

UCUENCA

Facultad de Ciencias Químicas

Carrera de Ingeniería Industrial

“Modelo de proyección de la demanda basado en Redes Neuronales Recurrentes de Memoria a Corto y Largo Plazo (LSTM) a través del análisis multivariante para la industria Láctea”

Trabajo de titulación previo a la
obtención del título de Ingeniero
Industrial

Autor:

Mónica Fernanda Guñanzaca Vallejo

CI:0106785496

Correo electrónico: vallejomonica914@gmail.com

Director:

Noe Rodrigo Guamán Guachichulca

CI:0105291595

Cuenca - Ecuador

11-enero-2023

Resumen:

Las industrias lácteas del Ecuador tienen un lento crecimiento en la utilización de nuevas tecnologías para el desarrollo de pronósticos multivariados, los mismos que proporcionan mayor confiabilidad en sus resultados. El uso de pronósticos tradicionales generalmente presenta altos niveles de error en las predicciones originando problemas como el sobre stock, la rotura de inventario, e incluso que muchos de sus productos tengan que ser descartados debido a que el tiempo de vida útil llegó a su fin. Hay que resaltar que la elaboración de estos pronósticos es el primer paso para la planificación del abastecimiento, producción, mano de obra e incluso la cantidad de vehículos necesarios para el transporte del producto. Por tal razón se propone el uso de Deep Learning para la elaboración de pronósticos de demanda, integrando en el modelo de Redes Neuronales Recurrentes – Memoria a Corto y Largo Plazo (LSTM) algunas variables independientes que afectan al comportamiento de las ventas, estableciendo una comparativa entre el uso del modelo univariado frente al modelo multivariado, mediante la evaluación de las métricas de desempeño. Teniendo como objetivo demostrar la reducción del error en MAPE haciendo uso del modelo multivariado.

La metodología utilizada se divide en tres fases generales que corresponden a la revisión de trabajos relacionados y marco teórico, recolección y preparación de datos y por último el análisis de los resultados obtenidos en cada modelo.

Se utilizaron variables independientes tanto endógenas como exógenas, logrando obtener una reducción del MAPE que va desde el 1% al 8% frente al modelo univariado, esto mejora las predicciones y aumenta el nivel de confianza, ya que se toma en cuenta características que interfieren en la demanda. Sin embargo, hay que tomar en cuenta que estas predicciones pueden mejorar haciendo uso de una cantidad de datos mayor a lo utilizado en este estudio.

Palabras clave: Deep learning. MAPE. Pronósticos. Redes neuronales recurrentes. LSTM.

Abstract:

The dairy industries of Ecuador have a slow growth in the use of new technologies for the development of multivariable forecasts, the same ones that provide greater reliability in their results. The use of traditional forecasts generally presents high levels of error in the predictions, causing problems such as overstock, inventory breaks, and even many of its products having to be discarded due to the end of their useful life. It should be noted that the preparation of these forecasts is the first step for planning supply, production, labor and even the number of vehicles needed to transport the product. For this reason, the use of Deep Learning is proposed for the elaboration of demand forecasts, integrating in the model of Recurrent Neural Networks - Short and Long Term Memory (LSTM) some independent variables that affect the behavior of sales, establishing a comparative between the use of the univariate model versus the multivariate model, by evaluating performance metrics. With the objective of demonstrating the reduction of the error in MAPE using the multivariate model.

The methodology used is divided into three general phases that correspond to the review of related works and theoretical framework, data collection and preparation, and finally the analysis of the results obtained in each model.

Both endogenous and exogenous independent variables were used, achieving a reduction in the MAPE ranging from 1% to 8% compared to the univariate model, this improves the predictions and increases the level of confidence, since characteristics that interfere in the analysis are taken into account. the demand. However, it must be taken into account that these predictions can be improved by using a larger amount of data than that used in this study.

Key words: Deep learning. MAPE. Forecasts. Recurrent neural networks. LSTM.

Índice

1.	Introducción	7
2.	Objetivo general y específico.....	9
2.1.	Objetivo general.....	9
2.2.	Objetivos específicos	9
3.	Revisión de la literatura	9
3.1.	Pronósticos.....	9
3.2.	Machine Learning y Deep Learning	10
3.2.1.	Industrias que usan Machine Learning y Deep Learning.....	10
3.2.2.	Redes Neuronales.....	12
3.2.3.	Redes Neuronales Recurrentes LSTM.....	12
3.3.	Métricas de evaluación.....	12
4.	Materiales y métodos	13
4.1.	Entorno de trabajo.....	14
4.2.	Recolección de datos.....	15
4.3.	Preparación de datos	15
4.4.	Serie de tiempo multivariante	19
4.4.1.	Variables exógenas	20
4.4.2.	Variables endógenas	20
4.5.	Diseño de la Red Neuronal Recurrente (LSTM).....	21
5.	Resultados y discusión	23
5.1.	Coefficiente de correlación de las variables.....	23
5.2.	Variabilidad de las ventas	24
5.3.	Automatización del pronóstico	25
5.4.	Entrenamiento de la Red Neuronal Recurrente LSTM.....	27
5.5.	MAPE generado en el modelo univariado frente al modelo multivariable.....	34
6.	Conclusiones	36
7.	Agradecimientos	37
8.	Bibliografía	37

Cláusula de Propiedad Intelectual

Mónica Fernanda Guñanzaca Vallejo, autora del trabajo de titulación "Modelo de proyección de la demanda basado en Redes Neuronales Recurrentes de Memoria a Corto y Largo Plazo (LSTM) a través del análisis multivariante para la industria Láctea", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autora.

Cuenca, 11 de enero de 2023



Mónica Fernanda Guñanzaca Vallejo

C.I: 0106785496

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Mónica Fernanda Guñanzaca Vallejo en calidad de autora y titular de los derechos morales y patrimoniales del trabajo de titulación "Modelo de proyección de la demanda basado en Redes Neuronales Recurrentes de Memoria a Corto y Largo Plazo (LSTM) a través del análisis multivariante para la Industria Láctea", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 11 de enero de 2023



Mónica Fernanda Guñanzaca Vallejo

C.I: 0106785496

1. Introducción

Con base en datos del año 2021 tomado del Instituto Ecuatoriano de estadísticas y Censos (INEC), el 12,5% del Producto Interno Bruto (PIB) del país es representado por la industria manufacturera, recuperándose en casi 6% con respecto al año anterior, dentro de la cual está bien posicionado el sector de alimentos que alcanza el 40% de la actividad económica. Cabe resaltar que en este sector incluye también a la industria de bebidas lácteas, estos sectores aportan con más del 10% de plazas de empleo. En cuanto a la industria de elaboración de productos lácteos, según datos tomados de la Superintendencia de Compañías, Valores y Seguros con corte al año 2021, tiene una participación en el PIB del 0.36%, el cual representa el 1.3 % superior a la cifra del 2020. Con respecto a los primeros 4 meses del 2022, presenta un incremento de ingresos del 0.6% en comparación a los mismos meses del año 2021. En términos generales, las ventas y exportaciones de este sector aumentaron en un 5 % con relación al 2020. Hay que mencionar que en el Ecuador existen alrededor de 70 industrias lácteas activas, la gran parte de empresas se encuentran en la sierra en un 65% y el otro 35% se ubica en la costa. En la provincia del Guayas se localiza el 33%, seguido de Pichincha con el 28% y la provincia del Azuay con al menos el 10%, datos que corresponden a las 3 provincias con mayor porcentaje de representación.

Siendo las ventas un pilar fundamental para el aporte de ingresos a las empresas, es importante que se ejecute los pronósticos con un mínimo porcentaje de error. Hay que tener presente que la planificación de la demanda no solo interviene en las ventas, puesto que es un paso previo para la planificación de la producción, planificación del abastecimiento y detalles importantes de la cadena de suministro tales como el nivel de inventario, el punto de pedido o la flota de camiones a contratar para cierta temporada.

Burgaenztle (2016) menciona que las micro, pequeñas y gran parte de las denominadas mediana empresa, llevan su pronóstico de demanda basados en métodos tradicionales con modelos de series de tiempo, afirma que se basan en datos de ventas históricos recabados en periodos de tiempo regulares para predecir la demanda futura. Estos datos incluyen el número de ventas y el tiempo en el cual se produjo la misma, sin contemplar otras variaciones inherentes de la dinámica del mercado, como por ejemplo la presencia de un producto sustituto nuevo, el uso de promociones, apertura de otros locales, variación de precios, nueva presentación del producto, variaciones que no están presentes dentro de un pronóstico tradicional Hanke y Wichern (2010) afirman que “estas variaciones hacen difícil pronosticar las ventas ya que el futuro no siempre es como el pasado”.

Debido a la importancia que tiene esta industria en el país y sobre todo en la región, es necesario analizar su proyección de la demanda y considerando un factor importante como es el tiempo de vida útil de los productos, su modelo de proyección no puede ser el mismo que se utilizan para productos no perecederos, ya que una mala proyección puede ocasionar en general, los siguientes problemas:

según indica un colaborador de la empresa anónima que forma parte de este estudio, “uno de los inconvenientes es la gestión de un sobre inventario, lo cual repercute en los costos excesivos de una empresa, ya que el producto no puede ser movilizadado, debido a que su porcentaje de vida útil llegó al máximo aceptable para que pueda ser recibido en cualquier centro de distribución o supermercado. Estos centros de distribución o supermercados necesitan un periodo de tiempo extenso hasta que el producto expire mientras es consumido por el cliente final. Otro de los problemas a tomar en cuenta es la rotura de stock, tiene un gran efecto en entregas de pedidos incompletos o por el contrario con la falta de abastecimiento a compradores, lo cual mina la confianza y satisfacción del cliente, y en algunos casos esto ocasiona que el consumidor opte por productos de la competencia” (W. Zumba, comunicación personal, 23 de junio de 2022).

Considerando estas problemáticas se propone realizar la proyección de la demanda utilizando un modelo de pronóstico que se adapte a las necesidades de la industria láctea, ajustándose a las ventas reales, es decir, que considere el tiempo de vida útil, promociones e incluso el nivel de abastecimiento de materia prima ya que el mismo se ve disminuido en temporadas de escasez. Estas condiciones especiales, son consideradas en modelos de pronósticos desarrollados mediante algoritmos que tuvieron auge en la nueva fase de la revolución industrial, denominada Industria 4.0, la cual, haciendo uso de la inteligencia artificial, ha desarrollado algoritmos que permitan el aprendizaje de los datos mediante el Machine Learning o de forma más profunda mediante el Deep Learning (Ganchozo, Vera, 2019, pg.35), para tales casos, teniendo como objetivo el mayor ajuste del pronóstico de ventas, es necesario considerar más de una variable independiente.

Según Garcés (2020) existen numerosas investigaciones que aplican el uso del Machine Learning en múltiples sectores industriales como ejemplos se puede mencionar en las áreas de logística, producción y manufactura, transporte, telecomunicaciones, entre otros, sin embargo, se encuentran muy pocos estudios referentes a las proyecciones de venta de productos alimenticios y que además tengan un tiempo de corta vida útil (pg. 72). Sin embargo, hay que recalcar que no fue posible encontrar estudios realizados en la industria láctea que apliquen inteligencia artificial como herramienta para la elaboración de los pronósticos y aún menos que incluya un análisis en el que intervenga más de una variable.

Al comprender que con el uso de algoritmos de Deep Learning frente a pronósticos tradicionales es posible mejorar el nivel de confianza, se incluye además que con el uso de variables exógenas los resultados pueden aportar niveles de confianza mayores. Así lo recomienda Garcés (2020) puesto que menciona que con el uso de un mayor número de variables descriptivas el algoritmo ofrece mejores resultados.

Aucapiña (2018) concluye que resulta eficiente el uso de variables exógenas tales como indicadores macro y microeconómicos a los cuales les asigna niveles de importancia diferentes,

indicando que la proyección, mediante Redes Neuronales Artificiales, se vuelve más sensible a las variables que presenten un mayor coeficiente en el algoritmo de Garson.

El uso de variables exógenas se ha visto necesario para modelos de Machine Learning tales como Ridge y SVR según menciona Medina y Galván (2007), estas variables requieren ser de tipo numéricas ya que el uso de variables categóricas genera predicciones con bajo nivel de confianza ya que, además, el ruido de todo el conjunto de datos produce un gran impacto en el modelo.

2. Objetivo general y específico

2.1.Objetivo general

Demostrar la reducción del error en el pronóstico, utilizando dos o más variables independientes en el modelo basado en Redes Neuronales Recurrentes - LSTM, en contraste a la utilización de una sola variable con el mismo modelo, para las industrias lácteas.

2.2.Objetivos específicos

1. Identificar los factores relevantes de la empresa o el mercado que corresponden a las variables independientes.
2. Establecer mediante el modelo de Redes Neuronales Recurrente de tipo Memoria a Corto y Largo Plazo (LSTM) la proyección de la demanda para la industria láctea.
3. Validar el modelo establecido a través de métricas de desempeño utilizando el error porcentual absoluto medio (MAPE)
4. Descripción del análisis de resultados obtenidos del modelo de Deep Learning (Redes Neuronales Recurrentes - LSTM).

3. Revisión de la literatura

3.1.Pronósticos

Según Montemayor (2013) los pronósticos permiten determinar el volumen de ventas en el futuro mediante el análisis de los datos históricos. Esto permite a las empresas que establezcan metas a corto, mediano y largo plazo, identificando las estrategias más adecuadas según el objetivo planteado.

El pronóstico de las ventas a un corto plazo facilita la decisión en cuanto al abastecimiento de materia prima y materiales, la mano de obra y el nivel de producción que se requiera. A mediano plazo el pronóstico es de gran ayuda para la elaboración de presupuestos de acuerdo a la estimación de las ventas.

3.2. Machine Learning y Deep Learning

El Machine Learning es una rama de la inteligencia artificial basada en algoritmos, los cuales aprenden de ciertos patrones que se encuentran en un conjunto de datos con el objetivo de realizar predicciones. Dorado y Velásquez (2020) mencionan que estos algoritmos se pueden aplicar a series de tiempo obteniendo una mayor precisión en los resultados comparado a los pronósticos tradicionales. La estructura algorítmica se basa en el conjunto sistemático de instrucciones que se programen, que a su vez están basados en modelos matemáticos sencillos.

El Deep Learning es un subcampo del Machine Learning que procesa los datos mediante modelos de Redes Neuronales Artificiales, los cuales extraen patrones de las capas sin necesidad de que el programador lo indique, desarrollando un aprendizaje profundo.

3.2.1. Industrias que usan Machine Learning y Deep Learning

En el Ecuador se han utilizado algoritmos tanto de Machine Learning como de Deep Learning para predecir la demanda de varias empresas ya sea del sector textil, sector hotelero o para la predicción de alimentos de consumo masivo, sin embargo, hay muy pocas investigaciones para la predicción de la demanda dentro de la industria láctea, la mayor parte de información se puede encontrar en el ámbito de las predicciones con pronósticos tradicionales como media móvil o Winters. A continuación, se revisa investigaciones que se han hecho utilizando Machine Learning y Deep Learning.

En la ciudad de Ambato, para la empresa Impactex, organización con fines de lucro que se dedica a la producción y comercialización de prendas de vestir, se realizó la proyección de su demanda mediante el uso de Machine Learning, utilizando algoritmos de Redes Neuronales, los datos se tomaron de ventas mensuales desde el año 2008 hasta el año 2019, esta predicción determinó que para un menor nivel de error es recomendable el uso de tres capas ocultas para los 4 productos proyectados, los mismos que se determinaron mediante un análisis ABC, obteniendo un error entre el 2 y 4%, la herramienta que se adaptó a este proceso es el lenguaje de programación Python, a través de librerías como Tensor Flow y Keras (Llumitasig Galarza, 2021).

En la ciudad de Guayaquil se realizó la proyección de ventas para la empresa Tiosa S.A. tomando como producto el Pan de Molde blanco, el modelo utilizado fue el de redes neuronales con una capa de entrada y una capa oculta para la correlación de las variables, el porcentaje de error obtenido es del 1% en torno a la cantidad de bolsas que se fabrican, los datos analizados comprenden desde el periodo del año 2014 hasta el año 2018 con un total de 1053 datos (Garcés Camacho, 2020).

De igual manera, en la ciudad de Guayaquil se realizó un análisis de la proyección de ventas en un laboratorio farmacéutico (el autor proteja la identidad del laboratorio). Molina (2020) indica que se ha generado sobre stock en productos digestivos que tienen muy baja rotación y por otro lado existe falta de inventario en medicamentos que tienen un alto consumo perjudicando totalmente los indicadores de

gestión. Además, encontró productos con grandes cantidades en el inventario y que están cerca de caducarse y por lo tanto deberán ser destruidos. Estas situaciones generan pérdidas económicas que ponen en riesgo la rentabilidad de estos productos. También mencionó que se pudieron atender solo el 89% de órdenes de compra ya que para el resto no hubo stock (pg. 18).

Este análisis fue de gran utilidad ya que aportó con una correcta planificación de ventas y optimización de los recursos de producción. Al no contar con un sistema de pronósticos adecuado, su gestión de inventario se basó simplemente en una evaluación subjetiva de las unidades vendidas determinando así, el nivel de stock. El autor utilizó los datos de 11 productos correspondientes a 24 meses y mediante la elaboración de modelos de pronósticos tales como: series temporales, regresiones y redes neuronales, logró reducir el error en promedio de un 5%, cabe resaltar que el modelo con menor error corresponde al de Redes Neuronales seguido del Arimax y por último el de Winters (Molina Rea, 2020).

En otro caso de estudio, realizado a una franquicia de farmacias del Ecuador, presenta ventas semanales de productos durante un periodo de 4 años. Para la predicción se utilizó algunos algoritmos, presentando mejores resultados el de redes neuronales convolucionales, el cual se usa generalmente para procesamiento de imágenes. Con los otros algoritmos la precisión disminuye, ya que sufren el problema de la desaparición del gradiente al usar demasiadas capas ocultas. Debido a la gran cantidad de pesos, este modelo de pronóstico fue diseñado y programado en el lenguaje denominado Matlab, además se demuestra que los modelos de predicción basados en Redes Neuronales Convolucionales (CNN) alcanzan una alta precisión utilizando menos cantidad de pesos en su estructura (Velasco Sandoval, 2020).

En el transcurso del año 2017, Corporación Favorita lanzó un proyecto denominado: Corporación Favorita Grocery Sales Forecasting a través de la plataforma Kaggle, el mismo que fue resuelto mediante diferentes modelos y autores, se cita uno de ellos.

En el año 2021, se determinó que la concentración de ventas de la Corporación Favorita se encuentra en las ciudades de Quito y Guayaquil. Para la predicción se tomó una base de datos desde el año 2016 hasta el mes de agosto del 2017 con un total de 592 días de ventas. El modelo combinado de Redes Neuronales Recurrentes con LSTM aportó un mejor desempeño con un 0.5 de error RMSE en la etapa de validación, comparado con los otros tres modelos: Perceptrón Multicapa (MLP), Memoria a Corto y Largo Plazo (LSTM) y Redes Neuronales Convolucionales (CNN), en este caso la herramienta utilizada es Google Colab (Jaramillo Mira & Gutierrez Muriel, 2021).

En México se realizó un análisis de proyección de la demanda para productos de consumo masivo, en el cual utilizaron métodos tradicionales y modelos Machine Learning, de acuerdo al menor porcentaje de error obtenido, los resultados indicaron un mejor ajuste con el modelo ARIMA y Random Forest debido a la dinámica de la demanda en el sector (López Zúñiga & Jaimes Campos, 2021).

3.2.2. Redes Neuronales

Las redes neuronales procesan información, entrenándolas para que adquieran conocimiento el cual es almacenado en pesos de las conexiones de cada neurona, además pueden responder a nuevos patrones que antes no se vieron, lo que permite desarrollar mejores predicciones.

El modelo de la red neuronal está formado por varias capas: la de entrada, la de salida y las capas ocultas, cada una contiene varias neuronas, las cuales pueden trabajar simultáneamente. Las neuronas de las capas ocultas necesitan de funciones de activación para entregar una salida diferente a un conjunto de datos de entrada. (Eguez, 2020)

Una red neuronal artificial posee una memoria a corto plazo, por lo que la secuencia que se procese debe ser corta para que la red no pierda la memoria.

3.2.3. Redes Neuronales Recurrentes LSTM

La particularidad de las redes recurrentes se encuentra en la existencia de un proceso de retroalimentación o bucles que se desarrolla en cada neurona, es decir, la neurona recibe datos de entrada con los cuales produce una salida, la cual es enviada nuevamente a la entrada de la neurona, esto le permite crear una memoria a largo plazo, seleccionando los patrones más relevantes que deben recordarse, o añadir nuevos patrones y eliminar otros. Este tipo de redes son diseñadas especialmente para predicción de secuencias.

La neurona se activa mediante algunas funciones que deciden si dejan pasar la información o si la misma es importante por lo cual debe ser recordada o en caso contrario debe ser eliminada (Bueno, 2019).

3.3. Métricas de evaluación

Son sistemas de medición que evalúan el desempeño de los modelos con base en los errores generados dentro de los resultados, esto permite medir la precisión que tiene los pronósticos utilizados. Sin embargo, un pronóstico se puede ajustar bien a los datos de entrenamiento, no así, a los datos de prueba, por esta razón existen algunas medidas que se usan para seleccionar el mejor modelo de pronóstico.

Root Mean Square Error (RMSE) es la raíz del promedio de los cuadrados del error en un periodo determinado, es útil para determinar la precisión de diferentes métodos de pronósticos y seleccionar el adecuado ya que esta penaliza de manera fuerte a los errores generados (Egüez, 2020).

Median Absolute Deviation (MAD) permite observar la diferencia de la predicción en comparación con los datos reales en las mismas unidades en valores absolutos.

Mean Absolute Percentage Error (MAPE) es la media de los errores porcentuales en valor absoluto, esta métrica no es utilizada como una medida de selección ya que presenta sesgos que pueden favorecer a predicciones que están por debajo de los valores reales. (Sanchez, 2017)

Bias o sesgo es la diferencia de la predicción en comparación con los datos reales en las mismas unidades.

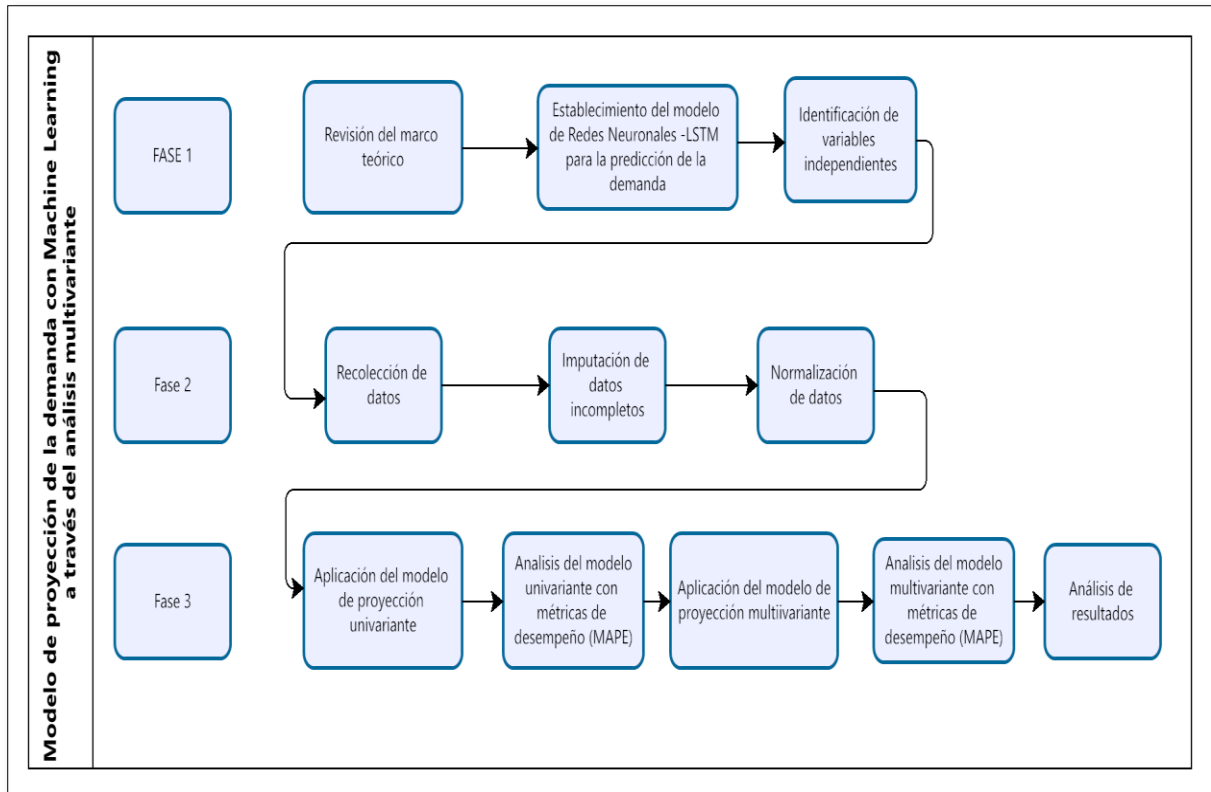
4. Materiales y métodos

El planteamiento del presente trabajo se enfocó en demostrar la reducción del error en el pronóstico, utilizando dos o más variables independientes en el modelo basado en Redes Neuronales Recurrentes - LSTM, en contraste a la utilización de una sola variable con el mismo modelo, para las industrias lácteas.

La población para este estudio pertenece al sector de bebidas Lácteas de la Sierra, de una empresa en particular. Se utilizó muestreo no probabilístico por conveniencia. Hay que resaltar que debido a un acuerdo convenido con quienes proporcionaron la información, se mantiene bajo estricta confidencialidad el nombre de esta organización.

En la Ilustración 1, se presenta un esquema metodológico, el cual se llevó a cabo para esta investigación.

Ilustración 1.



4.1. Entorno de trabajo

Para el desarrollo del modelo de Redes Neuronales Recurrentes (LSTM) se usó el lenguaje de programación denominado Python, ya que es considerado como uno de los lenguajes más sencillos de aprender y fácil de entender, además es muy utilizado para el desarrollo de algoritmos de Machine Learning y Deep Learning. Sin embargo, es necesario tener conocimiento previo sobre el uso del mismo.

Para la escritura del código se puede utilizar varias aplicaciones de código abierto como Jupyter Notebook, Spyder, Google Colab, entre otros. Para este trabajo el código se ejecutó en Jupyter Notebook ya que es adecuado para tareas de aprendizaje automático y se puede acceder desde cualquier ordenador sin la necesidad de conexión a internet.

Se utilizó librerías enfocadas en Deep Learning como Tensor Flow y Keras, librerías para procesamiento de datos como Pandas y Numpy, y para la visualización de los mismos mediante la generación de gráficos con Matplotlib.

Además del uso de Python se trabajó en conjunto con el Software Microsoft Excel, el cual permitió realizar cálculos sencillos, sin la necesidad de incluir programación.

4.2. Recolección de datos

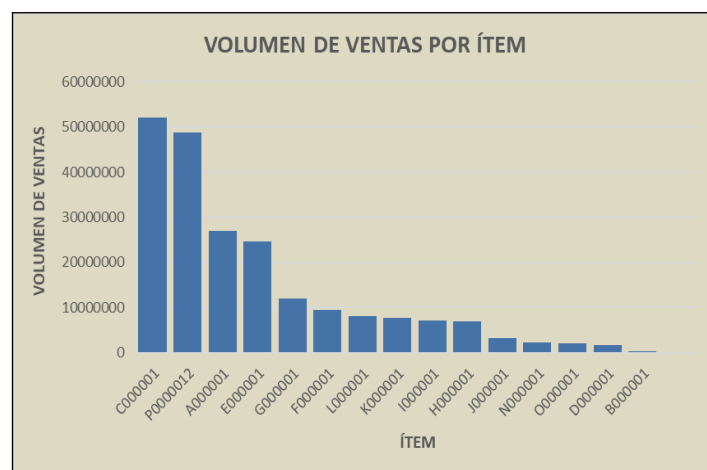
Para la recolección de datos se usó fuentes secundarias, puesto que los mismos fueron extraídos de la plataforma Qlick Sense, que es una aplicación que permite al usuario crear visualizaciones, gráficos, entre otros, una herramienta muy útil para la inteligencia de negocios. Esta plataforma se alimenta con los datos que proporciona el Sistema SAP de la empresa. Los datos se encuentran registrados mediante una codificación correspondiente a cada ítem de cierto producto lácteo, debido a la protección de información se trabajó solamente con la codificación de los ítems.

El conjunto de datos pertenece al periodo correspondiente desde enero del año 2019 hasta el mes de julio del año 2022, agrupados en semanas, por lo tanto, se tiene 188 datos.

La empresa oferta al mercado más de 50 ítems o Stock Keeping Unit (SKU), por lo tanto, el estudio se basó en los 2 productos con mayor volumen de ventas, en la ilustración 2 se puede visualizar los 15 productos más demandados dentro de todo el registro.

Ilustración 2.

Volumen de ventas



4.3. Preparación de datos

Antes del análisis de datos se realizó la limpieza de los mismos, es decir, se encontró las anomalías presentes y se procedió a corregirlas. Esto requirió invertir una gran cantidad de tiempo.

Este paso ofrece información de calidad para la posterior toma de decisiones con base en los resultados obtenidos del análisis (Ridzuan & Wan Zainon, 2019).

Para la preparación de los datos se ejecutó los siguientes pasos.

Filtrado de datos: El conjunto de datos que proporcionó la plataforma Qlick Sense, permitió observar el volumen de ventas de todos los productos. Luego de que se realizó el análisis se tuvo que

los 2 productos más vendidos corresponden a la codificación C000001(C) Y P000012 (P), los cuales fueron seleccionados para todo este estudio. Además, por cada ítem se desglosó una nueva segmentación que corresponde a las ventas del canal moderno y canal tradicional.

Ya que la base estuvo segmentada de acuerdo al tipo de canal, se agrupó cada valor para obtener el total del volumen de ventas correspondientes al canal tradicional y moderno.

El canal moderno presenta la codificación 10 y el canal tradicional corresponde a cualquier otro número, excepto 10.

La codificación numérica de los canales se muestra en la ilustración 3.

Ilustración 1.

Ventas con canales tradicional y moderno

CANAL	AñoMesFactura	2018-sep		2018-sep		2018-sep		2018-sep		2018-oct		2018-oct		2018-oct	
		2018-36	2018-37	2018-38	2018-39	2018-40	2018-41	2018-42	2018-43	2018-44	2018-45	2018-46	2018-47	2018-48	2018-49
13	G000001	-	72,00	72,00	48,00	96,00	-	-	-	-	-	-	-	-	36,00
15	G000001	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	H000001	56.088,00	26.401,00	42.019,00	107.365,00	36.759,00	25.644,00	46.236,00	60.792,00	-	-	-	-	-	-
11	H000001	4.830,00	12.186,00	11.841,00	14.620,00	9.236,00	6.982,00	6.191,00	9.625,00	-	-	-	-	-	-
12	H000001	-	-	-	49,00	-	-	24,00	72,00	-	-	-	-	-	-
13	H000001	-	12,00	96,00	24,00	-	24,00	48,00	36,00	-	-	-	-	-	-
15	H000001	-	12,00	12,00	-	-	-	12,00	-	-	-	-	-	-	-
10	I000001	53.228,00	14.029,00	35.780,00	104.203,00	38.598,00	31.092,00	53.100,00	64.752,00	-	-	-	-	-	-
11	I000001	-	-	24,00	-	-	24,00	-	-	-	-	-	-	-	-
12	I000001	-	-	120,00	-	180,00	180,00	-	-	-	-	-	-	-	-
13	I000001	-	-	120,00	60,00	36,00	84,00	36,00	72,00	-	-	-	-	-	-
15	I000001	-	-	-	12,00	-	-	-	-	-	-	-	-	-	-

Tomado de: Empresa láctea, 2022

Eliminación de variables y selección de datos: Para un mejor manejo de la base de datos se adaptó las columnas a un formato de fecha segmentado por año y mes, además, se incluyó el número de semana ya que la proyección se realizó de forma semanal. La selección y eliminación de variables se pueda observar en la Ilustración 4.

Ilustración 4.

Formatos de fecha y semana

AÑO	MES	N° SEMANA	TOTAL C1	CANAL MODERNO C1	CANAL TRADICIONAL C1	TOTAL P1	CANAL MODERNO P1	CANAL TRADICIONAL P1
2019	enero	1	\$ 283.767,00	\$ 16.019,00	\$ 267.748,00	\$ 100.260,00	\$ -	\$ 100.260,00
2019	enero	2	\$ 351.488,00	\$ 42.193,00	\$ 309.295,00	\$ 76.203,00	\$ -	\$ 76.203,00
2019	enero	3	\$ 369.551,00	\$ 23.472,00	\$ 346.079,00	\$ 215.267,00	\$ -	\$ 215.267,00
2019	enero	4	\$ 400.329,00	\$ 44.609,00	\$ 355.720,00	\$ 310.076,00	\$ -	\$ 310.076,00
2019	enero	5	\$ 365.881,00	\$ 22.037,00	\$ 343.844,00	\$ 390.692,00	\$ -	\$ 390.692,00
2019	febrero	6	\$ 346.903,00	\$ 39.487,00	\$ 307.416,00	\$ 120.518,00	\$ -	\$ 120.518,00
2019	febrero	7	\$ 360.253,00	\$ 34.078,00	\$ 326.175,00	\$ 156.133,00	\$ -	\$ 156.133,00
2019	febrero	8	\$ 366.584,00	\$ 63.082,00	\$ 303.502,00	\$ 169.161,00	\$ -	\$ 169.161,00

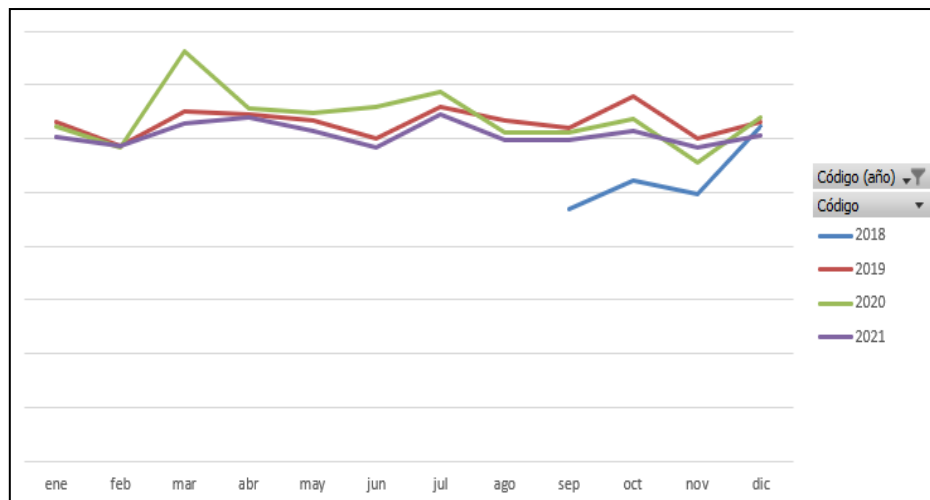
Los datos tuvieron la siguiente peculiaridad:

- Semanas repetidas: ya que, debido a la estructura del calendario semanal, se observó semanas en las cuales termina y comienza un mes, se tuvo como ejemplo la semana 5 del año 2019, puesto que el mes de enero termina el jueves y febrero comienza el viernes de esta misma semana. En estos casos la aplicación entrega los datos duplicando las semanas que tengan esta misma peculiaridad. Por esta razón se vio la necesidad de sumar el volumen de ventas correspondiente a la semana 5 que va desde el 28 al 31 de enero y la misma semana 5 que va desde el 1 al 2 de febrero. Ya que estos valores corresponden a la misma semana. Para los datos que no se realizó esta suma se observó ventas por debajo del promedio real.

Tomando en cuenta que para desarrollar modelos de Deep Learning es preferible una extensa base de datos, cabe mencionar que en cuanto a los registros existentes se trabajó con datos desde el mes de septiembre del año 2018. Sin embargo, se observó que las ventas de este año están por debajo de los registrados de los años siguientes. Esto generó ruidos en el desarrollo del algoritmo, por tal motivo se eliminaron los datos de este año. Las ventas se muestran en la Ilustración 5.

Ilustración 5.

Comportamiento de ventas



Detección de valores faltantes y outliers: según Medina y Galván (2007) no se recomienda imputar datos si los valores faltantes superan el 20% del total de la base, puesto que perdería confiabilidad estadística y, por consiguiente, esta base debe ser descartada. Sin embargo, para los casos académicos este porcentaje incrementa hasta el 25 % (pg. 12). Por tal motivo, no se puede realizar predicciones para el canal moderno del ítem con la codificación P, ya que tiene el 90% de datos faltantes.

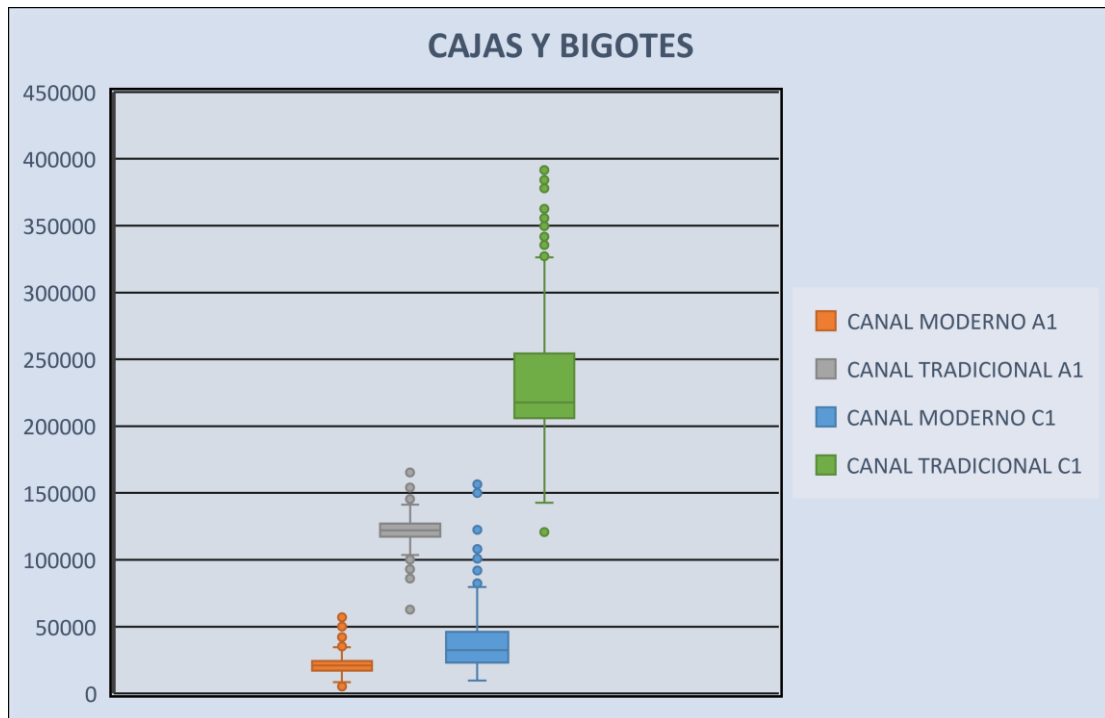
En el caso de los valores faltantes, se realizó una imputación de datos usando medias móviles con el número de periodos igual a seis. Según Medina y Galván (2007) se puede utilizar medias móviles, puesto que es el método más usado para imputar los datos.

Según Soler (2008) los outliers se consideran como datos que son inconsistentes y con un comportamiento diferente con respecto a todo el registro. Para determinar si existe datos atípicos se hizo uso de uno de los métodos clásicos denominado Cajas y Bigotes, además permitió observar la distribución de los datos.

Para el conjunto de datos se presenta el diagrama, ilustración 6, en la cual se puede observar la asimetría positiva y la dispersión de los datos del conjunto que pertenece al canal tradicional C1, al contrario del resto de cajas que tienen los datos concentrados. Además, se evidencia que todo el registro de datos presenta outliers que están por debajo o encima de los bigotes.

Ilustración 2.

Outliers de base de ventas en Cajas y Bigotes



Para el tratamiento de outliers, estadísticamente es necesario eliminar estos datos, ya que genera ruidos en el modelo, sin embargo, la naturaleza de los mismos no provino de cifras erróneas sino de ventas reales que se dieron en ciertas semanas, por esta razón no se manipuló los registros por ningún método.

Sin embargo, para los meses que corresponden al inicio de la pandemia, específicamente desde la semana 12, se utilizó el método de las medias móviles para eliminar los outliers, ya que estas cifras corresponden a un evento en particular que se generó dentro de un determinado periodo y tuvo incidencia en el aumento de la demanda, no obstante, al día de hoy las ventas se encuentran normalizadas y no se ven influenciadas por este hecho.

4.4. Serie de tiempo multivariante

Para el análisis del modelo multivariable es necesario incluir variables que influyan en el comportamiento de las ventas ya sean exógenas o endógenas, esto se realiza con el objetivo de mejorar las predicciones y reducir el error comparándolo con un modelo que contenga solamente los datos de ventas.

Según Arroyo (2014), las variables exógenas cambian su valor de forma externa al modelo, éstas pueden influir en las variables endógenas (pg.18).

Aquellas variables que influyen en el comportamiento de la demanda pudiendo generar un aumento en las ventas y que sobre todo son generadas por la empresa, se las denomina variables endógenas.

Para este estudio el análisis de las ventas se basó en estos dos tipos de variables. Para el caso de las variables exógenas el cambio de las ventas ya sea un incremento o decremento puede verse afectado por varios motivos que la empresa no puede controlar y son de relación directa con el producto en cuestión.

4.4.1. Variables exógenas

Las variables exógenas que se usaron en esta investigación son las siguientes:

- **Escasez de leche:** Para las industrias lácteas, existe periodos en donde es frecuente encontrarse desabastecidos de leche, esto responde a varios motivos, como por ejemplo la escasez en el mercado o que la competencia haya subido el precio de pago a los productores por lo que deja desabastecida a otras industrias, lo cual influye directamente en el total de la producción y stock de las empresas. Para esta variable se designó que los meses de escasez corresponden al periodo comprendido entre enero a julio, sin embargo, en la actualidad esta frecuencia puede tener alguna variación (V. Pérez, comunicación personal, 23 de junio de 2022).
- **Escasez de cartón:** Otra característica importante que se generó a partir la crisis logística mundial fue la escasez de cartón, que a su vez influyó en la cantidad de ventas al no contar con el material de empaquetado. Al limitar la producción esta variable influye en la oferta de producto y por lo tanto en las ventas, hecho que ocurre desde el mes de noviembre del año 2021.
- **Feriados y**
- **Pago de nómina (quincenal o a fin de mes):** estos eventos generalmente influyen en el consumidor incrementando la demanda.

4.4.2. Variables endógenas

La variable endógena que se utilizó corresponde a las promociones del Black Friday que generan las empresas lácteas, el histórico corresponde a las promociones generadas en el mes de noviembre desde el año 2019. Estas estrategias se realizan con el objetivo de aumentar las ventas atrayendo la atención de los consumidores, ya que debido a la disminución de precios se incrementan significativamente las ventas.

Para las variables endógenas y exógenas se utilizó el código Booleano, en el cual el dígito 1 representa la validez del enunciado y el dígito 0 indica que en tales semanas el enunciado no es verdadero. Las variables se muestran en la ilustración 7.

Ilustración 7.

Variables endógenas y exógenas

AÑO	MES	Feriados	quincena y fin de mes (1)	temporada a escasez de leche	escasez de cartón	promociones (black)	N° SEMANA	TOTAL VENTAS
2021	octubre	0	1	0	0	0	43	158138,00
2021	octubre	1	0	0	0	0	44	108414,00
2021	noviembre	0	0	0	1	0	45	146485,00
2021	noviembre	0	1	0	1	0	46	131960,00
2021	noviembre	0	0	0	1	1	47	140172,00
2021	noviembre	0	1	0	1	0	48	140777,00
2021	diciembre	1	0	0	1	0	49	135796,00

4.5. Diseño de la Red Neuronal Recurrente (LSTM)

Para la elaboración del modelo de proyección se usó del lenguaje de programación Python. Para diseñar la estructura del algoritmo, se instaló las siguientes librerías:

- Tensor Flow
- Keras
- Sklearn
- Matplotlib

Luego de que se importó las librerías necesarias, se cargó los datos con la ayuda de la librería Pandas y se realizó un análisis estadístico a través de los datos dados, el mismo se muestra en la ilustración 8.

Ilustración 8.

Análisis descriptivo

	N° SEMANA	TOTAL A1	CANAL MODERNO A1	CANAL TRADICIONAL A1	TOTAL C1	CANAL MODERNO C1	CANAL TRADICIONAL C1
count	187.000000	187.000000	187.000000	187.000000	187.000000	187.000000	187.000000
mean	24.877005	143843.550802	21808.165775	122035.385027	278269.836364	39436.743316	238833.093048
std	14.887783	15898.786358	8241.274740	11263.529881	58557.053357	25422.586086	53560.259565
min	1.000000	68050.000000	5306.000000	62744.000000	147293.000000	9660.000000	120839.000000
25%	12.000000	136476.500000	17043.500000	117232.000000	235935.500000	23228.000000	206114.500000
50%	24.000000	143505.000000	20861.000000	121988.000000	259635.000000	32447.000000	217760.000000
75%	37.000000	150775.000000	24246.500000	127100.000000	306067.100000	45806.500000	253941.000000
max	53.000000	215427.000000	56930.000000	165251.000000	424366.000000	156528.000000	391640.000000

Para iniciar, los datos fueron normalizados, ya que generalmente se encuentran en varias dimensiones y ocasiona problemas durante la proyección dando resultados erróneos y poco precisos.

Se estandarizó los datos mediante la función Escala estándar (StandardScaler), ya que las otras funciones como por ejemplo Mínimos y máximos (Min Max) son muy sensibles cuando encuentran datos atípicos.

La función `StandardScaler` resta la media y divide entre la desviación estándar por lo que se ve diluido el efecto del outlier en la media (Ganapathi, Prasanna, Mahesh, Kalindindi, & Padma, 2020).

Para que el modelo reconozca los patrones, sea capaz de aprender de ellos y sepa que información eliminar o mantener, se le asignó el 75 % de los datos para el entrenamiento y el 25% restante es utilizado como datos de prueba, con los cuales se evaluó la eficacia del modelo, sin embargo, esta opción puede ser modificada como mejor convenga, observando los resultados de cada entrenamiento.

Para la creación del modelo se utilizó la función Secuencial (`Sequential`) de la librería de Keras. Ya que los modelos necesitan una función de activación, se usó de la función `Leaky Relu` (`Rectified Linear Unit`). Esta es una función no lineal muy utilizada para algoritmos de Deep Learning, la cual toma los valores negativos y los multiplica por un factor rectificativo, evitando que los valores negativos se hagan cero. En el caso de presentar valores iguales a cero la neurona deja de aprender, no obstante, cuando los valores son positivos la salida es igual a la entrada

Esta función ya no tiene los problemas de saturación como las demás funciones (sigmoide o tangente hiperbólica), lo que permite que el algoritmo converja mucho más rápido y la red siga entrenando, esto posibilita el rápido aprendizaje (Mora, 2018).

Para este modelo se designó 5 neuronas en cada una de las 2 capas para el uso de todas las variables. Cabe mencionar que mientras mayor es el número de capas y de neuronas mayor es el número de mínimos locales. Hay que tener cuidado en el caso de que se indique una gran cantidad de neuronas, la red neuronal puede aprender bastante bien, sin embargo, pierde la capacidad de extrapolación, es decir no se podrá ajustar ante nuevos patrones ocasionando que el error de predicción sea mayor, este fenómeno se denomina como *overfitting*, aprendizaje de alta precisión en los patrones vistos (Ying, 2019). Se puede aceptar como tal cuando se tiene un error de entrenamiento muy pequeño en comparación al error producido con los datos de prueba. (Villa, Velásquez, & Sánchez, 2014)

Para tener un mejor control del entrenamiento y evitar que se produzca *overfitting*, se incluyó una gráfica en la que se observó la curva de aprendizaje de la red neuronal. La cual indicó en qué época el modelo empezó a incrementar el resultado del error. Se usó la función `Detención Anticipada` (`EarlyStopping`) de Keras para detener el entrenamiento cuando el error de validación empezó a subir.

Contrario a lo mencionado se puede tener otro caso, en el que, al ajustar el algoritmo con pocas neuronas, la red no es capaz de aprender nuevos patrones ya que las neuronas están saturadas de información y esto es conocido como *Underfitting*,

Para evitar estos casos se recomienda comenzar con una capa oculta, entrenar el modelo y seguido de esto, aumentar el número de neuronas, en el caso de que el error siga siendo alto, se aumenta el número de capas y luego el número de neuronas, hay que tener presente también que el número de

iteraciones también indica la cantidad estimada de entrenamientos necesarios para que los errores sean mínimos.

Para compilar el código se determinó las métricas de evaluación y un método de optimización el cual ajusta los pesos para que la neurona sepa que información tiene mayor importancia y permita el ajuste de forma automática de los hiperparámetros. Por tal razón se utilizó el algoritmo de Adam, este obtiene los beneficios de los algoritmos RMSprop y Momentum ya que puede calcular la tasa de aprendizaje de los hiperparámetros generando un mejor rendimiento en aquellos gradientes que son dispersos basándose en el gradiente del peso, permite tener un buen rendimiento en regresiones no estacionarias (Dinamarca, 2018).

Además, para crear la estructura de la red recurrente con memoria a corto y largo plazo, la misma que crea bucles y que permite que se recuerde información dependiendo el tipo de importancia, se incluyó la función LSTM de la librería de keras y Tensor Flow.

Se entrenó el modelo mediante la función Fit, para lo cual se asignó a 20 como el número de iteraciones, 30 la cantidad de datos por cada bloque a entrenar y falso a la función Shuffle que aleatoriza el conjunto de datos, esta última función no es requerida que se active como verdadera ya que las ventas siguen una secuencia en el tiempo, razón por la cual no pueden desordenarse los datos.

Luego de realizar las predicciones, se transformó a la escala original para poder realizar las respectivas evaluaciones de los modelos a través de las métricas correspondientes y ajustar el número de capas, neuronas e iteraciones en el caso de no alcanzar el valor adecuado en las métricas de evaluación.

Para el caso del análisis multivariable, la matriz de datos permitió escoger las variables necesarias durante el estudio, ya que el diseño de la red en forma general, no varió para este análisis. En este caso la matriz de datos aceptó la incorporación de nuevas variables y por consiguiente pudo indicar el número de variables añadidas dentro de la estructura de la red.

5. Resultados y discusión

5.1. Coeficiente de correlación de las variables

Para las variables endógenas y exógenas que fueron seleccionadas se aplicó el análisis de correlación. Esto permitió determinar qué tan relacionadas se encuentran las variables predictoras con la variable a predecir.

Los resultados obtenidos indicaron una correlación negativa muy débil tanto para el canal tradicional y moderno en las siguientes variables:

- Feriados

- Pago de quincena y fin de mes
- Promociones (Black Friday)

Así mismo, para la variable “Escasez de cartón” se observó una correlación negativa, sin embargo, en el Canal Tradicional presentó una correlación negativa débil en comparación con el Canal Moderno el cual tuvo una correlación negativa muy débil.

En el caso de la variable “Temporada de escasez de leche” se tuvo una correlación positiva. Hay que recalcar que para el Canal Moderno este resultado aumentó, indicando que su correlación se acerca a un nivel medio en comparación con el canal tradicional que se encontró en una correlación positiva muy débil. Los resultados se presentan en la tabla 1:

Tabla 1.

Coefficiente de correlación de las variables

	Feriados	Quincena y fin de mes	Temporada escasez leche	Escasez de cartón	Promociones (Black Friday)
CANAL TRADICIONAL	-0,18	-0,09	0,13	-0,07	-0,06
CANAL MODERNO	-0,18	-0,08	0,26	-0,18	-0,09

5.2. Variabilidad de las ventas

Es importante observar el comportamiento que tienen las ventas. El canal moderno evidenció una alta variación con respecto al canal tradicional en los dos productos analizados, generalmente se debe a nuevas estrategias de las grandes cadenas de supermercados y que no son notificadas a sus proveedores, por lo tanto, modifican cualquier pronóstico establecido con anterioridad sin previo aviso.

Para evaluar el nivel de variación que tienen las ventas, el coeficiente de variación fue útil para indicar que tan dispersos están los datos. Los resultados obtenidos se presentan en la tabla 2.

Tabla 2.

Coefficiente de variación de las ventas

CANAL MODERNO A1	CANAL TRADICIONAL A1	CANAL MODERNO C1	CANAL TRADICIONAL C1
------------------	----------------------	------------------	----------------------

Coefficiente de variación	27%	7%	34%	22%
---------------------------	-----	----	-----	-----

5.3. Automatización del pronóstico

Además de que se desarrolló el algoritmo de Redes Neuronales Recurrentes LSTM, mediante el lenguaje de programación Python, se creó una interfaz amigable con el usuario para realizar las predicciones de cualquier conjunto de datos, puesto que durante el entrenamiento se observó mejores resultados modificando los hiperparámetros de acuerdo a cada variable.

Esta interfaz permitió escoger el archivo que contiene la base de datos para realizar la predicción, además, fue posible seleccionar el tipo de producto necesario e inclusive se eligió manualmente la cantidad de datos que se destinaron a entrenamiento.

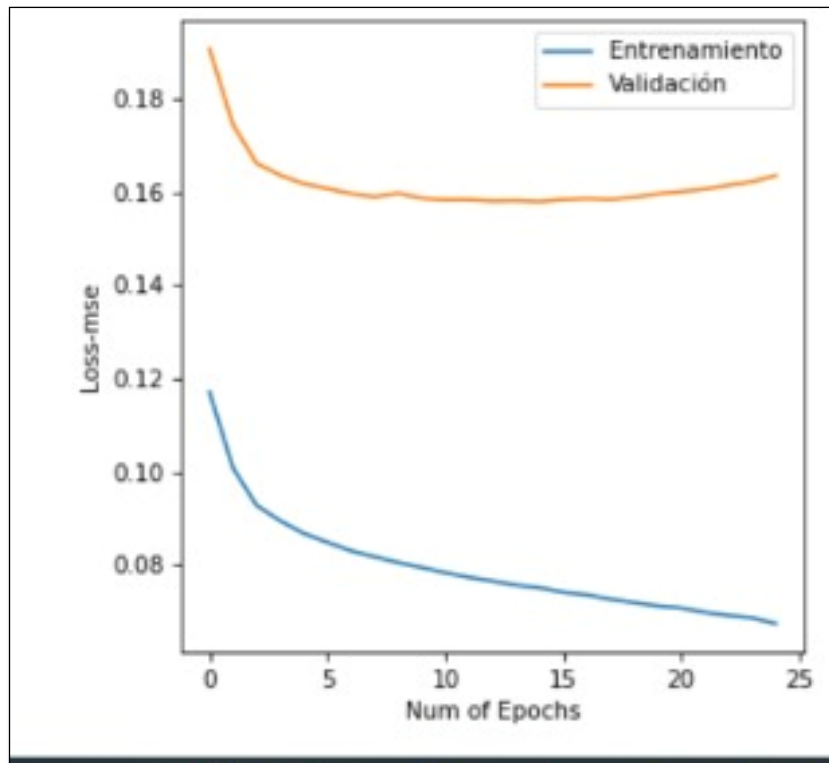
En cuanto a la modificación de hiperparámetros, permitió designar la cantidad de capas y neuronas por cada capa según corresponda a cada variable. La interfaz también permitió visualizar las métricas de evaluación del modelo correspondientes al RMSE, MAPE, MAD y Bias o sesgo. También se visualizó las gráficas en las cuales se aprecia el error generado según el número de iteraciones, muy importante a la hora de entrenar el modelo, ya que permitió estimar el número de entrenamientos necesarios según el error requerido. En una segunda gráfica se observó el ajuste de los datos de entrenamiento y prueba con los valores que pertenecen a la predicción. Es importante resaltar que, al tener los datos en una frecuencia semanal, fue posible digitar la cantidad de semanas pronosticadas. En los casos necesarios, también se hizo uso de un número mayor de capas.

Para verificar la brecha de aprendizaje entre el conjunto de entrenamiento y validación e identificar en qué época el modelo dejó de aprender y comenzó a incrementar el error, se analizó a través de la curva de aprendizaje del algoritmo el error generado según el número de épocas para cada conjunto.

La curva de aprendizaje se observa en la Ilustración 9.

Ilustración 9.

Curva de aprendizaje del modelo entrenado

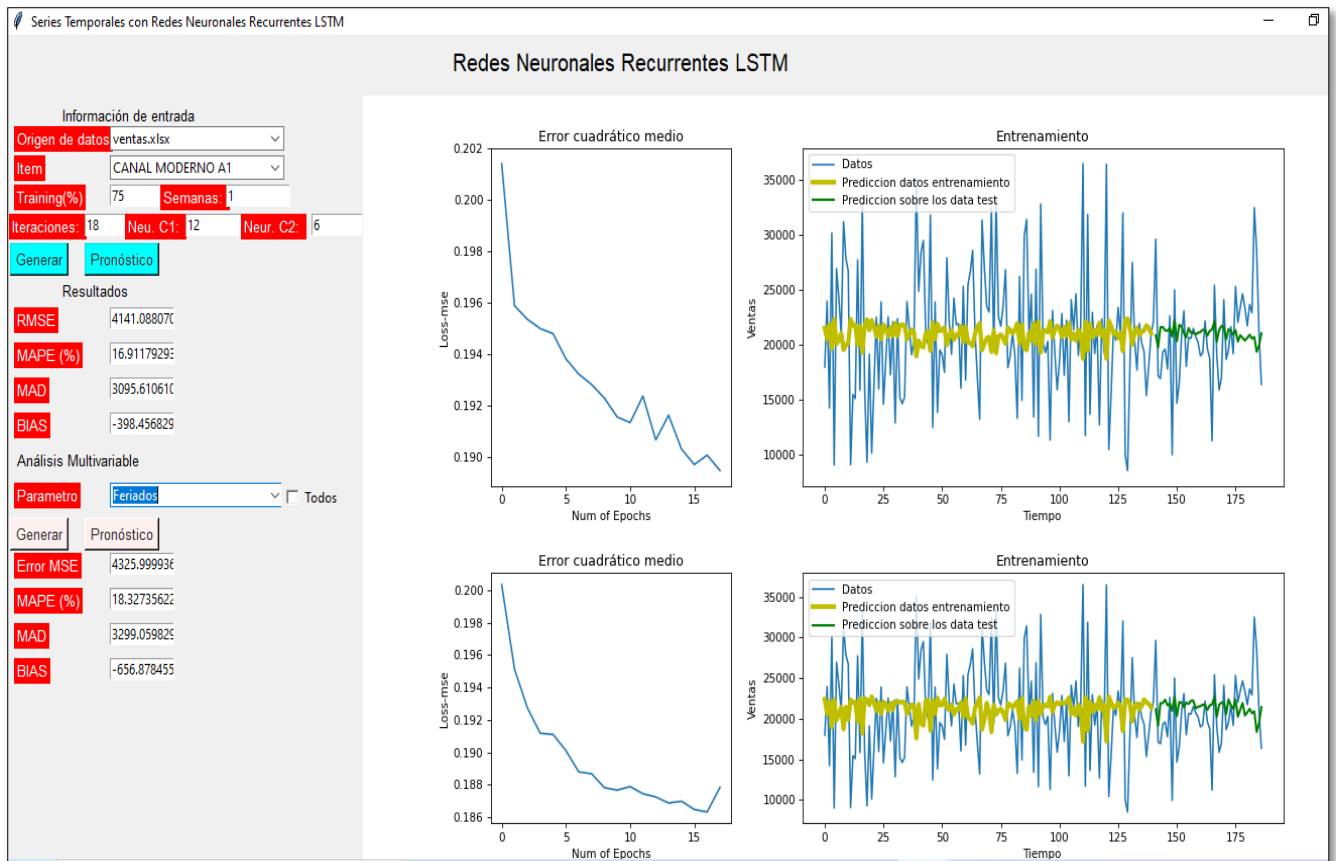


En este caso se puede observar que aproximadamente desde la época 20 en adelante el error en el conjunto de validación empezó a aumentar a diferencia de los datos de entrenamiento que disminuyeron considerablemente, esto muestra el número de épocas a las cuales se recomienda entrenar el modelo

Adicional, se realizó la predicción multivariable, se seleccionaron todas y cada una de las variables mencionadas con anterioridad, de igual forma se obtuvo los mismos datos de salida que para el análisis univariado. La interfaz se visualiza en la Ilustración 10.

Ilustración 10.

Interfaz para el entrenamiento del modelo



Este diseño facilitó a que se entrene nuevamente el conjunto de datos, hasta adaptarlo a los hiperparámetros requeridos tomando en cuenta las métricas de evaluación en cada entrenamiento.

5.4. Entrenamiento de la Red Neuronal Recurrente LSTM

Al realizar el entrenamiento hay que recordar que no se debe producir Overfitting, lo que generalmente sucede cuando hay un número alto de neuronas. Esto se evita disminuyendo el número de neuronas y aumentando el número de patrones. Por otro caso, cuando el número de neuronas es relativamente pequeño, sucede Underfitting, en este caso se tiene una red muy general con conceptos genéricos y cada neurona se satura de información, por lo que tampoco podrá aprender nuevos patrones. (Villa, Velásquez, & Sánchez, 2014)

Luego de que se realizó algunas iteraciones se obtuvieron varios datos de entrada que corresponden a los hiperparámetros necesarios para la creación del modelo de la red neuronal LSTM, en los cuales presentó un mínimo error en MAPE. Los hiperparámetros que se utilizaron se muestran en la tabla 3.

Tabla 3.

Hiperparámetros de los entrenamientos del modelo de proyección

	CANAL MODERNO A1					CANAL TRADICIONAL A1				
	HIPERPARÁMETROS					HIPERPARÁMETROS				
	N° de épocas	CAPA 1	CAPA 2	CAPA 3	capa 4 a 5	N° de épocas	CAPA 1	CAPA 2	CAPA 3	CAPA 4
		N° de neuronas	N° de neuronas	N° de neuronas	N° de neuronas		N° de neuronas	N° de neuronas	N° de neuronas	N° de neuronas
univariable	20	12	4			14	8	6		
Todas	13	10	9	7	5	10	8	4		
Ferriados	18	9	9	6	6	10	6	4		
quincena y fin de mes (1)	18	12	6			10	6	4		
temporada escasez leche	18	12	6			12	8	4		
escasez de carton	18	12	4			12	8	4		
promociones (black Friday) (1)	18	12	4			12	8	4		
	CANAL MODERNO C1					CANAL TRADICIONAL C1				
	HIPERPARÁMETROS					HIPERPARÁMETROS				
	N° de épocas	CAPA 1	CAPA 2	CAPA 3	capa 4 a 6	N° de épocas	CAPA 1	CAPA 2	CAPA 3	CAPA 4
		N° de neuronas	N° de neuronas	N° de neuronas	N° de neuronas		N° de neuronas	N° de neuronas	N° de neuronas	N° de neuronas
univariable	15	15	10	8	5	12	8	5		
Todas	13	8	5	3	2	10	6	4		
Ferriados	13	10	7	5	5	10	6	4		
quincena y fin de mes (1)	13	6	5	5	3	10	6	4		
temporada escasez leche	13	5	4	3	3	10	6	4		
escasez de carton	13	5	5	5		10	6	4		
promociones (black Friday) (1)	13	9	7	5	4	10	6	4		

Como se puede observar en el canal moderno, en el producto C1, fue necesario incluir un mayor número de capas con respecto al canal tradicional. Ya que el algoritmo necesita de más capas y por consiguiente más neuronas para almacenar la variedad de los patrones, aparentemente debido a que el coeficiente de variación supera el 30%. Además de esto, se debe también a la pequeña cantidad de datos con los que se entrenó el modelo, puesto que las redes neuronales están diseñadas para trabajar con una extensa base de datos. Estos resultados también tienen su origen en las variables predictoras utilizadas, ya que para algunas de ellas como la variable que corresponde a feriados se tuvo que aumentar la cantidad de neuronas y capas, algo similar sucede en el canal moderno A1.

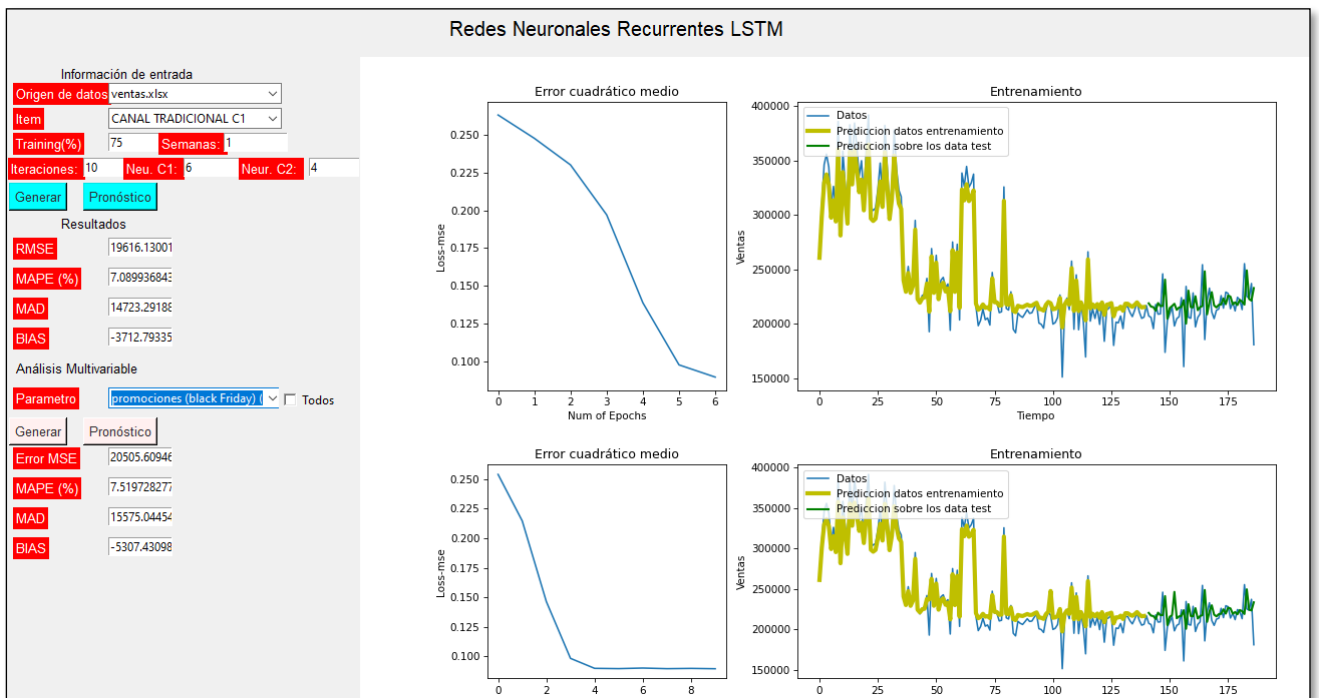
Además, en el canal moderno requirió de un mayor ajuste de hiperparámetros para la variable que corresponde a la temporada de escasez de leche y al pago de sueldos en quincena y fin de mes.

De forma general, también se observó que el número de épocas, capas y neuronas es mayor en el modelo univariado, en el cual la variable corresponde a la cantidad de ventas.

El análisis se realizó obteniendo los resultados de las métricas de evaluación con base en los hiperparámetros mencionados para los dos casos, es decir, para el caso donde se estudió las proyecciones con una sola variable (volumen de ventas) y para el caso donde se realizó las proyecciones con cada una de las variables endógenas y exógenas. La interfaz permitió observar los resultados para el análisis univariable y multivariable. Como se puede observar en la Ilustración 11, la primera gráfica sirve de apoyo, para observar el número de iteraciones necesarias con las cuales el entrenamiento logre el mínimo error. En la segunda grafica de la ilustración mencionada anteriormente se visualiza el ajuste que tiene la predicción realizada por el modelo sobre los datos de entrenamiento y prueba. Estos hiperparámetros permitieron obtener resultados para cada una de las variables. En la ilustración 11 se puede visualizar el entrenamiento que se realizó para el Canal Tradicional C1, para iterar el modelo se tiene los datos correspondientes al número de neuronas, número de capas, ítem, porcentaje de datos de entrenamiento, los resultados de cada métrica y las gráficas. En cada entrenamiento se escogió la variable a utilizar en la sección del análisis multivariable. El entrenamiento se realizó para cada ítem y a su vez con cada una de las variables independientes.

Ilustración 11

Resultados obtenidos del entrenamiento



Los datos obtenidos de cada ítem fueron recopilados para su mejor comprensión en tablas según el tipo de variable evaluada. En la tabla 4 se muestra los resultados correspondientes a los dos ítems de

los dos tipos de canales a través del análisis univariable, es decir, a la variable que corresponde a la cantidad de ventas.

Tabla 4.

Resultados de las métricas de evaluación para el modelo univariado

UNIVARIABLE				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4133	16,83	3092	-370
CANAL TRADICIONAL A1	10233	6,4	7431	-3135
CANAL MODERNO C1	12405	41,8	10801	-4499
CANAL TRADICIONAL C1	20702	7,6	15800	-4314

En la tabla 5 se presenta los resultados que se obtuvieron en el análisis de todas las variables endógenas y exógenas dentro del mismo entrenamiento. No representa el promedio de los demás entrenamientos.

Tabla 5.

Resultados de las métricas de evaluación para el modelo multivariado.

VARIABLE: TODAS				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4130	16,56	3086	-329
CANAL TRADICIONAL A1	10169	6,2	7265	-2953
CANAL MODERNO C1	12819	41,09	10659	-3953
CANAL TRADICIONAL C1	19952	7	17409	-4974

En la tabla 6 se observa los resultados correspondientes al entrenamiento que se realizó con la variable de ventas y la variable independiente “Feriados” la misma que corresponde a una de las variables exógenas.

Tabla 6.

Resultados de las métricas de evaluación para el modelo que incluye la variable "Feriados".

VARIABLE: FERIADOS				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4127	16,6	3080	-295
CANAL TRADICIONAL A1	10178	6,2	7295	-2837
CANAL MODERNO C1	12314	40,92	10636	-4411
CANAL TRADICIONAL C1	19592	6,9	14495	-2647

En la tabla 7 se observa los resultados correspondientes al entrenamiento que se realizó con la variable de ventas y la variable independiente “Quincena y fin de mes” la misma que corresponde a una de las variables exógenas.

Tabla 7.

Resultados de las métricas de evaluación para el modelo que incluye la variable "Pago de quincena y fin de mes".

VARIABLE: QUINCENA Y FIN DE MES				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4062	16,09	2339	-439
CANAL TRADICIONAL A1	10046	6,1	7165	-2944
CANAL MODERNO C1	12319	41,02	10657	-4452
CANAL TRADICIONAL C1	19732	7,1	15040	-4070

En la tabla 8 se observan los resultados correspondientes al entrenamiento que se realizó con la variable de ventas y la variable independiente “Temporada de escasez de leche” la misma que corresponde a una de las variables exógenas.

Tabla 8.

Resultados de las métricas de evaluación para el modelo que incluye la variable "Temporada de escasez de leche".

VARIABLE: TEMPORADA ESCASEZ DE LECHE				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4126	16,58	3078	-270
CANAL TRADICIONAL A1	10209	6,2	7326	-3,066
CANAL MODERNO C1	12401	41,54	10765	-4605
CANAL TRADICIONAL C1	21372	7,5	16110	-3616

En la tabla 9 se observa los resultados correspondientes al entrenamiento que se realizó con la variable de ventas y la variable independiente “Temporada de escasez de cartón” la misma que corresponde a una de las variables exógenas.

Tabla 9.

Resultados de las métricas de evaluación para el modelo que incluye la variable "Temporada de escasez de cartón".

VARIABLE: ESCASEZ DE CARTÓN				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4127	16,7	3085	-322
CANAL TRADICIONAL A1	10133	6,1	7221	-2711
CANAL MODERNO C1	12650	41,25	10721	-4220
CANAL TRADICIONAL C1	19616	7	14762	-2943

En la tabla 10 se observa los resultados correspondientes al entrenamiento que se realizó con la variable de ventas y la variable independiente “Promociones (Black Friday)” la misma que corresponde a una de las variables endógenas, la cual es provocada por la organización.

Tabla 10.

Resultados de las métricas de evaluación para el modelo que incluye la variable "Promociones (Black Friday)".

VARIABLE: PROMOCIONES (BLACK FRIDAY)				
PRODUCTO	MÉTRICAS DE EVALUACIÓN			
	RMSE	MAPE (%)	MAD	BIAS
CANAL MODERNO A1	4122	16,5	3071	-211
CANAL TRADICIONAL A1	10195	6,2	7326	-2896
CANAL MODERNO C1	12376	41,24	10702	-4489
CANAL TRADICIONAL C1	20505	7,4	15575	-5307

Como se pudo observar, los resultados obtenidos para todos los casos expuestos atribuyen un valor alto en MAPE para el canal moderno comparado con el canal tradicional, esto supuso su origen en el nivel de variación de los datos, ya que como se mostró anteriormente el canal moderno tiene un coeficiente de variación de 34% para el ítem C1 y 27% para el ítem A1. Esta variación generó la necesidad de aumentar el número de capas y así mismo el número de neuronas, e inclusive esto incurrió en un aumento en el tiempo del procesamiento de los datos.

Añadir un número mayor de capas, neuronas e iteraciones al que fue presentado en la tabla 3 genera un mayor consumo de tiempo de entrenamiento sin obtener mejoras significativas en las métricas de evaluación, sin embargo, si se extiende el conjunto de datos estas predicciones pueden mejorar. Inclusive se puede mencionar que las variables predictoras que se analizaron no presentaron una asociatividad fuerte con respecto a las ventas, por lo que no se logró reducir aún más el error en MAPE.

Hay que tener en cuenta que si bien la métrica de evaluación MAPE, al ser universal es la más utilizada, no siempre puede ofrecernos el mejor resultado en cuanto a las evaluaciones del pronóstico ya que puede favorecer a ciertas predicciones que estén por debajo de los valores reales. Al momento

de comparar el canal moderno y tradicional del mismo ítem, el resultado de la métrica MAPE es menor en el canal tradicional en contraste al canal moderno.

La métrica que corresponde al RMSE se registró de manera contraria, es decir, el valor es mayor para el canal tradicional que para el canal moderno. Esta métrica penaliza fuertemente los errores con valores grandes al elevarlos al cuadrado, por lo que se puede verificar la precisión del modelo. De igual manera se evidenció este hecho con la métrica correspondiente a la Desviación Media Absoluta (MAD) y con el sesgo o Bias.

5.5. MAPE generado en el modelo univariado frente al modelo multivariable

De manera semejante se realizó un consolidado de los valores obtenidos en MAPE, se observa en la tabla 11. Aquí se indican los resultados de cada canal, diferenciados con el uso de una sola variable (volumen de ventas) y por el contrario con el entrenamiento multivariable, en los cuales se especifica los datos de las variables endógenas y exógenas, así como también, el entrenamiento que se realizó con todas las variables independientes.

Tabla 11.

Resultados del MAPE del modelo univariado frente al modelo multivariable

	UNIVARIABLE	MULTIVARIABLE		
		TODAS	VARIABLES ENDÓGENAS	VARIABLES EXÓGENAS
	MAPE	MAPE	MAPE	MAPE
CANAL MODERNO A1	16,83	16,56	16,5	16,49
CANAL TRADICIONAL A1	6,4	6,2	6,2	6,15
CANAL MODERNO C1	41,8	41,09	41,24	41,18
CANAL TRADICIONAL C1	7,6	7	7,4	7,13

Para el caso del canal tradicional C1, al utilizar todas las variables en el mismo entrenamiento el error se redujo de un 7.6% a un 7%, lo cual representa una diferencia del 0.6 o una disminución del 8% con respecto al 7.6%. Para el evento en el cual se entrenó con las variables endógenas el error llega al 7.4%, lo que indica una reducción del 3%. En cuanto a las variables exógenas, el error que se generó es de 7.1%, esto registra una reducción del 3%. En términos porcentuales, los modelos que incluyen variables independientes tanto endógenas como exógenas, redujeron el error en MAPE entre un 2% a

un 4%, tomando en cuenta el promedio de los errores tanto para el canal moderno y el tradicional. De manera similar se obtuvo con el resto de ítems. Este análisis se observa en la tabla 12.

Tabla 12.

Reducción porcentual del error generado en MAPE en los dos modelos, univariado vs multivariado

	MULTIVARIABLE		
	TODAS	VARIABLES ENDÓGENAS	VARIABLES EXÓGENAS
	MAPE	MAPE	MAPE
CANAL MODERNO A1	2%	2%	2%
CANAL TRADICIONAL A1	3%	2%	2%
CANAL MODERNO C1	2%	1%	1%
CANAL TRADICIONAL C1	8%	3%	6%
PROMEDIO	4%	2%	3%

Es importante resaltar que, para el canal moderno, el cual tuvo mayor variabilidad en su demanda, la reducción del error representó un 2%, lo cual evidenció que existe variables independientes que pueden tener mayor injerencia en este tipo de ventas comparadas con las variables utilizadas en este estudio, ya que para el canal tradicional la reducción del error va desde un 2% a un 6%, lo que es un indicio que las variables utilizadas tuvieron mucha más asociatividad con el comportamiento de este tipo de ventas y por lo tanto lograron mejores resultados en cuanto a la disminución del error por lo que se generó una mayor precisión en el pronóstico. Lo mencionado se observa en la tabla 13.

Tabla 13.

Reducción del MAPE por cada canal (moderno y tradicional)

CANAL MODERNO				CANAL TRADICIONAL			
	MULTIVARIABLE				MULTIVARIABLE		
	TODAS	VARIABLES ENDÓGENAS	VARIABLES EXÓGENAS		TODAS	VARIABLES ENDÓGENAS	VARIABLES EXÓGENAS
	MAPE	MAPE	MAPE		MAPE	MAPE	MAPE
CANAL MODERNO A1	2%	2%	2%	CANAL TRADICIONAL A1	3%	2%	2%
CANAL MODERNO C1	2%	1%	1%	CANAL TRADICIONAL C1	8%	3%	6%
PROMEDIO	2%	2%	2%	PROMEDIO	6%	2%	4%

Como se mencionó anteriormente, el modelo de redes neuronales trabaja con una extensa cantidad de datos, esto facilita el reconocimiento de la mayor cantidad de patrones o comportamiento de las ventas, no obstante, con la cantidad de datos que se trabajó se demuestra la reducción del MAPE con la utilización de un modelo multivariado, y en este caso según lo observado en la Tabla 13, las variables que se utilizaron se ajustan mejor al canal tradicional.

6. Conclusiones

Se ha demostrado que el uso de variables endógenas y exógenas en los modelos de predicción de ventas mejora los resultados obtenidos con respecto a los pronósticos univariados, en este caso aporta con reducciones del error que van desde un 1% hasta un 8%, tomando en cuenta las condiciones de evaluación en las cuales interfiere la cantidad de datos usados para el entrenamiento. En este caso, a pesar de tener 180 datos, se evidenció la reducción del error con el uso de este modelo. Sin embargo, un aumento de la cantidad de datos, que permita que el modelo reconozca más patrones, puede reducir el error y mejorar las predicciones considerablemente.

Al mismo tiempo, se ha identificado que las variables independientes utilizadas en este estudio no presentan una clara asociación con la variable a predecir, específicamente en el canal moderno, el cual tiene una alta variabilidad. No obstante, estas mismas variables predictoras tienen mayor injerencia en el canal tradicional ya que reduce el MAPE en promedio de un 2% a un 6%. No sucede lo mismo en el canal moderno, ya que a pesar de que se logre una reducción en el MAPE, en promedio de un 2%, este porcentaje no aumenta.

Además, se observó que los resultados obtenidos correspondientes a las métricas de evaluación del MAPE, indican que no necesariamente es la medida más adecuada para evaluar el modelo, ya que para porcentajes pequeños se obtuvo resultados que indican un sesgo demasiado amplio, e inclusive valores para RMSE mucho más pequeños en comparación con valores del MAPE que sobrepasan el 15% contrastando dos tipos de canales del mismo producto, el canal tradicional y el canal moderno.

También se identificó que se requiere un mayor ajuste de hiperparámetros en el entrenamiento univariado con respecto al modelo multivariado. La variable que corresponde a los feriados del país requiere una mayor cantidad de épocas, capas y neuronas en los dos productos que corresponden al canal moderno, de la misma manera se realiza el ajuste para el canal moderno en contraste con el canal tradicional que requiere un menor procesamiento por la cantidad que se asigna a estos hiperparámetros.

En síntesis, se cumplió con los objetivos propuestos demostrando la reducción del error en MAPE, al incluir variables predictoras independientes en el modelo de la Red Neuronal Recurrente LSTM, sin embargo, se recomienda para posteriores investigaciones, analizar variables que tengan una mayor

asociación con la variable a predecir para el canal moderno con ventas que tienen un coeficiente de variación alto.

7. Agradecimientos

Extiendo mis más sinceros agradecimientos al Ingeniero Zhunio, por la apertura con la información necesaria para la elaboración del presente estudio.

Así como también un agradecimiento muy especial a mi director de trabajo de titulación, el Ingeniero Rodrigo Guamán y al Ingeniero Mario Peña por su apoyo brindado para el desarrollo y culminación de este trabajo.

Por último, agradezco a mis padres y amistades que estuvieron presentes durante todo este trayecto y que fueron una pieza fundamental para alcanzar mis metas propuestas.

8. Bibliografía

- Arroyo, M. (2014). *Crítica Económica*. Recuperado el 22 de 07 de 2022, de <https://campusvirtualdocente.continental.edu.pe/sitios/GestionDocentes/CV000020022437/112ActividadesInvestigacionPublicadas/190-articulo%20marco%20critica%20economica%202014r.pdf>
- Aucapiña, J. (2018). *Modelo de elaboración de pronóstico de ventas mediante el uso de*. Recuperado el 15 de 08 de 2022, de <https://dspace.ucuenca.edu.ec/bitstream/123456789/31256/1/Trabajo%20de%20titulaci%C3%B3n.pdf>
- Bueno, F. (2019). *Redes Neuronales: Entrenamiento y Comportamiento*. Recuperado el 25 de 09 de 2022, de <https://eprints.ucm.es/id/eprint/64564/1/BUENOPASCUALFERNANDO.pdf>
- Burgaentzle Jarrín, F. (Diciembre de 2016). *Pronósticos y modelos de inventarios en las industrias de alimentos: caso de estudio de una empresa láctea ecuatoriana*. Obtenido de <https://repositorio.usfq.edu.ec/bitstream/23000/6226/1/128660.pdf>
- Dinamarca, A. (2018). *Aprendizaje y Análisis de Redes*. Recuperado el 28 de 09 de 2022, de https://bdigital.uncu.edu.ar/objetos_digitaes/13989/dinamarca-agustina-tesina.pdf

- Dorado, S., & Velásquez, V. (Junio de 2020). *Pronóstico de demanda utilizando inteligencia artificial*. Recuperado el 20 de 08 de 2022, de https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/87548/1/TG03014.pdf
- Eguez, G. (2020). *Análisis del desempeño de redes neuronales artificiales en la reconstrucción de datos pluviométricos de la ciudad de Quito*. Recuperado el 25 de 09 de 2022, de <https://repositorio.uasb.edu.ec/bitstream/10644/7461/1/T3258-MCCNA-Eg%c3%bcez-Analisis.pdf>
- Ganapathi, N., Prasanna, k., Mahesh, V., Kalindindi, A., & Padma, V. (2020). *Study the influence of normalization/transformation process on the accuracy of supervised classification*. Recuperado el 01 de 10 de 2022, de [https://sci-hub.hkvisa.net/10.1109/icssit48917.2020.9214160](https://scihub.hkvisa.net/10.1109/icssit48917.2020.9214160)
- Ganchozo Morcillo, R. J., & Vera Mera, A. X. (Agosto de 2019). *Análisis del impacto del modelo de Industria 4.0 aplicada a la producción de la línea marrón*. Obtenido de http://repositorio.ug.edu.ec/bitstream/redug/44390/1/ANALISIS_DEL_IMPACTO_DEL_MODEL_DELO_DE_INDUSTRIA_40_APLICADA_A_LA_PRODUCCION_DE_LA_LINEA_MARRON_DE_LA_INDUSTRIA_ECUATORIANA-convertido.pdf
- Garcés Camacho, H. S. (2020). *Aplicación de Machine Learning en la demanda de productos*. Obtenido de <http://201.159.223.180/bitstream/3317/14264/1/T-UCSG-PRE-ECO-GES-621.pdf>
- Hanke, J., & Wichern, D. (2010). *Pronósticos en los negocios*. México: Pearson Educación.
- Jaramillo Mira, A., & Gutierrez Muriel, J. P. (2021). *Pronóstico de ventas de comestibles de Corporación Favorita, Reto Kaggle*. Obtenido de https://bibliotecadigital.udea.edu.co/bitstream/10495/25073/12/JuanPabloGutierrezMuriel_2021_PrediccionVentasUnidades.pdf
- Llunitasig Galarza, M. C. (2021). *Simulación de pronósticos de ventas en la empresa Impactex mediante Redes Neuronales*. Obtenido de <https://repositorio.uta.edu.ec/bitstream/123456789/33778/1/t1892mma.pdf>
- López Zúñiga, E., & Jaimes Campos, D. L. (2021). *Modelo de forecast para predecir la demanda semanal de alimentos y bebidas de consumo masivo*. Obtenido de <https://repositorio.uniandes.edu.co/bitstream/handle/1992/55819/25416.pdf?sequence=1&isAllowed=y>
- Medina, F., & Galván, M. (2007). *Imputación de datos: teoría y práctica*. Recuperado el 15 de 07 de 2022, de https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590_es.pdf

- Molina Rea, K. G. (2020). *Implementación de un modelo analítico para la predicción de la venta del portafolio de productos OTC de un Laboratorio Farmacéutico*. Obtenido de <http://repositorio.espe.edu.ec/xmlui/bitstream/handle/21000/22561/T-ESPE-043875.pdf?sequence=1&isAllowed=y>
- Montemayor, J. E. (2013). *Métodos de pronósticos para negocios*. Recuperado el 10 de 08 de 2022, de <http://prod77ms.itesm.mx/podcast/EDTM/P196.pdf>
- Mora, D. (2018). *Utilización de Redes Neuronales Artificiales (RNA) como alternativa para*. Recuperado el 30 de 07 de 2022, de <http://dspace.ucuenca.edu.ec/bitstream/123456789/31730/1/Tesis.pdf>
- Ridzuan, F., & Wan Zainon, W. M. (2019). *Science Direct*. Recuperado el 15 de 07 de 2022, de <https://reader.elsevier.com/reader/sd/pii/S1877050919318885?token=3ED581A3D611CF78F2C23A91192D4BB5F73A79D18D8872DDB154A26B34F062F2B5D9E8FA99AA7C299A47DE8A5F24308B&originRegion=us-east-1&originCreation=20220815032422>
- Sanchez, C. (01 de 2017). *"Medición para la precisión de los pronósticos de ventas y su efecto en los costos de la empresa"*. Recuperado el 15 de 08 de 2022, de https://repositorio.unab.cl/xmlui/bitstream/handle/ria/3167/a117570_Sanchez_C_Medicion_en_la_precision_2017_Tesis.pdf?sequence=1
- Soler Flores, F. J. (2008). *Tratamiento de outliers en los modelos de predicción de*. Recuperado el 15 de 07 de 2022, de https://oa.upm.es/4524/1/INVE_MEM_2008_61058.pdf
- Velasteguí Sandoval, R. X. (2020). *Sales Forecast by using Convolutional Neural Networks*. Obtenido de <https://repositorio.yachaytech.edu.ec/bitstream/123456789/118/1/ECMC0014>
- Villa, F., Velásquez, J., & Sánchez, P. (2014). *Control del sobreajuste en redes neuronales tipo cascada correlación aplicado a la predicción de precios de contratos de electricidad*. Recuperado el 01 de 10 de 2022, de <http://www.scielo.org.co/pdf/rium/v14n26/v14n26a11.pdf>
- Ying, X. (2019). *An Overview of Overfitting and its Solutions*. Recuperado el 16 de 09 de 2022, de <https://iopscience.iop.org/article/10.1088/1742-6596/1168/2/022022/pdf>