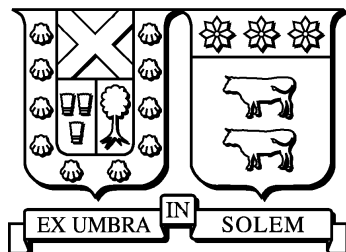


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INFORMÁTICA
SANTIAGO – CHILE



“COMPROBAR LA FACTIBILIDAD DE
PREDICCIÓN DE ESTACIONALIDAD DE VENTA
DE PRODUCTOS EN UNA TIENDA RETAIL.”

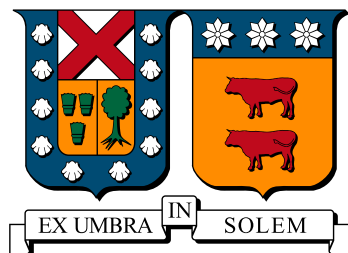
CATHERIN LISSET VARGAS QUINTRIQUEO

MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INFORMÁTICO

PROFESOR GUÍA: LIOUBOV DOMBROVSKAIA

DECIEMBRE 2022

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO – CHILE



**“COMPROBAR LA FACTIBILIDAD DE
PREDICCIÓN DE ESTACIONALIDAD DE
VENTA DE PRODUCTOS EN UNA TIENDA
RETAIL.”**

CATHERIN LISSET VARGAS QUINTRIQUEO

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INFORMÁTICO**

PROFESOR GUÍA: LIOBOV DOMBROVSKAIA

PROFESOR CORREFERENTE: RICARDO ÑANCULEF

DECIEMBRE 2022

MATERIAL DE REFERENCIA, SU USO NO INVOLUCRA RESPONSABILIDAD DEL AUTOR O DE LA INSTITUCIÓN

Agradecimientos

Quiero agradecer especialmente a mi madre María Quintriqueo. Gracias por la entrega y paciencia. Siempre apoyarme y preocuparse por mí y de mi educación.

También agradecer a mis amigos, por soportar las ausencias y el mal humor, darme ánimo cuando lo necesitaba y también llamarme la atención si era necesario.

Quiero agradecer la ayuda y apoyo de la profesora Liubov Dombrovskaia, por siempre insistir en que termine esta etapa, disculpar las ausencias y altos y bajos durante este proceso. Siempre con palabras de aliento y presionar si era necesario.

Finalmente y no menos importante agradecer al jefe de carrera, el profesor José Luis Martí, ya que desde que entré a la universidad lo he ido a molestar con dudas o consultas, encontrando siempre respuestas y soluciones.

Resumen

Dentro de cualquier negocio es necesario saber cuánta será la demanda, ya que esto permite preparar la cantidad adecuada y logística de transporte para poder proveerla. Predecir cuándo y qué producto se va a vender en una determinada fecha, se denomina la estacionalidad de productos o forecasting.

El objetivo del presente trabajo es analizar la factibilidad de la predicción de las ventas diarias de categorías de productos ya sea sin y con presencia de eventos especiales. Los datos de ventas diarias de 30 categorías de productos fueron recuperados y agregados a partir de las ventas de productos particulares.

Cuatro modelos predictivos fueron evaluados usando accuracy y error RMSE: regresión lineal, series de tiempo, redes neuronales convolucionales y redes neuronales LSTM. Cada uno de los modelos fue entrenado y evaluado en cada una de las categorías.

En la gran mayoría de las categorías, el mejor resultado ha sido logrado con modelos de redes neuronales LSTM en el caso donde se eliminan las ventas correspondientes a eventos especiales de cyber days, aunque su precisión ronda el 60 %.

Keywords: Forecasting, Métodos predictivos, Redes neuronales.

Abstract

Within any business it is necessary to know how much the demand will be, as this allows to prepare the adequate quantity and transportation logistics to provide it. Predicting when and which product will be sold on a certain date is called the seasonality of products or forecasting.

The objective of the present work is to analyze the feasibility of forecasting daily sales of product categories either without and with the presence of special events. Daily sales data for 30 product categories were retrieved and aggregated from the sales of particular products.

Four predictive models were evaluated using accuracy and RMSE error: linear regression, time series, convolutional networks and LSTM neural networks. Each of the models was trained and evaluated in each of the categories.

In the vast majority of categories, the best result has been achieved with LSTM neural network models in the case of eliminating sales corresponding to special cyber days events, although their accuracy is around 60 %.

Keywords: Forecast, Predictive methods, Neural networks.

Índice de Contenidos

| | |
|--|-------------|
| Agradecimientos | III |
| Resumen | IV |
| Abstract | V |
| Índice de Contenidos | VI |
| Lista de Tablas | VIII |
| Lista de Figuras | X |
| Glosario | XI |
| Introducción | 1 |
| 1. Definición del Problema | 3 |
| 2. Estado del Arte | 5 |
| 2.1. Regresión lineal | 5 |
| 2.1.1. Tipos de regresión lineal | 6 |
| 2.2. Series de tiempo | 6 |
| 2.3. Redes neuronales | 8 |

| | | |
|-----------|--|-----------|
| 2.3.1. | Redes neuronales artificiales | 9 |
| 2.3.2. | Redes neuronales convolucionales | 14 |
| 2.3.3. | Redes neuronales recurrentes | 15 |
| 2.4. | Comparación de modelos | 18 |
| 2.5. | CRISP-DM | 20 |
| 3. | Propuesta de Solución | 22 |
| 3.1. | Metodología | 22 |
| 3.1.1. | Comprensión de los datos | 23 |
| 4. | Resultados | 25 |
| 4.1. | Regresión lineal | 26 |
| 4.2. | Series de tiempo | 29 |
| 4.3. | Red convolucional | 32 |
| 4.4. | Red neuronal LSTM | 34 |
| 4.5. | Análisis de RMSE | 37 |
| 4.6. | Análisis de <i>Forecast accuracy</i> | 38 |
| | Conclusiones | 39 |
| 4.7. | Trabajo futuro | 41 |
| | Anexo | 42 |
| | Bibliografía | 47 |

Índice de cuadros

| | |
|---|----|
| 4.1. RMSE promedio de las 4 categorías estudiadas (para revisar la tabla completa ver el anexo 4.4). | 37 |
| 4.2. <i>Forecast accuracy</i> promedio de las 4 categorías estudiadas (para revisar la tabla completa ver el anexo 4.7) | 38 |
| 4.3. <i>Forecast accuracy</i> y RMSE promedio para los 4 modelos. | 39 |
| 4.4. Tabla completa de RMSE de la categoría refrigeración y calefacción. | 42 |
| 4.5. Tabla completa de RMSE de la categoría electrodomésticos. | 43 |
| 4.6. Tabla completa de RMSE de la categoría box spring. | 44 |
| 4.7. Tabla completa de <i>forecast accuracy</i> de la categoría refrigeración, calefacción y box spring. | 45 |
| 4.8. Tabla completa de <i>forecast accuracy</i> de la categoría electrodomésticos. | 46 |

Índice de figuras

| | |
|--|----|
| 2.1. Ilustración de una neurona biológica. | 9 |
| 2.2. Neurona artificial como modelo idealizado y simplificado de una neurona biológica. | 10 |
| 2.3. Arquitectura de una red neuronal forward. | 13 |
| 2.4. Esquema de convolución unidimensional, donde el kernel de tamaño 3 (borde rojo) realiza una operación de convolución a cada secuencia de 3 valores de la entrada. | 14 |
| 2.5. Al comienzo una neurona recurrente, luego la neurona recurrente desarrollada en el tiempo. | 15 |
| 2.6. Al comienzo una capa recurrente, luego la capa recurrente desarrollada en el tiempo. | 16 |
| 2.7. Fases del modelo CRISP-DM. | 20 |
| 4.1. Regresión lineal sobre los datos de venta de refrigeradores de más de 400L. | 26 |
| 4.2. Regresión lineal sobre los datos de venta de refrigeradores de más de 400L sin datos <i>cyber</i> | 27 |
| 4.3. Regresión lineal sobre los datos de venta de refrigeradores de más de 400L con y sin datos <i>cyber</i> | 28 |

| | |
|--|----|
| 4.4. Serie de tiempo sobre los datos de venta de refrigeradores de más de 400L. | 29 |
| 4.5. Serie de tiempo sobre los datos de venta de refrigeradores de más de 400L sin datos <i>cyber</i> . | 30 |
| 4.6. Serie de tiempo sobre los datos de venta de refrigeradores de más de 400L con y sin <i>cyber</i> . | 31 |
| 4.7. Red neuronal convolucional sobre los datos de venta de refrigeradores de más de 400L. | 32 |
| 4.8. Red neuronal convolucional sobre los datos de venta de refrigeradores de más de 400L sin datos <i>cyber</i> . | 33 |
| 4.9. Red neuronal convolucional sobre los datos de venta de refrigeradores de más de 400L con y sin datos <i>cyber</i> . | 33 |
| 4.10. Red neuronal LSTM sobre los datos de venta de refrigeradores de más de 400L. | 34 |
| 4.11. Red neuronal LSTM sobre los datos de venta de refrigeradores de más de 400L sin <i>cyber</i> . | 35 |
| 4.12. Red neuronal LSTM sobre los datos de venta de refrigeradores de más de 400L con y sin <i>cyber</i> . | 36 |

Glosario

- **ANN:** Artificial neural networks.
- **CNN:** Convolutional Neural Networks.
- **RNN:** Recurrent neural networks
- **LSTM:** Long Short Term Memory

Introducción

Un elemento fundamental para una buena planeación en cualquier tipo de negocio consiste en poder pronosticar con la mayor exactitud posible la demanda de un producto.

El pronóstico de la demanda, también llamado forecast, consiste en estimar las ventas de un producto en un determinado periodo futuro. En general, se calcula la demanda de toda la industria o mercado para luego predecir las ventas de los productos de la compañía.

El pronóstico de la demanda da origen a varios tipos de proyecciones. Por ejemplo, uno puede referirse a una industria, a una línea o categoría de productos o bien a una marca particular.

Se dice que las ventas de un determinado producto muestran estacionalidad cuando se produce una variación periódica y predecible de su demanda en un periodo de tiempo o estación. Este suele contener un patrón que se repite anual, semanal, o diariamente.

La estacionalidad es común relacionarla con las estaciones del año. Por ejemplo, en navidad aumenta la demanda de productos que serán regalados, en vacaciones aumenta la cantidad de huéspedes en los hoteles, en San Valentín aumenta la demanda de rosas y chocolates. Las festividades son factores que pueden resultar determinantes en la estacionalidad y temporalidad de un producto o servicio.

En este trabajo se desarrolla la comparación de distintos modelos de predicción para la venta de productos y así estudiar su estacionalidad, entre estos modelos se encuentra regresión lineal, series de tiempo y redes neuronales.

El documento a continuación presenta la definición del problema y objetivo principal a desarrollar. Posteriormente se expone el marco conceptual, en el cual se describen diferentes modelos que ayudarán a tener una visión global, además se encontrarán las herramientas para alcanzar el objetivo propuesto. Luego se expone la propuesta y desarrollo de este trabajo, donde se realizan pruebas con cada uno de los modelos elegidos, se realiza la evaluación y comparación de los modelos. Finalmente se entregan las conclusiones del trabajo realizado.

Capítulo 1

Definición del Problema

Este trabajo se enfoca en el estudio y predicción de la demanda de diferentes productos de un *retail*, ésta se ve afectada por la época o estación del año, incluidas las festividades, también por la cantidad de canales de ventas que existen hoy en día, como tienda física o internet.

Las ventas de productos a través de internet van en aumento, el conocido *e-commerce*, que significa llevar a cabo transacciones (compra/venta) de productos o servicios mediante medios electrónicos. Junto con esto surge un nuevo concepto para incrementar las ventas, el *Marketplace*.

Al *marketplace* se le puede entender como un local de compras, una especie de vitrina virtual que presenta a los clientes productos de diversas marcas o empresas donde pueden llevar a cabo una transacción comercial.

La diferencia entre el *marketplace* con el concepto de *e-commerce* tradicional, es que la plataforma de la empresa que lo implementa, no sólo venderá los productos que pudo haber adquirido al por mayor, sino que servirá como plataforma para que otros vendedores puedan ofertar sus productos, dando respaldo (confianza), prestándole sus servicios logísticos en algunos casos, como también servicios de *marketing*.

Dado que un *retail* puede tener todas las características definidas anteriormente, es importante conocer la demanda de un producto, una marca o categoría en específico. Entre las cosas

que se pueden mejorar dentro de la compañía o negocio se encuentran:

- Campañas de marketing más efectivas según la fecha y canal de compra.
- Conseguir que el cliente que ya haya comprado un producto con anterioridad, siga comprando, que se vuelva un cliente habitual dado que la compañía tiene los productos de estación. A esto se le conoce como fidelización.
- Tener un inventario controlado, debe entrar la cantidad de mercancía necesaria para la cantidad de ventas que se tendrá, para así evitar pérdida de ventas. Si se tiene exceso de productos se incurre en gastos de almacenamiento y transporte por un producto que sólo se mantendrá en bodega.

Otro factor importante por el cual es necesario conocer la estacionalidad, es que dentro del *retail*, la publicación o disponibilidad de un producto incluye plazos, ya que según el tipo de producto o medio por el cual se disponibilizará, se debe revisar la información provista por el proveedor, certificaciones, si cumple con las regularizaciones vigentes en el país, si las fotografías cumplen con los estándares, etc.

Por todo lo anterior es que se define al objetivo principal de esta memoria como:

”Comprobar la factibilidad de predicción de estacionalidad de venta de productos en una tienda *retail*.”

Además, se definen los siguientes objetivos secundarios:

- Recolectar y comprender los datos disponibles por parte del negocio.
- Clasificar, limpiar y transformar los datos para construir el dataset final.
- Diseñar y aplicar técnicas de modelado que sean pertinentes para descubrir la estacionalidad de los productos.
- Evaluar y ajustar el modelo para obtener resultados sobre la estacionalidad de productos.

Capítulo 2

Estado del Arte

En este capítulo se expone un repaso de los métodos más relevantes y utilizados para realizar pronósticos, además se incluirán propuestas que tienen relación con el método propuesto.

Existen varios modelos y formas de realizar *forecast*, estos son aplicables a cualquier tipo de industria, se revisarán dichas técnicas a continuación.

2.1. Regresión lineal

La regresión lineal es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictoras. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros y biológicos [8].

Las técnicas de regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente Y (también conocida como la respuesta) como una función de una o varias variables independientes X_i (denominadas predictores). La ecuación general correspondiente a un modelo de regresión lineal es:

$$Y = \beta_0 + \sum \beta_i X_i + \epsilon_i$$

2.1.1. Tipos de regresión lineal

- **Regresión lineal simple:** modelos que utilizan un único predictor. La ecuación general es:

$$Y = \beta_0 + \beta_i X + \epsilon_i$$

- **Regresión lineal múltiple:** modelos que utilizan múltiples predictores. Esta regresión tiene múltiples X_i para predecir la respuesta, Y . Este es un ejemplo de la ecuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

- **Regresión lineal multivariante:** modelos para varias variables de respuesta. Esta regresión tiene múltiples Y_i que derivan de los mismos datos Y . Se expresan con fórmulas diferentes. Este es un ejemplo del sistema con 2 ecuaciones:

$$Y_1 = \beta_{01} + \beta_{11} X_1 + \epsilon_1$$

$$Y_2 = \beta_{02} + \beta_{12} X_1 + \epsilon_2$$

- **Regresión lineal múltiple multivariante:** modelos que utilizan varios predictores para múltiples variables de respuesta. Esta regresión tiene múltiples X_i para predecir varias respuestas Y_i . Esta es una generalización de las ecuaciones:

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}) + \epsilon_i$$

2.2. Series de tiempo

Las series de tiempo pueden ser utilizadas en diversas áreas, tales como ingeniería, ciencia, sociología y economía. Esta técnica consiste en una secuencia de observaciones o mediciones, las cuales son registradas en determinados momentos del tiempo, ordenadas cronológicamente y, espaciadas entre sí de manera uniforme, así los datos usualmente son dependientes entre sí [11].

Se puede hacer uso sólo de los datos de series temporales para generar el pronóstico y pueden identificar y extrapolar patrones de tendencia, estacionalidad y autocorrelación con éxito, pero no tienen en cuenta factores externos. Por lo que antes de que se pueda hacer uso de una serie de tiempo, es recomendable establecer un modelo de probabilidad hipotético para representar los datos. Una vez que se ha elegido una familia adecuada de modelos, es posible estimar los parámetros, verificar si los datos se adecuan y, posiblemente, utilizar el modelo ajustado para mejorar la comprensión del mecanismo que genera la serie. Una vez que se ha desarrollado un modelo satisfactorio, se puede emplear de varias maneras dependiendo del campo particular de aplicación.

Las series temporales poseen ciertas características las que explican su comportamiento en el tiempo:

- **Tendencia o Tendencia secular:** es el resultado de factores a largo plazo. Caracteriza un patrón consistente y gradual de la propia serie. Se puede decir que algunas series decrecen o se mueven hacia abajo, otras crecen continuamente y otras permanecen constantes en cierto periodo de tiempo.
- **Variación estacional:** son los movimientos de la serie que ocurren año tras año en los mismos meses y con similar intensidad, por lo que este componente representa la variabilidad debido a la influencia de las estaciones.
- **Variación cíclica:** luego de que se han eliminado variaciones o tendencias estacional o irregular, esta variación se mantiene, dura más de un año y representa una secuencia alterna de puntos bajo y sobre la línea de tendencia. Un ejemplo de esto son los ciclos de comerciales ya que dependes de la recesión, prosperidad, depresión y recuperación.
- **Variación irregular:** Estos son factores a corto plazos, inesperados y que no son frecuentes que afectan a la serie de tiempo. Este componente es impredecible y explica la variabilidad aleatoria. Existen dos tipos:
 1. Las variaciones provocadas por acontecimientos especiales, hechos puntuales.
 2. Las variaciones aleatorias o por casualidad, estas no se pueden señalar de manera exacta, pero tienden a equilibrarse con el tiempo.

- **Tendencia lineal:** Esta tendencia muestra que algo aumenta o disminuye a un ritmo constante, se aproxima a una línea recta. El método que se utiliza para obtener la línea recta de mejor ajuste es el método de mínimos cuadrados.
- **Tendencia no lineal:** Si la serie de tiempo tiene un comportamiento curvilíneo, se le llama no lineal. Dentro de esta se encuentran polinomial, logarítmica, exponencial y potencial, entre otras.

2.3. Redes neuronales

El aprendizaje automático (*machine learning*) es parte de la inteligencia artificial (AI) que proporciona a los computadores la capacidad de aprender, sin ser programadas explícitamente [1]. El aprendizaje automático se centra en el desarrollo de programas informáticos que pueden cambiar cuando se exponen a nuevos datos. Los computadores aprenden de los datos que se les proporcionan.

Los algoritmos de *machine learning*, se basan en el aprendizaje a partir de una muestra de datos de patrones y relaciones funcionales entre distintas variables. Mitchell ofrece la definición "Se dice que un programa de computador aprende de la experiencia E con respecto a una clase de tareas T y una medida de rendimiento P , si su rendimiento en las tareas de T , medido por P , mejora con la experiencia E " [17].

Esta técnica utiliza algoritmos para obtener datos, aprender y luego analizar los datos. Los resultados llegan en forma de predicciones.

En [19] se realiza la comparación de diferentes técnicas de *machine learning* para realizar pronósticos y se llega a la conclusión que los modelos avanzados de redes neuronales funcionan mucho mejor cuando se enfrentan a tendencias dinámicas no lineales y patrones de estacionalidad.

2.3.1. Redes neuronales artificiales

Las redes neuronales artificiales o ANN (*Artificial neural networks*) son un modelo computacional inspirado en el comportamiento básico de las redes neuronales biológicas que constituyen los cerebros [9]. Las neuronas en nuestro cerebro, están compuestas de dendritas, el soma y el axón, cada neurona recibe impulsos eléctricos de otras neuronas a través de las dendritas. Los impulsos son procesados por el soma, si estos superan un umbral, serán transmitidos a través del axón que emite un impulso nervioso a las neuronas contiguas con las que está conectada mediante estructuras denominadas sinapsis [6].

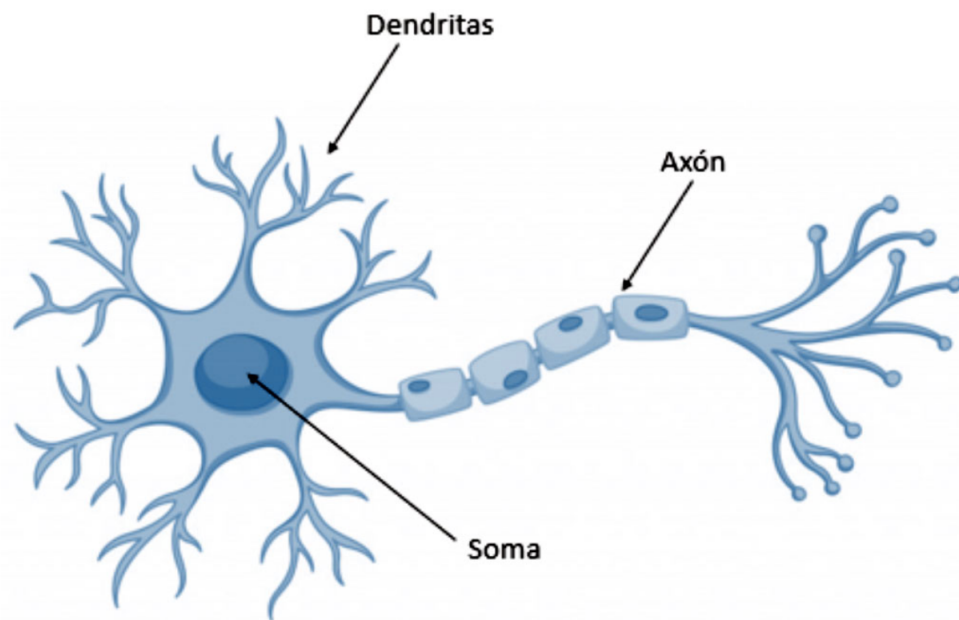


Figura 2.1: Ilustración de una neurona biológica.

Las ANN están compuestas por una colección de unidades o nodos llamados neuronas artificiales (los cuales representan las neuronas biológicas). Las neuronas conectadas entre sí y agrupadas en diferentes niveles se denominan capas, cada capa de la ANN está conectada a una siguiente capa. Esta conexión permite simular, como las sinapsis en un cerebro, el proceso de transmisión de información. Una neurona artificial recibe una señal, luego la procesa y puede enviar señales a las neuronas conectadas a ella, permitiendo a las ANN aprender y

especializarse automáticamente a una tarea determinada.

Para imitar el comportamiento descrito anteriormente se plantea un modelo idealizado de neurona. En términos generales se dice que una neurona artificial es una función no lineal parametrizada:

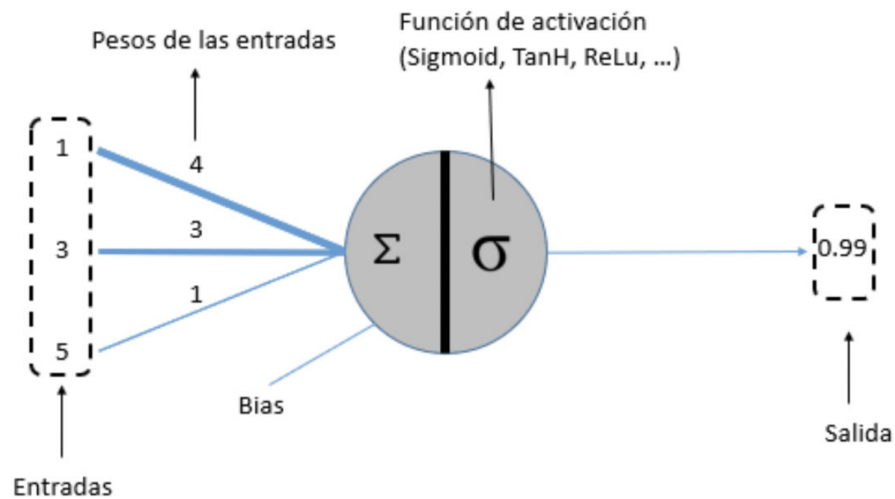


Figura 2.2: Neurona artificial como modelo idealizado y simplificado de una neurona biológica.

donde se tienen variables de entradas (*inputs*), existen parámetros los que son denominados pesos (*weights*). Existe una función de activación que tiene una salida (*output*), por lo regular, es una combinación no lineal entre las entradas y pesos. Por su parte la activación o no de una neurona viene dado por una función de pre-activación, la cual corresponde a una combinación lineal de entradas y pesos. Esta función a veces recibe el nombre de potencial debido a su símil con el modelo biológico.

Formalmente, se puede expresar como:

$$N = \sigma \left(\sum_{i=1}^d W_i X_i - b \right)$$

donde los elementos de la función son:

- σ = Función de activación.
- X = Vector de datos de entrada (*input*), representado de forma que la red neuronal pueda aceptarlo.
- W = Vector de pesos (*weights*). El peso W_i se asocia a una conexión entre dos neuronas y regula la importancia de la transmisión de información entre las capas de la red.
- b = Valor de sesgo, que permite mover la función de activación hacia la izquierda o hacia la derecha.

Función de activación

La función de activación tiene como objetivo principal convertir una señal de entrada de una neurona, en una señal de salida que se usa como una señal de entrada en la siguiente capa de la red, por lo tanto es la encargada de calcular el valor que será transmitido entre neuronas conectadas. A continuación se presenta un listado con las funciones de activación más comunes:

- **Sigmoid – Sigmoide:** se caracteriza por mantener un valor 0 hasta que supera cierto umbral de activación desde el cual crece rápidamente hasta su valor máximo. Una de las grandes ventajas de esta función es que es derivable en todo su dominio lo que facilita la tarea de aprendizaje.

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

- **Tangente hiperbólica:** transforma los valores introducidos a una escala $(-1, 1)$, donde los valores altos tienden de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a -1 .

$$\sigma(x) = \frac{2}{1 + e^{-2x}} - 1$$

- **ReLU:** esta función asigna 0 a todos los valores de entrada menores a 0, anulando los valores negativos y es lineal cuando los valores de entrada son mayores a 0

$$\sigma(x) = \text{Max}(0, x) = \begin{cases} 0, & \text{cuando } x < 0 \\ x, & \text{cuando } x \geq 0. \end{cases}$$

- **Softmax**: esta función transforma las salidas a una representación en forma de probabilidades, de tal manera que la sumatoria de todas las probabilidades de las salidas sea 1. Se utiliza para normalizar los tipo multiclase.

$$\sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}}$$

Arquitectura de redes neuronales

El concepto de arquitectura en redes neuronales describe la conexión entre neuronas o capas, al tipo de neuronas presentes e incluso a la forma en la que son entrenadas. Esta organización determina lo que la red neuronal es capaz de hacer.

Redes Neuronales *Feedforward*

Fueron las primeras redes en desarrollarse por lo que son las más comunes. Constan de varias capas de neuronas, "*fully connected*", es decir, un nodo está conectado con todos los nodos de la siguiente capa, hay al menos una capa oculta, la información circula de izquierda a derecha "*feed forward*" de forma ordenada, las neuronas no tienen ciclos entre sí y sólo tienen conexión entre capas vecinas.

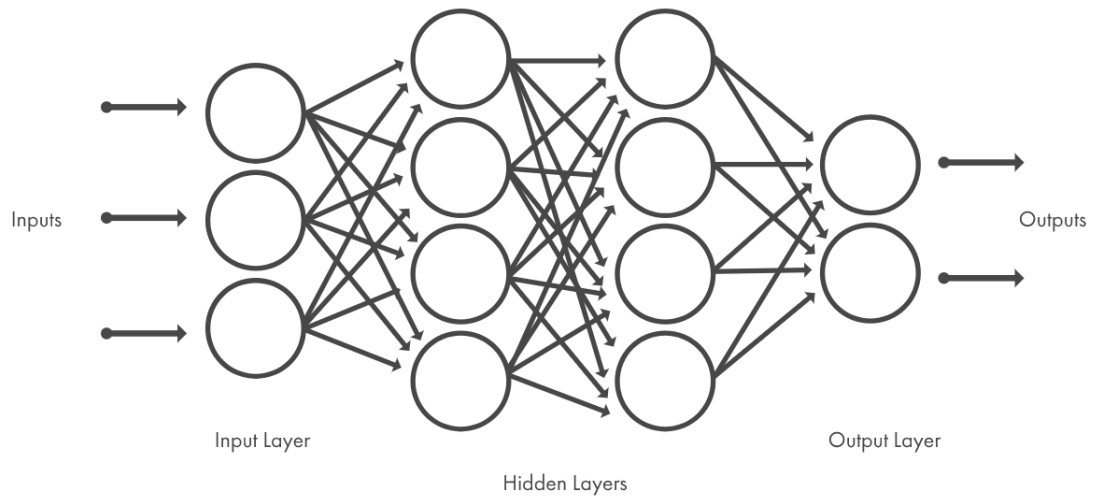


Figura 2.3: Arquitectura de una red neuronal forward.

Este tipo de red se compone de los siguientes tipos de capas:

- Capa de entrada (*input layer*): es la encargada de recibir los datos a procesar por la red como un vector de información. No realiza ningún tipo de transformación sobre la entrada.
- Capa oculta (*hidden layer*): cantidad N de capas, las cuales se encargan de transformar la entrada con el fin de producir algo a la capa de salida.
- Capa de salida (*output layer*): es donde se le da sentido al resultado obtenido por las capas ocultas.

Entrenamiento

Para utilizar una ANN debe pasar por una fase de entrenamiento, este es el proceso de aprendizaje, donde se utilizan los ejemplos proporcionados (conjunto de entrenamiento) para ajustar los pesos y sesgo de la red. Con estos ajustes la red podrá resolver el problema para el cual fue construida. El algoritmo más utilizado es *backpropagation*.

Backpropagation busca modificar los pesos de la red de modo que mejore su desempeño. Se define una función de error para medir la diferencia entre el resultado obtenido y el resultado deseado, se realizan varias ejecuciones para reducir este error y que sea un valor aceptable.

2.3.2. Redes neuronales convolucionales

Las redes neuronales convolucionales o CNN (*Convolution neural networks*) son un tipo de redes neuronales multicapa o profundas. Su característica principal es que utiliza en al menos una de sus capas, la operación matemática de convolución [7].

Estos modelos surgen gracias a LeCun [18], ya que introdujo un método de aprendizaje basado en la propagación hacia atrás para poder entrenar el sistema correctamente.

Las CNN han mostrado buenos resultados en el área de reconocimiento de imágenes [2], en muchos casos obtienen mejores resultados que los alcanzados por los seres humanos.

Gracias al patrón de conectividad de sus neuronas es que obtienen buenos resultados, este está inspirado en la organización de la corteza visual animal [5]. Se observaron distintos patrones (ej líneas horizontales o verticales) que activaron diferentes rutas neuronales. Las CNN replican este comportamiento usando *kernels*, este es un marco móvil, que escanea la imagen y pondera cada píxel de acuerdo con los pesos del *kernel*. Al utilizar pesos fijos, el *kernel* extrae el mismo resultado cuando un patrón se repite en la imagen, independiente en donde este se encuentre.

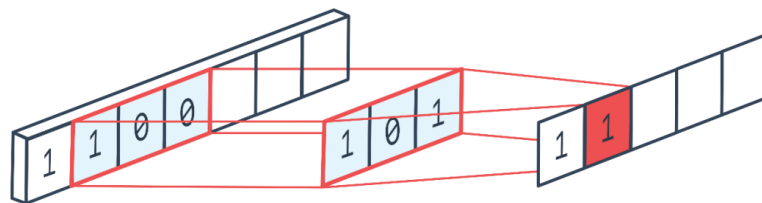


Figura 2.4: Esquema de convolución unidimensional, donde el kernel de tamaño 3 (borde rojo) realiza una operación de convolución a cada secuencia de 3 valores de la entrada.

2.3.3. Redes neuronales recurrentes

Las redes neuronales recurrentes o RNN (*Recurrent neural networks*) son un modelo que tienen la capacidad de analizar secuencias temporales de datos de tamaño variable y predecir cuál será el siguiente valor de la serie, por lo que resultan útiles en problemas en donde existe una dependencia temporal, como por ejemplo series de tiempo o procesamiento de audio/video [3].

Las RNN pueden considerarse un caso específico de las redes neuronales convencionales, ya que se le añade una conexión extra que va desde la salida de la neurona hacia la entrada, tal como se muestra en la figura a continuación:

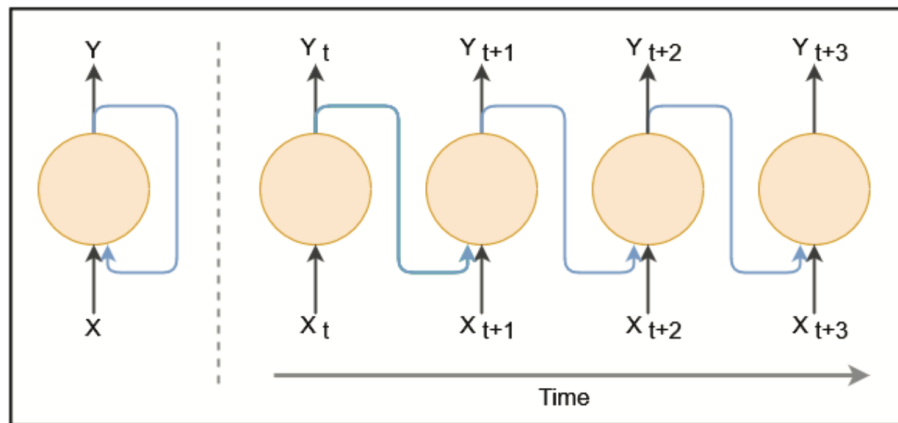


Figura 2.5: Al comienzo una neurona recurrente, luego la neurona recurrente desarrollada en el tiempo.

Para entender de mejor maneja el funcionamiento de este tipo de redes, se necesitan desarrollar o extender en el tiempo, como se muestra en la parte derecha de la imagen 2.5, en cada instante de tiempo la neurona recibe la entrada que le corresponde, pero además recibe la salida de la neurona en el instante anterior, de esta forma la red puede recordar información de secuencias anteriores cuando se encuentra procesando los datos en el instante actual.

También se pueden construir capas de neuronas recurrentes, en tal caso cada neurona recibirá un vector como entrada y tendrá un vector de salida que será de entrada para la neurona en

un instante siguiente, como se muestra en la figura 2.6.

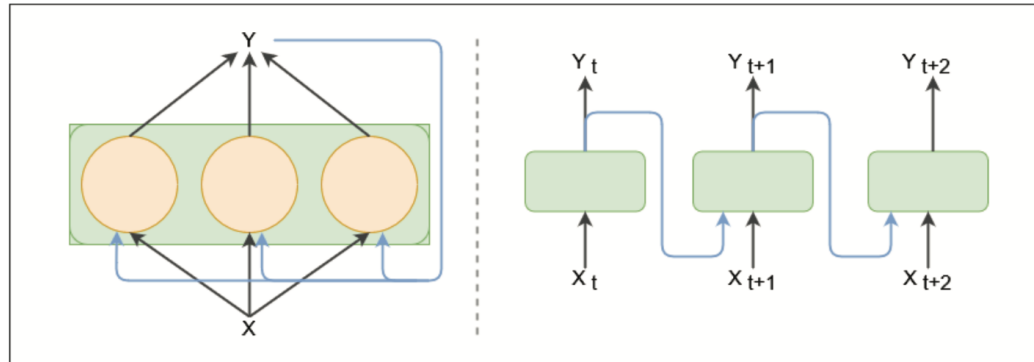


Figura 2.6: Al comienzo una capa recurrente, luego la capa recurrente desarrollada en el tiempo.

Formalmente, se puede expresar como:

$$y(t) = \sigma(X(t)W_x + Y_{(t-1)}W_y + b)$$

donde los elementos de la función corresponden a:

- σ = Función de activación.
- $X_{(t)}$ = Vector de datos de entrada (*input*), representado de forma que la red neuronal pueda aceptarlo.
- W_x = Vector de pesos (*weights*).
- $Y_{(t-1)}$ = Vector de datos de entrada (*input*) en el instante actual, pero que también representa la salida de la neurona en el instante anterior.
- W_y = Vector de pesos (*weights*) en el instante anterior.
- b = Valor de sesgo, que permite mover la función de activación hacia la izquierda o hacia la derecha.

Si se apilan varias capas de neuronas recurrentes, se habla de redes neuronales recurrentes profundas o deep RNN.

El concepto de celda de memoria se utiliza para referirse a capas de neuronas recurrentes, ya que su funcionamiento se basa en recordar información de instantes pasados, de esta forma la salida de un instante concreto, está determinado por los datos de entrada que la capa ha visto hasta ese momento.

Un problema que presentan las RNN es el *short-term memory* [14], esto deriva del desvanecimiento del gradiente. Como se ha explicado, las redes neuronales recurrentes funcionan procesando secuencias de información, en cada operación se agrega al computo actual el análisis del instante anterior. De esta forma el resultado final de la RNN será una función que agregará el procesamiento de todos los elementos de la secuencia, pero a medida que la red procesa más elementos comienza a tener problemas para recordar información pasada.

Este problema es por el algoritmo por el cual las redes neuronales recurrentes son capaces de entrenar, el algoritmo *back-propagation throug time* o BPTT [13]. A medida que la RNN se desarrolla, el gradiente del error se propaga hacia atrás del mismo modo que en una red normal, es por esto que cuando el gradiente alcanza las capas cercanas al *input*, en este caso a los instantes iniciales, decae exponencialmente. Por esto las RNN más sencillas no son capaces de aprender patrones muy extendidos en el tiempo, sino que sólo rangos cortos.

Para mitigar el problema de *short-term memory*, aparece una celda más compleja *Long-Short Term Memory* o LSTM, la cual sí es capaz de extraer patrones de secuencias de mayor longitud.

Las redes *Long-Short Term Memory* o LSTM son un tipo especial de redes recurrentes. La característica principal de las redes recurrentes es que la información puede persistir introduciendo bucles en el diagrama de la red, por lo que, básicamente, pueden «recordar» estados previos y utilizar esta información para decidir cuál será el siguiente. Esta característica las hace muy adecuadas para manejar series cronológicas. Mientras las redes recurrentes estándar pueden modelar dependencias a corto plazo (es decir, relaciones cercanas en la serie cronológica), las LSTM pueden aprender dependencias largas, por lo que se podría decir que tienen una «memoria» a más largo plazo.

Long-Short Term Memory (LSTM) son una extensión de las redes neuronales recurrentes, que básicamente amplían su memoria para aprender de experiencias importantes que han pasado hace mucho tiempo. Las LSTM permiten a las RNN recordar sus entradas durante un largo período de tiempo. Esto se debe a que LSTM contiene su información en la memoria, que puede considerarse similar a la memoria de un computador, en el sentido que una neurona de una LSTM puede leer, escribir y borrar información de su memoria.

2.4. Comparación de modelos

Es importante destacar que la simplicidad del modelo ayuda a los gerentes a comprender y aprobar o guiar la modificación de los modelos, y a medida que se vuelven más conocedores y notan que las predicciones son una ayuda para la toma de decisiones, estos pueden estar listos para implementar modelos más sofisticados y complejos, ya que esto generalmente significa un aumento en los costos.

Existen dos enfoques para guiar y desarrollar los pronósticos, por un lado se tiene un enfoque causal donde se puede involucrar una hipótesis concreta, por ejemplo, el marketing, donde se hipotetizan modelos de respuesta de ventas particulares basados en la teoría del comportamiento del consumidor, y los datos se usan para estimar sus parámetros [12]. Por otro lado, las técnicas de *machine learning*, como árboles, SVM o redes neuronales, no asumen una relación particular entre las variables e implican una búsqueda a través del espacio de forma funcional, así como la estimación de parámetros.

Se tienen las siguientes medidas de error:

- **RMSE:** llamado así por su sigla en inglés (Raíz Cuadrada de Error Medio). Es la raíz cuadrada del promedio de los errores al cuadrado:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (R_t - A_t)^2}{n}}$$

donde n es el número de observaciones, R_t es el valor real y A_t el valor ajustado del pronóstico.

- **MAPE**: llamado así por su sigla en inglés (Error Porcentual Absoluto Medio). Es el promedio porcentual de los valores absolutos de los errores:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{R_t - A_t}{R_t} \right|$$

donde n es el número de observaciones, R_t es el valor real y A_t el valor ajustado del pronóstico.

- **Forecast accuracy**: es la precisión de la predicción en porcentaje, expresada por la fórmula:

$$FA = 1 - MAPE$$

Dado que la fórmula de cálculo de error de MAPE tiene dificultades o problemas, esto debido a que se pueden dar casos de división por cero. En [15] se propone utilizar como métrica MAAPE. La cual se define como:

$$MAAPE = \frac{100}{n} \sum_{t=1}^n \arctang\left(\frac{R_t - A_t}{R_t}\right)$$

donde n es el número de observaciones, \arctang es la función arco-tangente (inversa de la tangente), R_t es el valor real y A_t el valor ajustado del pronóstico.

2.5. CRISP-DM

La metodología CRISP-DM es una forma probada y ampliamente utilizada para guiar el trabajo en minería de datos. Como metodología tiene fases típicas de un proyecto, las tareas involucradas en cada fase y una explicación de las relaciones entre estas tareas [10].

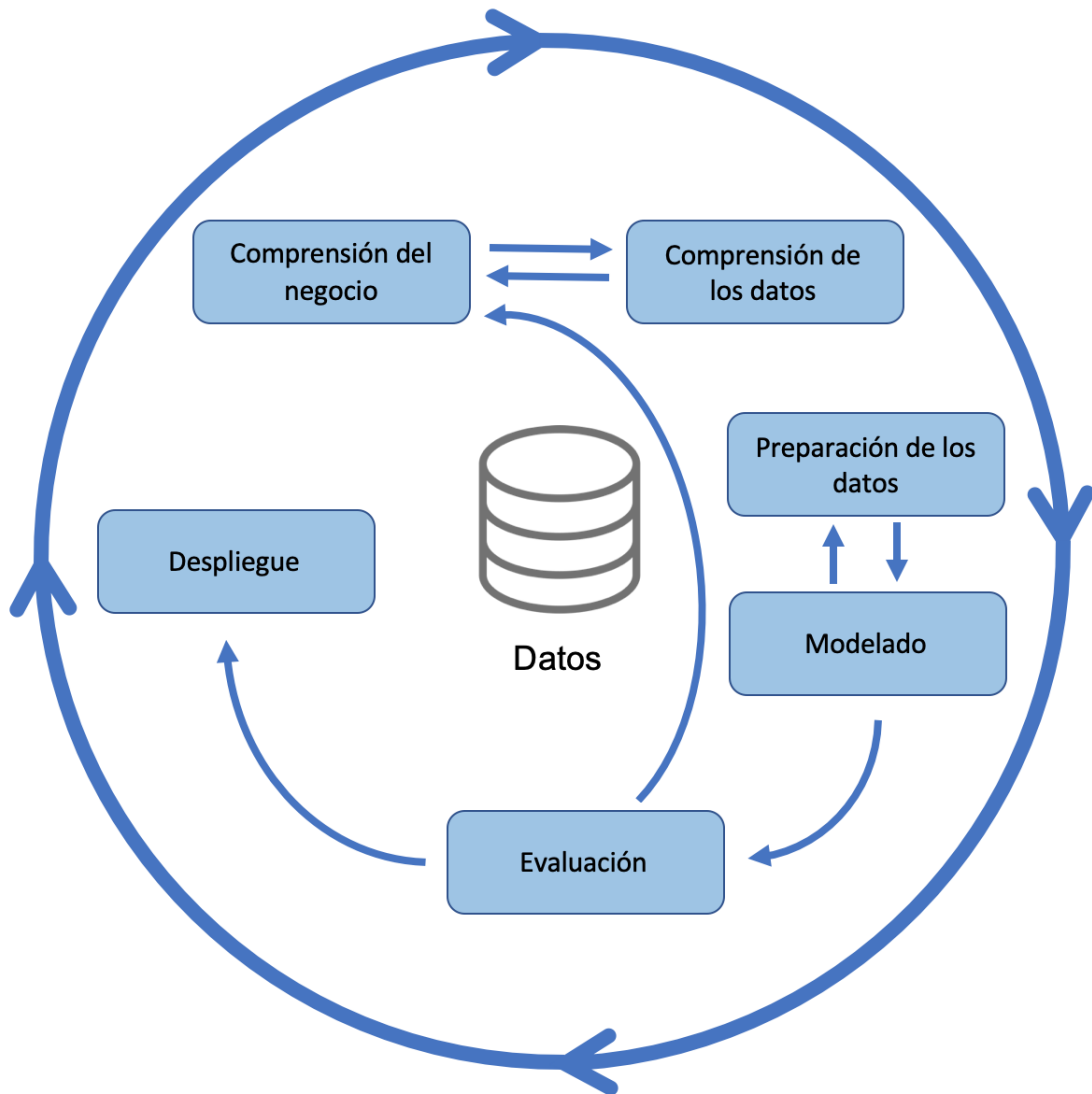


Figura 2.7: Fases del modelo CRISP-DM.

Dentro de las fases que se identifican se encuentran:

- **Comprensión del negocio:** Es la fase inicial del proyecto, se enfoca en entender los requisitos y objetivos del proyecto desde el punto de vista de la empresa para posteriormente convertir dicho conocimiento en la definición de un problema de minería de datos y diseñar un plan de trabajo para lograr los objetivos fijados
- **Comprensión de los datos:** Esta fase empieza con la recolección de datos y comprende actividades como la familiarización con ellos, poder identificar problemas de calidad, detectar conjuntos o subconjuntos interesantes para luego formular hipótesis sobre estos.
- **Preparación de los datos:** En esta fase se realizan todas las tareas necesarias para limpieza, selección y transformación de los datos que constituirán el conjunto final a utilizar. Estas tareas se pueden realizar varias veces.
- **Modelado:** En esta fase se seleccionan las técnicas de modelado a utilizar en el proyecto, además, se calibran sus parámetros para que estos tengan los valores óptimos. Como existen varias técnicas para un mismo tipo de problema es necesario estudiarlos y elegir el más adecuado. Algunas técnicas necesitan datos de forma específica por lo que es necesario volver a la fase de preparación.
- **Evaluación:** En esta fase del proyecto ya se tiene uno o más modelos, pero es necesario revisarlos y evaluarlos para asegurarse que el modelo cumple con los objetivos inicialmente fijados. También es importante revisar los pasos ejecutados y los parámetros utilizados por si existe algún tipo de error que haya que corregir.
- **Implementación:** No basta con la creación del modelo, esta fase contempla desde la presentación de un informe al cliente o empresa hasta la implementación de un proceso de minería de datos repetible en toda la empresa. Es importante que el cliente comprenda el conocimiento que está adquiriendo con el trabajo realizado y pueda traducirlo en acciones y/o decisiones.

Capítulo 3

Propuesta de Solución

En este capítulo se presenta la propuesta de esta memoria que consiste en el uso de modelos de predicción estadísticos clásicos y también de redes neuronales convolucionales y recurrentes.

3.1. Metodología

Dado que se está trabajando en minería de datos se utiliza la metodología CRISP-DM, la cual está explicada en la sección 2.5. A continuación se explica como se aplicó cada una de las etapas:

- **Comprensión del negocio:** Esta etapa está aplicada en el capítulo 1, donde se explica el problema y se definen los objetivos que se necesitan cumplir en este trabajo. También en el capítulo 2 ya que se explican los conceptos necesarios para entender la solución.
- **Comprensión y preparación de los datos:** Esta etapa queda expuesta en la sección 3.1.1, se muestran los datos que se tienen a disposición, cual fue su tratamiento y cuales son los finalmente seleccionados y utilizados.

- **Modelado e implementación:** Esta fase está presente en el capítulo 4 donde se explican los modelos de predicción utilizados.
- **Evaluación:** Esta fase se encuentra en la sección 4 donde se explica qué métricas se utilizaron y también en dicho capítulo se muestran los resultados obtenidos de estas comparaciones.

3.1.1. Comprensión de los datos

Estos modelos se aplican a los datos de ventas recolectados de una tienda de retail, específicamente datos de los años 2017 y 2018.

Se recolectaron datos de diferentes tiendas, en un principio sólo se seleccionaron las tiendas con mayor cantidad de ventas, siendo estas 5 tiendas ubicadas en la región metropolitana, finalmente se decide utilizar los datos de ventas de productos realizados a través de internet, los cuales corresponden a aproximadamente 2,2 millones de ventas.

Los datos de ventas fueron ordenados de manera que se tuviera la información del producto comprado, fecha de compra, cantidad de compra y precio pagado. Por otro lado se tiene información del producto como tal, así como la categoría a la que pertenece, si es un producto *marketplace* o no, precio de venta, talla o color, según corresponda.

Este trabajo se centra en el estudio de ciertas categorías, esto se debe al interés por parte del *retail* en mejorar flujos y procesos logísticos. De esta manera se estudiarán categorías donde se encuentran productos como refrigeradores y camas, ya que son productos de gran tamaño y requieren camiones o vehículos grandes para realizar las entregas. Por otra parte se tiene categorías con productos como calefacción o electrodomésticos, los cuales son más pequeños y son fáciles de transportar, pero su volumen de venta es alto, por lo que a pesar de no necesitar un camión, de igual manera necesitan un vehículo para la realizar las entregas.

A continuación se encuentra en detalle las categorías seleccionadas y productos dentro de éstas:

- Electrodomésticos: con productos como hervidores eléctricos, planchas a vapor, hornos eléctricos, licuadoras, minipimer, ollas arroceras, batidoras manuales, cafeteras eléctricas con filtro, cafeteras eléctricas express, picadores eléctricos, sandwicheras, parrillas eléctricas, tostadores eléctricos, exprimidores eléctricos, procesadores de alimento, saca jugo, maquina para hacer pan, freidores eléctricos, varios electro menor y batidoras de pedestal
- Box Spring: con productos como box spring 1.5 plazas, box spring 2 plazas y box spring king
- Refrigeración: con productos como: refrigeradores menor a 300L, refrigeradores de 300L a 400L y refrigeradores mayor a 400L
- Calefacción con productos como estufas eléctrica, estufas a gas, estufas a parafina y estufa convección directa

Con todo esto en mente se organizan y agrupan los datos en una matriz, donde la cantidad de filas serán los días, como se tienen 2 años serán 730 y la cantidad de columnas es cada producto a estudiar, quedando una matriz de 730x30.

Capítulo 4

Resultados

En esta sección se explicará como se entrenó y desarrolló cada modelo utilizado, describiendo sus parámetros. En cada caso o modelo se mostrará tanto lo obtenido considerando todos los valores disponibles, como también quitando los valores correspondientes a ventas especiales, específicamente días *cyber*.

Los eventos *Cyber Day* o *Cyber Monday* corresponden a 3 días especiales de ventas con diferentes ofertas y descuentos, estos se realizan dos veces en el año. Durante estos días las visitas a los sitios web participantes se multiplican y con esto también aumentan la cantidad de transacciones y ventas, cada uno de estos eventos tiene un aumento de al menos el 20 % en ventas con respecto al año anterior [4].

La fecha de estos eventos que estan dentro del set de datos corresponden a:

- Cyber Day 2017: 29, 30 y 31 de Mayo
- Cyber Monday 2017: 6, 7 y 8 de Noviembre
- Cyber Day 2018: 28, 29 y 30 de Mayo
- Cyber Monday 2018: 8, 9 y 10 de Octubre

Para poder llevar a cabo la comparación de los distintos modelos, se seleccionan como métrica la medición de *RMSE* y de *Forecast accuracy*, por lo que se considerará al mejor modelo, el que sea capaz de minimizar estos valores.

Los modelos serán aplicados a todas las categorías mencionadas en el estudio, pero se ejemplificará específicamente con la de refrigeradores sobre 400 litros.

4.1. Regresión lineal

Dentro de los primeros modelos realizados se encuentra la regresión lineal, este modelo suele ser el tradicional, en este caso se realiza utilizando sólo una variable, de esta forma se obtiene lo siguiente:

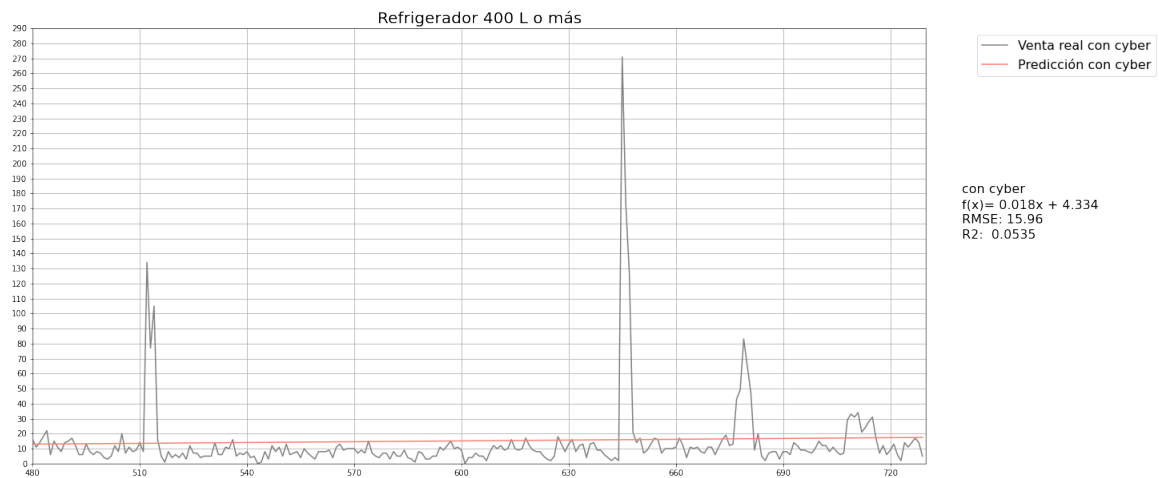


Figura 4.1: Regresión lineal sobre los datos de venta de refrigeradores de más de 400L.

En la figura 4.1 se observa el comportamiento de ventas de refrigeradores de sobre 400L, las ventas no se comportan de manera lineal, además se observan *peak's* los cuales corresponden a los días de eventos especiales, estando estos ubicados en los días 510 y 640 aproximadamente.

En este caso es posible notar que la regresión lineal no se ajusta correctamente al comportamiento de las ventas, ya que la línea de color rojo, que es la que representa las predicciones,

se aleja de la línea de color gris, que es la que representa a las ventas reales, cuando un ajuste es bueno el coeficiente de determinación (R^2) es cercano a 1, lo cual no se logra en este caso, dada su cercanía a 0 nos dice que no hay relación entre las dos variables representadas. Por otra parte, lo que se logra obtener es que la pendiente, es positiva, dando a entender un comportamiento futuro que va en aumento.

Se repite el experimento manteniendo las mismas condiciones, pero ahora se quitan los datos correspondiente a *cyber*, de esta forma se tiene:

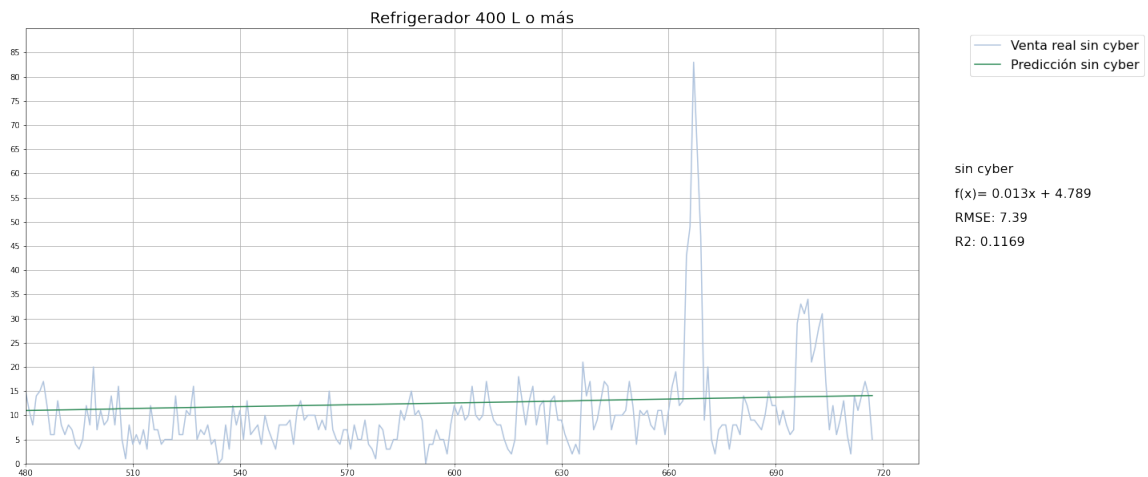


Figura 4.2: Regresión lineal sobre los datos de venta de refrigeradores de más de 400L sin datos *cyber*.

Ahora, es posible ver de manera un poco más clara el comportamiento de ventas y de las predicciones realizadas por la regresión lineal. Nuevamente se obtiene un coeficiente de determinación (R^2) cercano a 0, por lo cual no existe relación entre las variables utilizadas y se obtiene una pendiente positiva, indicando nuevamente un aumento en las ventas.

Para poder comparar ambas predicciones de mejor manera se genera el siguiente gráfico:

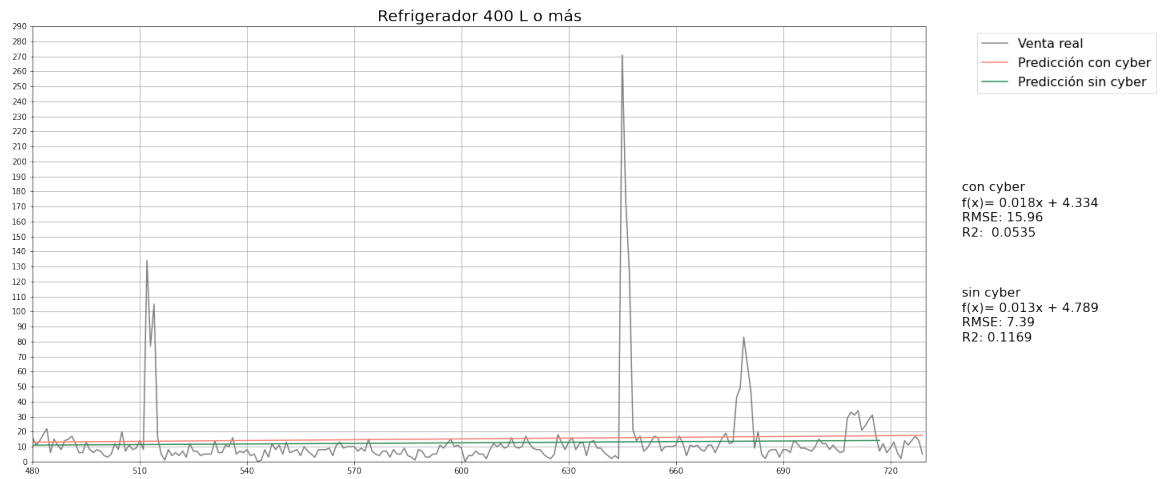


Figura 4.3: Regresión lineal sobre los datos de venta de refrigeradores de más de 400L con y sin datos *cyber*.

Como se muestra en la figura 4.3, los días de ventas especiales o *cyber*, a pesar de sólo corresponder a 6 días en el año, dado su volumen de venta, son capaces de cambiar el resultado del modelo. En las predicciones donde no se consideraron estos datos, el error RMSE mejora, reduciéndose a la mitad y el coeficiente de determinación (R^2) mejora, pero no lo suficiente para considerarse bueno.

Pese a que se iteró y se intentó mejorar este modelo, no se logra ajustar su comportamiento al comportamiento de ventas.

4.2. Series de tiempo

Dentro de los siguientes modelos realizados se encuentran las series de tiempo, en este caso se utilizó un modelo de SARIMA ya que tiene una componente estacional, este modelo se expresa de la forma $SARIMA(p, d, q)(P, D, Q)[m]$, de esta forma se expresa m , que corresponde a la cantidad de periodos, en 30 días. Las demás componentes (p, d, q) y (P, D, Q) se han escogido mediante ensayo y error, quedando en $SARIMA(2, 0, 1) \times (2, 1, 0)[30]$.

Dentro de los primeros resultados se tiene:

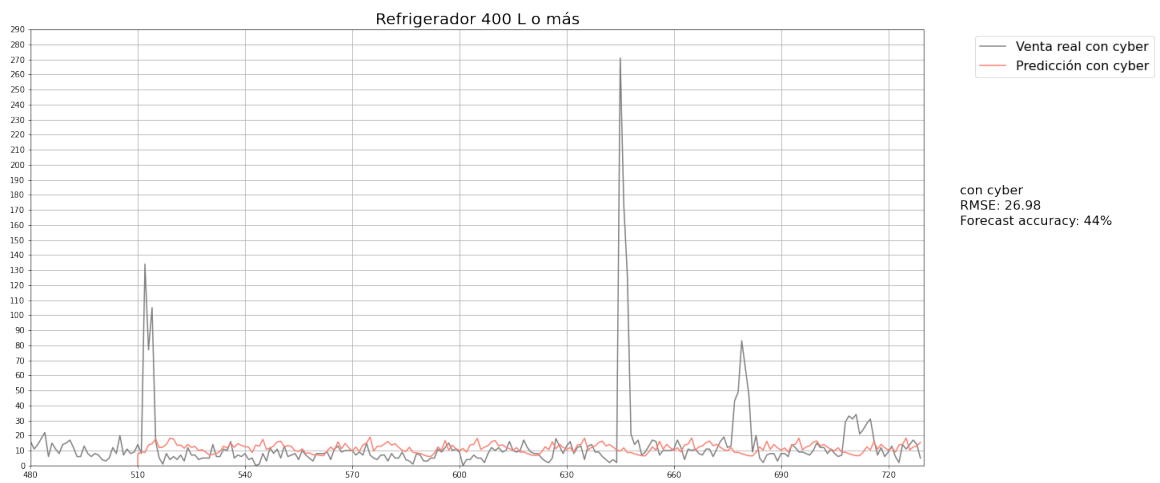


Figura 4.4: Serie de tiempo sobre los datos de venta de refrigeradores de más de 400L.

En la figura 4.4 se puede ver que sólo en algunos tramos podrían ser certeras las predicciones, pero se deben realizar ajustes para lograr esto.

No se obtiene un comportamiento lineal, pero el RMSE es más alto en comparación con el modelo de regresión lineal, siendo este de 27 unidades y en la regresión lineal de 16 unidades.

Nuevamente se repite el experimento, manteniendo las mismas condiciones, pero se quitan los valores de *cyber*, mejorando el *forecast accuracy* en 7% y reduciendo el error a 10 unidades, tal como se muestra en la figura 4.5:

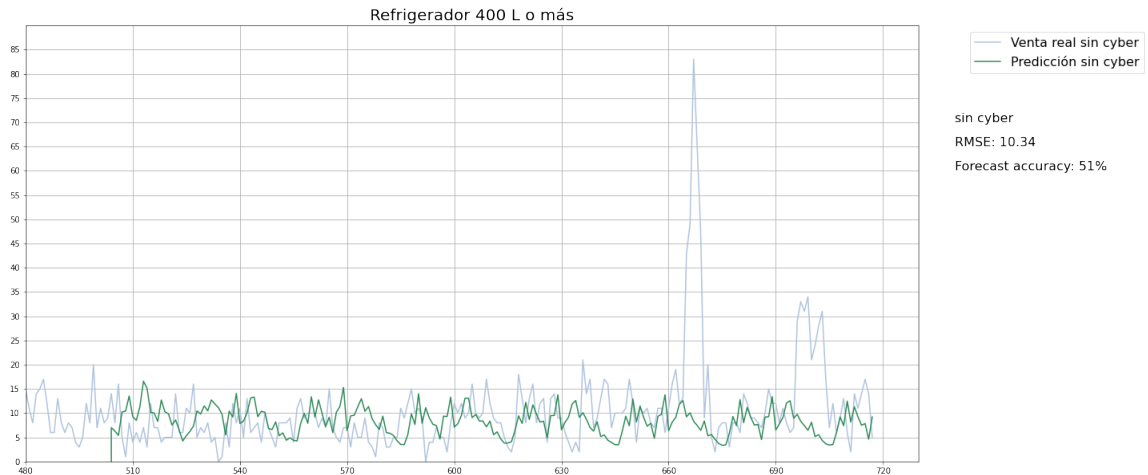


Figura 4.5: Serie de tiempo sobre los datos de venta de refrigeradores de más de 400L sin datos *cyber*.

Dado que se quitan los datos *cyber*, se ve de mejor manera el comportamiento de las ventas y de las predicciones, se ve que en algunos tramos la predicción sí se acerca a la venta, también es posible ver que las predicciones tienen un comportamiento repetitivo esto se debe a que al definir el experimento se utilizó que el largo del periodo correspondería a 30 días.

Por lo anterior es que se realizaron pruebas extendiendo el periodo a 60 días, con esto algunos casos mejoran las predicciones, pero no todas. Se intentó extender aun más estos periodos, pero por limitaciones del computador estos experimentos no se lograron concluir.

Para poder comparar ambas predicciones, se genera el siguiente gráfico:

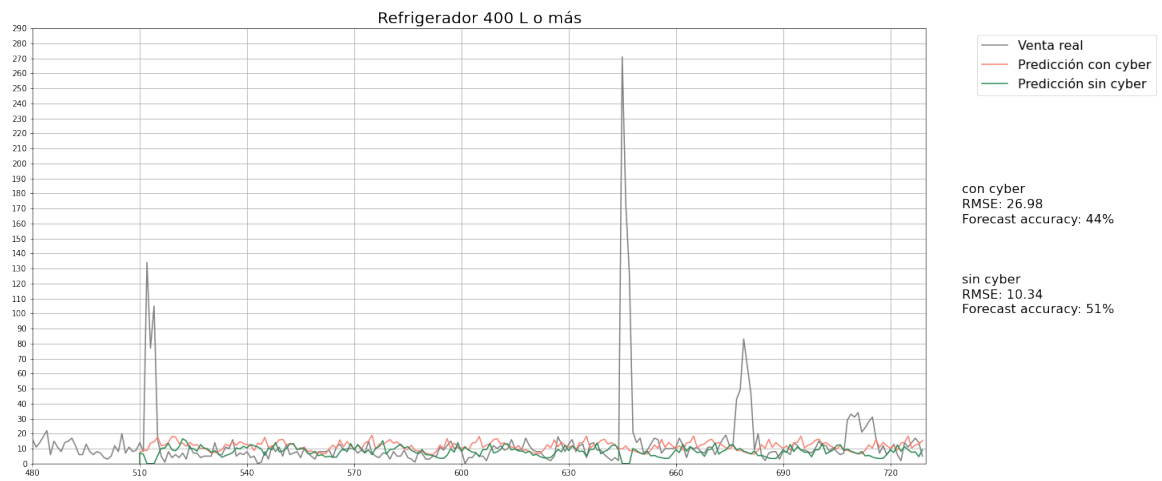


Figura 4.6: Serie de tiempo sobre los datos de venta de refrigeradores de más de 400L con y sin *cyber*.

Al comparar ambas predicciones en la figura 4.6, se logra ver que al quitar los días *cyber*, las predicciones mejoran, ya que el RMSE da un valor de 10,34 unidades y en este caso es más bajo que en el primer experimento con series de tiempo. Igualmente mejora el *forecast accuracy*, pero no lo suficiente para considerar que es un buen ajuste.

Aún cuando las series de tiempo se utilizan en el día a día como predictores [16], estas requieren de largas horas de estudio y entendimiento del negocio donde son aplicadas. En este caso no se continuó iterando este modelo, ya que se encontraron modelos con mejores resultados y con menos esfuerzo.

4.3. Red convolucional

Siguiendo con los experimentos, el siguiente fue utilizar redes neuronales convolucionales. Para llevar a cabo este modelo es necesario preparar vectores de datos a partir de la secuencia de observaciones, estos vectores serán utilizados como entradas dando como resultado una única salida.

Se divide el set de datos en dos grupos, 70 % para entrenamiento y el resto para probar el modelo. Se generan vectores de largo 30 con los datos de entrenamiento, como función de activación se utiliza reLU, kernel de tamaño 2 y 1000 epochs.

Al realizar las pruebas con la red neuronal, se consigue el ajuste mostrado a continuación:

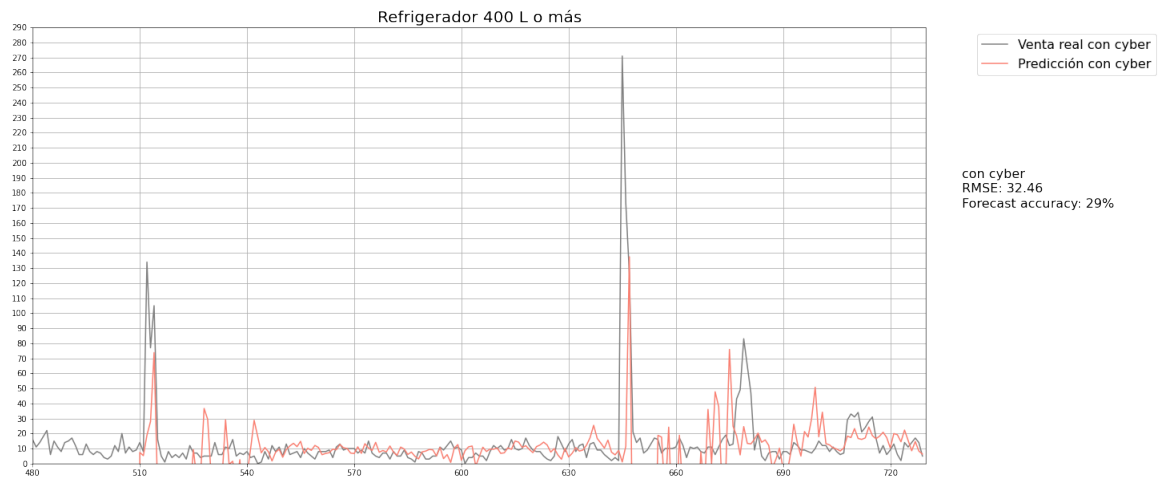


Figura 4.7: Red neuronal convolucional sobre los datos de venta de refrigeradores de más de 400L.

En la figura 4.7 es posible ver que la red convolucional no se ajusta del todo al comportamiento de ventas, pero en algunos tramos se acerca, es capaz de conseguir los *peak's* en varios casos y en otros los consigue, pero con un desfase en los días, desfase el cual se podría identificar y corregir para mejorar el experimento.

Al repetir el experimento manteniendo las mismas condiciones y sólo quitando los datos de *cyber*, mejora el ajuste en 13 % como se ve en la figura 4.8:

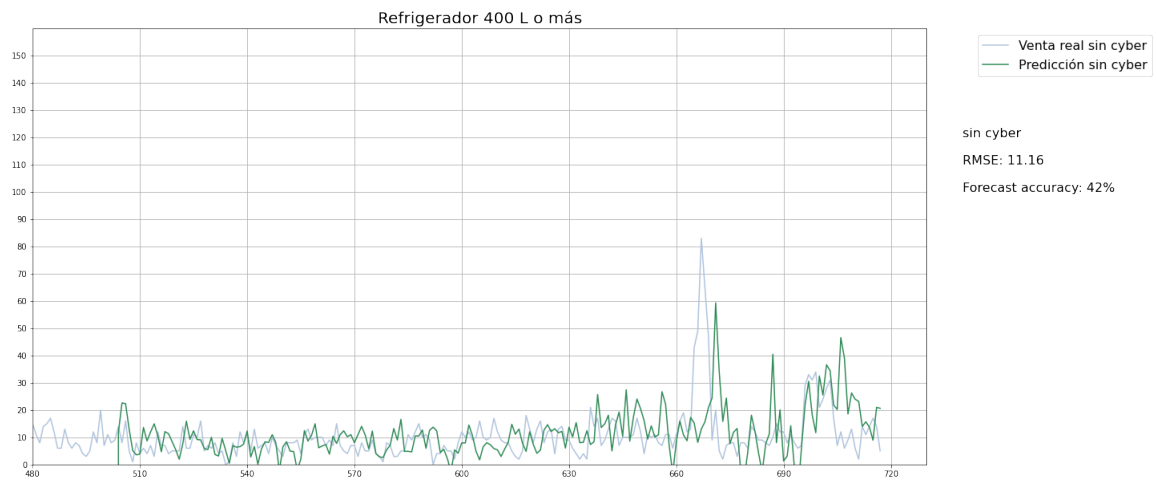


Figura 4.8: Red neuronal convolucional sobre los datos de venta de refrigeradores de más de 400L sin datos *cyber*.

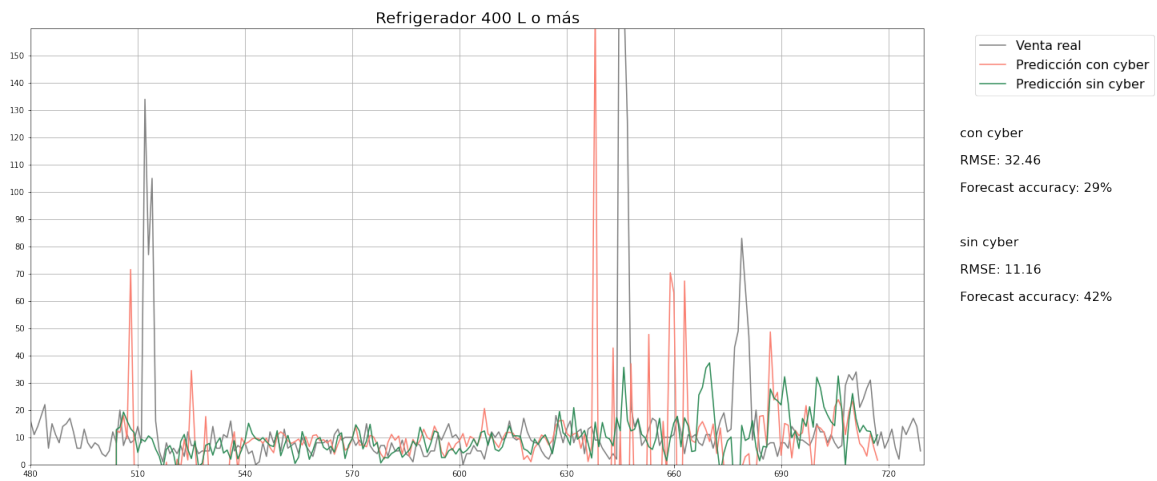


Figura 4.9: Red neuronal convolucional sobre los datos de venta de refrigeradores de más de 400L con y sin datos *cyber*.

Al comparar ambas predicciones tenemos que mejora el error y el *forecast accuracy* del modelo, pero no lo suficiente como para utilizarlo en la realidad, tal como se grafica en la figura 4.9.

En estos experimentos se intentó variar algunas componentes como el largo de los vectores, pero no se consiguieron mejoras considerables.

4.4. Red neuronal LSTM

Finalmente en los últimos modelos se encuentran las redes LSTM, en este caso se utiliza la función de activación sigmoide, se normalizan los datos, pasando los valores de ventas a un rango de 0 a 1. Se divide el set de datos, utilizando el 70 % de los datos para entrenamiento del modelo y el 30 % restante para probar el modelo.

La red tiene una capa visible con 1 entrada, una capa oculta con 4 bloques LSTM y una capa de salida que hace una predicción de valor único, este valor está en el rango entre 0 y 1 por lo cual se debe deshacer la normalización y el valor resultado corresponde a la cantidad que se venderá.

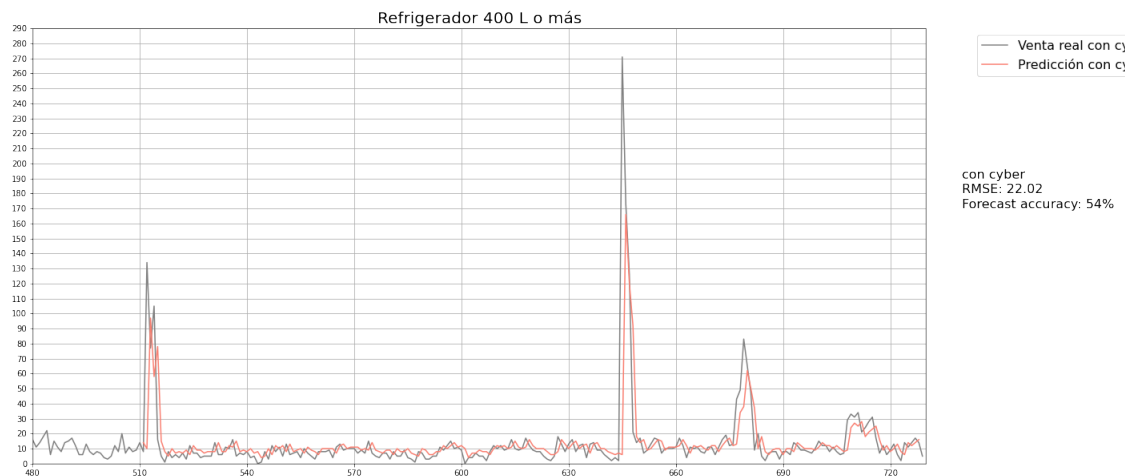


Figura 4.10: Red neuronal LSTM sobre los datos de venta de refrigeradores de más de 400L.

En este caso es notorio que la red neuronal se acerca al comportamiento de ventas, tiene desfase respecto al número de unidades, pero la tendencia ya es similar, como se puede ver en la figura 4.10.

Se consigue un error RMSE de 22,02 unidades y un *forecast accuracy* de 54 %.

Se repite el experimento anterior, donde la única variación es que se quita los días *cyber* y se consigue lo graficado en la figura 4.11:

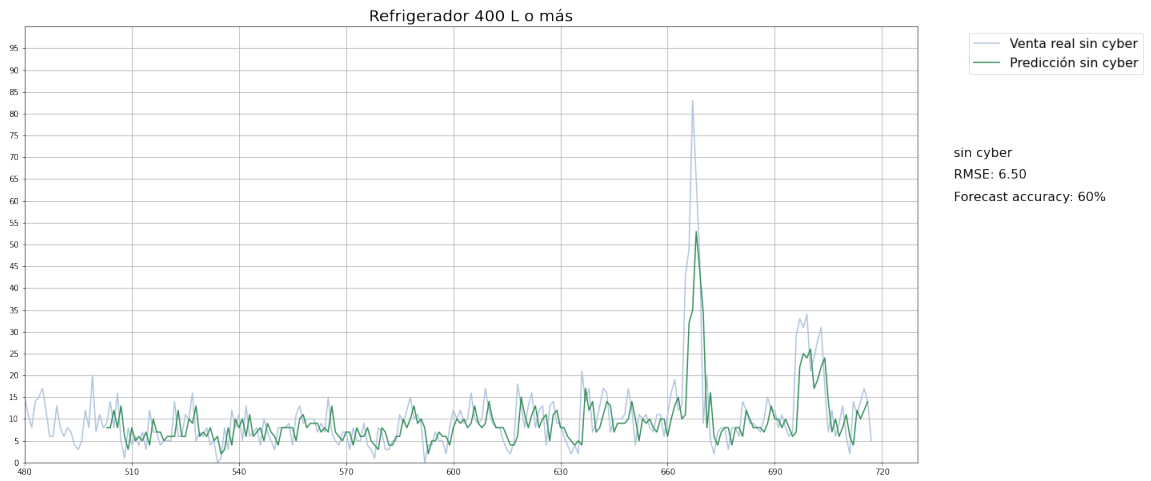


Figura 4.11: Red neuronal LSTM sobre los datos de venta de refrigeradores de más de 400L sin *cyber*.

Se logra apreciar con mayor detalle el comportamiento de las ventas y el comportamiento de las predicciones, se ve que el ajuste en esta oportunidad es mucho mejor. Por otra parte, el error RMSE disminuye, quedando en 6,50 unidades y el *forecast accuracy* aumenta quedando en 60 %.

Finalmente, se superponen las predicciones para poder comparar ambas predicciones en la figura 4.12:

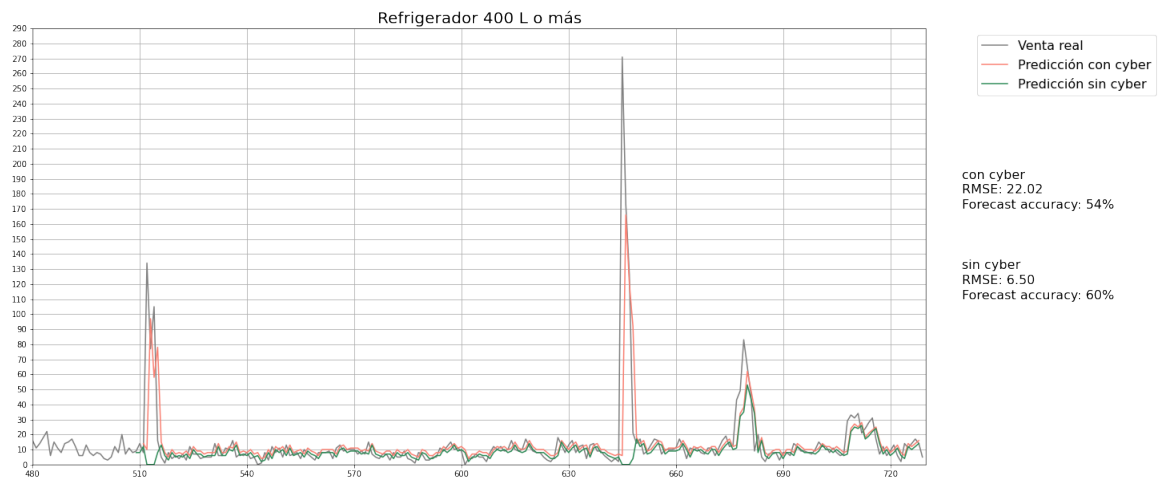


Figura 4.12: Red neuronal LSTM sobre los datos de venta de refrigeradores de más de 400L con y sin *cyber*.

Se ve que ambas predicciones son similares entre sí, ambas ajustándose al comportamiento de ventas real.

Se observa que al quitar los *peaks's* de ventas ayuda al modelo a disminuir su error y aumentar su *accuracy*.

Los experimentos con este tipo de modelo son aquellos de mejor desempeño a lo largo de todo el estudio, con comportamientos similares a las ventas, consiguiendo en muchos casos los errores RMSE más bajos y *forecast accuracy* más altos, llegando en algunos casos al 70 %.

4.5. Análisis de RMSE

Se calcula el error cuadrático medio en 4 categorías y se presenta en la tabla 4.1 cómo se comporta cada modelo y de manera diferenciada si contiene o no los datos de *cyber*.

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|---------------|------------------|--------------|------------------|-------|----------|-------|-------|--------------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Refrigeración | 26.64 | 14.15 | 44.80 | 18.61 | 74.76 | 15.58 | 36.35 | 9.60 |
| Calefacción | 50.39 | 25.76 | 88.80 | 46.66 | 134.09 | 32.95 | 69.40 | 21.86 |
| Electro | 47.21 | 24.23 | 70.98 | 28.23 | 99.49 | 29.84 | 56.60 | 17.46 |
| Box spring | 56.93 | 11.71 | 91.53 | 15.88 | 101.20 | 20.82 | 76.95 | 14.22 |

Cuadro 4.1: RMSE promedio de las 4 categorías estudiadas (para revisar la tabla completa ver el anexo 4.4).

Se observa que la regresión lineal es la que obtiene los menores valores de error al comparar los 4 modelos utilizados, pero si se quita del set de datos los valores correspondientes a días de ventas especiales, todos los modelos mejoran, disminuyendo el error.

Si se considera que lo que se quiere predecir son los días sin *peak's* de ventas y se toma en cuenta la segunda columna de cada experimento, se ve que las redes neuronales mejoran considerablemente, disminuyendo el error, específicamente con las redes LSTM.

En el caso de la categoría de box-spring es donde mejor rendimiento tuvo la regresión lineal, esto debido a que las variaciones de venta son más constantes, moviéndose en un intervalo y en general con muy pocos valores que se escapen de este, favoreciendo así a este modelo.

4.6. Análisis de *Forecast accuracy*

Así como se calculó el RMSE para cada categoría, posteriormente se calcula el *forecast accuracy* para cada categoría, mostrado en el cuadro 4.2:

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|---------------|------------------|------|------------------|------|----------|------|-------|-------------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Refrigeración | 46 % | 51 % | 48 % | 46 % | 25 % | 52 % | 60 % | 66 % |
| Calefacción | 7 % | 6 % | 4 % | 5 % | 14 % | 19 % | 40 % | 51 % |
| Electro | 50 % | 55 % | 49 % | 48 % | 29 % | 46 % | 58 % | 59 % |
| Box spring | 51 % | 61 % | 62 % | 62 % | 38 % | 55 % | 70 % | 71 % |

Cuadro 4.2: *Forecast accuracy* promedio de las 4 categorías estudiadas (para revisar la tabla completa ver el anexo 4.7)

Se obtiene *accuracy* bajo cuando se consideran los datos de ventas de días *cyber*, pero aun considerando estos, las redes neuronales LSTM tienen mejor ajuste. Una vez retirados los valores de estos días y volver a ejecutar los experimentos, los valores mejoran excepto en las series de tiempo. Nuevamente, el modelo y experimento con mejores resultados, son las redes LSTM.

Cabe destacar que la categoría de calefacción es la que tiene peor desempeño, teniendo bajo un 50 % de precisión en todos los modelos. La venta de este tipo de productos se genera durante el invierno, no hubo suficientes datos para poder entrenar estos modelos de mejor manera.

Conclusiones

En esta memoria se recolectaron y se agregaron los datos de ventas diarias de 30 categorías de productos durante dos años en un tienda de retail chilena. Las categorías correspondieron a equipamiento doméstico: refrigeración, calefacción, box-spring y electrodomésticos. La elección de estas categorías reviste en la necesidad de disminuir el tiempo de bodegaje y simplificación de la logística de transporte para la tienda. Los datos contenían las ventas de eventos especiales Cyber Days, que aumentaban las ventas debido a las múltiples ofertas, por lo que uno de los objetivos de este trabajo era evaluar el efecto de estas promociones en los algoritmos de predicción.

Se seleccionaron y evaluaron cuatro modelos: regresión lineal, series de tiempo, redes convolucionales y redes LSTM. Las series de tiempo y las redes convolucionales no produjeron resultados esperados. En la tabla 4.3 se pueden ver valores promedios de accuracy y RMSE para los cuatro modelos.

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|----------|------------------|------|------------------|------|----------|------|-------|------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Accuracy | 43.7 | 48.7 | 44.0 | 43.6 | 27.6 | 43.7 | 56.9 | 59.7 |
| RMSE | 46.5 | 22.2 | 72.8 | 28.5 | 101.8 | 27.9 | 58.3 | 16.9 |

Cuadro 4.3: *Forecast accuracy* y RMSE promedio para los 4 modelos.

Se puede notar que independiente del modelo aplicado, las predicciones tienen menor error y mayor accuracy sin tomar en cuenta los datos de venta de cyber days. O sea, los métodos no se alcanzan a ajustar a un súbito aumento de volumen de ventas. Por lo tanto, las predicciones

para estos días particulares se deben realizar de forma separada de las predicciones de ventas diarias.

Los dos métodos más eficientes para predecir las ventas son: Redes LSTM y Regresión lineal, aunque ninguno de los dos llega a los niveles aceptables. Cabe mencionar que para categorías de productos de menor dinamismo de ventas es más adecuada la regresión lineal, mientras que en caso contrario es mejor emplear las redes LSTM.

El desfase de la predicción de un par de días puede ser arreglado fácilmente y probablemente mejorará bastante los resultados del modelo LSTM, queda para el desarrollo en los trabajos futuros.

Por eso, el objetivo general de comprobar la factibilidad de predicción de estacionalidad de venta de productos en una tienda retail queda comprobado en forma parcial, ya que los modelos logrados tienen una precisión limitada.

Se logró recolectar los datos suficientes para llevar a cabo el trabajo presente. También se logró entender el funcionamiento del *retail* y el impacto de predecir la venta.

Se consiguió ordenar, limpiar y transformar los datos con ayuda de una base de datos local, de tal manera que fue posible generar un dataset, con el que se llevaron a cabo los 4 modelos aquí expuestos, sin realizar muchos cambios para pasar de uno a otro.

Se utilizaron modelos y técnicas presentes hoy en día en las áreas encargadas de control de inventario y procesos logísticos. También modelos de redes neuronales, los cuales a pesar de que pueden ser nuevos para estos negocios, están siendo impulsados por diferentes áreas dado su prometedor rendimiento.

Al utilizar la metodología CRISP-DM fue posible entender el problema y el negocio en el que nos encontramos, se hizo notoria la influencia de los datos cyber dando la oportunidad de iterar los modelos variando sus parámetros como iterar sobre los datos.

Durante este trabajo se tuvo varias limitaciones, principalmente por el computador utilizado, inicialmente se planteó la idea de usar redes neuronales de tipo *Encoder/Decoder*, las cuales inicialmente tuvieron mejor rendimiento que las redes convolucionales, pero la falta de un

computador con tarjeta gráfica dedicada impidieron concluir los experimentos. Por otra parte, al realizar variaciones en parámetros de las series de tiempo o redes convolucionales, como tomar periodos superiores a 30 días, varios experimentos no lograban terminar su ejecución, por lo que debieron descartarse. Por último, las series de tiempo son utilizadas actualmente, pero la falta de conocimientos más avanzados en este campo, llevaron a realizar un modelo menos complejo y por ello menos exacto.

4.7. Trabajo futuro

Considerando que eliminar los datos de ventas de los eventos cyber ayuda a los modelos, se podría estudiar e iterar los modelos quitando más eventos como estos, que tienen un alto volumen de venta aunque inferior a un cyber, como lo son *black friday*, días sin IVA, etc

Estudiar el comportamiento de ventas durante los días cyber, debe constituir un nuevo trabajo y puede llegar a generar un nuevo modelo. En una primera instancia se pueden aplicar los mismos modelos presentes en este trabajo e ir iterando, de esta manera se lograrán mejores resultados.

Como trabajo futuro queda el estudio del comportamiento de las ventas durante la pandemia, al principio todo el volumen de venta presencial, o gran parte de éste, se movió a internet, colapsando los centros de distribución, posteriormente con la apertura o vuelta a la normalidad, se tendería a pensar que las ventas volvieron a como estaban antes, pero mucho de ese volumen se quedó en internet.

Anexo

- Tabla completa de errores RMSE de cada subcategoría estudiada

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|------------------------------|------------------|--------------|------------------|-------|----------|-------|--------|--------------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Refrigeradores (<300 L) | 15.96 | 19.75 | 38.80 | 26.50 | 71.92 | 17.07 | 30.21 | 10.95 |
| Refrigeradores (300 a 400 L) | 48.01 | 15.30 | 68.61 | 18.99 | 112.13 | 17.66 | 56.82 | 11.34 |
| Refrigeradores (>400 L) | 15.96 | 7.39 | 26.98 | 10.34 | 40.22 | 12.01 | 22.02 | 6.50 |
| Estufa eléctrica | 81.35 | 47.29 | 149.04 | 93.27 | 202.58 | 47.79 | 115.96 | 40.16 |
| Estufa a gas | 65.07 | 34.04 | 115.87 | 58.45 | 141.92 | 65.22 | 92.56 | 34.62 |
| Estufa parafina | 45.61 | 17.85 | 73.33 | 28.93 | 171.85 | 12.95 | 53.11 | 6.97 |
| Estufa conv directa | 9.52 | 3.86 | 16.95 | 6.00 | 20.00 | 5.85 | 15.98 | 5.68 |

Cuadro 4.4: Tabla completa de RMSE de la categoría refrigeración y calefacción.

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|------------------------|------------------|--------------|------------------|-------|----------|-------|--------|--------------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Hervidor eléc. | 87.75 | 35.85 | 142.43 | 48.75 | 198.34 | 44.76 | 117.07 | 28.62 |
| Plancha vapor | 59.36 | 14.49 | 89.96 | 20.49 | 121.00 | 24.42 | 71.98 | 17.17 |
| Horno eléctrico | 70.41 | 20.84 | 107.61 | 37.45 | 132.11 | 29.17 | 91.05 | 18.54 |
| Licuadora | 142.08 | 142.08 | 207.47 | 73.49 | 215.27 | 65.80 | 163.30 | 45.01 |
| Minipimer | 42.17 | 18.88 | 51.71 | 29.62 | 68.76 | 48.95 | 41.65 | 25.63 |
| Ollas arroceras | 9.91 | 5.12 | 14.58 | 6.02 | 16.18 | 6.43 | 12.33 | 5.04 |
| Batidora manual | 12.73 | 5.90 | 16.04 | 7.80 | 17.04 | 7.91 | 13.19 | 5.82 |
| Cafetera eléc. filtro | 16.47 | 14.63 | 23.85 | 23.24 | 35.98 | 56.78 | 19.79 | 18.08 |
| Cafetera eléc. express | 102.00 | 77.75 | 136.06 | 98.66 | 251.87 | 70.46 | 90.66 | 27.49 |
| Picador eléc. | 5.88 | 4.93 | 8.80 | 7.28 | 9.85 | 6.32 | 6.15 | 4.71 |
| Sandwichera | 22.32 | 10.56 | 32.66 | 17.42 | 33.24 | 15.14 | 23.83 | 10.68 |
| Parrilla eléc. | 41.99 | 29.02 | 66.90 | 48.98 | 114.55 | 66.96 | 52.29 | 35.45 |
| Tostador eléc. | 20.03 | 6.51 | 32.16 | 7.58 | 40.13 | 9.29 | 27.78 | 6.49 |
| Exprimidor | 3.65 | 2.65 | 5.19 | 3.34 | 6.01 | 3.00 | 4.13 | 2.62 |
| Procesador alimento | 8.98 | 4.99 | 14.35 | 5.04 | 27.68 | 6.88 | 11.58 | 4.41 |
| Saca jugo | 16.26 | 6.80 | 19.00 | 7.05 | 22.50 | 9.31 | 15.30 | 6.48 |
| Máquina hacer pan | 21.22 | 9.85 | 34.58 | 17.19 | 36.98 | 13.07 | 26.04 | 8.81 |
| Freidor eléc. | 31.82 | 7.16 | 57.84 | 11.79 | 65.70 | 13.61 | 47.56 | 8.31 |
| Varios electro menor | 174.98 | 52.92 | 271.87 | 71.88 | 487.41 | 78.10 | 225.86 | 56.78 |
| Batidoras de pedestal | 54.10 | 13.75 | 86.61 | 21.47 | 89.16 | 20.36 | 70.42 | 13.11 |

Cuadro 4.5: Tabla completa de RMSE de la categoría electrodomésticos.

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|--------------------------|------------------|--------------|------------------|-------|----------|-------|--------|-------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Box spring 1.5 plazas | 21.64 | 6.54 | 36.65 | 8.28 | 46.28 | 11.12 | 31.32 | 6.75 |
| Box spring 2 plazas | 98.90 | 20.61 | 162.59 | 26.72 | 159.42 | 33.17 | 138.55 | 23.68 |
| Box spring king | 50.24 | 7.98 | 75.36 | 12.63 | 97.89 | 18.16 | 60.98 | 12.23 |

Cuadro 4.6: Tabla completa de RMSE de la categoría box spring.

- Tabla completa de *Forecast accuracy* de cada subcategoría estudiada

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|------------------------------|------------------|------|------------------|------|----------|------|-------------|-------------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Refrigeradores (<300 L) | 48 % | 50 % | 54 % | 40 % | 22 % | 51 % | 58 % | 66 % |
| Refrigeradores (300 a 400 L) | 50 % | 58 % | 46 % | 48 % | 32 % | 59 % | 69 % | 72 % |
| Refrigeradores (>400 L) | 40 % | 44 % | 44 % | 51 % | 22 % | 46 % | 54 % | 60 % |
| Estufa eléctrica | 2 % | 4 % | 3 % | 4 % | 20 % | 45 % | 62 % | 59 % |
| Estufa a gas | 11 % | 10 % | 1 % | 1 % | 10 % | 17 % | 51 % | 49 % |
| Estufa parafina | 7 % | 4 % | 12 % | 14 % | 13 % | 11 % | 32 % | 18 % |
| Estufa conv directa | 7 % | 5 % | 1 % | 1 % | 11 % | 1 % | 14 % | 79 % |
| Box spring 1.5 plazas | 58 % | 62 % | 59 % | 59 % | 35 % | 53 % | 67 % | 67 % |
| Box spring 2 plazas | 56 % | 66 % | 68 % | 68 % | 46 % | 64 % | 77 % | 79 % |
| Box spring king | 40 % | 56 % | 59 % | 58 % | 34 % | 47 % | 65 % | 66 % |

Cuadro 4.7: Tabla completa de *forecast accuracy* de la categoría refrigeración, calefacción y box spring.

| | Regresión lineal | | Series de tiempo | | Red conv | | LSTM | |
|------------------------|------------------|-------------|------------------|------|----------|------|-------------|-------------|
| | cyber | - | cyber | - | cyber | - | cyber | - |
| Hervidor eléc. | 54 % | 58 % | 55 % | 50 % | 33 % | 62 % | 74 % | 77 % |
| Plancha vapor | 55 % | 63 % | 57 % | 58 % | 30 % | 51 % | 67 % | 66 % |
| Horno eléctrico | 61 % | 67 % | 64 % | 54 % | 45 % | 64 % | 76 % | 77 % |
| Licuadaora | 54 % | 61 % | 55 % | 54 % | 38 % | 60 % | 61 % | 71 % |
| Minipimer | 48 % | 58 % | 56 % | 45 % | 24 % | 27 % | 45 % | 53 % |
| Ollas arroceras | 44 % | 48 % | 48 % | 47 % | 25 % | 43 % | 49 % | 42 % |
| Batidora manual | 53 % | 53 % | 54 % | 54 % | 45 % | 50 % | 55 % | 60 % |
| Cafetera eléc. filtro | 51 % | 53 % | 51 % | 52 % | 31 % | 33 % | 64 % | 64 % |
| Cafetera eléc. express | 52 % | 58 % | 41 % | 45 % | 32 % | 53 % | 66 % | 69 % |
| Picador eléc. | 50 % | 51 % | 40 % | 41 % | 32 % | 44 % | 49 % | 49 % |
| Sandwichera | 49 % | 54 % | 50 % | 53 % | 35 % | 49 % | 63 % | 62 % |
| Parrilla eléc. | 56 % | 59 % | 56 % | 55 % | 35 % | 51 % | 68 % | 66 % |
| Tostador eléc. | 56 % | 62 % | 57 % | 58 % | 27 % | 49 % | 60 % | 60 % |
| Exprimidor | 38 % | 40 % | 40 % | 39 % | 28 % | 38 % | 45 % | 45 % |
| Procesador alimento | 48 % | 51 % | 47 % | 49 % | 9 % | 40 % | 49 % | 49 % |
| Saca jugo | 45 % | 51 % | 40 % | 46 % | 28 % | 36 % | 28 % | 36 % |
| Máquina hacer pan | 40 % | 43 % | 13 % | 10 % | 27 % | 36 % | 53 % | 51 % |
| Freidor eléc. | 37 % | 48 % | 45 % | 46 % | 12 % | 34 % | 58 % | 56 % |
| Varios electro menor | 63 % | 70 % | 61 % | 64 % | 30 % | 55 % | 65 % | 64 % |
| Batidoras de pedestal | 39 % | 53 % | 42 % | 45 % | 18 % | 41 % | 62 % | 59 % |

Cuadro 4.8: Tabla completa de *forecast accuracy* de la categoría electrodomésticos.

Bibliografía

- [1] A. L. Samuel: *Some studies in machine learning using the game of checkers*. IBM Journal of Research and Development, 1959.
- [2] Alex Krizhevsky, Ilya Sutskever y Geoffrey E Hinton: *Imagenet classification with deep convolutional neural networks*. Communications of the ACM, 60, 2017.
- [3] Aurélien Géron: *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor-Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'reilly & Assoc Inc, 2019.
- [4] Comité de comercio electrónico: *CYBERDAY 2017*, 2017. <https://www.ecommerceccs.cl/comunicado-final-cyberday-2017-2>.
- [5] D. H. Hubel y T. N. Wiesel: *Receptive fields and functional architecture of monkey striate cortex*. The Journal of physiology, 1968.
- [6] David Kriesel: *A Brief Introduction to Neural Networks*. 2007. <http://www.dkriesel.com>.
- [7] Ian Goodfellow, Yoshua Bengio y Aaron Courville: *Deep learning*. The MIT Press, 2016.
- [8] John Neter, Michael H. Kutner, Christopher J. Nachtsheim y W. Wasserman: *Applied linear statistical models*. McGraw-Hill/Irwin, 1996.
- [9] Martin T. Hagan, Howard B. Demuth y Mark H. Beale: *Neural Network Design*. Martin Hagan, 2014.
- [10] Peter Chapman, Janet Clinton, Randy Kerber, Tom Khabaza, Thomas P. Reinartz, Colin Shearer y Richard Wirth: *CRISP-DM 1.0: Step-by-step data mining guide*. 2000.
- [11] Peter J. Brockwell y Richard A. Davis: *Introduction to Time Series and Forecasting*. Springer, 2016.
- [12] Peter S. Fader y Bruce G. S. Hardie: *Modeling consumer choice among SKUs*. Journal of Marketing Research, 1996.

- [13] Ronald J. Williams y David Zipsere: *Gradient-based learning algorithms for recurrent. Backpropagation: Theory, architectures, and applications*. L. Erlbaum Associates Inc, 1995.
- [14] Sepp Hochreiter y Jürgen Schmidhuber: *Long Short-Term Memory*. Neural Computation, 1997.
- [15] Sungil Kim y Heeyoung Kim: *A new metric of absolute percentage error for intermittent demand forecasts*. International Journal of Forecasting, 2016.
- [16] Taha Falatouri, Farzaneh Darbanian, Patrick Brandtner y Chibuzor Udokwu: *Predictive Analytics for Demand Forecasting - A Comparison of SARIMA and LSTM in Retail SCM*. Procedia Computer Science, 2022.
- [17] Tom M. Mitchell: *Machine Learning*. McGraw-Hill Education, 1997.
- [18] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner: *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 1998.
- [19] Özden GürAli, Serpil Sayın, Tom van Woensel y Jan Fransoo: *SKU demand forecasting in the presence of promotions*. Expert Systems with Applications, 2009.