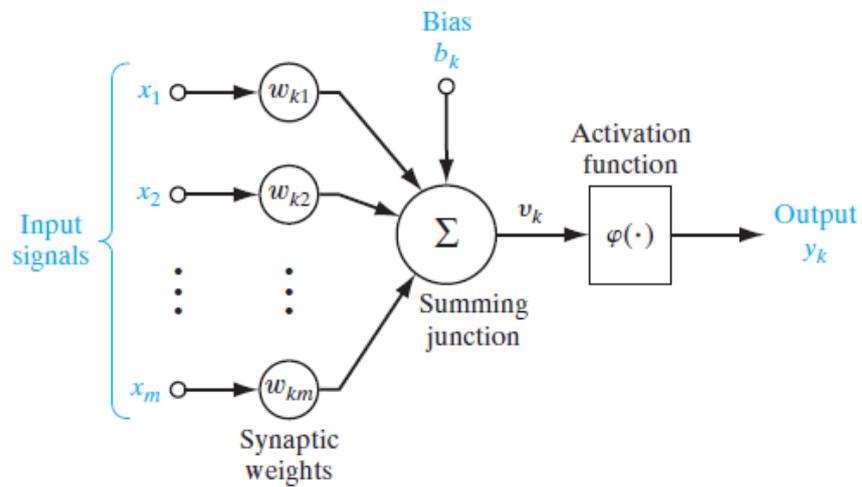
Nota: Adaptado de Ponce (2010)

2.3.9 Neurona Artificial

Martín, Emilio, & Antonio (2010), afirman que “las RNA se encuentran explicadas a modo de una estructura similar al sistema nervioso de los seres vivos, constituidos en base a una gran cantidad de unidades de procesamiento ensamblados, cada uno de ellos con un determinado peso. Cada unidad se denomina neurona, la cual recibe señales de las otras unidades y envía una salida escalar simple, la cual está sujeta a la información que se encuentre disponible en dicha unidad, almacenada de forma doméstica o que proviene de las uniones y sus respectivos pesos. La cantidad de conexiones determina la realización de funciones complejas o simples”.

Figura 4

Modelo no lineal de una neurona



Nota: Adaptado de Martín, Emilio & Antonio (2010)

Haykin (2009) en la misma línea afirma que “el modelo neuronal de la figura anterior también incluye un sesgo aplicado externamente, indicado por b_k , que posee la habilidad de aumentar o disminuir la entrada neta de la función de activación, en función de si el valor es positivo o negativo”.

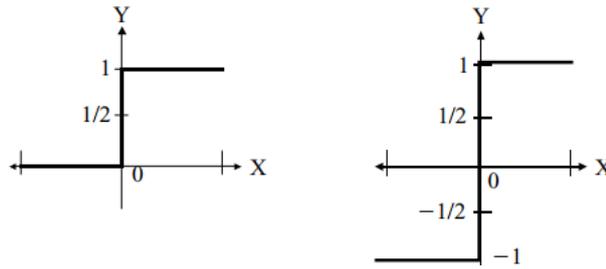
2.3.10 Funciones de activación

Existen una variedad de funciones, a continuación, citamos las más importantes:

Función de activación escalón

Figura 5

Función de activación escalón.



$$F_k(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

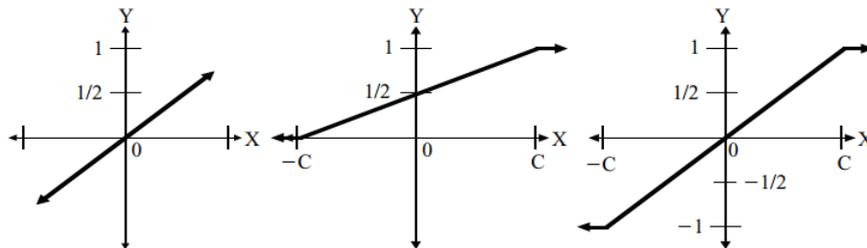
$$F_k(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

Nota: Adaptado de Matich (2001)

Función lineal o identidad y la función mixta

Figura 6

Función lineal y función mixta.



$$F_k(x) = x$$

$$F_k(x) = \begin{cases} 0, & x < -c \\ 1, & x > c \\ \frac{x}{2c} + \frac{1}{2}, & \text{en } OC \end{cases}$$

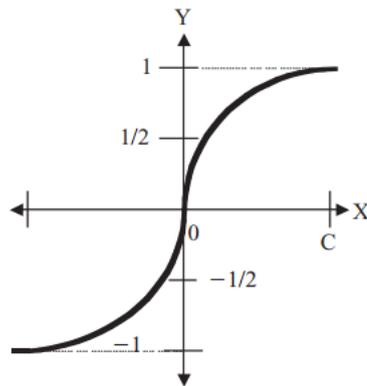
$$F_k(x) = \begin{cases} -1, & x < -c \\ 1, & x > c \\ ax, & \text{en } OC \end{cases}$$

Nota: Adaptada Matich (2001)

Función de activación tangente hiperbólica

Figura 7

Función tangente hiperbólica.



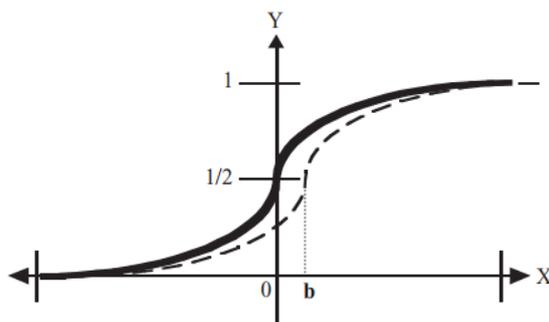
$$F_k(x) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$$

Nota: Adaptada de Matich (2001)

Función de activación sigmoideal

Figura 8

Función sigmoideal



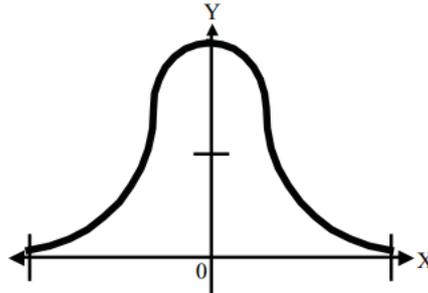
$$F_k(x) = \frac{1}{1 + e^{-u}}$$

Nota: Adaptado de Matich (2001)

Función de activación de Gauss

Figura 9

Función de activación de Gauss



$$F_k(x) = \frac{1}{1 + \vartheta\sqrt{2\pi}} e^{-\left(\frac{x^2}{2\vartheta^2}\right)}$$

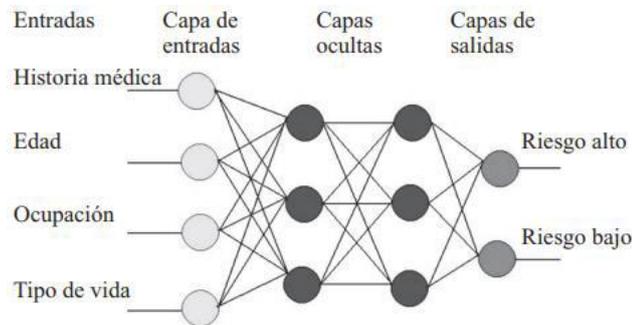
Nota: Adaptado de Matich (2001)

2.3.11 Tipología de las RNA

Haykin (2009) En su libro menciona “el concepto de *Network Architectures* o conexión entre neuronas. Las neuronas son los componentes principales de una RNA que se conectan entre ellas por medio de sus uniones para procesar y compartir información. En la siguiente figura se observa un croquis simple de una RNA, que presenta una estructura de 3 capas y está relacionada con la topología denominada feed-forward debido a que la información fluye hacia delante”.

Figura 10

Diagrama del esquema básico de una Red Neuronal Artificial (RNA).



Nota: Adaptado de Haykin (2009)

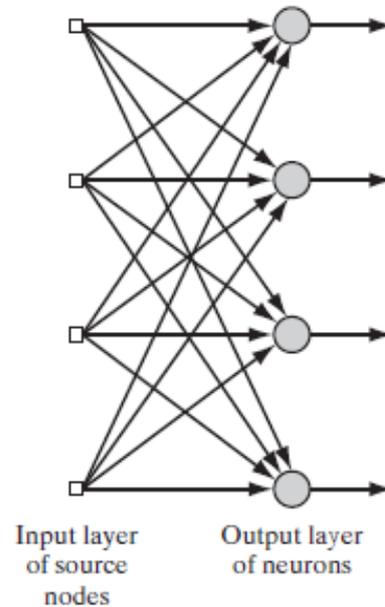
Las condiciones en las cuales se estructuran las neuronas de una RNA están estrechamente ligado al algoritmo de aprendizaje utilizado en el entrenamiento de ella. Por lo tanto, podemos hablar de algoritmos de aprendizaje (reglas) utilizados en el diseño de redes neuronales artificiales como estructurados. En general, podemos identificar tres clases fundamentalmente diferentes de arquitecturas de red (estructuras):

- Redes feedforward de una sola capa (*Single-Layer Feedforward Networks*)

Haykin (2009) afirma que en “una RNA estratificada, las neuronas se organizan formando niveles o capas. Para el croquis más básico de una red estratificada, deberíamos tener un nivel de entrada que se proyecta directamente en el nivel de salida, pero no al revés. Dicho de otra manera, es una RNA de tipo *feedforward* rigurosamente y se designa como red de una capa, y cuyo significado de "capa única" hace referencia a la capa de salida. No debemos contar el nivel de entrada porque allí no se ejecuta ningún cálculo”.

Figura 11

Red feedforward con 1 capa de neuronas, 4 neuronas tanto en la capa de entrada como en la capa de salida.



Nota: Adaptado de Haykin (2009)

- Redes feedforward de múltiples capas (*Multilayer Feedforward Networks*)

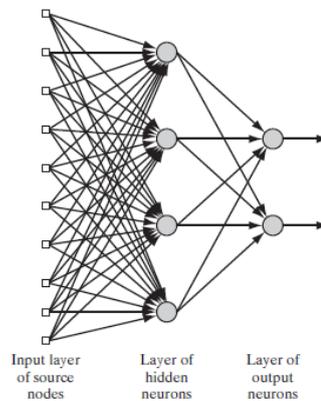
Haykin (2009) afirma que este tipo de neuronas se distingue por estar representado por una o más capas ocultas, y los nodos de cálculo denominados como neuronas o unidades ocultas; el término "oculto" se refiere al hecho de que esta parte de la red neuronal artificial no se ve directamente desde la entrada o la salida de la red. Las unidades ocultas tienen la obligación de meterse entre la entrada externa y la salida de red de alguna manera útil. En un sentido bastante flexible, la red adquiere una perspectiva global a pesar de su conectividad local, debido al conjunto adicional de conexiones sinápticas y la dimensión adicional de las interacciones neuronales.

Las neuronas o nodos en la capa de entrada proporcionan los componentes para el vector de entrada (patrones de activación), dicho vector conforma la señal de entrada

que se aplica a las neuronas (nodos de cálculo) de la capa segunda (en otras palabras, la capa oculta primera). Al producirse las señales de salida de esta capa se utilizarán como entradas en la tercera capa, y así sucesivamente. En general, cada capa de neuronas recibe de entradas únicamente las señales de salida de la anterior capa. El conjunto de señales de salida de las neuronas de la última capa significa la respuesta de la RNA a los componentes del vector de entrada (capa primera). La red *feedforward* con m nodos de origen, h_1 neuronas para la primera capa oculta, h_2 neuronas en la segunda capa escondida y q neuronas para la capa de respuesta (salida), se conoce como una red $m-h_1-h_2-q$. Como ejemplo, la figura siguiente se conoce como una red 10-4-2 porque tiene 10 nodos de origen, 4 neuronas ocultas y 2 neuronas de salida. (Haykin, 2009)

Figura 12

Red feedforward de múltiples capas totalmente acopladas, teniendo una capa de salida y una capa oculta.



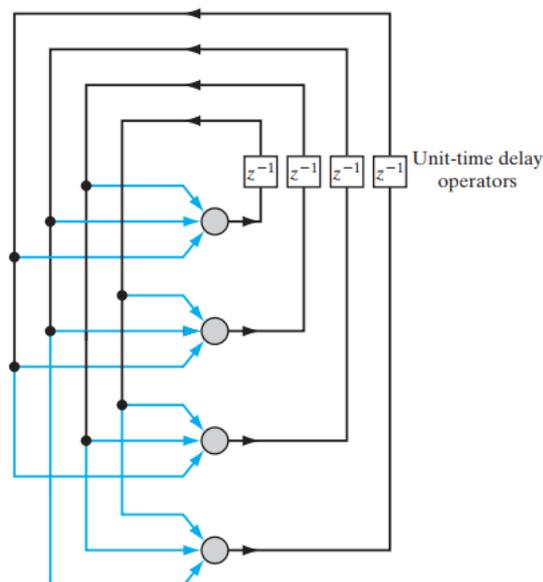
Nota: Adaptado de Haykin (2009)

- Redes recurrentes (*Recurrent Networks*)

Haykin (2009) “La RNA recurrente se diferencia de la RNA predictiva porque posee por lo menos un circuito de retroalimentación. Como ejemplo tenemos la RNA recurrente que consiste en una sola capa de neuronas donde cada neurona entrega su señal de salida a las entradas del resto de neuronas, como se ilustra en la siguiente figura. EN dicha estructura no hay bucles de auto-retroalimentación en la red, adicionalmente tampoco tiene neuronas ocultas; *self-feedback* describe un contexto donde la salida de la neurona alimenta su propia entrada”.

Figura 13

Red recurrente sin bucles de auto-retroalimentación y sin ninguna neurona oculta.

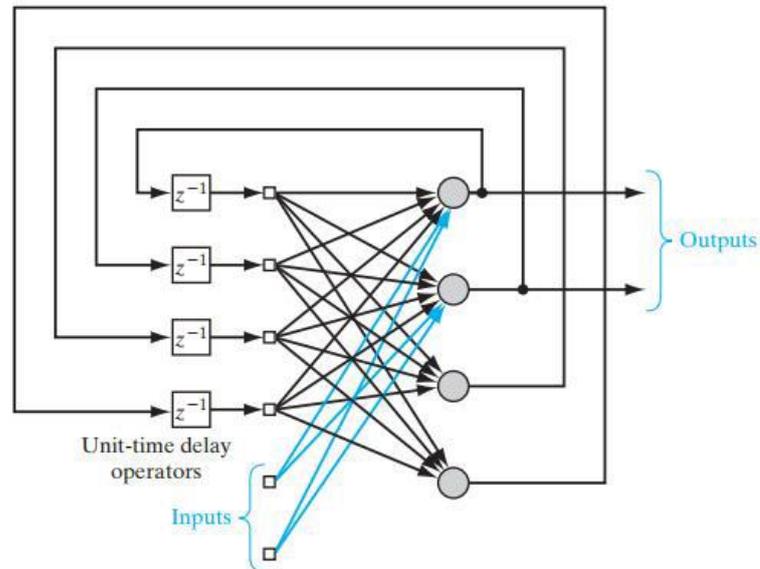


Nota: Adaptado de Haykin (2009)

En la figura siguiente ilustramos otra clase de redes recurrentes con neuronas ocultas. Las conexiones de retroalimentación que se muestran se originan en las neuronas de salida y en las escondidas.

Figura 14

Red recurrente con bucles de auto-retroalimentación y con una capa de neuronas ocultas



Nota: Adaptado de Haykin (2009)

La disposición de los bucles de retroalimentación, en cualquiera de las 2 figuras anteriores, tiene un profundo impacto en la capacidad de aprendizaje de la red y en su rendimiento. Además, los bucles de retroalimentación implican el uso de ramas particulares compuestas de elementos de retraso en la unidad de tiempo (denotados por z^{-1}), que dan como resultado una conducta dinámico no lineal, suponiendo que en la red neuronal artificial hay presencia de unidades no lineales.

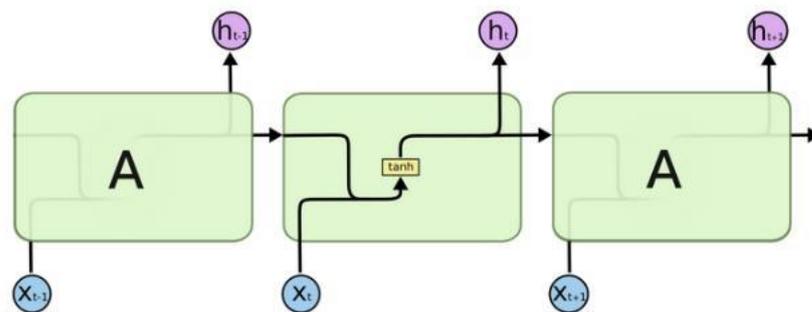
2.3.12 Redes recurrentes Profundas

Núñez (2017) El proceso en la mayoría de las redes neuronales recurrentes puede ser descompuesto en tres simples bloques de parámetros y correspondiente transformación:

- Desde la entrada al estado oculto
- Desde el estado oculto al siguiente estado oculto.
- Desde el estado oculto a la salida

Figura 15

Simple red recurrente.



Nota: Adaptado de Haykin (2009)

2.3.13 Problema de dependencia a largo plazo

Núñez (2017) “Como las redes neuronales recurrentes hacen uso de información secuencial se supone que todas las entradas (y salidas) son independientes entre sí. Pero para muchas tareas no se obtienen buenos resultados ya que se pierde información a medida que avanza. Otra forma de pensar en las redes neuronales recurrentes es que tiene una “memoria” que captura la información sobre lo que se ha calculado anteriormente. En teoría las redes neuronales recurrentes pueden hacer uso de la información en

secuencias arbitrariamente largas, pero en la práctica se limitan a mirar hacia atrás unos pocos pasos perdiendo el contexto de la predicción”.

2.3.14 Gradiente descendente

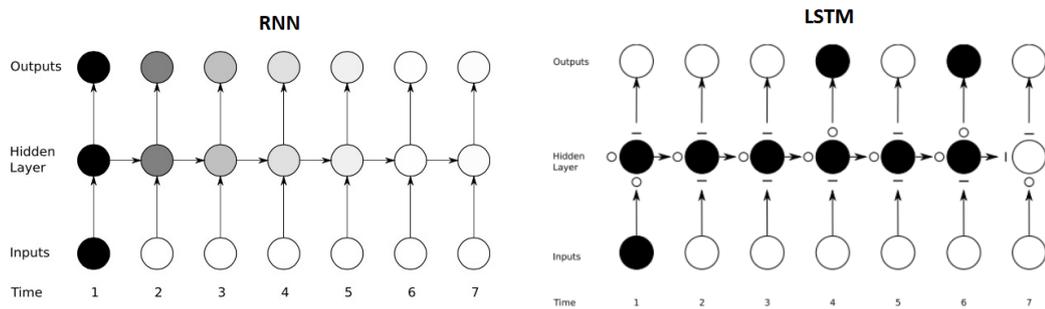
Introducido por Bengio et al.(1994) “el problema de los gradientes de explosión se refiere al gran aumento en la norma del gradiente durante el entrenamiento. Tales eventos se deben a la explosión de los componentes a largo plazo, los cuales pueden crecer más exponencialmente que los términos a corto plazo. El problema de gradiente descendente se refiere a lo opuesto del comportamiento descrito, cuando los componentes a largo plazo crecen exponencialmente rápido a la norma 0, haciendo imposible para el modelo aprender entre eventos temporalmente distantes”.

2.3.15 LSTM

Núñez (2017) afirma que “las redes de memoria de largo a corto plazo o Long Short-Term Memory (LSTM) es un sistema de aprendizaje profundo y son un tipo especial de red neuronal recurrente que es capaz de lidiar con el problema de la dependencia a largo plazo, estas redes fueron explícitamente diseñadas para recordar información por largos periodos de tiempo. Este tipo de red es ideal para trabajar con datos que vengan en secuencias, ya sea datos discretos o series temporales”.

Figura 16

Problema gradiente descendiente

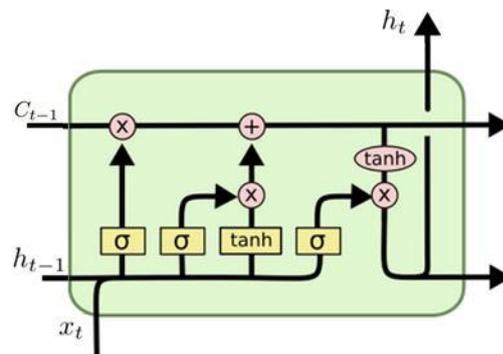


Nota: Adaptado de Nuñez(2017)

La unidad de larga memoria a corto plazo o *Long Short-Term Memory* (LSTM) fue inicialmente propuesta por Hochreiter y Schmidhuber (1997), desde entonces, unos mínimos de modificaciones fueron hechas a la unidad original de LSTM.

Figura 17

Arquitectura de red LSTM



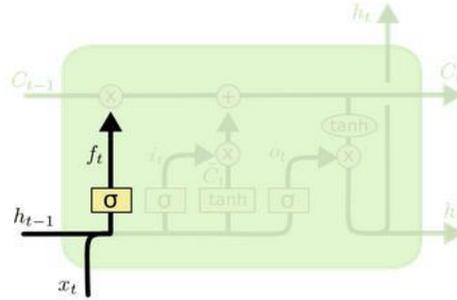
Nota: Adaptado de Hochreiter & Schmidhuber (1997)

2.3.16 Funcionamiento de la red LSTM

El primer paso de la red LSTM es decidir qué información va a desechar de la memoria, esta decisión es tomada por la compuerta “Forget Gate”.

Figura 18

Ingreso de datos



Nota: Adaptado de Hochreiter y Schmidhuber (1997)

Donde h_{t-1} y x_t se concatenan y el resultado será la entrada a la red que conforma la *Forget Gate* f_t^j , para cada j -ésimo elemento de la unidad LSTM, este resultado servirá para decidir si el estado de la memoria se dejará o se alterará eliminando algún elemento. El alcance en que se olvida la memoria existente se modula mediante la compuerta de olvido f_t^j .

$$f_t^j = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1})^j$$

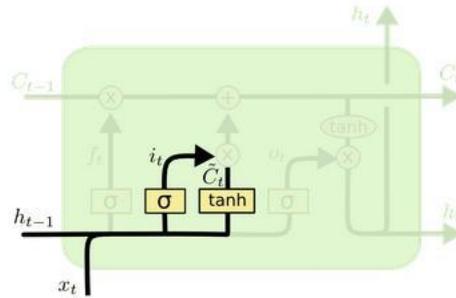
El siguiente paso que realiza la LSTM es decidir qué información nueva almacenará en la memoria, primero en una compuerta llamada “Input Gate” i_t decide que valores actualizar, luego una pequeña red con \tanh crea un vector de valores nuevos candidatos, C_t^j , y luego se combinan los dos resultados para crear una actualización del estado. El grado en el cual el nuevo contenido de memoria es agregado a la celda de memoria es modulada por la compuerta de entrada i_t^j . Donde V_f y V_i son matrices diagonales.

$$i_t^j = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1})^j$$

$$c_t^j = \tanh(W_c x_t + U_c h_{t-1})^j$$

Figura 19

Almacenamiento de datos

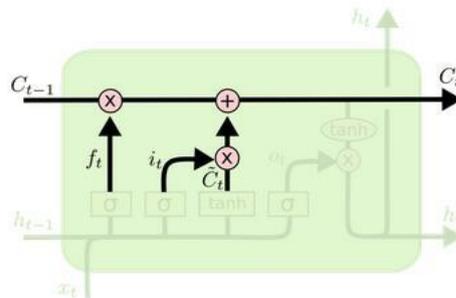


Nota: Adaptado de Hochreiter y Schmidhuber (1997)

Luego es posible actualizar el estado de la memoria cambiando C_{t-1}^j por el nuevo estado \tilde{C}_{t-1}^j , debido a que las decisiones ya fueron realizadas solo se aplican las operaciones correspondientes, es decir, se multiplica el estado anterior por i_t^j olvidando una cierta cantidad de información en función de f_t^j , se suma $i_t^j * C_t^j$ a la memoria, actualizando esta con nuevos valores que pueden servir a futuro.

Figura 20

Actualización de memorias



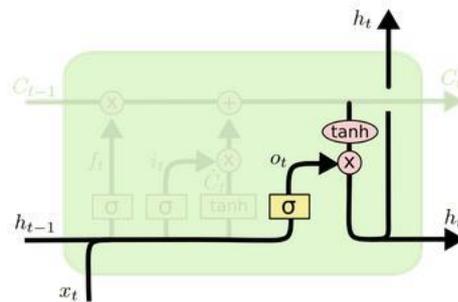
Nota: Adaptado de Hochreiter y Schmidhuber (1997)

$$C_t^j = f_t^j * C_{t-1}^j + i_t^j \tilde{C}_{t-1}^j$$

Adaptado de Hochreiter y Schmidhuber (1997) “Finalmente se decide la salida de la celda LSTM mediante la salida de la red con el sigmoide, que servirá para decidir qué elementos de la memoria se combinarán y un segundo elemento será filtrado de datos desde la memoria \tanh para “empujar” los valores entre -1 y 1. Estos dos elementos se multiplicarán dando como resultado la nueva salida de la celda”.

Figura 21

Salida a siguiente neurona



Nota: Adaptado de Hochreiter y Schmidhuber (1997)

Adaptado de Hochreiter y Schmidhuber (1997), “A diferencia de la unidad recurrente que simplemente calcula una suma ponderada de la señal de entrada y aplica una función no lineal, cada j -ésimo elemento de la unidad LSTM mantiene una memoria C_t^j al tiempo t . Donde o_t^j es una compuerta de salida que modula la cantidad de contenido de la memoria que es expuesta. La salida h_t^j , o activación, de la unidad LSTM es entonces”

$$\sigma_t^j = \sigma(W_0 x_t + U_0 h_{t-1} + V_0 c_t)^j$$

$$h_t^j = \sigma_t^j * \tanh(C_t^j)^j$$

Donde σ es una función lógica Sigmoide y V_0 es una matriz diagonal.

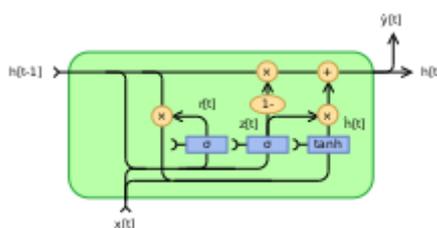
2.3.17 GRU

Calin (2020) Menciona que “las unidades recurrentes cerradas (GRU) son un mecanismo de compuerta en redes neuronales recurrentes, introducido en 2014 por Kyunghyun Cho. El GRU es como una memoria a corto plazo (LSTM) con una puerta de olvido, pero tiene menos parámetros que LSTM, ya que carece de una puerta de salida. Se encontró que el desempeño de GRU en ciertas tareas de modelado de música polifónica, modelado de señales de voz y procesamiento del lenguaje natural es similar al de LSTM. Se ha demostrado que las GRU presentan un mejor rendimiento en determinados conjuntos de datos más pequeños y menos frecuentes. Sin embargo, como lo demostraron Gail Weiss, Yoav Goldberg y Eran Yahav, el LSTM es "estrictamente más fuerte" que el GRU, ya que puede realizar fácilmente un conteo ilimitado, mientras que el GRU no. Es por eso que GRU no logra aprender lenguajes simples que se pueden aprender con LSTM. Hay varias variaciones en la unidad con compuerta completa, con la compuerta realizada utilizando el estado oculto anterior y el sesgo en varias combinaciones, y una forma simplificada llamada unidad de compuerta mínima”.

- Unidad completamente cerrada

Figura 22

Red recurrente con unidad completamente cerrada

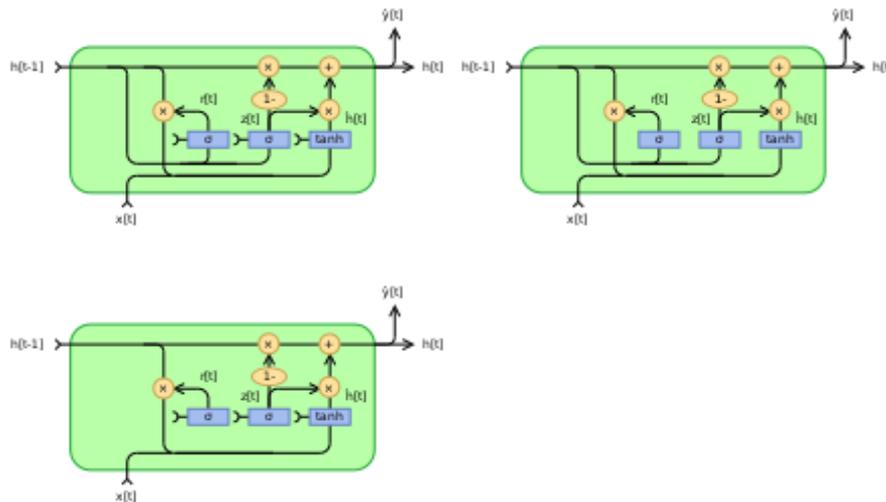


Nota: Adaptado de Calin (2020)

- Unidad cerrada mínima

Figura 23

Red recurrente con unidad cerrada mínima



Nota: Adaptado de Calin (2020)

2.3.18 Series Temporales

Prabhudesai et al.(2020) “Una serie temporal o cronológica es una sucesión de datos medidos en determinados momentos y ordenados cronológicamente. Los datos pueden estar espaciados a intervalos iguales (como la temperatura en un observatorio meteorológico en días sucesivos al mediodía) o desiguales (como el peso de una persona en sucesivas mediciones en el consultorio médico, la farmacia, etc.). Para el análisis de las series temporales se usan métodos que ayudan a interpretarlas y que permiten extraer información representativa sobre las relaciones subyacentes entre los datos de la serie o de diversas series y que permiten en diferente medida y con distinta confianza extrapolar o interpolar los datos y así predecir el comportamiento de la serie en momentos no observados, sean en el futuro (extrapolación pronóstica), en el pasado (extrapolación retrógrada) o en momentos intermedios (interpolación)”.

2.3.19 Descomposición de una serie de tiempo

Tendencia y nivel

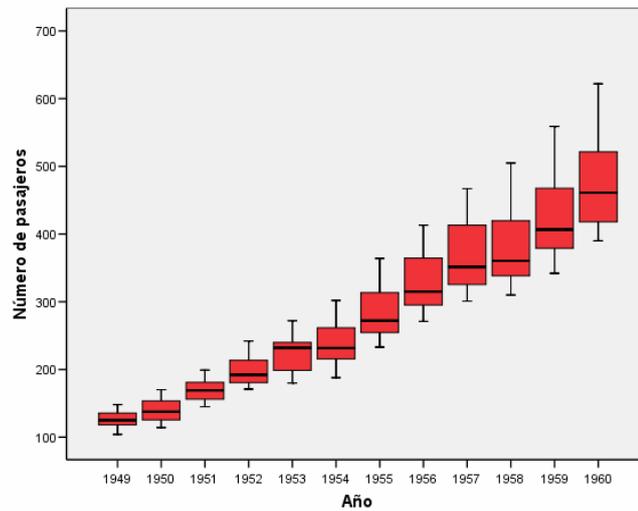
“El nivel de una serie es una medida local de tendencia central, como por ejemplo la media, de cada periodo de tiempo que consideremos. Cuando trabajamos con un calendario (tiempo representado en días, meses o años), no es recomendable establecer periodos de tiempo antinaturales para estudiar esta característica”.

“Una vez calculado el nivel de la serie debemos observar su estabilidad, debemos ver si la medida de tendencia central elegida tiene valores similares en los periodos de tiempo en los que dividimos el calendario. También tendremos que observar su tendencia, es decir, si presenta una dirección constante de cambio de nivel”.

“Una vez representado el nivel de la serie lo estudiamos para determinar si es o no estable. En el caso de la serie Airline vemos que no lo es, la media del número de pasajeros contabilizados cada año (periodo elegido para dividir el calendario) va aumentando progresivamente, de manera que presenta tendencia creciente”.

Figura 24

Diagrama de cajas de la serie Airline



Nota: Adaptado de Catalan (2010)

Estacionalidad

Una serie es estacional cuando podemos observar en ella un patrón sistemático que se repite periódicamente (cada año, cada mes, etc., dependiendo de las unidades de tiempo en que vengán recogidos los datos).

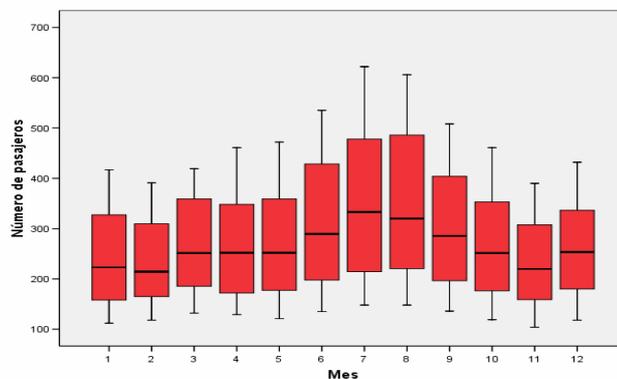
Existen muchos ejemplos de series con comportamiento estacional. El hecho de que las vacaciones laborales se concentren en los meses de verano condiciona los valores de muchas series. Un claro ejemplo es el de las series relacionadas con el turismo, tales como número mensual de pernoctaciones hoteleras, número de viajeros en avión registrado por meses, etc. La pauta sistemática que caracteriza la estacionalidad no tiene por qué ser anual, algunas series tienen una estacionalidad cuyo periodo es de un mes, una semana o incluso un día. Un ejemplo de este último caso podría ser el de las ventas por horas de un determinado producto en un centro comercial. Es muy probable que cada día se repita un patrón de comportamiento muy similar; habrá horas punta con gran

número de ventas y otras en las que apenas hay gente y por lo tanto se registran muchas menos ventas.

Si una serie tiene un comportamiento estacional muy claro se suele detectar fácilmente en el gráfico de secuencia, tal y como ocurre con la serie Airline. En esta serie se observa claramente que hay un patrón similar que se repite año tras año, con evidentes aumentos en el número de pasajeros en los meses de verano e incrementos más moderados en los meses correspondientes a las vacaciones de Navidad. Sin embargo, para estudiar el comportamiento estacional de una serie es más adecuado representar el diagrama de box-plot de cada una de las unidades de tiempo que componen los periodos en los que se agrupan los patrones de repetición. Es decir, en el caso de las series con observaciones mensuales representaremos el diagrama de box-plot de cada uno de los años.

Figura 25

Diagrama de cajas de la serie Airline



Nota: Adaptado de Catalan (2010)

2.3.20 Evaluación de Modelos de Pronóstico

Ramasubramanian y Singh (2019) mencionan “que la distribución de variables dependientes es una consideración importante al elegir los métodos para evaluar los modelos. Intuitivamente, terminamos comparando lo residual de la distribución (valor real versus valor predicho) con distribución normal (es decir, ruido aleatorio) o alguna otra distribución basada en las métricas que elegimos. Para el caso de las series temporales, las mediciones para evaluar el desempeño de un modelo son el RMSE (raíz del error cuadrático medio), MAPE (Porcentaje de error absoluto medio) y MAE (error absoluto medio)”.

Porcentaje de error absoluto medio (MAPE)

$$Mape = \frac{\sum_{i=1}^n 100 |y_{ireal} - y_{ipronostico}|}{y_{ireal} n}$$

Raíz del error cuadrático medio (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{ireal} - y_{ipronostico})^2}{n}}$$

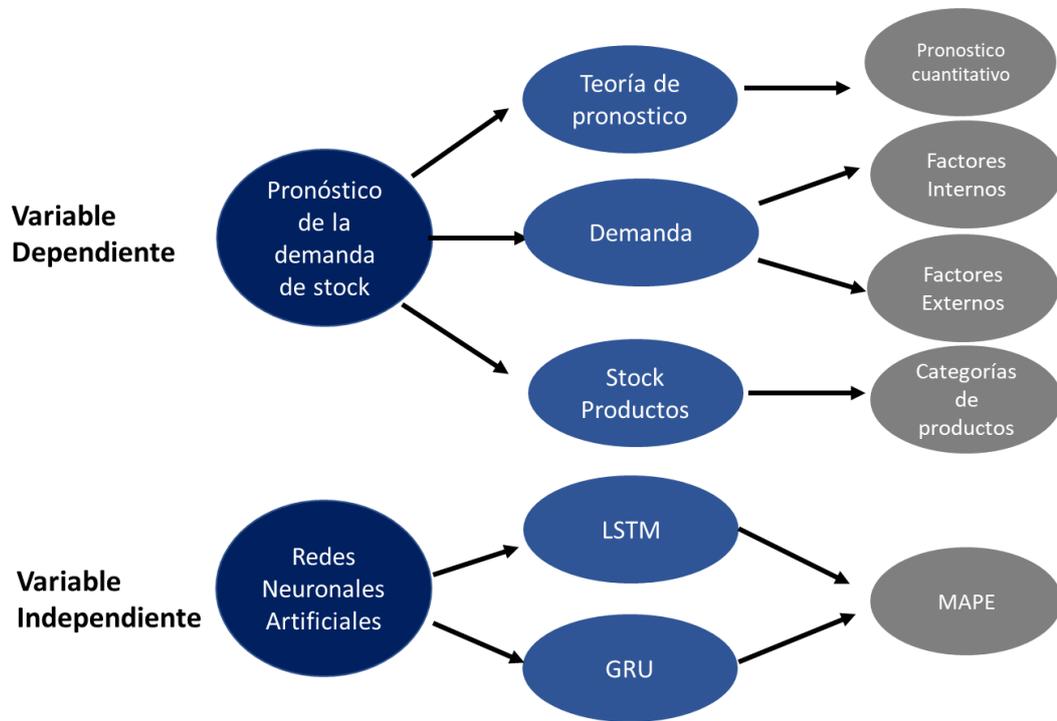
Error absoluto medio (MAE)

$$MAE = \frac{1}{N} \sum_i^n |(y_{ireal} - y_{ipronostico})|$$

2.3.21 Mapa conceptual de las teorías que sustentan el estudio

Figura 26

Diagrama de los fundamentos teóricos que sustentan el estudio



Nota: La representación gráfica de fundamento teóricos es una elaboración propia

El diagrama mostrado anteriormente hace referencia a la descomposición conceptual de las variables usadas en la investigación y es una adaptación a la teoría brindada por: Prabhakar(2018), Lanham, et al.(2017), Chase et al.(2009)

2.4 Definición de términos básicos

- Consumo masivo:

Fernie y Sparks (2019), “lo definen como la acción y efecto de consumir o gastar, sean productos, bienes o servicios, como por ejemplo la energía, entendiendo por consumir, como el hecho de utilizar estos productos y servicios para satisfacer necesidades primarias y secundarias. El consumo masivo ha dado lugar al consumismo y a la denominada sociedad de consumo. En términos puramente económicos se entiende por consumo la etapa final del proceso económico de producción, definida como el momento en que un bien o servicio

produce alguna utilidad al consumidor. En este sentido hay bienes y servicios que directamente se destruyen en el acto del consumo, mientras que con otros lo que sucede es que su consumo consiste en su transformación en otro tipo de bienes o servicios diferentes”.

- Productos:

Fernie y Sparks (2019), lo definen como “una cosa o un objeto producido o fabricado, algo material que se elabora de manera natural o industrial mediante un proceso, para el consumo o utilidad de los individuos. La palabra producto deriva del latín productus y posee diferentes significados según el área en el cual se emplee. Los productos son todos aquellos objetos u artefactos, fabricados en industrias, empresas siguiendo una línea de producción o de manera artesanal por las personas”.

- Demanda:

Fernie y Sparks (2019), la definen como “la solicitud para adquirir algo. En economía, la demanda es la cantidad total de un bien o servicio que la gente desea adquirir. El significado de demanda abarca una amplia gama de bienes y servicios que pueden ser adquiridos a precios de mercado, bien sea por un consumidor específico o por el conjunto total de consumidores en un determinado lugar, a fin de satisfacer sus necesidades y deseos. Estos bienes y servicios pueden englobar la práctica totalidad de la producción humana como la alimentación, medios de transporte, educación, ocio, medicamentos y un largo etcétera. Por esta razón, casi todos los seres humanos que participan de la vida moderna son considerados como ‘demandantes’”.

- Stock:

Thomopoulos (2015), menciona que “es un término anglosajón el cual indica la cantidad de productos o materias primas que posee un comercio en su almacén a la espera de su venta o comercialización. El stock es una inversión que debe de realizar el comercio para asegurarse de atender el flujo de demanda de manera adecuada para el buen funcionamiento del mismo, y así poder llevar a cabo sus actividades de manera normal sin necesidad de paralizar sus ventas por falta del producto”.

- Algoritmo:

Ponce (2010), define en su libro que el termino algoritmo “es parte de las matemáticas, lógica, ciencias de la computación y disciplinas relacionadas, un algoritmo (del latín, dixit algorithmus y este del griego arithmos, que significa «número», quizá también con influencia del nombre del matemático persa Al-Juarismi)¹ es un conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que permite, típicamente, solucionar un problema, realizar un cómputo, procesar datos y llevar a cabo otras tareas o actividades.² Dados un estado inicial y una entrada, siguiendo los pasos sucesivos se llega a un estado final y se obtiene una solución. Los algoritmos son el objeto de estudio de la algoritmia”.

- Pronostico:

Wei (2019), “El pronóstico es el proceso de estimación en situaciones de incertidumbre. El término predicción es similar, pero más general, y usualmente se refiere a la estimación de series temporales o datos instantáneos. El pronóstico ha evolucionado hacia la práctica del plan de demanda en el pronóstico diario de los negocios. La práctica del plan de demanda también se refiere al pronóstico de la cadena de suministros”.

- Aprendizaje Supervisado:

Ponce (2010), define aprendizaje como “una técnica para deducir una función a partir de datos de entrenamiento. Los datos de entrenamiento consisten de pares de objetos (normalmente vectores): una componente del par son los datos de entrada y el otro, los resultados deseados. La salida de la función puede ser un valor numérico (como en los problemas de regresión) o una etiqueta de clase (como en los de clasificación). El objetivo del aprendizaje supervisado es el de crear una función capaz de predecir el valor correspondiente a cualquier objeto de entrada válida después de haber visto una serie de ejemplos, los datos de entrenamiento. Para ello, tiene que generalizar a partir de los datos presentados a las situaciones no vistas previamente”.

- Ciclos:

Buteikis (2019) lo define como “la exhibición de datos aumenta y disminuye no son de plazo fijo. La duración promedio de los ciclos es más larga que la longitud de un patrón estacional. En la práctica, el componente de tendencia es se supone que incluye también el componente cíclico. A veces los componentes de tendencia y cíclicos juntos se denominan ciclo de tendencia”.

- Tendencias:

Buteikis (2019), “lo define como un aumento o disminución a largo plazo en los datos que puede no ser lineal a veces la tendencia puede cambiar dirección a medida que aumenta el tiempo”.

- Estación:

Buteikis (2019), “existe cuando una serie exhibe regular fluctuaciones basadas en la temporada (por ejemplo, cada mes / trimestre / año). La estacionalidad es siempre de un período fijo y conocido”.

2.5 Hipótesis

2.6.1 Hipótesis General

El modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock es más preciso que el modelo actualmente implementado en la empresa de consumo masivo.

2.6.2 Hipótesis Específicas

- El modelo de Redes Neuronales Artificiales LSTM es más eficiente que GRU en el pronóstico de demanda de stock usando los datos de la empresa de consumo masivo.
- La topología adecuada de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock es obtener un modelo de pronóstico menor al 10% en error cuadrático medio.
- Las variables que tienen más influencia en el modelo de redes neuronales artificiales recurrentes para el pronóstico de stock son las relacionadas a factores internos de la empresa de consumo masivo.

2.6 Variables

En esta sección colocamos la definición operacional de las variables, dimensiones e indicadores necesaria para realizar el modelo de pronóstico de demanda con redes neuronales recurrentes.

- **La variable dependiente**

Demanda de stock de productos

- **Las variables independientes o drivers:**

Transacciones de clientes por productos y día

Órdenes de compra de clientes por día

Segmentos de clientes

Inventario de productos

Número búsquedas de palabras clave relacionados a alimentos en *Google Trends*

Tabla 6

Matriz de operacionalización de variables

Variab les	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Item	Instrumento
<p>Independiente:</p> <p>Redes Neuronales Artificiales</p> <p>Dependiente:</p> <p>Pronóstico de demanda de stock</p>	<p>Independientes:</p> <p>Las redes neuronales artificiales (RNA) son modelos de la inteligencia artificial que se inspiran en el comportamiento de las neuronas y las conexiones cerebrales para resolver problemas</p> <p>Dependientes:</p> <p>Es un proceso que consiste en utilizar datos pasados para determinar acontecimientos futuros, los cuales nos dará las cantidades de stock</p>	<p>Independientes:</p> <p>Los algoritmos de redes neuronales artificiales se utilizan para la predicción en datos históricos.</p> <p>Dependientes:</p> <p>El pronóstico de demanda de stocks se da en niveles de toneladas, las cuales se tratan de estimar por factores externos e internos.</p>	<p>Independientes:</p> <p>Redes Neuronales LSTM</p> <p>Redes Neuronales GRU</p> <p>Dependientes:</p> <p>Demanda generada por factores internos</p> <p>Demanda generada por factores externos</p>	<p>Independientes:</p> <p>Error Porcentual Absoluto Medio (MAPE)</p> <p>Dependientes:</p> <p>Transacciones de los clientes a nivel de producto y día.</p> <p>Órdenes de compra los clientes a nivel de producto y día.</p> <p>Inventarios de stock en la cadena de suministro</p> <p>Segmento de Clientes</p> <p>Búsquedas de palabras de categoría alimentos en Google Trends</p>	<p>Independientes:</p> <p>Observaciones pronosticadas</p> <p>Observaciones reales</p> <p>Dependientes:</p> <p>I1: Código único de la transacción por producto</p> <p>I2: Código único de la transacción por cliente</p> <p>I3: Código único de ordenes por producto</p> <p>I4: Código único de la ordenes por cliente</p> <p>I5: Marca o nombre del segmento</p> <p>I6: Numero búsquedas de nombre de productos de Google Trends</p>	<p>Independientes:</p> <p>Arquitectura de modelos para Redes Neuronales Artificiales LSTM</p> <p>Arquitectura de modelos para Redes Neuronales Artificiales GRU</p> <p>Dependientes:</p> <p>Base de datos del ERP de la empresa de consumo masivo</p>

CAPÍTULO III: MARCO METODOLÓGICO

3.1 Enfoque, Tipo, nivel, diseño y método de investigación

- La investigación tiene un **enfoque de tipo cuantitativo**, según lo definido por Hernandez et al.(2014), dado que se empleo la recolección de datos para probar hipotesis con base a la definición numérica y el analisis estadístico con el fin de establecer pautas y comportamientos.
- Al tratarse de resolver un problema de una empresa peruana de consumo masivo, específicamente para la gerencia de Demanda, **el nivel es de tipo aplicada**; según lo definido por Hernandez et al.(2014).
- La investigación tiene un **diseño de tipo no experimental**, que va en concordancia con lo siguiente: “un estudio no experimental no se genera ninguna situación, sino que se observan situaciones ya existentes, no provocadas intencionalmente en la investigación por quien la realiza. En la investigación no experimental las variables independientes ocurren y no es posible manipularlas, no se tiene control directo sobre dichas variables ni se puede influir en ellas, porque ya se produjeron, de la misma forma que las respectivas consecuencias”; según lo definido por Hernandez et al.(2014).
- Y por el **método de investigación es hipotético-deductivo** porque a partir del comportamiento de datos históricos de la demanda de productos es posible tomar características que puedan predecir la demanda futura; según lo definido por Hernandez et al.(2014).

3.2 Población y Muestra

3.2.1 Población de estudio

La población de estudio estará conformada por clientes de la empresa de consumo masivo y el detalle se sus compras en toneladas a un nivel de detalle de producto, en el periodo de tiempo enero 2016 a febrero 2020.

La unidad de estudio estará conformada por la combinación de cliente, producto y día para la empresa peruana de consumo masivo. La cantidad de registros para la combinación es de 5.6 millones en el periodo 2016 a febrero 2020.

3.2.2 Diseño muestral

Para el pronóstico de la demanda de stock de productos, no se optó por extraer una muestra. Para este caso específico, se seleccionó a toda la población para el análisis, por lo tanto, no hubo un diseño muestral. Sin embargo, el diseño metodológico para los modelos de Machine Learning, nos indica que, la población de estudio se debe dividir en 2 partes aleatorias y representativas: una llamada “train” que tiene como finalidad elaborar un modelo (80% de la población); y otra llamada “test” para realizar las pruebas de performance del modelo obtenido (20% de la población).

3.3 Técnicas e instrumentos de recolección de datos

La recolección de datos se realizará de fuentes de información internas y externas. Para el caso de las fuentes internas, la empresa peruana de consumo masivo trabaja con un sistema ERP, donde almacena la información de ventas de productos, clientes, producción, distribución y logística.

Para el caso de fuentes externas tomaremos información de búsquedas en Google Trends de forma evolutiva para cada categoría de productos.

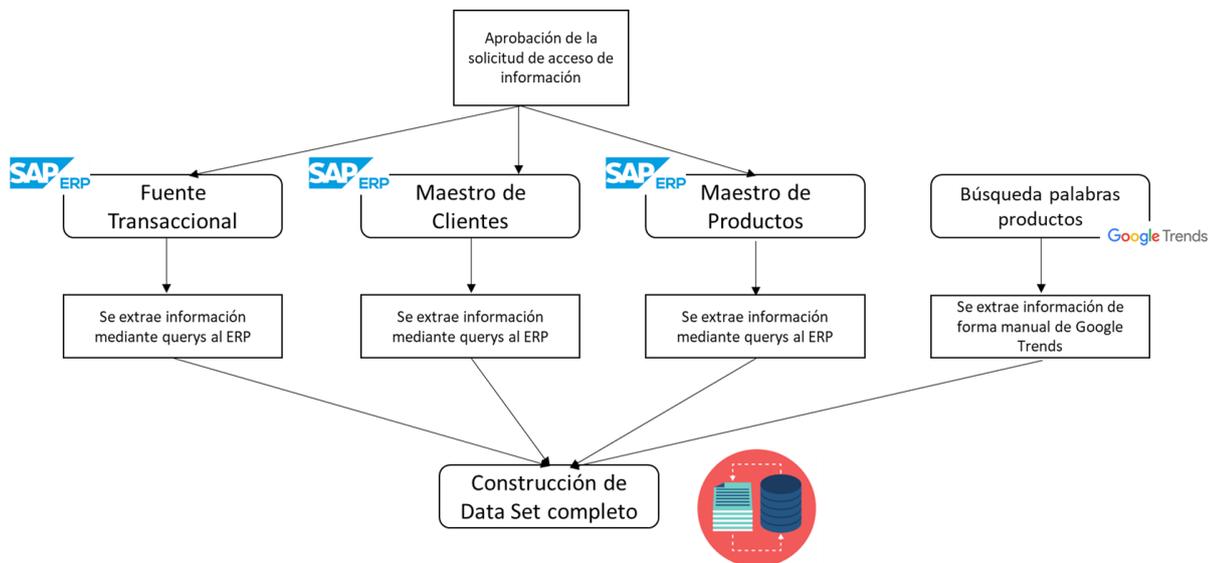
3.3.1 Procedimientos para la recolección de datos

Para obtener los datos se realizará los siguientes pasos:

1. Firmar un acuerdo interno entre las gerencias de Advance Analytics, gerencias de Demanda y gerencia de Tecnología de Información de la empresa de consumo masivo. Dado que para acceder a los datos se necesita de permisos especiales y generación de accesos hacia su ERP.
2. Luego de contar con los accesos, se realizarán queries vía SQL a las bases de datos, considerando los filtros de periodo. El resultado de las descargas se almacenará en archivos planos.
3. Para recolectar los datos de fuentes externas lo realizaremos de google trends, los resultados se almacenarán en archivos planos.
4. Consolidamos toda la información extraída en un solo archivo plano.

Figura 27

Diagrama de proceso de recolección de datos



Nota: Adaptado del flujo de información para empresas de SAP

3.4 Descripción de procedimiento de análisis de datos

Los datos serán analizados utilizando técnicas de redes neuronales artificiales. Para implementar el modelo se siguió las etapas mostradas en la siguiente secuencia:

Figura 28

Diagrama de proceso de recolección de datos



Nota: Modificado del libro de Mechelli & Vieira, 2019.

3.5.1 Formulación del problema

Esta sección tiene como objetivo entender la serie temporal, para ello se tendrá que realizar una descomposición en sus tres componentes tales como la tendencia, estacionalidad y ruido blanco. Para este punto usaremos librerías de python como Matplotlib, Seaborn, Numpy y Pandas. Las técnicas para usar se citan a continuación:

- Análisis espectral
- Descomposición de la serie (Ciclos, Estacionalidad, Tendencia y residuo)
- Estadísticos descriptivos: media, mediana, dispersión y correlaciones
- Scatterplot de serie contra retrasos

3.5.2 Preparación de los datos

Esta fase cubre todas las actividades necesarias para construir el conjunto final de datos, para llegar a este punto hay que desarrollar mucho tratamiento de los datos y entenderlos. Las tareas incluyen la selección de tablas, transformación de atributos, limpieza de variables, identificación y tratamiento de datos atípicos. Para este punto usaremos librerías de Python como Matplotlib, Seaborn, Numpy y Pandas.

- Imputación de valores faltantes

- Escalar o Normalizar variables

3.5.3 Feature engineering

En esta sección generamos las variables relevantes que explican adecuadamente a nuestra variable respuesta que es la demanda por producto, día y cliente. Las variables generadas corresponden a indicadores como la suma, media, mediana, moda y desviación de la demanda y algunas variables que son combinación de otras que están correlacionadas con la variable respuesta.

3.5.4 Entrenamiento del Modelo

Los modelos aplicados al conjunto final de datos serán los siguientes:

- Redes neuronales LSTM
- Redes neuronales GRU

Las librerías usadas en python para el desarrollo de los modelos serán Sklearn, Keras

3.5.4 Evaluación del modelo

En esta etapa en el proyecto, se han construido uno o varios modelos que parecen alcanzar calidad suficiente desde la una perspectiva de análisis de datos. Para la evaluación usaremos el error cuadrático medio, el cual nos ayudará a detectar que modelo desarrollado tiene el mejor desempeño.

CAPITULO IV: RESULTADOS Y ANÁLISIS DE

4.1 Resultados

Para un mejor entendimiento y a fin de que se conozcan las características de las series temporales que queremos predecir, iniciamos con un análisis descriptivo y posteriormente predictivo a nivel de categorías y centros de distribución.

4.1.1 Análisis descriptivo

Para el desarrollo del modelo de pronóstico de demanda contamos con las siguientes bases de datos proporcionados por la empresa de consumo masivo: Información transaccional, información de cotizaciones internacionales, información de producto bruto interno por sector, información de tipo de cambio.

Presentamos una tabla resumen con las estadísticas descriptivas principales de los 1,441 registros a nivel de día, algunas características resumen:

Las bases de datos no contienen valores vacíos o nulos

Las variables que presentan mayor desviación son el número de clientes y el monto de venta.

Las variables que presentan un valor máximo son el número de clientes y el monto de venta.

Las variables que presentan la menor desviación son las relacionadas al tipo de cambio.

Contamos con un total de 43 variables, 3 corresponden a información transaccional, 16 a información de cotizaciones internacionales, 20 a Producto bruto interno por sector y 4 a información de tipo de cambio.

Tabla 7*Estadísticas descriptivas de las variables cuantitativas a nivel de día de venta*

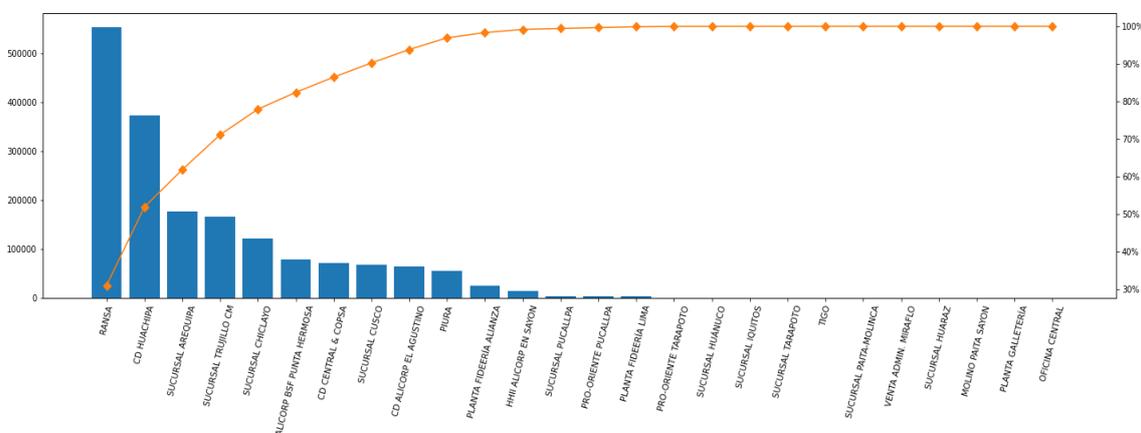
variable	count	media	desv. est.	min	max	mediana
volumen_sum	1441	1,159.8	690.8	-9.7	4,772.9	1,163.3
amount_sum	1441	5,382.6	3,203.2	-53.5	23,561.8	5,290.6
num_customers_sum	1441	4,294.7	2,651.6	2.0	16,403.0	4,076.0
Cotizacio_internacionales_aceitesoya_USD	1441	680.1	53.5	576.1	836.2	672.2
Cotizacio_internacionales_maiz_USD	1441	132.8	10.6	109.4	176.0	131.9
Cotizacio_internacionales_petroleo_barril_USD	1441	54.0	9.7	26.2	77.4	53.5
Cotizacio_internacionales_trigo_USD	1441	187.5	27.2	128.2	246.6	188.9
PBI_DI_Electricidad_Agua_day	1441	35.6	1.7	32.5	39.5	35.4
PBI_DI_Agricola_day	1441	30.1	9.1	19.8	52.5	26.1
PBI_DI_Agropecuario_day	1441	31.5	6.1	24.0	47.0	29.9
PBI_DI_Construcción_day	1441	38.9	7.6	28.4	63.8	37.0
PBI_DI_Derechos de Importación_day	1441	34.0	2.0	30.1	39.6	34.2
PBI_DI_Manufactura no Primaria_day	1441	25.6	1.3	22.5	29.0	25.6
PBI_DI_Manufactura_Procesadores Recursos Primarios_day	1441	26.7	5.1	19.6	44.3	24.7
PBI_DI_Mineria_Hidricarburos_day	1441	30.7	1.8	25.6	34.0	31.1
PBI_DI_Mineria_Mineria_Hidrocarburos_day	1441	29.6	2.5	19.9	33.4	29.9
PBI_DI_Mineria_Mineria_Manufactura_day	1441	25.9	1.7	22.7	31.0	25.7
PBI_DI_Mineria_Mineria_Metalica_day	1441	30.9	1.9	25.9	34.4	31.1
PBI_DI_Otros_Servicios_day	1441	37.9	2.5	33.3	46.0	37.5
PBI_DI_Pecuario_day	1441	34.0	2.1	29.6	38.0	34.0
PBI_DI_Pesca_day	1441	19.7	10.8	7.2	53.9	15.1
PBI_Demanda_Interna_day	1441	35.8	1.9	31.1	39.8	35.5
PBI_Demanda_Interna_sin_inventarios_day	1441	36.7	3.0	31.4	44.7	36.4
PBI_total_day	1441	33.9	2.1	29.3	39.4	33.7
Precio_ton_usd_arroz_day	1441	95.7	14.1	72.4	110.9	106.5
Precio_ton_usd_azucar_day	1441	83.1	16.9	64.3	123.9	76.1

Precio_ton_usd_leche_polvo_day	1441	597.3	77.5	436.1	695.0	627.7
Precio_ton_usd_maiz_day	1441	35.0	2.4	31.5	41.2	34.6
Precio_ton_usd_pizo_arroz_day	1441	112.0	16.6	83.5	128.6	123.3
Precio_ton_usd_pizo_azucar_day	1441	95.7	12.8	79.5	127.3	92.9
Precio_ton_usd_pizo_leche_polvo_day	1441	658.5	59.3	604.0	807.0	628.3
Precio_ton_usd_pizo_maiz_day	1441	41.8	8.2	34.8	60.8	37.7
Precio_ton_usd_techo_arroz_day	1441	127.3	16.1	98.2	143.3	137.8
Precio_ton_usd_techo_azucar_day	1441	113.2	17.9	94.3	157.2	105.9
Precio_ton_usd_techo_leche_polvo_day	1441	834.9	68.5	715.6	985.4	834.5
Precio_ton_usd_techo_maiz_day	1441	49.7	9.9	37.2	73.0	47.4
ProPBI_DI_Comercio_day	1441	35.9	2.3	30.3	41.0	36.2
TC_interbancario_compra	1441	3.3	0.1	3.2	3.5	3.3
TC_interbancario_venta	1441	3.3	0.1	3.2	3.5	3.3
TC_sbs_compra	1441	3.3	0.1	3.2	3.5	3.3
TC_sbs_venta	1441	3.3	0.1	3.2	3.5	3.3
PBI_DI_sectores_no_primarios_day	1441	35.2	2.3	30.8	41.8	35.0
PBI_DI_sectores_primarios_day	1441	29.9	2.8	24.8	38.5	29.5

Nota: Elaboración propia

Figura 29

Distribución de la demanda por centros de distribución

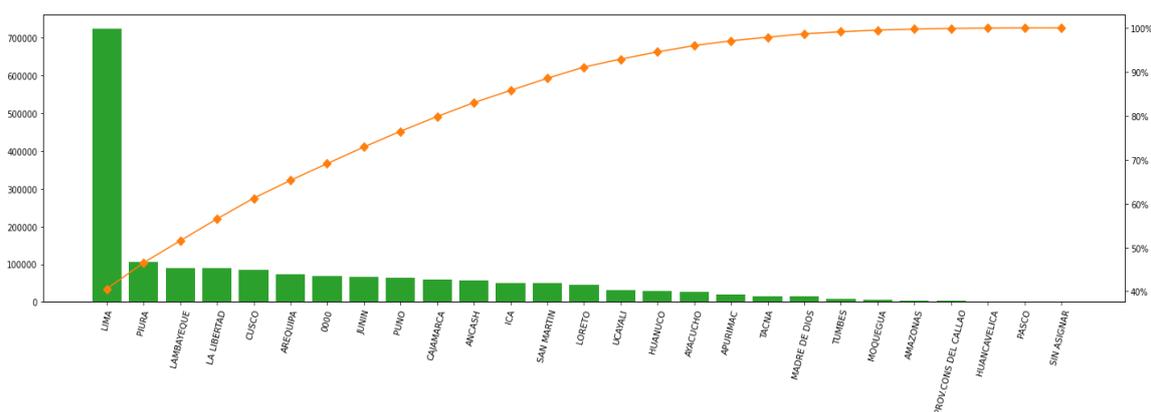


Nota: Elaboración propia

Existen 26 centros de distribución que almacenan el producto terminado en toneladas, el 80% de las ventas se concentra en los 6 de mayor volumen, siendo los más importantes en ventas Ransa y Huachipa con ubicación en Lima, estos 2 centros venden el 50% de todo el volumen, por lo cual amerita realizar un análisis especial.

Figura 30

Distribución de ventas por departamentos del Perú



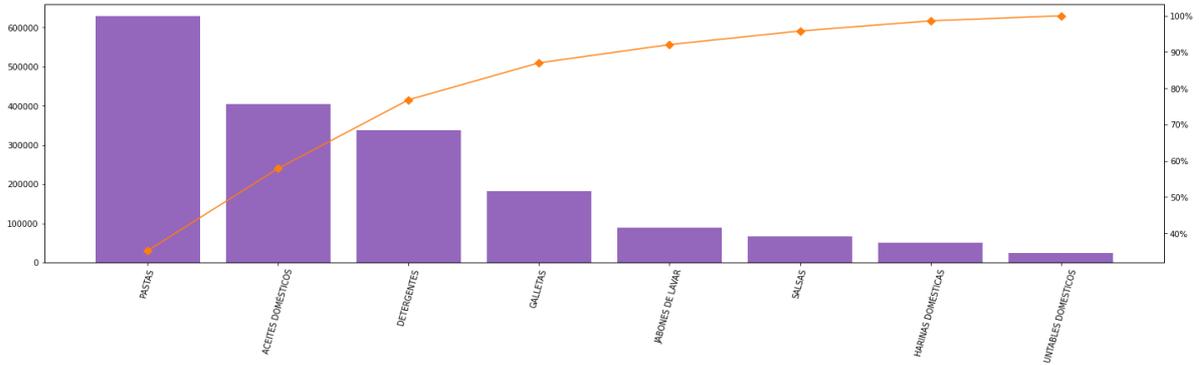
Nota: Elaboración propia

Cuando analizamos las zonas de ubicación de los centros de distribución, el 80% de las toneladas que venden se concentra en 10 ciudades del Perú, siendo Lima el principal

departamento para almacenar y distribuir los productos seguida de Piura. Lo cual nos da indicios de realizar un análisis más profundo en el departamento de Lima.

Figura 31

Distribución de ventas por categorías

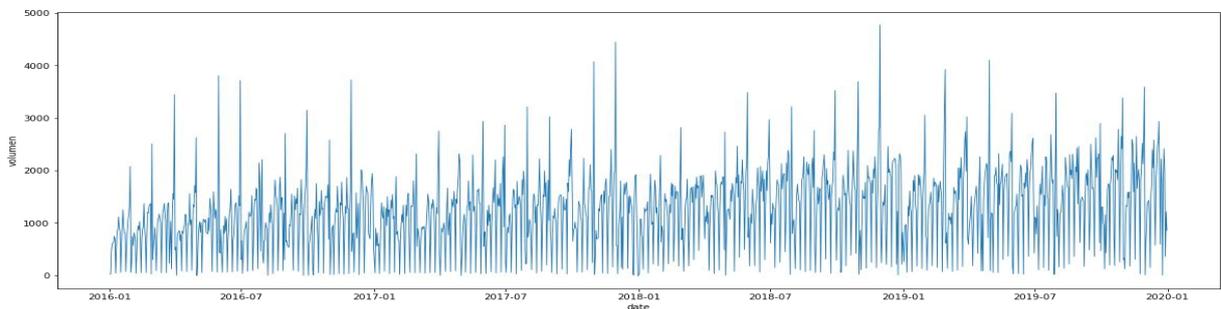


Nota: Elaboración propia

La empresa de consumo masivo tiene en su portafolio 30 categorías, de los cuales el 90% de sus ventas se concentra en 8 categorías de la figura 31, los análisis sucesivos se desarrollarán en base a ese universo de categorías. Resalta la categoría de pastas, aceites domésticos y detergentes quienes concentran el top 3 de la venta en toneladas.

Figura 32

Evolutivo de las ventas en toneladas de la empresa de consumo masivo



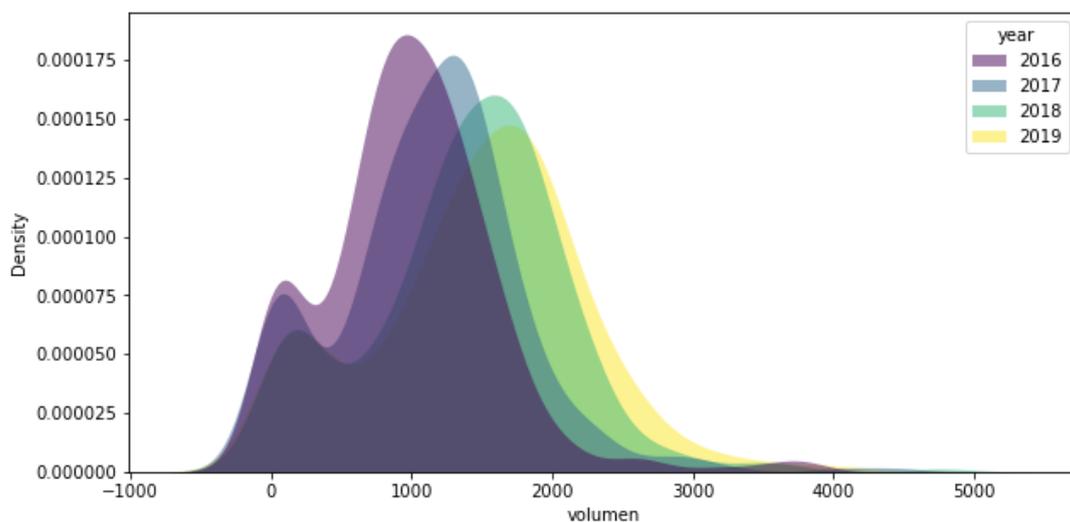
La figura 32 muestra la evolución de las ventas en toneladas para las 8 categorías principales y para los 10 centros de distribución. La evolución de las ventas ha sido creciente desde el año 2016, con picos de venta marcados por la coyuntura de las

celebraciones en Perú, adicionalmente se puede apreciar que la venta es cíclica lo cual nos da indicios de que si es posible estimar un modelo de pronóstico de demanda y que podría ser estable.

Posteriormente se tendría que realizar un análisis de descomposición aditiva de la serie de tiempo de forma general y a nivel de centros de distribución y categorías, este método nos dará algunas hipótesis las cuales se podrán probar al momento de ver los resultados del modelamiento predictivo.

Figura 33

Distribución de las ventas en toneladas por año

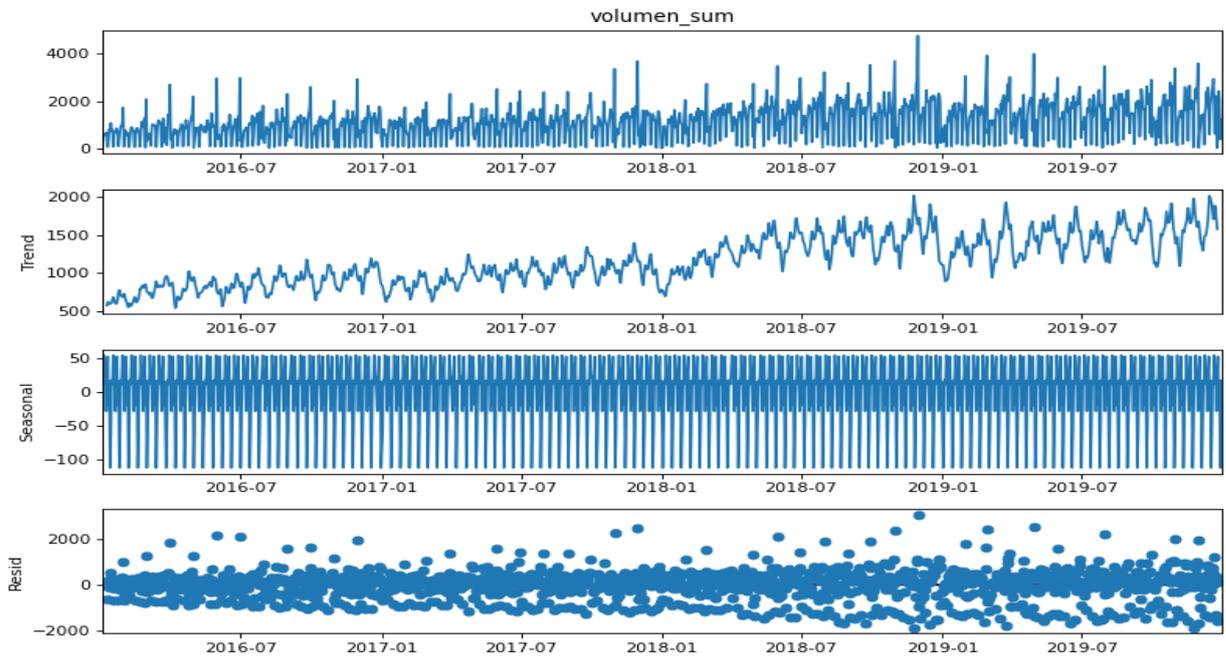


Nota: Elaboración propia

En la figura 33 observamos que mientras los años pasen, la cantidad de ventas promedio se incrementa, lo cual también nos lo confirmó el gráfico de tendencia. Adicionalmente existe un patrón de distribución que se mantiene a lo largo de todos los años, lo cual hace posible crear un modelo estructurado que pueda capturar los efectos de la media y varianza en el tiempo y se pueda realizar una predicción futura para los siguientes periodos.

Figura 34

Descomposición aditiva de la serie volumen de venta



Nota: Elaboración propia

En la figura 34, se muestra la descomposición aditiva del evolutivo de ventas en toneladas para las 8 categorías y 10 centros de distribución:

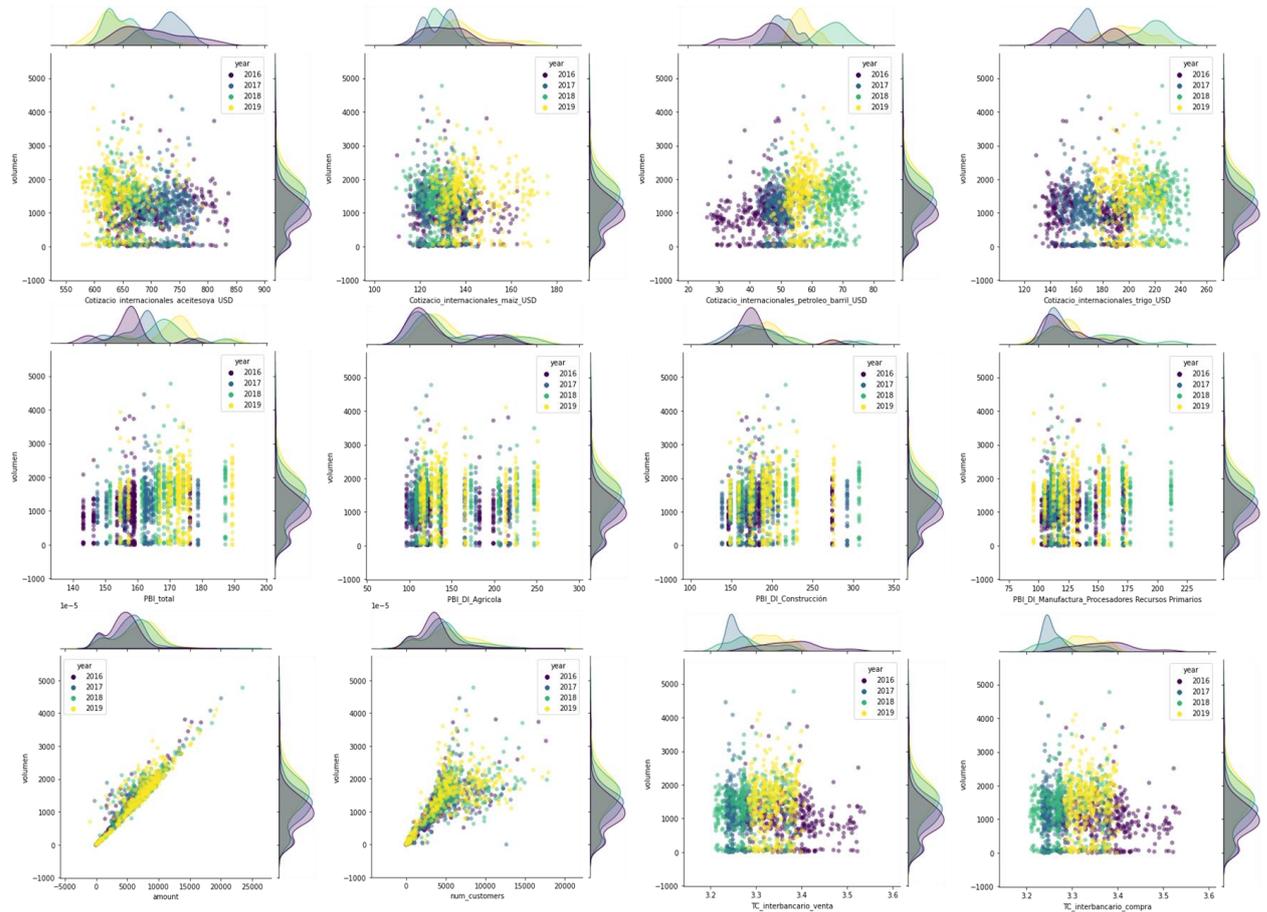
Tendencia: Consiste en la evolución a largo plazo de la serie, claramente la tendencia de la serie tiene un componente creciente a lo largo de los años, este crecimiento es impulsado internamente por la demanda de las categorías y los centros de distribución de forma individual.

Estacionalidad: Está formada por fluctuaciones alrededor de la tendencia que se repiten de forma más o menos periódica y de amplitud superior al año. La serie claramente presenta una estacionalidad la venta generalmente tiene una tendencia baja a inicio del mes y a fin de mes una tendencia alta, adicionalmente tiene algunas características marcadas como festividad de año nuevo donde la venta es 0.

Residuo: Se comporta de manera aleatoria a partir del año 2019, antes de ese periodo presenta intermitencia lo cual se puede dar por coyuntura del país o el consumo interno.

Figura 35

Gráficos de dispersión de la venta versus las covariables



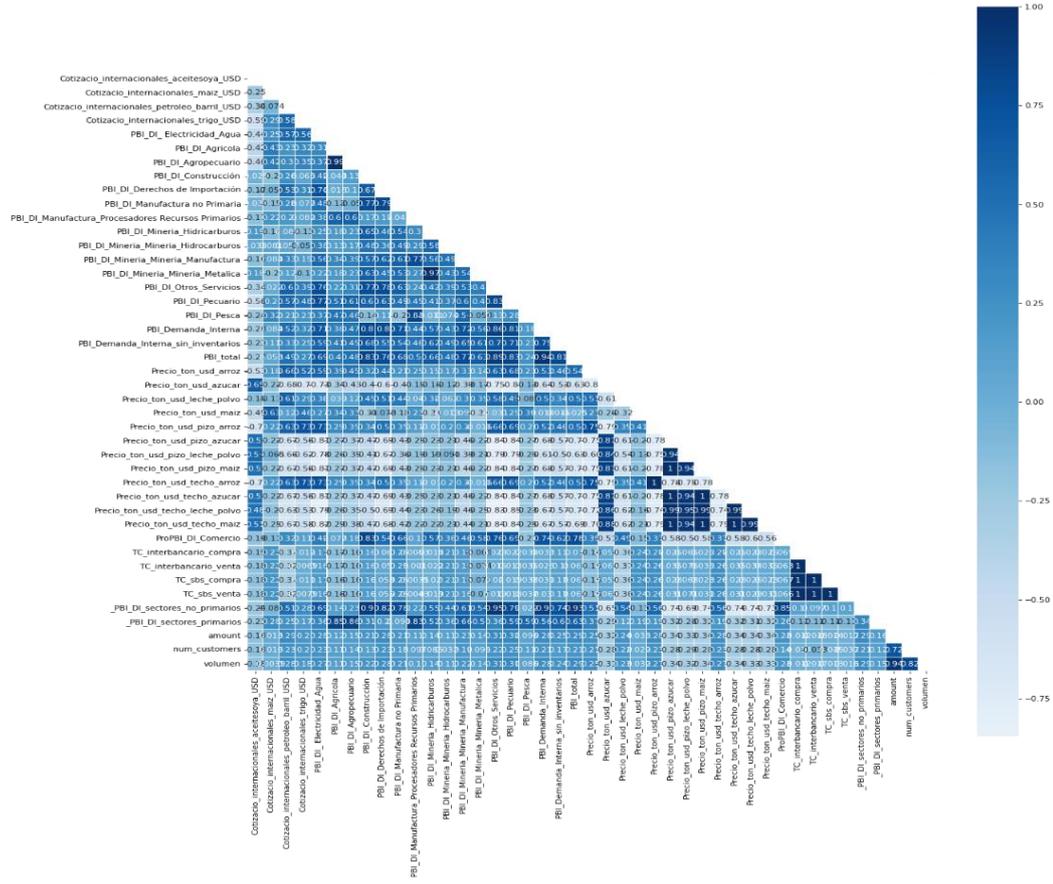
Nota: Elaboración propia

La figura 35 muestra cómo se relaciona la variable ventas en toneladas con las demás variables, se ha dividido por años para encontrar patrones de correlación y dispersión. Lo más destacable es la relación cuasi lineal que existe entre la venta en toneladas, la cantidad de clientes y el monto vendido. Por otro lado, las cotizaciones internacionales y los precios de tipo de cambio tienen una relación aleatoria con las ventas en toneladas y no hay un patrón entre los años. Respecto al PBI del Perú versus las ventas, la relación no es

tan clara, para descubrir una relación cuantitativa y no visual, se presentará una matriz de correlaciones que se muestra a continuación.

Figura 36

Mapa de calor de las correlaciones de las variables cuantitativas



Nota: Elaboración propia

El gráfico de correlación mostrado en la figura 36 nos muestra en un panorama completo la relación de todas las variables que logramos consolidar en una base de datos ya estructurada. Lo que más resalta es que las variables que explican más a nuestra variable dependiente es la venta en monto y el número de clientes, las otras covariables tienen poca correlación entre un 0.3 a 0.4, por lo cual una primera hipótesis podría ser que las variables externas no están afectando a la venta en toneladas directamente; una opción válida y que probaremos en la sección de resultados de modelos es desarrollar un

modelo puro, sin covariables externas es decir solo tomando como regresores a modificaciones de la serie.

4.1.2 Tratamiento y construcción de las variables

Según la metodología CRISP-DM antes de iniciar con el modelamiento predictivo se tiene que realizar una construcción o transformación de variables, las cuales serán las adecuadas para realizar el pronóstico, a continuación, se muestran los cambios que se realizaron en las variables y el grupo de variables finales:

- Corrección de nombres de las variables centros de distribución y de Categorías.
- Selección de los centros de distribución que son constantes en las ventas a lo largo de todo el periodo analizado.
- Las ventas en toneladas que tienen valor negativo se deben a notas de crédito realizadas en una fecha, centro y categoría específica serán completadas con valor 0.
- Construcción de una llave única considerando centro, fecha y producto.
- Para las bases de datos que no están a nivel de día o fecha, se replicó el valor mensual en la casilla de días, esto para lograr una unión adecuada entre las bases de datos.
- Se consolidan todas las bases de datos citadas al inicio de la sección 4.1.1, con la unión se construye una sola base de datos limpia con la cual se puede construir el modelo de pronóstico de demanda.

4.1.3 Entrenamiento del modelo

En esta sección mostraremos los resultados del entrenamiento de los modelos de redes neuronales con la arquitectura LSTM y GRU. Para lo cual se han desarrollado 3 tipos de

experimentos para compararlos y ver el rendimiento predictivo que tienen cada uno de ellos:

- Experimento 2: Comparación de arquitecturas a nivel de día y centro de distribución.
- Experimento 3: Comparación de arquitectura a nivel de día y categoría.

Una vez seleccionado el modelo y el experimento ganador, se desarrollará un análisis de extrapolación para llegar a la predicción de toneladas por SKU, el cual se tendrá que comparar con los resultados del modelo implementado por la empresa de consumo masivo.

Resultados del experimento 1: GRU general

A continuación, se muestran los resultados de la arquitectura GRU y los parámetros usados en cada capa de la red neuronal recurrente. Se ha logrado utilizar un total de 176,321 parámetros. La validación cruzada se desarrolló utilizando 300 épocas, con un *batch size* de 32 y una *validation split* de 0.2, el indicador de *loss* utilizado es el error absoluto medio, según la literatura observada es el más usado en los problemas de pronóstico de demanda. Observamos en la figura 38 que el indicador *loss validation* se llega a estabilizar alrededor de 0.08 a partir de la época 100, mientras que el valor del *training loss* se estabiliza alrededor de las 200 a 300 épocas.

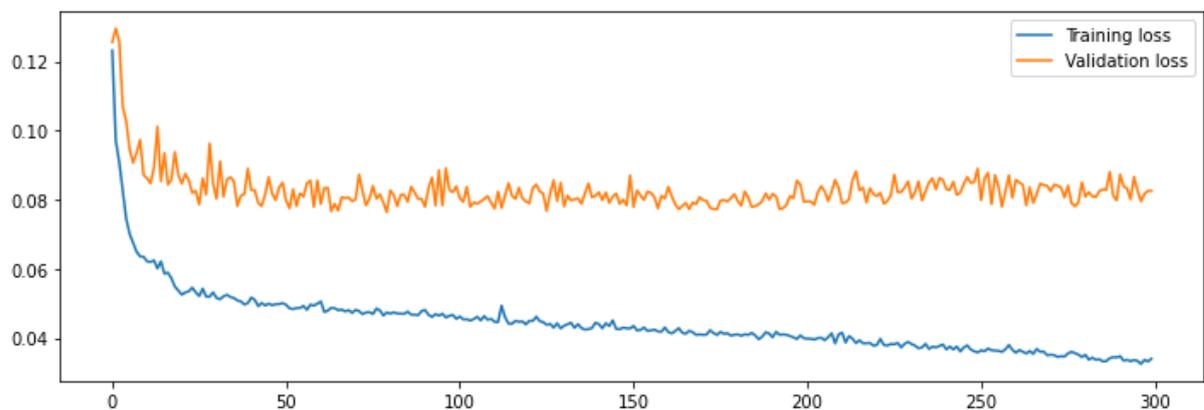
Figura 37

Parámetros usados en arquitectura GRU

```
Model: "GRU-Model"
-----
Layer (type)                Output Shape                Param #
-----
gru (GRU)                    (None, 7, 64)              14592
dropout (Dropout)           (None, 7, 64)              0
Hidden-LSTM-Encoder-Layer ( (None, 128)                49920
Bidirectional)
Repeat-Vector-Layer (Repeat (None, 1, 128)             0
Vector)
Hidden-LSTM-Decoder-Layer ( (None, 1, 128)             74496
Bidirectional)
dropout_1 (Dropout)         (None, 1, 128)             0
gru_3 (GRU)                  (None, 1, 64)              37248
Output-Layer (TimeDistribut (None, 1, 1)               65
ed)
-----
Total params: 176,321
Trainable params: 176,321
Non-trainable params: 0
```

Figura 38

Funciones de perdida para entrenamiento y validación arquitectura GRU general

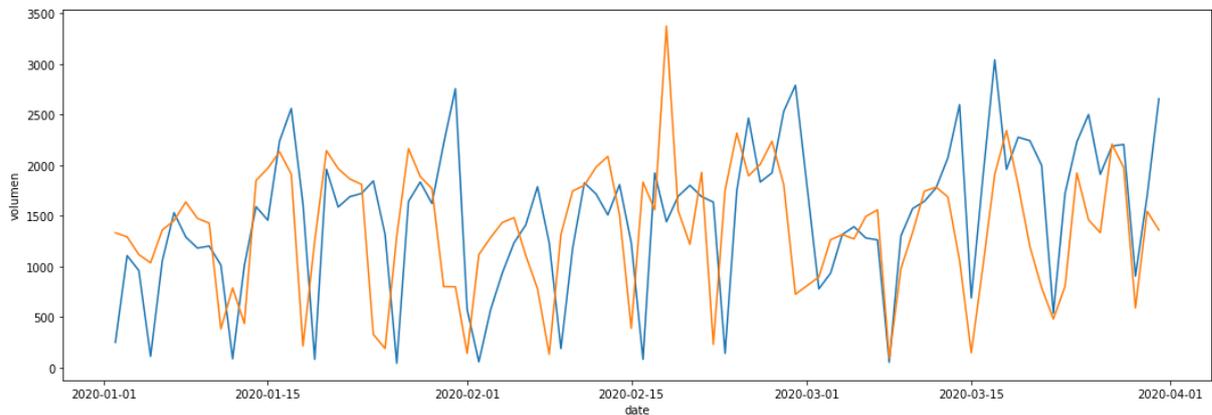


Luego del entrenamiento del modelo con arquitectura GRU, procedemos a desarrollar el pronóstico en los datos de prueba, que comprende el periodo del enero a febrero 2020, 90 días específicamente. Con lo cual el error absoluto obtenido es de 509.47 toneladas y un MAPE de 166% esto calculado de forma diaria, esto se da por la volatilidad que hay en la demanda a nivel diaria de forma general. Procedemos a calcular el MAPE a nivel mensual, tal y como lo calcula la empresa de consumo masivo en su gestión, para este

caso el resultado nos da Enero (1.2%) y Febrero (4.2%) , es lógico que la predicción tenga un error más alto mientras más nos alejemos del último mes con resultados reales.

Figura 39

Volumen real vs estimado para arquitectura GRU general



Resultados del experimento 1: LSTM general

A continuación, se muestran los resultados de la arquitectura LSTM y los parámetros usados en cada capa de la red neuronal recurrente. Se ha logrado utilizar un total de 233,537 parámetros. La validación cruzada se desarrolló utilizando 300 épocas, con un *batch size* de 32 y una *validation split* de 0.2, el indicador de *loss* utilizado es el error absoluto medio, según la literatura observada es el más usado en los problemas de pronóstico de demanda. Observamos en la figura 38 que el indicador *loss validation* se llega a estabilizar alrededor de 0.08 a partir de la época 100, mientras que el valor del *training loss* se estabiliza alrededor de las 200 a 300 épocas.

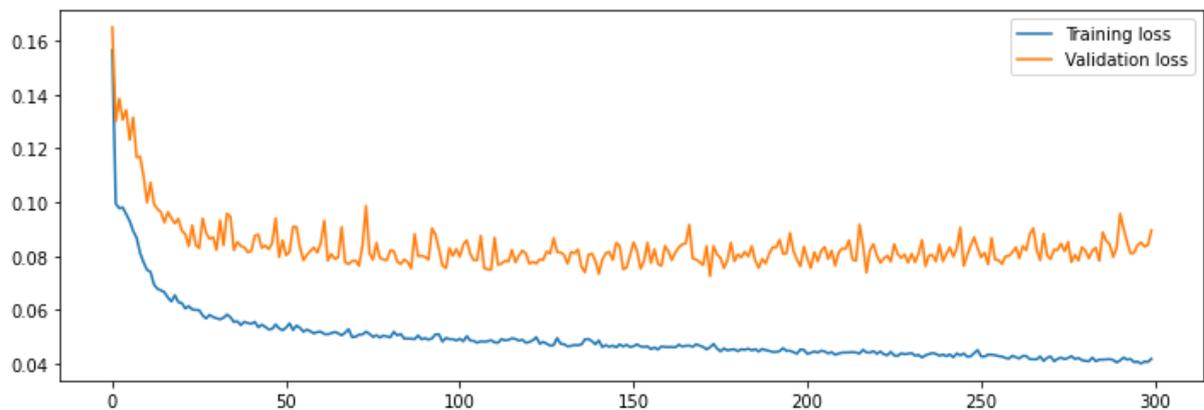
Figura 40

Parámetros usados en arquitectura LSTM

```
Model: "LSTM-Model"
-----
Layer (type)                Output Shape                Param #
-----
lstm (LSTM)                  (None, 7, 64)              19200
dropout (Dropout)           (None, 7, 64)              0
Hidden-LSTM-Encoder-Layer ( (None, 128)                66048
Bidirectional)
Repeat-Vector-Layer (Repeat (None, 1, 128)            0
Vector)
Hidden-LSTM-Decoder-Layer ( (None, 1, 128)            98816
Bidirectional)
dropout_1 (Dropout)         (None, 1, 128)            0
lstm_3 (LSTM)                (None, 1, 64)              49408
Output-Layer (TimeDistribut (None, 1, 1)              65
ed)
-----
Total params: 233,537
Trainable params: 233,537
Non-trainable params: 0
```

Figura 41

Funciones de perdida para entrenamiento y validación arquitectura LSTM

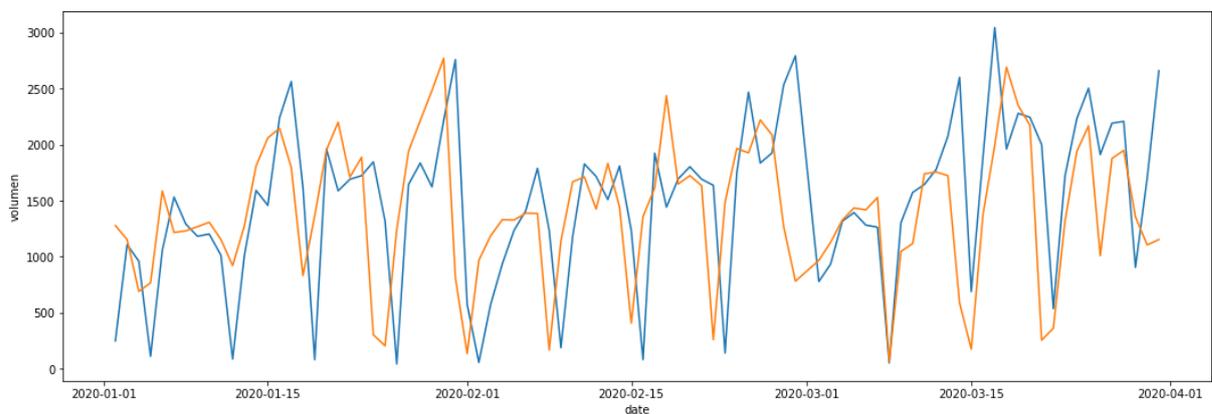


Luego del entrenamiento del modelo con arquitectura LSTM, procedemos a desarrollar el pronóstico en los datos de test, que comprende el periodo del enero a febrero 2020, 90 días específicamente. Con lo cual el error absoluto obtenido es de 540.37 toneladas y un MAPE de 48.8% esto calculado de forma diaria, esto se da por la volatilidad que hay en

la demanda a nivel diaria de forma general. Procedemos a calcular el MAPE a nivel mensual, tal y como lo calcula la empresa de consumo masivo en su gestión, para este caso el resultado nos da Enero (6.8%) y Febrero (2.7%) es lógico que la predicción tenga un error más alto mientras más nos alejemos del último mes con resultados reales.

Figura 42

Volumen real vs estimado para arquitectura LSTM general



Resultados del experimento 2: GRU y LSTM por centro

Se muestran los resultados de las arquitecturas GRU y LSTM para los parámetros usados en cada capa de la red neuronal recurrente. En la tabla 8 se detalla el resultado de los MAPE para cada uno de los modelos por centro de distribución. Considerando los 60 días (enero a febrero 2020) para predicción, el modelo que tiene el MAPE más bajo es el aplicado a Cusco (LSTM: 43% y GRU:45%) y el más alto es COPSA (LSTM: 121% y GRU:181%), con estos resultados preliminares muestran que el modelo LSTM es más eficiente en predicción.

Tabla 8*Resultados del MAPE por centro de distribución*

Centro de Distribución	LSTM	GRU
ALICORP BSF PUNTA HERMOSA	71%	87%
CD CENTRAL & COPSA	121%	181%
CD HUACHIPA	14%	87%
PIURA	87%	73%
RANSA	76%	83%
SUCURSAL AREQUIPA	82%	75%
SUCURSAL CHICLAYO	60%	55%
SUCURSAL CUSCO	45%	43%
SUCURSAL PUCALLPA	593%	391%
SUCURSAL TRUJILLO CM	42%	50%

Nota: elaboración propia

Resultados del experimento 3: GRU y LSTM por categoría

Se muestran los resultados de la arquitectura LSTM y GRU por categoría de producto, los parámetros usados en cada capa de la red neuronal recurrente. En la tabla 9 se detalla el resultado de los MAPE para cada uno de los modelos por categoría de producto. Considerando los 60 días (enero a febrero 2020) para predicción, el modelo que tiene el MAPE más bajo es el aplicado a Detergentes (LSTM: 87% y GRU: 87%) y el más alto es Aceite Doméstico (LSTM: 324% y GRU: 448%), con estos resultados preliminares muestran que el modelo LSTM es más eficiente en predicción.

Tabla 9*Resultados del MAPE por centro de distribución*

Categoría	LSTM	GRU
Aceites domésticos	324%	448%
Salsas	196%	211%
Detergentes	85%	87%
Pastas	149%	186%
Galletas	203%	181%
Harinas Domésticas	300%	326%
Jabones de lavar	284%	413%
Untables domésticos	256%	257%

Nota: elaboración propia

Resumen de los 3 experimentos

Se muestran un resumen de los resultados de las 2 arquitecturas entrenadas (GRU y LSTM). La evaluación se desarrolló para el periodo de predicción de 60 días (enero a febrero 2020). En general para los 3 experimentos y para los 3 meses de evaluación se obtiene un MAPE menor para el modelo LSTM, con lo cual podemos afirmar que este modelo con la arquitectura planteada se ajusta mejor a los datos de la empresa de consumo masivo.

Tabla 10

Resumen de resultados del MAPE para los 3 experimentos por mes

3 experimentos	MAPE de los modelos RNN		
	Mes	LSTM	GRU
General	Ene-20	1.2%	7.3%
	Feb-20	2.6%	4.4%
Centro	Ene-20	2.3%	5.3%
	Feb-20	1.3%	3.2%
Categoría	Ene-20	1.2%	4.2%
	Feb-20	2.3%	8.4%

Nota: elaboración propia

En general para los 3 experimentos y evaluación se obtiene un MAPE acumulado menor para el modelo LSTM, con lo cual podemos afirmar que este modelo con la arquitectura planteada se ajusta mejor a los datos de la empresa de consumo masivo.

Tabla 11

Resumen de resultados del MAPE para los 3 experimentos

3 experimentos	MAPE de los modelos RNN	
	LSTM	GRU
General	5.8%	7.5%
Centro	3.7%	6.2%
Categoría	3.4%	7.6%

Nota: elaboración propia

4.2 Análisis de resultados o discusión de resultados

La investigación demuestra un enfoque aplicado a técnicas de redes neuronales recurrentes para el pronóstico de demanda de una empresa de consumo masivo. Debido al volumen de la información contamos con 115,280 registros de combinaciones únicas a nivel de día, centro, categoría y las múltiples fuentes de información, 43 variables iniciales para desarrollar el pronóstico; elegimos las técnicas de LSTM y GRU.

Encontramos que existen 26 centros de distribución que almacenan el producto terminado medido en toneladas y el 80% de las ventas se concentra en los 6 centros de mayor volumen, siendo los más importantes en ventas Ransa, Huachipa con ubicación en Lima, estos 2 centros venden el 50% de todo el volumen. Similar análisis desarrollamos con las categorías de los cuales el 90% de sus ventas se concentra en 8 categorías, resaltan las categorías de pastas, aceites domésticos y detergentes quienes concentran el top 3 de la venta en toneladas. Por otro lado, cuando analizamos las zonas de ubicación de los centros de distribución, el 80% de las toneladas que venden se concentra en 10 ciudades del Perú, siendo Lima el principal departamento para almacenar y distribuir los productos seguida de Piura.

La evolución de las ventas en toneladas ha sido creciente desde el año 2016, con picos de venta marcados por los días festivos en Perú, un claro ejemplo son las épocas de año nuevo donde las ventas son cercanas a 0.

Se desarrolló un análisis de descomposición aditiva para la serie de ventas en toneladas que comprende las 8 categorías y 26 centros de distribución, encontrando resultados favorables en la tendencia y estacionalidad que nos permiten modelar la serie de forma adecuada.

Se desarrolló experimentos dividido en 3 casos para validar las hipótesis planteadas al inicio de la investigación. Para el caso general se desarrollaron 2 modelos, para el caso por centro de distribución se desarrollaron 20 modelos y para el caso de categorías se desarrollaron 10 modelos.

En el experimento 1, también llamado el caso general comparamos el MAPE resultante de las arquitecturas LSTM y GRU, evaluados en el periodo de prueba 60 días (enero a febrero 2020). En los resultados mensuales para los meses de enero y febrero quien tuvo un mejor desempeño fue el modelo LSTM. Adicionalmente notamos que el MAPE es menor en el mes de enero (1.2%) y va creciendo mientras transcurren los meses, dado que es más complicado para el modelo predecir meses más alejados.

En el experimento 2, también llamado el caso por centro de distribución comparamos el MAPE resultante de las arquitecturas LSTM y GRU, evaluados en el periodo de prueba 60 días (enero a febrero 2020). En este caso entrenaron 10 modelos para la arquitectura LST y 10 para GRU, dado que en el análisis descriptivo existían una alta variabilidad por centro decidimos separarlo; para este caso podemos notar que existen centros con un alto MAPE y otros con bajo, lo cual nos da indicios de que hay centros con venta más estable, pero para tomar una decisión concluyente y asegurar la estabilidad

de las predicciones se tomó un acumulado promedio por centro con lo cual nos da un MAPE de LSTM(3.7%) y GRU(6.2%), la misma tendencia se evidencia cuando se desarrolla la apertura por mes. Se puede concluir que para el experimento por centro la arquitectura LSTM tiene un mejor desempeño predictivo.

En el experimento 3, también llamado el caso por categorías de productos comparamos el MAPE resultante de las arquitecturas LSTM y GRU, evaluados en el periodo de prueba 60 días (enero a febrero 2020). En este caso se entrenaron 8 modelos para la arquitectura LSTM y 8 para GRU, dado que en el análisis descriptivo existían una alta variabilidad por categorías decidimos separarlo; para este caso podemos notar que existen categorías con un alto MAPE y otros con bajo, lo cual nos da indicios de que hay categorías con venta más estable, pero para tomar una decisión concluyente y asegurar la estabilidad de las predicciones se tomó un acumulado promedio por categoría con lo cual nos da un MAPE de LSTM(3.4%) y GRU(7.6%), la misma tendencia se evidencia cuando se desarrolla la apertura por mes. Se puede concluir que para el experimento por categoría la arquitectura LSTM tiene un mejor desempeño predictivo.

Finalmente se realizó una comparación de los 3 experimentos realizados para esta investigación en todos los casos nos muestra que el modelo LSTM tiene un MAPE menor que el modelo GRU, por lo cual podemos concluir que la arquitectura LSTM se ajusta mejor a la predicción de la demanda en toneladas para la empresa de consumo masivo.

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

Se ha desarrollado un modelo de redes neuronales artificiales de tipo GRU y LSTM para el pronóstico de la demanda de stock, la predicción se está haciendo con un enfoque de regresión y se están graficando los resultados en líneas de tiempo, para darle una vista de series temporales. Adicionalmente se concluye que el modelo es útil solo en etapas donde no hay pandemia o catástrofes mundiales, es decir donde no hay crisis como el caso del COVID-19.

El 80% de las ventas se concentra en los 6 centros de mayor volumen, siendo los más importantes en ventas Ransa, Huachipa con ubicación en Lima, estos 2 centros venden el 50% de todo el volumen. El 90% de las ventas de las categorías se concentra en 8, resaltan las categorías de pastas, aceites domésticos y detergentes. El 80% de las ventas se concentra en 10 ciudades del Perú, siendo Lima el principal departamento para almacenar y distribuir los productos seguida de Piura.

Contamos con un total de 43 variables, 3 corresponden a información transaccional, 16 a información de cotizaciones internacionales, 20 a Producto bruto interno por sector y 4 a información de tipo de cambio. Para la selección de variables que ingresan a los modelos de redes neuronales se aplicó un *Random Forest* usando su técnica de importancia de variables.

Se realizaron 3 experimentos para esta investigación, el primero planteando de manera general, el segundo desarrollado para centros y el ultimo para categorías, en todos los casos nos muestra que el modelo LSTM tiene un MAPE menor que el modelo GRU, por lo cual podemos concluir que la arquitectura LSTM se ajusta mejor a la predicción de la demanda en toneladas para la empresa de consumo masivo.

El MAPE resultante del modelo actualmente implementado en la empresa de consumo masivo es de 20%, por lo cual podemos concluir que los modelos aplicados en esta investigación tienen un mejor desempeño obteniendo un MAPE promedio global en los datos test que separamos (enero 2020 a febrero 2020) de LSTM (5.8%) y GRU (7.5%). Esto también lo confirma el MAPE en el proceso de entrenamiento MAPE promedio global de LSTM (4.9%), GRU (6.4%). Podemos confirmar lo mencionado para el caso de centros LSTM (3.3%), GRU (5.6%) y categorías LSTM (2.4%) y GRU (5.3%).

Y finalmente podemos concluir que las variables que más tienen impacto en la predicción son aquellas que están orientadas a la cadena de suministro interna, el valor de las variables externas no tiene un aporte significativo para el pronóstico de la demanda.

Tabla 12

Resumen de la importancia de variable, resultado de random forest

Variable	Importancia
Ventas	0.4628
numero_clientes	0.3025
PBI_DI_Electricidad_Agua_day	0.2021
PBI_DI_Manufactura no Primaria_day	0.0921
PBI_Demanda_Interna_day	0.0622
PBI_DI_Derechos de Importación_day	0.0409
PBI_total_day	0.0406
_PBI_DI_sectores_no_primarios_day	0.0335
PBI_DI_Pecuario_day	0.0315
Precio_ton_usd_pizo_leche_polvo_day	0.0273
PBI_DI_Otros_Servicios_day	0.0112
Precio_ton_usd_maiz_day	0.0110
PBI_DI_Mineria_Hidricarburos_day	0.0107
Precio_ton_usd_techo_leche_polvo_day	0.0098
PBI_DI_Mineria_Mineria_Metalica_day	0.0053
Precio_ton_usd_techo_arroz_day	0.0033
PBI_DI_Mineria_Mineria_Hidrocarburos_day	0.0033
PBI_DI_Mineria_Mineria_Manufactura_day	0.0029
ProPBI_DI_Comercio_day	0.0028
Precio_ton_usd_pizo_arroz_day	0.0022
Precio_ton_usd_arroz_day	0.0017
amount	0.0011
Precio_ton_usd_leche_polvo_day	0.0009

Nota: elaboración propia

5.2 Recomendaciones

Recomendamos actualizar los modelos con arquitectura LSTM y GRU tomando información reciente, es decir periodos de febrero 2020 en adelante. El objetivo sería probar cómo se comportan los modelos en los periodos donde está presente la pandemia.

Recomendamos probar técnicas de machine learning para retar a las arquitecturas LSTM y GRU, según la literatura métodos como XgBoost y LigthGBM también son eficientes para pronosticar la demanda de productos.

Recomendamos agregar más información interna del propio negocio, como ya se logró probar las variables internas ayudan a explicar mejor el comportamiento de la demanda. Internamente existe variables relacionado a la producción, quiebres, almacenes y transporte.

Recomendamos explorar los recursos para implementar el modelo, dado que para llevarlo a producción se necesitamos capacidades de tecnología y bigdata, dado que se a demostrado que el modelo tiene un mejor MAPE que el modelo actualmente implementado por la empresa de consumo masivo.

REFERENCIAS BIBLIOGRÁFICAS

Arunraj, N., & Ahrens, D. (2015). hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *International Journal of Production Economics* 170, 321-335.

- Ashwini Rekha, B., & Vijaykumar, B. (2019). Retail Giant Sales Forecasting using Machine Learning. *International Journal of Recent Technology and Engineering (IJRTE) Volume-8, Issue-2S11*.
- Barnett, W. A. (1988). Cuatro pasos para pronosticar la demanda total del mercado. *Harvard Business Review* , 1.
- Bell, J. (2020). *Hands-On for Developers and Technical Professionals*. Indianapolis, Indiana: John Wiley & Sons.
- Bengio, Y., Simard, Y., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE TRANSACTIONS ON NEURAL NETWORK*, vol. 5, n° 2.
- Buteikis, A. (19 de 08 de 2019). *03 Time series with trend and seasonality components*. Obtenido de <http://web.vu.lt/mif/a.buteikis/>
- Calin, O. (2020). *Deep Learning Architectures a Mathematical Approach*. Michigan: Springer Series in the Data Sciences.
- Cárdenas Garro, J. (2015). Pronósticos y comparación de una serie de tiempo con cambios estructurales mediante la red neuronal artificial de retropropagación resiliente y modelos no lineales. *Universidad Nacional Mayor de San Marcos - Facultad de ciencia matemática y estadística*.
- Catalan Oliva, C. (2012). *PRONÓSTICOS EN LOS NEGOCIOS*. El Salvador: Facultad de ciencias economicas, Maestria en administracion financiera.
- Catalan, C. (2010). *Series temporales* . Cataluña: Tesis.
- Chase, R., Jacobs, R., & Aquilano, N. (2009). *Administración de Operaciones* . Mexico: McGRAW-HILL.

- Chau de la Cruz, F., & Montero, C. (2018). *Desarrollo de algoritmos de pronóstico de demanda y planificación de compras para reducir excedentes y faltantes en la gestión de Stocks de un RETAIL farmacéutico*. Lima : Tesis universidad nacional de ingeniería.
- Cho, K., Bart van , M., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. EEUU - New York: Cornell University Coumputer Ciences -Computation and Language.
- Corsten, D., & Gruen, T. (2016). On Shelf Availability: An Examination of the Extent, the Causes, and the Efforts to Address Retail Out-of-Stocks. *Consumer Driven Electronic Transformation*, 131-149.
- Coug Pham, C. (2019). Developing a Demand-driven Forecasting in Purchasing. *Technology, communication and transport Degree program in Logistics Engineering*.
- Cuong Pham, C. (2019). Developing a Demand-driven Forecasting in Purchasing. *Technology, communication and transport Degree program in Logistics Engineering*.
- Dalkey, N. (1969). The Delphi Metod: An experiment study of group opinion. *Rand*, 14-17.
- Di Pillo, G. (2013). An application of learning machines to sales forecasting under promotions. *Tech. rep. Sapienza, Universiteta di Roma*.
- Fernie, J., & Sparks, L. (2019). *Logistic and Retail Managment*. Great Britain: Kogan Page Limited.

- Fildes, R., Ma, S., & Kolassa, S. (2018). Retail forecasting: research and practice. *Lancaster University Management School & Nanjing Audit University.*
- Gil Zavaleta, E., & Rodríguez Collas, E. (2010). Sistema de pronóstico de la demanda de productos farmacéuticos basado en redes neuronales. *Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática .*
- Goodwin, B., Ford, R., & Chvosta, J. (2018). *Applied Econometrics with SAS: Modeling Demand, Supply, and Risk.* New York: SAS Institute Inc.
- Guzmán Valle, M. (2018). *Aplicación de métodos determinísticos y dinámicos para calcular los pronósticos de la demanda en la cadena de suministro de la Empresa Chiclayana Sipán Distribuciones S.A.C. utilizando Minitab y Oracle Crystal Bal.* Lambayeque.
- Haykin, S. (2009). *Neural networks and learning machines- 3rd ed.* New Jersey, USA: Prentice Hall.
- Hernandez, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación.* McGraw-Hill Education.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *MIT Press Cambridge, 1735-1780.*
- James A, C. (2014). *Protean Supply Chain ten dynamycs of supply and demand alignment.* New Jersey: John Wiley & Sons.
- Kasun , B., Peibei , S., Christoph , B., & Hansika , H. (2018). Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology. *Faculty of Information Technology, Monash University, Melbourne, Australia.*

- Koottatep, P. (2006). Promotional Forecasting in the Grocery Retail Business. *Master of Engineering in Logistics at the Massachusetts Institute of Technology.*
- Krause-Traudes, M. (2008). Spatial data mining for retail sales forecasting. *Tech. rep. Fraunhofer-Institut Intelligente Analyse- und Informationssysteme.*
- Lanham, M., Gujela, P., Soni, A., & Mupparaju, K. (2017). A Comparative Study of Machine Learning Frameworks for Demand Forecasting. *Purdue University Krannert School of Management.*
- Levis, A., & Papageorgiou, L. (2005). Customer Demand Forecasting via Support Vector Regression Analysis. *Chemical Engineering Research and Design*, 1009-1018.
- M. Ozhegov, E., & Teterina, D. (2018). Ensemble Method for Censored Demand Prediction. *National Research University Higher School of Economics.*
- Martín, J., Emilio, S., & Antonio, S. (20210). *Redes neuronales artificiales*. Valencia: Universidad de Valencia.
- Matich, D. J. (2001). *Redes Neuronales: Conceptos básicos y aplicaciones* . Rosario, Argentina: Universidad Tecnológica Nacional – Facultad Regional Rosario.
- Moskalev, A. (2019). Demand forecasting for fast-moving products in grocery retail. *Master's Thesis Programme in Computer, Communication and Information Sciences.*
- Nahmias, S., & Lennon Olsen, T. (2017). *Production and Operation Analysis*. EEUU: Waveland Press, Inc.
- Núñez Castro, J. (2017). *Aprendizaje automático en fusión nuclear con Deep Learning*. Valparaiso: Pontificia universidad católica de Valparaiso.

- Pavlyshenko, B. (2018). Machine-Learning Models for Sales Time Series Forecasting. *IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, 21-25.
- Perez, J. (2019). *Investigacion Estadistica*. Lima: Planeta.
- Pilinkienė, V. (2019). Selection of Market Demand Forecast Methods: Criteria and Application. *ISSN 1392-2785 ENGINEERING ECONOMICS*. 2008. No No 3 (58), 19-20.
- Ponce Cruz, P. (2010). *Inteligencia Artificial con Apicaciones en la Ingenieria*. Bogota: Alfaomega Grupo Editor.
- Prabhakar, V., & Sayiner, D. (2018). Demand forecasting for a large grocery chain in Ecuador. *Purdue University, Department of Management*.
- Prabhudesai, K., Valsaraj, V., Woo, D., Yi, J., & Baldrige, R. (2020). Using Machine Learning and Demand Sensing to Enhance. *SAS4405-2020*, 1-4.
- Quinto, B. (2020). *Next-Generation Machine Learning with Spark*. Carson, CA, USA: Apress.
- Ramasubramanian, K., & Singh, A. (2019). *Machine Learning using R With Time Series and Industry-based Use Cases in R*. New Delhi, India: Apress.
- Ramesh Vasudev, S. (2018). *Demand forecasting using statistical and machine learning algorithms*. Dublin: Dublin Business School.
- Rosa, A. C.-d. (2003). La salud de adolescentes en cifras. *Salud Publica De Mexico*, 45, 153-166. Recuperado el 6 de 7 de 2020, de http://scielo.org.mx/scielo.php?script=sci_arttext&pid=s0036-36342003000700019

- Rusell, S., & Norving, P. (2010). *Artificial Intelligence a Modern Approach*. New Jersey: Pearson Education.
- Sandivar Rosas, J. (2008). *Pronóstico de la demanda de postulantes de un centro superior universitario, basado en redes neuronales*. Lima: Universidad Nacional de Ingeniería Tesis Maestría.
- Siwerz, R., & Dahlén, C. (2017). *Predicting sales in a food store department using machine learning*. Stockholm: Degree project in Computer Science.
- Thomopoulos, N. (2015). *Demand Forecasting for Inventory Control*. Switzerland: Springer International.
- Wei, W. (2019). *Multivariate Time Series Analysis and Applications*. UK: John Wiley & Sons.
- Witten, I., Mark, A., Hall, M., & Pal, C. (2017). *Data Mining Practical Machine Learning Tools and Techniques*. United States: Morgan Kaufmann is an imprint of Elsevier.

ANEXOS

Anexo 1: Declaración de Autenticidad



UNIVERSIDAD RICARDO PALMA ESCUELA DE POSGRADO

ANEXO N° 1

DECLARACIÓN DE AUTENTICIDAD Y NO PLAGIO

DECLARACIÓN DEL GRADUANDO

Por el presente, el graduando:

Osman Dennys Rafel Taipe

En condición de egresado del Programa de Posgrado:

Maestría en Ciencia de Datos

Deja constancia que ha elaborado la tesis intitulada:

Modelo de Redes Neuronales Artificiales para el pronóstico de demanda de stock en una empresa de consumo masivo

Declara que el presente trabajo de tesis ha sido elaborado por el mismo y no existe plagio/copia de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso o similar) presentado por cualquier persona natural o jurídica ante cualquier institución académica, de investigación, profesional o similar.

Deja constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no ha asumido como suyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o de la internet.

Asimismo, ratifica que es plenamente consciente de todo el contenido de la tesis y asume la responsabilidad de cualquier error u omisión en el documento y es consciente de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, el graduando se somete a lo dispuesto en las normas de la Universidad Ricardo Palma y los dispositivos legales vigentes.

Firma de graduando

02/12/2021

Fecha

Anexo 2: Autorización del consentimiento para realizar la investigación



UNIVERSIDAD RICARDO PALMA ESCUELA DE POSGRADO

ANEXO N° 2



AUTORIZACIÓN DE CONSENTIMIENTO PARA REALIZAR LA INVESTIGACIÓN

DECLARACIÓN DEL RESPONSABLE DEL ÁREA O DEPENDENCIA DONDE SE REALIZARÁ LA
INVESTIGACIÓN

Dejo constancia que el área o dependencia que dirijo ha tomado conocimiento del proyecto de tesis titulado:

Modelo de Redes Neuronales Artificiales para el pronóstico de demanda de stock en una empresa de
consumo masivo

El mismo que es realizado por el señor estudiante

Osman Dennys, Rafael Taipe

En condición de estudiante – investigador del Programa de:

Maestría en Ciencia de Datos

Asimismo, señalamos, que según nuestra normativa interna procederemos con el apoyo al desarrollo del proyecto de investigación, dado las facilidades del caso para aplicación de los instrumentos de recolección de datos.

En razón de lo expresado doy mi consentimiento para el uso de la información y/o la aplicación de los instrumentos de recolección de datos:

Nombre de la empresa: Alicorp SAA	Autorización para el uso del nombre de la Empresa en el Informe Final	SI <input checked="" type="checkbox"/>
--------------------------------------	---	---

Apellidos y nombres del Jefe/Responsable del área: Mariella Aoki Nakazaki	Cargo del Jefe/Responsable del área: Gerente Ejecutivo de Analítica Avanzada
--	---

Teléfono fijo (incluyendo anexo) y/o Celular:	Correo electrónico de la empresa: raokin@alicop.com.pe
---	---

Mariella Aoki Nakazaki

Firma de responsable de área

02/12/2021

Fecha

Anexo3: Matriz de consistencia de la investigación

Título: Modelo de Redes Neuronales Artificiales Recurrentes para el pronóstico de demanda de stock en una empresa de consumo masivo				
Problemas	Objetivos	Hipótesis	VARIABLES	Dimensiones
<p>General: ¿Es posible desarrollar un modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo?</p> <p>Específicos: ¿Cuál es el algoritmo de redes neuronales artificiales recurrentes más eficiente para el pronóstico de demanda de stock y a su vez que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo?</p> <p>¿Es posible definir la topología adecuada de las redes neuronales artificiales recurrentes para el pronóstico de demanda de stock, de tal forma que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo?</p> <p>¿Cuáles son las variables que tienen más influencia en el pronóstico de</p>	<p>General: Desarrollar un modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo.</p> <p>Específicos: Identificar el algoritmo de redes neuronales artificiales recurrentes más eficiente para el pronóstico de la demanda de stock y a su vez que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo.</p> <p>Definir la topología adecuada de las redes neuronales artificiales recurrentes para el pronóstico de demanda de stock, de tal forma que sea más preciso que el modelo actualmente implementado en la empresa de consumo masivo.</p> <p>Identificar las variables que tienen más influencia en el</p>	<p>General: El modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock es más preciso que el modelo actualmente implementado en la empresa de consumo masivo.</p> <p>Específicos: El modelo de Redes Neuronales Artificiales LSTM es más eficiente que GRU en el pronóstico de demanda de stock usando los datos de la empresa de consumo masivo.</p> <p>La topología adecuada de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock es obtener un modelo de pronóstico menor al 10% en error cuadrático medio.</p> <p>Las variables que tienen más influencia en el modelo de redes neuronales artificiales recurrentes para el pronóstico de</p>	<p>Independiente: Modelo de redes neuronales artificiales recurrentes</p> <p>Dependiente: Pronóstico de demanda de stock</p>	<p>Independiente: Redes Neuronales LSTM Redes Neuronales GRU</p> <p>Dependiente: Demanda generada por factores Internos Demanda generada por factores externos</p>

demanda de stock, usando los datos de la empresa de consumo masivo?	modelo de redes neuronales artificiales recurrentes para el pronóstico de demanda de stock, usando los datos de.	stock son las relacionadas a factores internos de la empresa de consumo masivo.
---	--	---

Anexo4: Cuadro de base de datos

Base	Nombre
Transaccional	per_sku_unicos_COMPRA_CMP
Transaccional	per_amount_ECONOMY
Transaccional	per_amount_MAINSTREAM
Transaccional	per_amount_PREMIUM
Transaccional	per_amount_SUPERECONOMY
Transaccional	per_amount_sin_tier
Transaccional	per_sku_count_ECONOMY
Transaccional	per_sku_count_MAINSTREAM
Transaccional	per_sku_count_PREMIUM
Transaccional	per_sku_count_SUPERECONOMY
Transaccional	per_amount_ACEITE A GRANEL
Transaccional	per_amount_ACEITES DOMÉSTICOS
Transaccional	per_amount_HARINAS DOMÉSTICAS
Transaccional	per_amount_HARINAS INDUSTRIALES
Transaccional	per_amount_INSUMOS PANADEROS
Transaccional	per_amount_MANTECAS INDUSTRIALE
Transaccional	per_amount_MARGARINAS INDUSTRIA
Transaccional	per_amount_PRE-MEZCLAS INDUSTRI
Transaccional	per_amount_RELLENOS Y COBERTURA
Transaccional	per_amount_UNTABLES DOMESTICOS
Maestro Productos	Codigo_producto
Maestro Productos	Nombre_Producto
Maestro Productos	Nombre_Familia
Maestro Productos	Nombre_Categoria
Maestro Productos	Codigo_Familia
Maestro Productos	Codigo_Categoria
Maestro Clientes	Codigo_cliente
Maestro Clientes	Nombre_Cliente
Maestro Clientes	Ubicacion_Cliente
Maestro Clientes	Zona_Cliente

Anexo5: Tablas y gráficos que complementan la investigación

Figura 43

Parámetros para el modelo general arquitectura LSTM

```
Model: "LSTM-Model"
-----
Layer (type)                Output Shape                Param #
-----
lstm (LSTM)                  (None, 7, 64)              19200
dropout (Dropout)           (None, 7, 64)              0
Hidden-LSTM-Encoder-Layer (Bidirectional) (None, 128)                66048
Repeat-Vector-Layer (Repeat Vector) (None, 1, 128)             0
Hidden-LSTM-Decoder-Layer (Bidirectional) (None, 1, 128)            98816
dropout_1 (Dropout)         (None, 1, 128)             0
lstm_3 (LSTM)               (None, 1, 64)              49408
Output-Layer (TimeDistributed) (None, 1, 1)               65
-----
Total params: 233,537
Trainable params: 233,537
Non-trainable params: 0
```

Figura 44

Parámetros para el modelo general arquitectura GRU

```
Model: "GRU-Model"
-----
Layer (type)                Output Shape                Param #
-----
gru (GRU)                   (None, 7, 64)              14592
dropout (Dropout)           (None, 7, 64)              0
Hidden-LSTM-Encoder-Layer (Bidirectional) (None, 128)                49920
Repeat-Vector-Layer (Repeat Vector) (None, 1, 128)             0
Hidden-LSTM-Decoder-Layer (Bidirectional) (None, 1, 128)            74496
dropout_1 (Dropout)         (None, 1, 128)             0
gru_3 (GRU)                 (None, 1, 64)              37248
Output-Layer (TimeDistributed) (None, 1, 1)               65
-----
Total params: 176,321
Trainable params: 176,321
Non-trainable params: 0
```

Figura 45

Parámetros para el modelo por centro de distribución arquitectura LSTM

```
Model: "LSTM-Center-Model"
-----
Layer (type)                Output Shape                Param #
-----
lstm_40 (LSTM)               (None, 7, 64)             19200
dropout_20 (Dropout)        (None, 7, 64)             0
Hidden-LSTM-Encoder-Layer ( (None, 128)               66048
Bidirectional)
Repeat-Vector-Layer (Repeat (None, 1, 128)           0
Vector)
Hidden-LSTM-Decoder-Layer ( (None, 1, 128)           98816
Bidirectional)
dropout_21 (Dropout)        (None, 1, 128)           0
lstm_43 (LSTM)              (None, 1, 64)             49408
Output-Layer (TimeDistribut (None, 1, 1)             65
ed)
-----
Total params: 233,537
Trainable params: 233,537
Non-trainable params: 0
```

Figura 46

Parámetros para el modelo por centro de distribución arquitectura GRU

```
Model: "GRU-Center-Model"
-----
Layer (type)                Output Shape                Param #
-----
gru_36 (GRU)                 (None, 7, 64)             14592
dropout_18 (Dropout)        (None, 7, 64)             0
Hidden-LSTM-Encoder-Layer ( (None, 128)               49920
Bidirectional)
Repeat-Vector-Layer (Repeat (None, 1, 128)           0
Vector)
Hidden-LSTM-Decoder-Layer ( (None, 1, 128)           74496
Bidirectional)
dropout_19 (Dropout)        (None, 1, 128)           0
gru_39 (GRU)                 (None, 1, 64)             37248
Output-Layer (TimeDistribut (None, 1, 1)             65
ed)
-----
Total params: 176,321
Trainable params: 176,321
Non-trainable params: 0
```

Figura 47

Parámetros para el modelo por centro de distribución arquitectura LSTM

```
Model: "LSTM-Categoria-Model"
-----
Layer (type)                Output Shape                Param #
-----
lstm_28 (LSTM)              (None, 7, 64)             19200
dropout_14 (Dropout)        (None, 7, 64)             0
Hidden-LSTM-Encoder-Layer (Bidirectional) (None, 128) 66048
Repeat-Vector-Layer (Repeat Vector) (None, 1, 128) 0
Hidden-LSTM-Decoder-Layer (Bidirectional) (None, 1, 128) 98816
dropout_15 (Dropout)        (None, 1, 128)            0
lstm_31 (LSTM)              (None, 1, 64)             49408
Output-Layer (TimeDistributed) (None, 1, 1) 65
-----
Total params: 233,537
Trainable params: 233,537
Non-trainable params: 0
-----
```

Figura 48

Parámetros para el modelo por centro de distribución arquitectura GRU

```
Model: "GRU-Categoria-Model"
-----
Layer (type)                Output Shape                Param #
-----
gru_28 (GRU)                (None, 7, 64)             14592
dropout_14 (Dropout)        (None, 7, 64)             0
Hidden-LSTM-Encoder-Layer (Bidirectional) (None, 128) 49920
Repeat-Vector-Layer (Repeat Vector) (None, 1, 128) 0
Hidden-LSTM-Decoder-Layer (Bidirectional) (None, 1, 128) 74496
dropout_15 (Dropout)        (None, 1, 128)            0
gru_31 (GRU)                (None, 1, 64)             37248
Output-Layer (TimeDistributed) (None, 1, 1) 65
-----
Total params: 176,321
Trainable params: 176,321
Non-trainable params: 0
-----
```

Figura 49

Volumen real vs estimado para arquitectura LSTM por centro

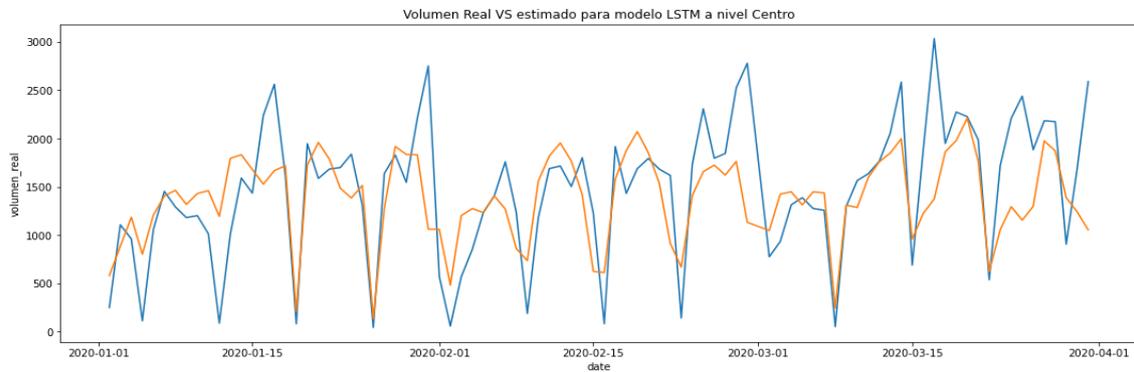


Figura 50

Volumen real vs estimado para arquitectura LSTM por categoría

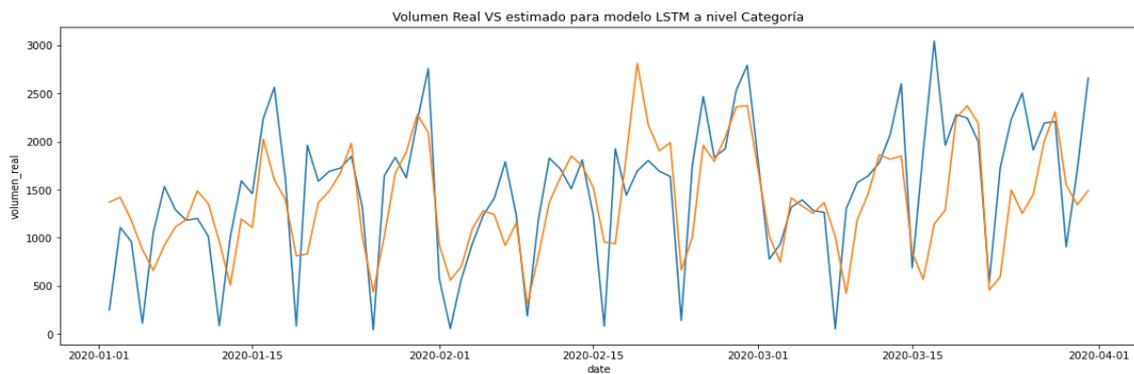


Figura 51

Volumen real vs estimado para arquitectura GRU por centro

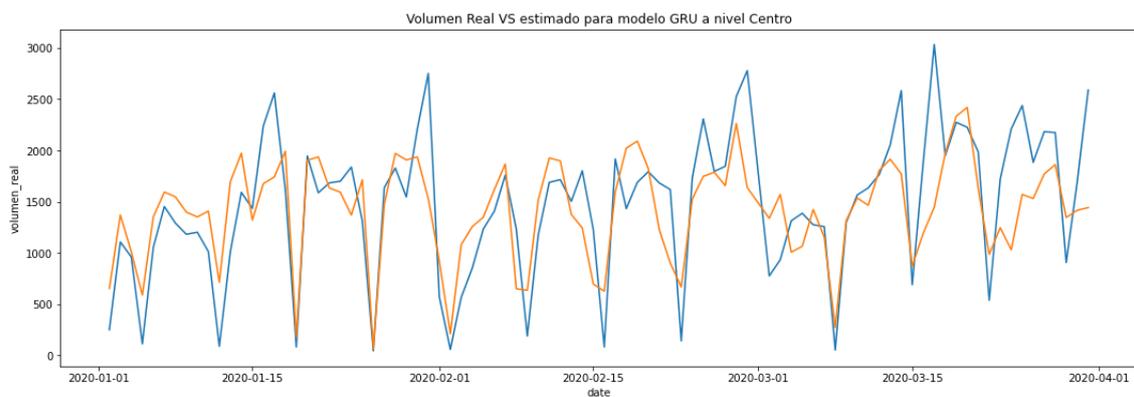


Figura 52

Volumen real vs estimado para arquitectura GRU por categoría

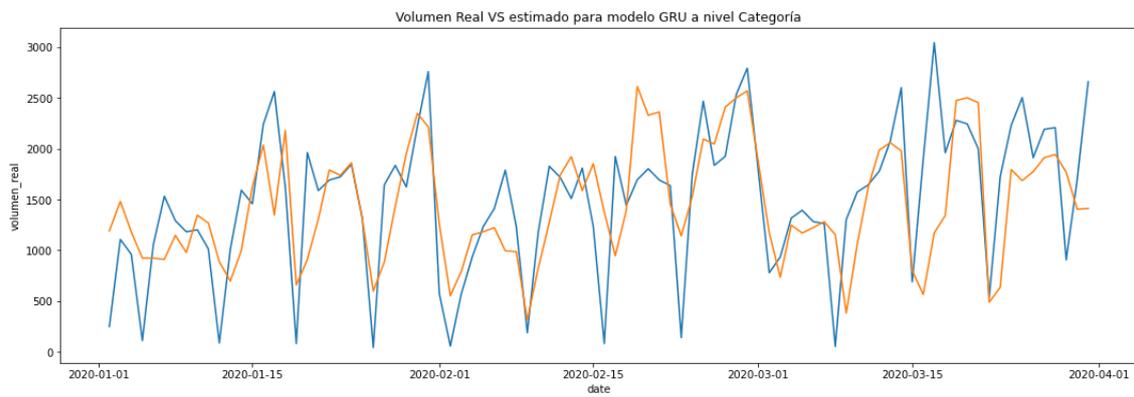


Figura 53

Funciones de perdida para entrenamiento y validación arquitectura LSTM por centro

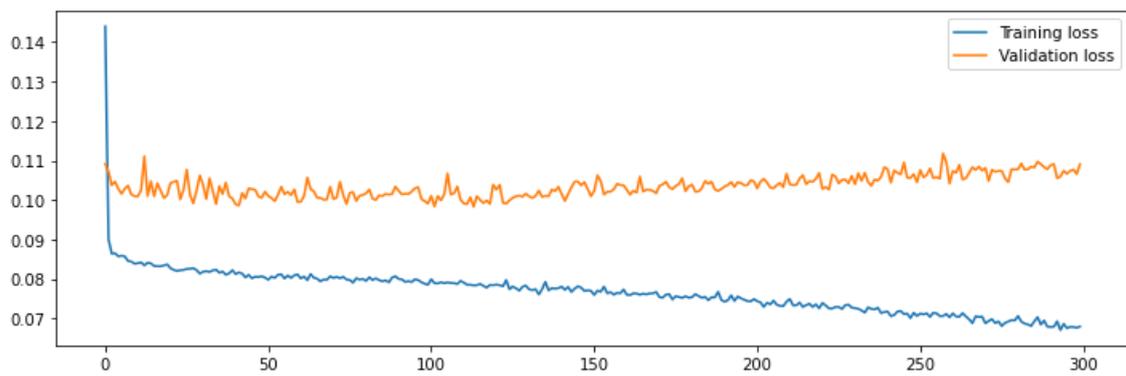


Figura 54

Funciones de perdida para entrenamiento y validación arquitectura LSTM por categoría

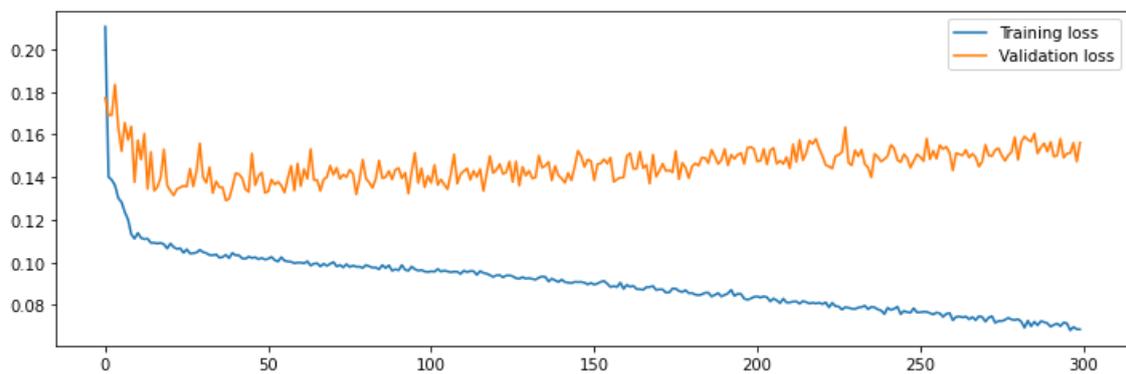


Figura 55

Funciones de pérdida para entrenamiento y validación arquitectura GRU por centro

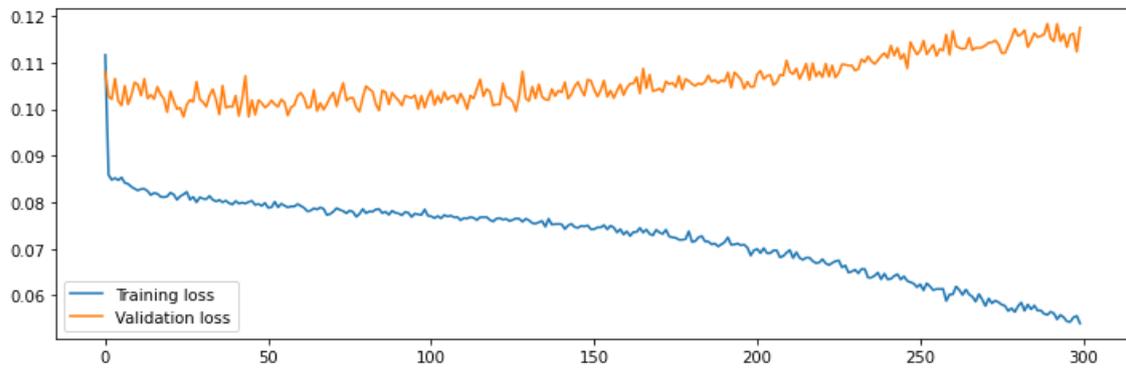


Figura 56

Funciones de pérdida para entrenamiento y validación arquitectura GRU por categoría

