



Universidad Politécnica  
de Madrid



**Escuela Técnica Superior de  
Ingenieros Informáticos**

Grado en Ingeniería Informática

Trabajo Fin de Grado

**Diseño de una Metodología de Cálculo  
del Valor del Cliente (CLTV) Basada en  
Redes Neuronales para la Predicción del  
Comportamiento de los Consumidores  
en un Entorno Empresarial**

Autor: Lucas Abad Bermejo

Tutora: Consuelo Gonzalo Martin,

Tutora: Rocío González Martínez

Madrid, Junio 2024

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

*Trabajo Fin de Grado*

*Grado en Ingeniería Informática*

*Título:* Diseño de una Metodología de Cálculo del Valor del Cliente (CLTV)  
basada en Redes Neuronales para la Predicción del Comportamiento  
de los Consumidores en un Entorno Empresarial Real

Junio 2024

*Autor:* Lucas Abad Bermejo

*Tutora:*

Consuelo Gonzalo Martin

Arquitectura y Tecnología de Sistemas Informáticos

ETSI Informáticos

Universidad Politécnica de Madrid

# Resumen

En el dinámico entorno empresarial actual, caracterizado por una competencia feroz y una evolución constante de las expectativas de los clientes, la capacidad que tiene una empresa para comprender y anticipar las necesidades y comportamientos de sus clientes se ha convertido en una piedra angular para el éxito sostenido. En este escenario, el cálculo del Valor del Cliente a lo Largo del Tiempo (CLTV, por sus siglas en inglés) emerge como una herramienta estratégica esencial, permitiendo a las empresas no solo entender el valor actual de sus clientes, sino también poder predecir su comportamiento de compra en el futuro.

En este trabajo de fin de grado se diseña una metodología de creación de modelo de Redes Neuronales para la predicción del CLTV. El objetivo principal es desarrollar una metodología innovadora que permita evaluar el valor monetario futuro de los clientes existentes y comparar sus resultados con los modelos tradicionales de predicción.

La metodología propuesta incluye la recopilación y análisis de los datos históricos proporcionados por la empresa, seguido del entrenamiento de cada uno de los valores a predecir necesarios para el cálculo del CLTV empleando la misma red neuronal.

El trabajo concluye con la implementación de una red densa capaz de predecir con efectividad y precisión el valor de cliente en un periodo de 6 meses con un mejor rendimiento que el cálculo generado con modelos tradicionales como son el modelo BetaGeoFitter y el modelo Gamma-Gamma. Además de nombrar las ventajas e inconvenientes que tiene este tipo de modelos frente a los modelos tradicionales.

Este estudio aporta una nueva perspectiva sobre el cálculo de valor de cliente a lo largo del tiempo empleando redes neuronales y ofrece recomendaciones de trabajos futuros para este campo.

**Palabras clave:** Customer Lifetime Value, Redes Neuronales, Análisis de Datos, Valor de Cliente, Inteligencia Artificial.

# Abstract

In today's dynamic business environment, characterized by fierce competition and the constant evolution of customer expectations, a company's ability to understand and anticipate its customers' needs and behaviors has become a cornerstone for sustained success. In this scenario, the calculation of Customer Lifetime Value (CLTV) emerges as an essential strategic tool, enabling companies not only to understand the current value of their customers but also to predict their future purchasing behavior.

This undergraduate thesis designs a methodology for creating a Neural Network model to predict CLTV. The main objective is to develop an innovative methodology that allows evaluating the future monetary value of existing customers and comparing its results with traditional prediction models.

The proposed methodology includes the collection and analysis of historical data provided by the company, followed by training each of the values to be predicted necessary for the calculation of CLTV using the same neural network.

The thesis concludes with the implementation of a dense network capable of effectively and accurately predicting customer value over a 6-month period, with better performance than calculations generated with traditional models such as the BetaGeoFitter model and the Gamma-Gamma model. Additionally, it names the advantages and disadvantages of this type of model compared to traditional models.

This study provides a new perspective on the calculation of customer lifetime value using neural networks and offers recommendations for future work in this field.

Keywords: Customer Lifetime Value, Neural Networks, Data Analysis, Customer Value, Artificial Intelligence.

# Tabla de contenidos

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Objetivos del TFG	2
1.2	Estructura del documento	3
1.2.1	Marco Teórico	3
1.2.2	Análisis y Transformación	3
1.2.3	Diseño del modelo	3
1.2.4	Comparativa de los Modelos Tradicionales	3
1.2.5	Metodología de cálculo de CLTV	3
1.2.6	Resultados	4
1.2.7	Conclusiones y Trabajos Futuros	4
1.2.8	Análisis de Impacto	4
<b>2</b>	<b>Marco teórico</b>	<b>4</b>
2.1	Definición de Customer Lifetime Value	4
2.2	Cómo calcular la métrica CLTV	4
2.3	Aplicaciones del CLTV	5
2.4	Modelo RFM	6
2.5	Modelos Tradicionales de Predicción de CLTV	6
2.5.1	Modelo de BetaGeoFitter	6
2.5.1.1	Fundamentos Teóricos	7
2.5.1.2	Parámetros del Modelo	7
2.5.1.3	Entrenamiento del Modelo	7
2.5.1.4	Implementación y Ventajas	7
2.5.1.5	Limitaciones	7
2.5.1.6	Uso Conjunto con el Modelo Gamma-Gamma	8
2.5.2	Modelo Gamma-Gamma	8
2.5.2.1	Fundamentos Teóricos	8
2.5.2.2	Parámetros del Modelo	8
2.5.2.3	Implementación y Ventajas	8
2.5.2.4	Limitaciones	9
2.5.2.5	Empleo para el Cálculo del CLTV	9
2.6	Redes Neuronales	9
2.6.1	Modelo de Red Convolutiva	9
2.6.1.1	Fundamentos Teóricos	9
2.6.1.2	Arquitectura de una CNN	10
2.6.1.3	Ventajas y Limitaciones	11
2.6.2	Modelo de Red Recurrente	12
2.6.2.1	Fundamentos Teóricos	12
2.6.2.2	Arquitectura de una RNN	12

2.6.2.3	Ventajas y Limitaciones .....	13
2.6.3	Modelo de Red Densa.....	14
2.6.3.1	Fundamentos Teóricos.....	14
2.6.3.2	Arquitectura de una DNN .....	14
<b>3</b>	<b>Desarrollo .....</b>	<b>15</b>
3.1	Comprensión del Negocio .....	16
3.2	Análisis Exploratorio .....	16
3.3	Sábana de Predicción .....	20
3.4	Modelos Anteriores.....	25
3.4.1	Modelo basado en Trabajos Previos.....	25
3.4.2	Modelo Convolutacional.....	26
3.4.3	Modelo DNN.....	27
3.5	Modelo DNN Implementado .....	29
3.5.1	Desarrollo del Modelo .....	29
3.6	Evaluación del Modelo.....	32
<b>4</b>	<b>Comparativa de los Modelos Tradicionales .....</b>	<b>36</b>
4.1	Modelo BetaGeoFitter .....	37
4.2	Modelo Gamma-Gamma.....	38
4.3	Evaluación con el Modelo DNN.....	39
<b>5</b>	<b>Metodología de cálculo de CLTV.....</b>	<b>42</b>
5.1	Entendimiento de Negocio .....	43
5.2	Análisis, Limpieza y Creación de Sábana de Datos .....	43
5.3	Desarrollo del Modelo.....	43
5.4	Evaluación de resultados .....	44
5.5	Implementación .....	44
5.6	Resumen de la Metodología .....	44
<b>6</b>	<b>Resultados .....</b>	<b>46</b>
6.1	Modelo basado en ZILN .....	46
6.2	Modelo CNN .....	46
6.3	Modelo DNN .....	46
6.4	Modelo BetaGeoFitter .....	47
6.5	Modelo Gamma-Gamma.....	47
6.6	Modelo Implementado .....	47
<b>7</b>	<b>Conclusión.....</b>	<b>49</b>
7.1	Trabajos Futuros.....	50
<b>8</b>	<b>Análisis de Impacto .....</b>	<b>51</b>
8.1	Ámbito Académico.....	51
8.2	Sector Empresarial.....	51
8.3	Impacto en la Comunidad Científica y Tecnológica .....	52

<b>9 Bibliografia .....</b>	<b>53</b>
-----------------------------	-----------

# 1 Introducción

En el actual mercado competitivo, donde las expectativas de los clientes están en constante aumento, las herramientas para el seguimiento y análisis del comportamiento de los clientes se han vuelto esenciales para el buen funcionamiento de las empresas. Una de las métricas más utilizadas para la comprensión de este comportamiento es el Valor de Vida del Cliente a lo Largo del Tiempo, denominado como CLTV debido a sus siglas en inglés (Customer Lifetime Value). Esta métrica tiene como objetivo principal evaluar el valor monetario que tendrá cada cliente en un tiempo determinado proporcionando información esencial para la toma de decisiones a corto, mediano y largo plazo. Esta métrica también sirve para poder determinar qué cliente es rentable, Kotler y Armstrong constituyeron la definición de cliente rentable como “persona, hogar o empresa cuyos ingresos a lo largo del tiempo superan, por una cantidad aceptable, los costos de la empresa en atraer, vender y atender al cliente” (Kotler y Armstrong, 1996).

La métrica CLTV fue definida en 1974 por el economista estadounidense Philip Kotler como “el valor actual del flujo de ganancias futuras esperado dado un horizonte temporal de transacciones con el cliente” (Kotler, 1974). A lo largo de las décadas, se han desarrollado diversos enfoques y modelos para la estimación del CLTV. Tradicionalmente, se han empleado modelos basados en el análisis de compras anteriores de clientes, como el modelo RFM (Recency, Frequency, Monetary) que analiza el comportamiento de los clientes basándose en tres dimensiones clave: la recencia de su última compra, la frecuencia de sus compras y el valor monetario de las mismas (Roberts y Berger, 1989).

En la actualidad los modelos más utilizados para el cálculo del CLTV incluyen el modelo Gamma-Gamma y la distribución Binomial Negativa. Sin embargo, con los avances en inteligencia artificial y el aprendizaje automático, han surgido nuevas metodologías que prometen mejorar la precisión de estas estimaciones.

El objetivo principal de este proyecto es el diseño de una metodología innovadora en la predicción del CLTV empleando Redes Neuronales, basada en los datos transaccionales de una empresa dedicada a la venta de productos alimenticios a domicilio, con una frecuencia de compra quincenal. Para ello, se llevará a cabo un estudio exhaustivo de diferentes algoritmos, modelos y enfoques aplicables al sector y a los datos proporcionados por la empresa.

Las preguntas de investigación que guían este proyecto son:

- 1- ¿Cuáles son los principales modelos empleados en el cálculo del CLTV?
- 2- ¿Cómo ha evolucionado el uso de modelos de inteligencia artificial en la predicción en la predicción del CLTV?
- 3- ¿Qué variables son las más significativas para la predicción del CLTV?

Este trabajo se centrará en tres áreas principales:

- La comprensión del negocio, que incluye la preparación, creación de variables y limpieza de los datos para la ingesta en el modelo.
- El diseño y entrenamiento del modelo de redes neuronales.
- La evaluación de resultados, comparando el modelo desarrollado con modelos tradicionales como el Gamma-Gamma y la distribución Binomial Negativa.

Este proyecto se ha desarrollado completamente en Python empleando bibliotecas como matplotlib, pyplot y seaborn para la visualización de los datos, pandas para el manejo y limpieza de datos, keras y lifetimes para el desarrollo de los modelos de predicción del Customer Lifetime Value.

## 1.1 Objetivos del TFG

Para el desarrollo de este trabajo se plantean los siguientes objetivos:

### Objetivo principal

El objetivo principal del documento es el diseño de una metodología de cálculo del Customer Lifetime Value adaptado a una empresa del sector alimentario con una frecuencia de compra quincenal mediante un modelo basado en redes neuronales y una comparativa con los modelos tradicionalmente empleados para la predicción del CLTV.

### Objetivos específicos

- Comprensión del negocio: Realizar un análisis de los datos proporcionados para entender los objetivos de la empresa y la razón por la cual necesitan una métrica como el Customer Lifetime Value.
- Análisis Exploratorio de los datos proporcionados: Implementar técnicas de análisis que ayuden al entendimiento de los datos y a un análisis exhaustivo.
- Aplicación de metodologías de análisis de datos: Buscar e implementar técnicas de análisis de datos
- Modelización del modelo: Diseño e implementación de modelo basado en redes neuronales para la predicción del CLTV.
- Explicación de los resultados obtenidos de la red neuronal: Interpretación de los resultados que se obtengan de la ejecución del modelo desarrollado.
- Comparativa con modelos tradicionales: Elaborar una comparativa de resultados del modelo desarrollado en este trabajo con modelos tradicionalmente usados en el entorno empresarial en la actualidad.
- Descripción de la metodología diseñada: Elaborar un esquema donde se describa cada paso para la implementación de la metodología replicable para otros casos de negocio.
- Documentación del desarrollo del trabajo: Recopilar la información de todo proceso desde el inicio del trabajo hasta la final redacción de este documento.

## **1.2 Estructura del documento**

En esta sección se detalla la estructura del documento junto a una breve introducción de los temas que se van a tratar en cada uno de los 6 capítulos.

### **1.2.1 Marco Teórico**

En este capítulo se redactará los fundamentos teóricos que se consideran relevantes para el entendimiento de este proyecto. Se definirá el valor de cliente, cómo se calcula y cómo se aplica en diferentes sectores, también se explicarán los modelos tradicionales que posteriormente en el documento se emplearán para la comparación de resultados del modelo de redes neuronales.

### **1.2.2 Análisis y Transformación**

En este capítulo se dedica al análisis de los datos proporcionados por la empresa, al entendimiento del negocio para comprender el trabajo que se va a realizar y a las diferentes transformaciones necesarias para la creación de la sabana de datos que se empleará en la predicción y un análisis posterior de esta sabana para asegurarnos el buen comportamiento a posteriori del modelo.

### **1.2.3 Diseño del modelo**

Este capítulo está dedicado a la explicación de los diferentes modelos de redes neuronales que se pueden aplicar para resolver el problema de negocio de la predicción de las variables necesarias para el cálculo del valor de cliente.

También se desarrollará una explicación de la estructura de la arquitectura de la red.

Por último, se analizarán los resultados predichos del modelo desarrollado.

### **1.2.4 Comparativa de los Modelos Tradicionales**

En este capítulo está dedicado a la explicación de la implementación de los modelos tradicionales que se van a comparar. Por último, los resultados de ambos modelos se compararán con los resultados predichos por el modelo desarrollado en proyecto.

### **1.2.5 Metodología de cálculo de CLTV**

En este capítulo se desarrollará y explicará cada uno de los pasos necesarios para implementar la metodología de cálculo de Customer Lifetime Value.

Por último, se pondrá como ejemplo la implementación en otra empresa real dedicada al sector del retail.

### **1.2.6 Resultados**

En este capítulo se realizará un resumen de los diferentes resultados extraídos de todos los modelos implementados en la realización de este proyecto proporcionando todas las variables empleadas para la extracción de los insights que se recogen en este documento

### **1.2.7 Conclusiones y Trabajos Futuros**

En el séptimo capítulo se contextualizará y se comentarán las conclusiones alcanzadas durante la elaboración del trabajo, también, se nombrarán algunos trabajos futuros que se podrán desarrollar a partir de este trabajo.

### **1.2.8 Análisis de Impacto**

En este último capítulo se empleará para realizar un análisis del impacto de este trabajo de fin de grado en diferentes sectores como en el académico o en el sector del negocio entre otros.

## **2 Marco teórico**

En esta sección se redactará los fundamentos teóricos que se consideran relevantes para el entendimiento de este proyecto. Se definirá el valor de cliente, cómo se calcula y cómo se aplica en diferentes sectores, también se explicarán los modelos tradicionales que posteriormente en el documento se emplearán para la comparación de resultados del modelo de redes neuronales.

### **2.1 Definición de Customer Lifetime Value**

La métrica Customer Lifetime Value (CLTV) representa el valor monetario de cada cliente o grupo de clientes. Esta medida evalúa la calidad de la relación entre el proveedor y el cliente, permitiendo determinar la rentabilidad de cada cliente y tomar decisiones estratégicas de negocio. Las empresas valoran esta métrica porque el CLTV agregado refleja el beneficio total aportado por los clientes a la empresa (Buttle & Groeger, 2015).

### **2.2 Cómo calcular la métrica CLTV**

El cálculo del Customer Lifetime Value puede ser diferente dependiendo de la complejidad, disponibilidad de los datos e intereses de la empresa. Una de las fórmulas más empleadas para su obtención fue definida por Peter Fader y Bruce Hardie (Fader et al., 2009):

$$CLTV = \sum_{t=1}^T \frac{R_t * M_t}{(1 + d)^t}$$

**Ecuación 1** Cálculo tradicional CLTV

Donde  $R_t$  es la tasa de retención durante el periodo  $t$ ,  $M_t$  es el margen de contribución durante el periodo  $t$  y siendo  $d$  la tasa de descuento.

Este tipo de cálculo está basado en la segmentación de clientes mediante cohortes, calculando su valor para cada una de ellas. Para obtener su CLTV total se ha de realizar una suma de todos los segmentos.

Otro enfoque de cálculo del CLTV que no se basa en la segmentación de cohortes es la ecuación propuesta por Sunil Gupta, Donald R. Lehmann y Jennifer A. Stuart (Gupta et al., 2004):

$$CLTV = \frac{V * F * M * G}{C}$$

### **Ecuación 2** Cálculo de CLTV por cohortes

Donde  $V$  es el valor promedio de una transacción,  $F$  es la frecuencia de compra,  $M$  el margen de beneficio,  $G$  la duración esperada de ese cliente y  $C$  es la tasa de descuento; esta última es opcional y puede no ser empleada si se quiere ajustar por el valor temporal del dinero.

Basado en las necesidades de la empresa, en nuestro caso el cálculo de CLTV se hará utilizando la **Ecuación 3**. Las diferentes fórmulas del CLTV las podemos definir por Valor de tiempo multiplicado por Valor monetario. Para las necesidades de la empresa se ha tomado la decisión que el valor de tiempo sea la probabilidad que se mantenga activo, considerando como activo a todo cliente que realice una compra en los próximos seis meses, dando más importancia a aquellos clientes que se vayan a mantener en el periodo estudiado:

$$CLTV = Prob_{vivo} \times Num_{compras} \times Mean_{compras}$$

### **Ecuación 3** Cálculo de CLTV del proyecto

Donde  $Prob_{vivo}$  es la probabilidad de que el cliente se mantenga activo durante el periodo que se quiere calcular,  $Num_{compras}$  es la cantidad de compras realizadas durante el periodo y  $Mean_{compras}$  es la media de compras del cliente.

Además de tener el cálculo del CLTV de esta manera también se dispone de la probabilidad de supervivencia y su complementario, probabilidad de abandono de cliente, ambas siendo métricas muy valiosas para la toma de decisiones.

## **2.3 Aplicaciones del CLTV**

La métrica del Customer Lifetime Value (CLTV) se puede utilizar en diversas aplicaciones, entre las más comunes se encuentran la segmentación de clientes y la personalización del marketing para distintos segmentos. Además, ayuda a medir el retorno de la inversión de las campañas de marketing.

Contar con una métrica de CLTV permite una mejor expansión, ya que se prioriza la adquisición de clientes con un posible CLTV más alto, lo que se traduce en un mayor beneficio para la empresa a mediano y largo plazo.

## 2.4 Modelo RFM

El modelo RFM es un modelo de segmentación que se nombra por las siglas de las tres componentes con los que se construye:

- **Recency (Recencia):** Tiempo transcurrido desde que se ha realizado la última compra de un cliente.
- **Frequency (Frecuencia):** Número de compras que ha realizado un cliente en un determinado periodo.
- **Monetary (Valor Monetario):** Cantidad de dinero gastado por un cliente durante un periodo determinado.

Estos componentes proporcionan información valiosa sobre el comportamiento del cliente por lo que se integran en modelos predictivos de Customer Lifetime Value. Por ejemplo, modelos de regresión logística o lineales emplean las variables RFM como predictores para estimar tanto la cuantía de compra como el número de compras futuras (Zhao et al., 2019).

Dentro del modelo RFM los clientes se agrupan por quintiles o deciles para cada una de las variables RFM, asignando puntuaciones del 1 a 5 o del 1 a 10, respectivamente. Por ejemplo, un valor de 5 en el caso de segmentar a los clientes por quintiles se refiere a un cliente con baja recencia y alta frecuencia y valor monetario.

En predicciones de CLTV basado en cohortes, se emplean técnicas de segmentación basadas en las puntuaciones RFM para definir dichas cohortes. Estos segmentos se emplean en modelos como Gamma-Gamma o BetaGeoFitter, modelos que se emplearán en la validación de resultados posteriormente.

## 2.5 Modelos Tradicionales de Predicción de CLTV

En la predicción del Customer Lifetime Value tradicionalmente se emplean diferentes tipos de modelos. Para la realización de este proyecto se ha empleado dos modelos para la predicción, el modelo BetaGeoFitter y el modelo Gamma-Gamma.

### 2.5.1 Modelo de BetaGeoFitter

El modelo BetaGeoFitter es una herramienta estadística para la predicción del CLTV. Este modelo es especialmente adecuado para empresas con bajo volumen de datos como minoristas o empresas de suscripción.

### 2.5.1.1 Fundamentos Teóricos

El modelo BetaGeoFitter se basa en el modelo Binomial Negativo que es representado por las siglas BN/NBD que fue desarrollado por Fader Hardie y Lee en 2005. Este modelo asume:

- **Tiempo hasta la compra:** La probabilidad de que un cliente realice una compra sigue una distribución geométrica.
- **Tasa de compra:** La tasa de los clientes sigue una distribución beta.
- **Periodo inter-compra:** Se denomina periodo inter-compra al número de días que pasan entre dos compras, esta medida sigue una distribución exponencial.

En estas distribuciones permiten el modelaje de la heterogeneidad en el comportamiento de compra de los clientes.

### 2.5.1.2 Parámetros del Modelo

El modelo BetaGeoFitter emplea los siguientes parámetros para la predicción de los valores:

- **$r$  y  $\alpha$ :** Parámetros de la distribución betas que modelan la heterogeneidad en la tasa de compra.
- **$a$  y  $b$ :** Parámetros de la distribución beta que modelan la heterogeneidad en la probabilidad de abandono.

### 2.5.1.3 Entrenamiento del Modelo

El modelo se entrena empleando el número de compras realizados por el cliente, el tiempo desde que se realizó la última compra, el *tenure* del cliente, es el tiempo total desde que el cliente ha realizado su primera compra hasta la fecha actual del estudio, y la edad del cliente.

### 2.5.1.4 Implementación y Ventajas

El modelo BetaGeoFitter es sencillo de implementar empleando bibliotecas estadísticas que están disponibles dentro de diferentes lenguajes de programación en el caso de Python se encuentra en la biblioteca ***lifetimes***.

Algunas de las ventajas que tiene este modelo frente a otros es su eficiencia a nivel computacional permitiendo su aplicación en grandes conjuntos de datos sin requerir recursos excesivos.

### 2.5.1.5 Limitaciones

Las suposiciones del modelo, aunque útiles, pueden no ser precisas en todos los sectores empresariales. La suposición de que la tasa de compra sigue una distribución beta puede no ser adecuada para negocios con frecuencias de compra muy irregulares como puede ser un negocio dedicado al sector del retail.

### 2.5.1.6 Uso Conjunto con el Modelo Gamma-Gamma

Para una predicción más precisa del Customer Lifetime Value, el BetaGeoFitter se utiliza junto al modelo Gamma-Gamma. Siendo el BetaGeoFitter para la predicción de la probabilidad de abandono del cliente y la frecuencia de compras estimadas y empleando el modelo Gamma-Gamma para predecir el valor monetario de las transacciones futuras.

## 2.5.2 Modelo Gamma-Gamma

El modelo Gamma-Gamma es una herramienta estadística utilizada para predecir el Customer Lifetime Value (CLTV). Este modelo es especialmente adecuado para datos transaccionales repetidos.

### 2.5.2.1 Fundamentos Teóricos

El modelo Gamma-Gamma se basa en la suposición de que tanto el gasto monetario por transacción de compra sigue una distribución Gamma al igual que la distribución que sigue la media de las transacciones de cada cliente.

### 2.5.2.2 Parámetros del Modelo

Los parámetros clave que emplea el modelo Gamma-Gamma son:

- **Alpha y Beta:** Parámetros que describen la variabilidad en el gasto monetario de las transacciones.
- **p y q:** Parámetros que describen la variabilidad en la media del gasto por transacción de cada cliente.

La función de verosimilitud del modelo se maximiza para estimar los parámetros **Alpha**, **Beta**, **p** y **q** que mejor se ajustan a los datos observados. Este proceso permite capturar la variabilidad en el comportamiento de los clientes relacionado con el gasto en sus transacciones.

### 2.5.2.3 Implementación y Ventajas

El Gamma-Gamma es sencillo de implementar empleando bibliotecas estadísticas que se encuentran disponibles en distintos lenguajes de programación en el caso del lenguaje de programación Python se emplea la biblioteca *lifetimes*.

El modelo es eficiente a nivel computacional, permitiendo su aplicación en grandes conjuntos de datos sin requerir recursos excesivos.

#### **2.5.2.4 Limitaciones**

Aunque el modelo Gamma-Gamma es muy flexible y puede manejar datos sesgados, su efectividad depende de que la distribución Gamma sea adecuada para los datos específicos que se están analizando. En algunos casos, los patrones de gasto no se pueden describir con una distribución Gamma, lo que reduce la efectividad del modelo.

#### **2.5.2.5 Empleo para el Cálculo del CLTV**

Del uso de estos dos modelos de manera conjunta, BetaGeoFitter y Gamma-Gamma, se obtiene una predicción integral del Customer Lifetime Value de forma simple y eficaz para el desarrollo de estrategias de marketing y retención de clientes más efectivas y precisas.

### **2.6 Redes Neuronales**

En este proyecto el modelo desarrollado está basado en redes neuronales por lo que en este apartado se va a dar una breve explicación de los diferentes modelos de redes neuronales aplicados a la predicción del Customer Lifetime Values.

#### **2.6.1 Modelo de Red Convolutiva**

Las Redes Neuronales Convolutivas (**CNN**) han surgido como una herramienta avanzada para la predicción del Customer Lifetime Value debido a su capacidad para capturar patrones complejos y no lineales en datos transaccionales y de comportamiento de cliente. Este enfoque es particularmente valioso en escenarios donde los métodos tradicionales, como los modelos Gamma-Gamma y BetaGeoFitter anteriormente explicados, no logran captar todas las interacciones y dependencias en los datos (Zhang et al., 2019).

##### **2.6.1.1 Fundamentos Teóricos**

Las CNN son un tipo de red neuronal profunda que emplean operaciones convolutivas para procesar datos con una estructura de cuadrícula, imágenes o secuencias temporales.

Dentro del marco del Customer Lifetime Value, los datos transaccionales de los clientes pueden ser representados como una serie temporal, donde cada entrada puede incluir variables como la fecha en la que se realizó la compra, el coste total de esta, tipología de los productos comprados, entre otras (Lecun et al., 1998).

### 2.6.1.2 Arquitectura de una CNN

Una red neuronal convolucional está formada por diferentes capas. Para un mejor entendimiento se ha añadido la **Figura 0.1** donde se muestra un ejemplo de estructura de red. Las capas con las que está formada una red convolucional son:

- **Capas de Entrada y Preprocesamiento**

- **Entrada de Datos:** Capa de ingesta de datos a la CNN. Los datos transaccionales de los clientes tendrán que ser representados en matrices donde las filas puedan representar cada uno de los clientes y las columnas representan las diferentes características de las transacciones.
- **Normalización y Escalado:** Se aplican técnicas de normalización y escalado para la estandarización de los datos, facilitando el aprendizaje de la red. Esto es crucial para manejar la variabilidad de los datos.

- **Capas Convolucionales**

- **Convolución 1D:** Para datos secuenciales se emplean este tipo de capas que aplican filtros a lo largo de la dimensión temporal de las secuencias. Estos filtros son capaces de captar patrones temporales y relaciones entre transacciones consecutivas.
- **Extracción de Características:** Los filtros convolucionales generan mapas de características que resaltan patrones relevantes en los datos. Estos mapas de características son esencialmente, representaciones de los datos originales donde las características importantes obtienen más peso mientras que las características más irrelevantes reducen su peso.

- **Capas de Pooling**

- **Max Pooling:** Esta capa se aplica para reducir la dimensionalidad de los mapas de características, manteniendo las características más importantes y reduciendo la cantidad de los parámetros a aprender. Se emplea para controlar el sobreajuste y para mejorar la generalización del modelo.
- **Average Pooling:** Esta capa, al contrario que la capa anterior, Max Pooling, se emplea para suavizar las representaciones, promediando las características dentro de cada región del mapa de características.

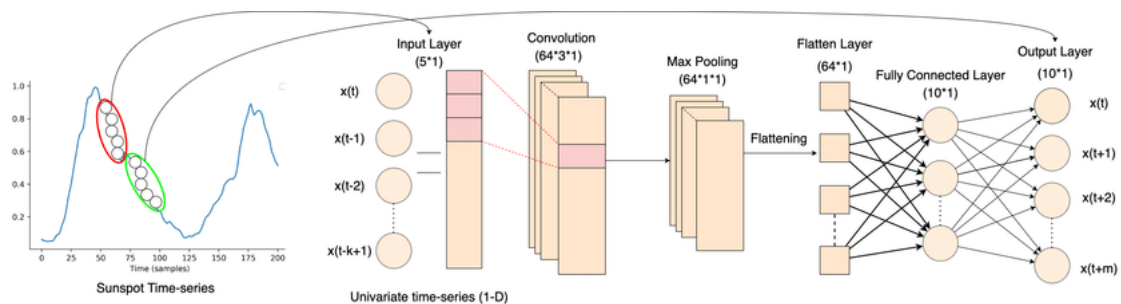
- **Capas Completamente conectadas**

- **Aplanamiento:** Normalmente después de una capa de pooling se aplanan con este tipo de capas en un vector unidimensional que alimenta a las capas completamente conectadas.
- **Capas Densas:** Estas capas combinan las características extraídas para formar representaciones más abstractas. Estas capas aprenden combinaciones no lineales de las características.
- **Salida:** Esta es la última capa densa que genera la salida del modelo.

- **Funciones de Activación y Optimización**

Dentro de las capas se encuentran este tipo de funciones que realizan diferentes transformaciones a los datos.

- **ReLU (Rectified Linear Unit):** La función de activación ReLU se emplea para introducir no linealidades en el modelo. Esto permite a la red aprender representaciones más complejas.
- **Algoritmo de Optimización Adam:** El algoritmo Adam ajusta los pesos del modelo minimizando una función de pérdida que mide el error entre las predicciones y los valores reales.



**Figura 0.1** Representación de Red CNN

Imagen extraída de [ResearchGate]([https://www.researchgate.net/figure/One-dimensional-convolutional-neural-network-for-multi-step-ahead-time-series-prediction\\_fig4\\_352013154](https://www.researchgate.net/figure/One-dimensional-convolutional-neural-network-for-multi-step-ahead-time-series-prediction_fig4_352013154))

### 2.6.1.3 Ventajas y Limitaciones

El uso de CNN para la predicción de CLTV ofrece diferentes ventajas frente a modelos tradicionales como la capacidad de capturar los patrones complejos que se encuentran en datos transaccionales como las tendencias de compra, la escalabilidad que tienen estos modelos comparados con el Gamma-Gamma o el BetaGeoFitter y además de que las CNN pueden adaptarse a diferentes tipos de datos y configuraciones de entrada.

Sin embargo, también presentan algunas limitaciones:

- **Requiere Gran Cantidad de Datos.**
- **Complejidad Computacional**
- **Interpretabilidad:** Las CNN son modelos de caja negra, es decir, que dificulta mucho la interpretación de sus predicciones.

## 2.6.2 Modelo de Red Recurrente

Las Redes Neuronales Recurrentes (**RNN**) también se han convertido en una herramienta capaz de predecir el Customer Lifetime Value, especialmente para el análisis de datos secuenciales y temporales. Las RNN son especialmente útiles para captar patrones y dependencias a lo largo del tiempo en los datos tradicionales de los clientes (Hochreiter & Schmidhuber, 1997).

### 2.6.2.1 Fundamentos Teóricos

Las RNN son un tipo de red neuronal diseñada para el procesamiento de secuencias de datos, como series temporales, donde las conexiones entre las neuronas forman un ciclo dirigido. Esto permite a las RNN mantener una “memoria” de las entradas anteriores en la secuencia, lo que es crucial para capturar dependencias temporales y secuenciales de los datos transaccionales de los clientes (Gers et al., 2002).

### 2.6.2.2 Arquitectura de una RNN

Una red neuronal recurrente está formada por diferentes capas. Para un mejor entendimiento se ha añadido la **Figura 0.2** donde se muestra un ejemplo de estructura de red. Las capas con las que está formada una red recurrente son:

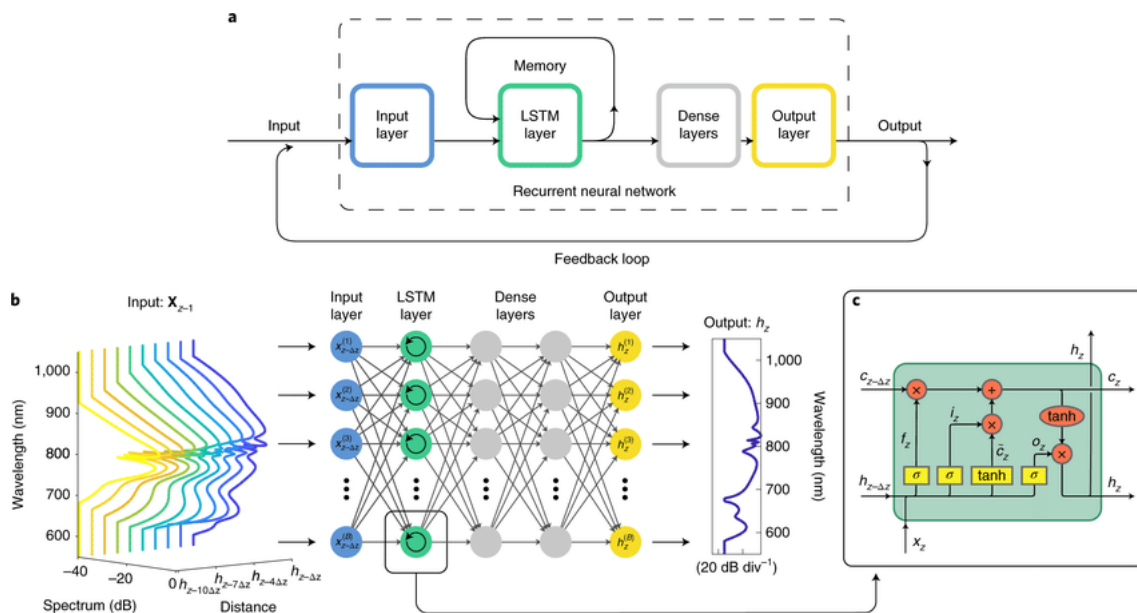
- **Capas de Entrada y Preprocesamiento**
  - **Entrada de Datos:** Los datos de entrada consisten en secuencias temporales de transacciones del cliente.
  - **Normalización y Escalado:** Los datos se normalizan y se aplica un escalado para facilitar el aprendizaje de la red y mejorar el proceso de entrenamiento
- **Capas Recurrentes**
  - **Celdas Recurrentes:** Las capas recurrentes están compuestas por celdas que pueden ser básicas (**SimpleRNN**), de memoria a corto y largo plazo (**LSTM**), o de unidad recurrente de puerta (**GRU**). Estas celdas permiten que la red mantenga y actualice un estado oculto que representa la memoria de la secuencia (Chung et al., 2014)

- **Propagación Temporal:** Las RNN procesan las secuencias de entrada un paso temporal a la vez, actualizando su estado oculto en cada fase. Esto permite que la red capture las dependencias a largo plazo en las secuencias de las transacciones de cada cliente.

- **Funciones de Activación y Optimización**

Dentro de las capas se encuentran este tipo de funciones que realizan diferentes transformaciones a los datos.

- **Funciones de Activación:** Se emplean en RNN funciones de activación como **ReLU** o **tanh** en las celdas recurrentes.
- **Algoritmo de Optimización Adam:** El algoritmo Adam ajusta los pesos del modelo minimizando una función de pérdida que mide el error entre las predicciones y los valores reales.



**Figura 0.2** Arquitectura red RNN

Imagen extraída de

[ResearchGate]([https://www.researchgate.net/figure/Recurrent-neural-networks-a-Schematic-of-the-RNN-architecture-used-showing-the-input\\_fig7\\_349426498](https://www.researchgate.net/figure/Recurrent-neural-networks-a-Schematic-of-the-RNN-architecture-used-showing-the-input_fig7_349426498))

### 2.6.2.3 Ventajas y Limitaciones

El uso de redes neuronales recurrentes enfocado a la predicción del CLTV tiene varias ventajas entre ellas se encuentra la flexibilidad de manejar datos de longitud variable y de diferentes tipos al tratarse de una red neuronal y la captura de dependencias temporales ya que las RNN son especialmente efectivas para capturar estos patrones.

Sin embargo, existen varias limitaciones. Principalmente se encuentran:

- **Requiere Gran Cantidad de Datos.**
- **Complejidad Computacional**
- **Desvanecimiento del Gradiente:** Las RNN pueden sufrir algunos problemas de desvanecimiento del gradiente que implica que dificulta el aprendizaje de dependencias a largo plazo. Variantes de celdas como **LSTM** o **GRU** están diseñadas para poder hacer frente a este problema.

### 2.6.3 Modelo de Red Densa

Al igual que las Redes CNN o DNN, las Redes Neuronales Densas (**DNN**) se han establecido como otra herramienta predictiva para el Customer Lifetime Value. Las DNN, también son conocidas como redes completamente conectadas, son útiles para capturar patrones complejos y no lineales en los datos transaccionales y en el comportamiento del cliente, característica que lo hace especialmente adecuado para la predicción del CLTV (Goodfellow et al., 2016).

#### 2.6.3.1 Fundamentos Teóricos

Las DNN están compuestas por múltiples capas densas donde cada neurona en una capa está conectada a todas las neuronas en la siguiente capa. Estas redes pueden aprender representaciones abstractas lo que favorece la captura de relaciones complejas entre características (Zhang et al., 2019).

#### 2.6.3.2 Arquitectura de una DNN

A continuación, se detallan las capas que forman la arquitectura de una red DNN:

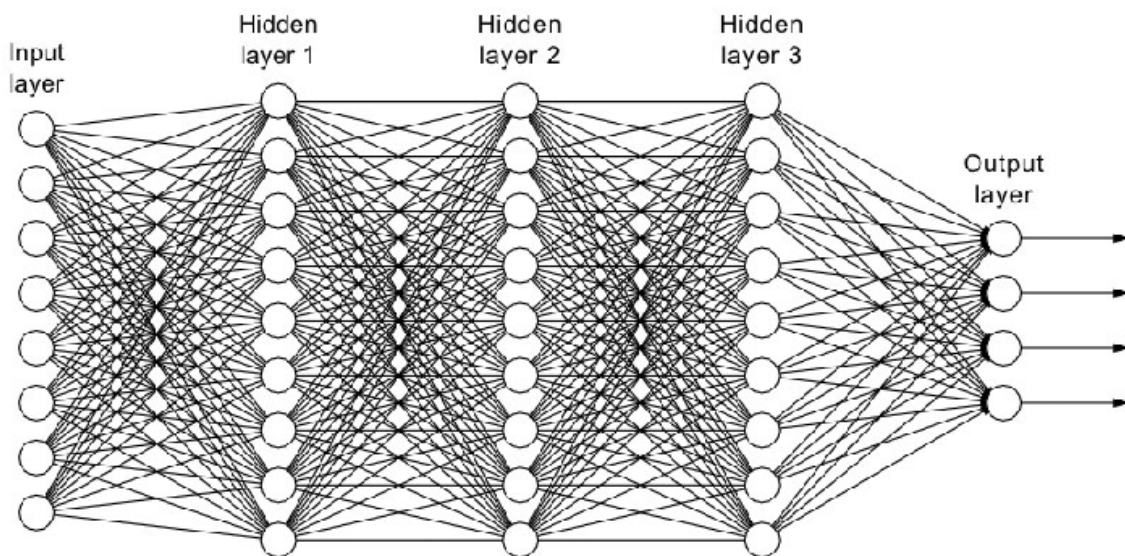
- **Capas de Entrada y Preprocesamiento**
  - **Entrada de Datos:** Los datos de entrada consisten en características agregadas de transacciones de clientes como son la frecuencia de compra, la cantidad de dinero gastado por transacción, la recencia o la tenencia entre otras.
  - **Normalización y Escalado:** Los datos se normalizan y se aplica un escalado para facilitar el aprendizaje de la red y mejorar el proceso de entrenamiento
- **Capas Densas (Completamente Conectadas)**
  - **Capas Ocultas:** Las capas ocultas están compuestas por neuronas densas. Cada neurona recibe entradas de todas las neuronas de la capa anterior y pasa su salida a todas las neuronas de la siguiente capa. Esto permite que la red aprenda las combinaciones complejas entre las características.

- **Funciones de Activación:** Se emplean las funciones de activación no lineales como la función **ReLU** para introducir la no linealidad en el modelo.

- **Capa de Salida**

La capa de salida es una capa densa con la salida y dependiendo del problema puede ser una única neurona en el que caso de que el modelo produzca un único valor o como en el caso de nuestro proyecto de múltiples neuronas, una neurona por cada output.

En la siguiente **Figura 0.3** se observa un ejemplo de arquitectura de una red densa para aclaración de la arquitectura.



**Figura 0.3** Arquitectura de red DNN

Imagen extraída de

[ResearchGate]([https://www.researchgate.net/figure/Schematic-of-an-ANN-comprised-of-a-series-of-dense-layers-Input-values-are-transformed\\_fig2\\_349194903](https://www.researchgate.net/figure/Schematic-of-an-ANN-comprised-of-a-series-of-dense-layers-Input-values-are-transformed_fig2_349194903))

### 3 Desarrollo

En esta sección se realizará un recorrido por las diferentes fases que conllevan la metodología para el cálculo del Customer Lifetime Value mediante el modelo predictivo desarrollado con redes neuronales. Siendo la primera fase (apartado 3.1) donde se evalúan los objetivos y las ventajas de la predicción del CLTV para un negocio en el sector alimenticio. Dentro de la segunda fase (apartado 3.2) corresponde al análisis exploratorio por los datos y la creación de la sabana de datos para el modelo basado en redes neuronales. En la tercera fase (apartado

3.3) se explica el uso de redes densas para la predicción de cada una de las variables necesarias para el cálculo de CLTV y se construye el modelo para la predicción. En la cuarta y última fase (apartado 3.4) se lleva a cabo la evaluación del modelo.

### **3.1 Comprensión del Negocio**

El negocio que ha proporcionado los datos para el desarrollo de la metodología se dedica a la venta de productos alimenticios. Este tipo de negocio se caracteriza por una logística precisa y una gestión eficiente del inventario, además, entender los patrones de consumo de los clientes es fundamental para poder ofrecer productos relevantes y tener una mejor optimización en cada viaje de venta.

En este contexto, el Customer Lifetime Value resulta especialmente útil, permite a la empresa identificar qué clientes generarán más ingresos a lo largo del tiempo, ayudándola a enfocar sus esfuerzos de marketing y recursos en aquellos clientes más valiosos. Con esta información, la empresa puede diseñar estrategias de retención más efectivas, como promociones personalizadas y programas de fidelización, lo que no solo mejora la satisfacción del cliente, sino que también incrementa el retorno de inversión (ROI) de sus campañas de marketing. En un mercado competitivo como el sector alimenticio, el uso estratégico del CLTV puede marcar la diferencia entre el éxito y el fracaso empresarial.

### **3.2 Análisis Exploratorio**

Los archivos que han sido proporcionados por la empresa son tres:

- Artículos: contiene información sobre el código del artículo, descripción, categoría y subcategoría para cada artículo.
- Clientes: contiene información sobre el cliente como la fecha de alta y de baja la dirección del cliente entre otras.
- Ventas: contiene la información de las ventas por la línea de ticket.

Dado que la empresa nos ha pedido realizar el estudio sin tener en cuenta el periodo durante la pandemia, los datos que van a ser empleados en el estudio comprenden desde enero de 2021 hasta abril de 2024.

De los ficheros dados por la empresa los únicos interesantes son los artículos y ventas ya que la información que proporciona el fichero de clientes no aporta nueva información para el cálculo del Customer Lifetime Value.

El conjunto de Ventas está compuesto por 54.320.266 líneas de tickets y 20 variables. A continuación, se muestra una visión estructural del fichero.

```

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54320266 entries, 0 to 54320265
Data columns (total 20 columns):
#   Column                Dtype
---  ---
0   ID_MOVIMIENTO         float64
1   FECHA_PEDIDO          object
2   FECHA_ENTREGA         object
3   ID_USUARIO            float64
4   USUARIO               float64
5   ID_CLIENTE           float64
6   ID_ARTICULO          float64
7   CODIGO_ARTICULO      int64
8   CODIGO_PADRE         float64
9   CTD_ARTICULOS_ORDENADOS float64
10  CTD_ARTICULOS_ENTREGADOS int64
11  ID_ORDINE            float64
12  IMPORTE_NETO_ORDENADO object
13  IMPORTE_BRUTO_ORDENADO object
14  IMPORTE_BRUTO_ENTREGADO object
15  IMPORTE_CATALOGO    object
16  CANAL_VENTA          object
17  IMPORTE_NETO_ENTREGADO object
18  DESCUENTOS_APLICADOS object
19  COD_TELEOP           float64
dtypes: float64(9), int64(2), object(9)
memory usage: 8.1+ GB

```

**Figura 1** Análisis de estructura interna de los datos correspondientes del fichero Ventas

Debido a que muchos de estos campos no proporcionan información para nuestro estudio se procede a describir cada uno de los atributos empleados en el estudio en la **Tabla 1**.

Atributo	Tipo	Descripción
ID_MOVIMIENTO	float	Identificador del ticket
FECHA_ENTREGA	object	Fecha de entrega de los productos al cliente
ID_USUARIO	float	Identificador del comercial que realiza la ven...
ID_CLIENTE	float	Identificador del cliente
CODIGO_ARTICULO	int	Identificador de producto
CTD_ARTICULOS_ENTREGADOS	int	Cantidad de productos entregados
IMPORTE_BRUTO_ENTREGADO	object	Cantidad de dinero que ha costado la compra al...

**Tabla 1** Descripción de los atributos seleccionados del fichero Ventas

Dado que los atributos dados no están limpios hay que realizarles una transformación para poder proceder a su análisis e interpretación. Quedando como resultado la siguiente **Figura 2**.

```
data_transform.info()
✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54320266 entries, 0 to 54320265
Data columns (total 7 columns):
#   Column                Dtype
---  -
0   ID_MOVIMIENTO         object
1   FECHA_ENTREGA         datetime64[ns]
2   ID_USUARIO            object
3   ID_CLIENTE            object
4   CODIGO_ARTICULO       int64
5   CTD_ARTICULOS_ENTREGADOS int64
6   IMPORTE_BRUTO_ENTREGADO float64
dtypes: datetime64[ns](1), float64(1), int64(2), object(3)
memory usage: 2.8+ GB
```

**Figura 2** Análisis de estructura de los atributos seleccionados

Con el dataset transformado ya se pudo realizar un análisis de los componentes que comprenden el fichero. Tenemos por una parte un identificador de ticket que identifica la transacción y puede estar n veces, como se puede ver representado en la **Figura 3**, haciendo referencia a que en la misma transacción se realizó la compra de diferentes productos. Realizando un descriptivo a la variable IMPORTE\_BRUTO\_ENTREGADO (**Figura 4**), observamos el comportamiento de precios que tiene cada producto y vemos además la tenencia de números negativos. Haciendo un análisis más detallado junto a la empresa nos pudieron aclarar que estos negativos pueden ser debidos a correcciones de pedidos por el comercial además de ajustes empresariales que se ven reflejados en el dataset y que tendrán que ser tratados en el modelo posteriormente.

```
print("Número de ids de ticket únicos:",data_transform.ID_MOVIMIENTO.nunique())
✓ 15.5s
```

Número de ids de ticket únicos: 16599930

```
print("Número de ids de ticket totales:",data_transform.ID_MOVIMIENTO.count())
✓ 0.0s
```

Número de ids de ticket totales: 54320266

**Figura 3** ID\_MOVIMIENTO únicos y totales

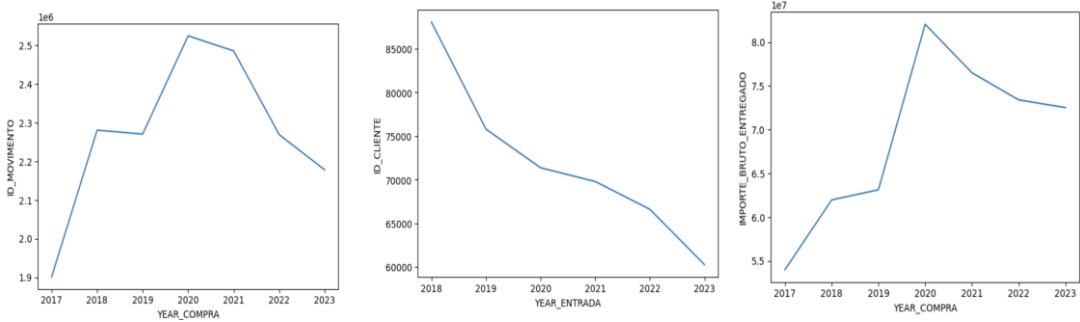
```
data_transform.IMPORTE_BRUTO_ENTREGADO.describe()  
✓ 1.7s
```

```
count    5.432027e+07  
mean     9.310195e+00  
std      6.471300e+00  
min      -3.300000e+02  
25%     5.500000e+00  
50%     8.950000e+00  
75%     1.195000e+01  
max      1.046500e+03  
Name: IMPORTE_BRUTO_ENTREGADO, dtype: float64
```

**Figura 4** IMPORTE\_BRUTO\_ENTREGADO descriptivo

También durante este proceso se pudo observar un comportamiento irregular en diferentes clientes que tenían un número de transacciones muy alto. La empresa funciona por un modelo de ventas quincenales por lo que el cliente como máximo puede realizar 2 transacciones al mes. Mediante este exploratorio se pudo observar que varios comerciales desagregaban las compras de clientes que compraban mucha cantidad de productos para poder llegar a número de tickets que les exigía la empresa por lo que además de desarrollar el modelo del cálculo del Customer Lifetime Value se le comunicó a la empresa la existencia de estos comportamientos y creamos una sábana de datos con la que eran capaces de monitorear a sus comerciales para subsanar este problema.

A continuación, vamos a analizar el número de tickets distintos por años (**Figura 5 a**) y el número de clientes distintos incorporados por año (**Figura 5 b**) para poder hacer una imagen de la situación actual del negocio.



**Figura 5 a)** Cantidad de transacciones por año **b)** Cantidad de clientes adquiridos por año **c)** Cantidad de dinero generada por año

En estas graficas podemos ver que tiene una tendencia negativa en captación de clientes y que el número de transacciones totales han bajado tras la época de la pandemia, donde experimentaron su mejor momento. Ahora vamos a compararlo con la cantidad de dinero obtenido por estos años (**Figura 5 c**).

En esta grafica podemos observar que, aunque estén teniendo menos clientes adquiridos y un menor número de tickets totales han aumentado la facturación en comparación a la época prepandemia. Esto nos indica que los clientes que han adquirido en los periodos anteriores no solo se han mantenido a lo largo del tiempo, sino que también han aumentado su media de compra por transacción que sitúa a la empresa en un muy buen punto de partida para la implementación de nuevas métricas como la segmentación de clientes o el CLTV.

En este punto ya tenemos información suficiente como para poder generar una sábana de datos que nos pueda ayudar a la predicción del CLTV

### 3.3 Sábana de Predicción

En este apartado vamos a comenzar a realizar todas las transformaciones necesarias para el desarrollo del modelo.

El primer paso es obtener toda la información correspondiente al comportamiento de cliente. Para ello empleamos las variables del modelo RFM, que nos proporciona diferentes métricas, las cuales nos ayudan a comprender el comportamiento de compra para cada cliente. Esto es de ayuda debido a que la métrica CLTV es una métrica individual para cada cliente por lo que es importante presentar los datos al modelo de manera más individualizada. Teniendo esto en cuenta se han creado las siguientes variables:

- **Length**, siendo esta variable la cantidad de días desde que hicieron su primera compra hasta la fecha de estudio, normalmente se puede encontrar como Tenure, debido a que hace referencia a la tenencia del cliente.
- **Recency**, siendo esta variable la cantidad de días que han pasado desde la última compra.
- **Frecuency**, siendo esta variable la cantidad de transacciones que ha realizado en su vida como cliente
- **PCV (Past Customer Value)**, esta es una variable que determina el valor de cliente que tiene en el momento del estudio.

Aunque normalmente en el modelo RFM se utiliza la variable Monetary para el cálculo de los segmentos, en este caso se ha decidido reemplazarla por el Past Customer Value. Esto se debe a que el uso de Past Customer Value en la predicción del Customer Lifetime Value proporciona una imagen más precisa en términos de beneficio monetario que la variable Monetary característica del modelo RFM.

El segundo paso consiste en obtener todos los productos comprados en cada categoría por cliente. En esta etapa, surgen diversos problemas debido a la forma en que la empresa categoriza sus productos. La categorización puede variar dependiendo de la funcionalidad que la empresa busque. Dado que nuestro objetivo final es evaluar el comportamiento del cliente, utilizamos las categorías que ellos tienen accesibles, es decir, la categorización web. Esto implica obtener todos los códigos de los artículos y asignarles cada categoría correspondiente según la web de la empresa.

Como resumen de la sabana de datos para la predicción quedaría como:

**Periodo de entrenamiento: 2021-01-01 -> 2023-10-01**

**Número de clientes entrenamiento: 279022**

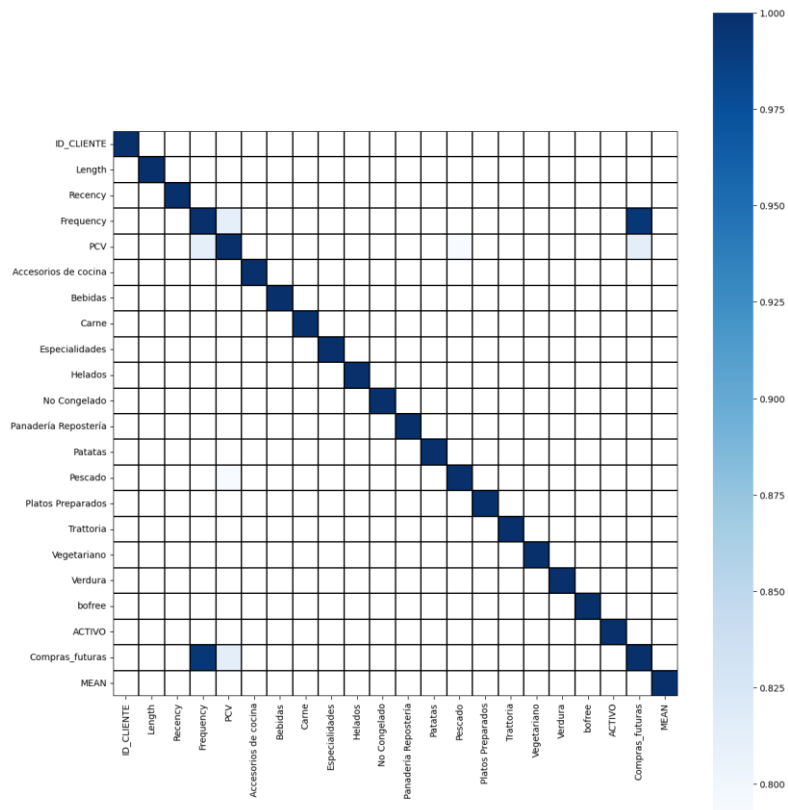
**Periodo de validación: 2023-10-02 -> 2024-04-01**

**Número de clientes de validación: 119581**

Para poder realizar el entrenamiento es necesario añadir los valores reales de las variables que se va a predecir. Por este motivo se añaden las siguientes tres variables:

- **ACTIVO:** siendo esta la variable que indica si está activo en los próximos 6 meses, es decir, si ha realizado alguna compra en ese periodo.
- **Compras futuras:** siendo esta variable el número total de transacciones que van a tener cada cliente en 6 meses.
- **MEAN:** siendo esta variable el valor de ticket medio por cada cliente.

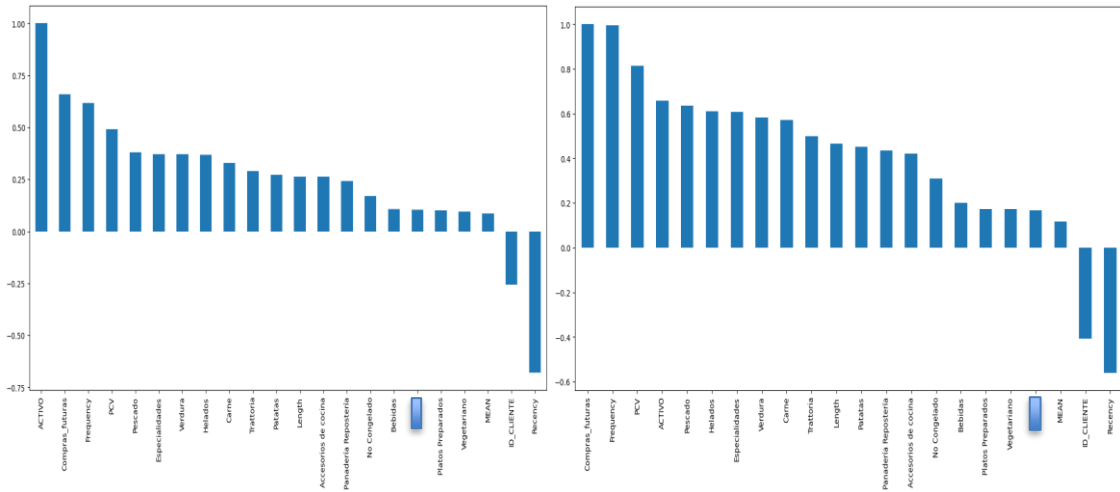
A continuación, vamos a hacer un análisis de este dataset para poder comprender mejor los atributos elegidos para la predicción del Customer Lifetime Value, comenzando con un estudio de correlación de las variables (**Figura 6**).



**Figura 6** Análisis de correlación de variables dataset de predicción

Podemos observar que Frequency está muy correlacionado con Compras\_futuras dado que ambas variables representan lo mismo, pero en 2 periodos diferentes. Es interesante ver la correlación entre PCV y Frequency dado que es normal que cuanto mayor sea el Frequency del cliente mayor es su PCV.

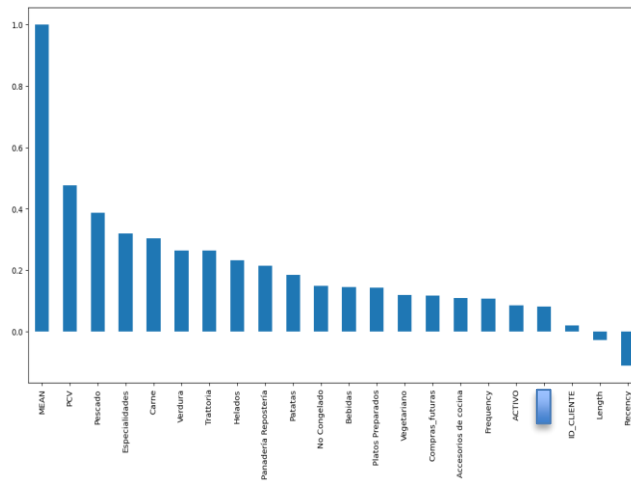
Continuamos con las siguientes figuras que nos proporcionan un análisis de correlación para cada variable que se va a predecir. Siendo **Figura 7 a)** el análisis de correlación para la variable ACTIVO, **Figura 7 b)** el análisis de correlación para la variable COMPRAS\_FUTURAS, **Figura 7 c)** el análisis de correlación para la variable MEAN.



**Figura 7.**

**a)**

**b)**



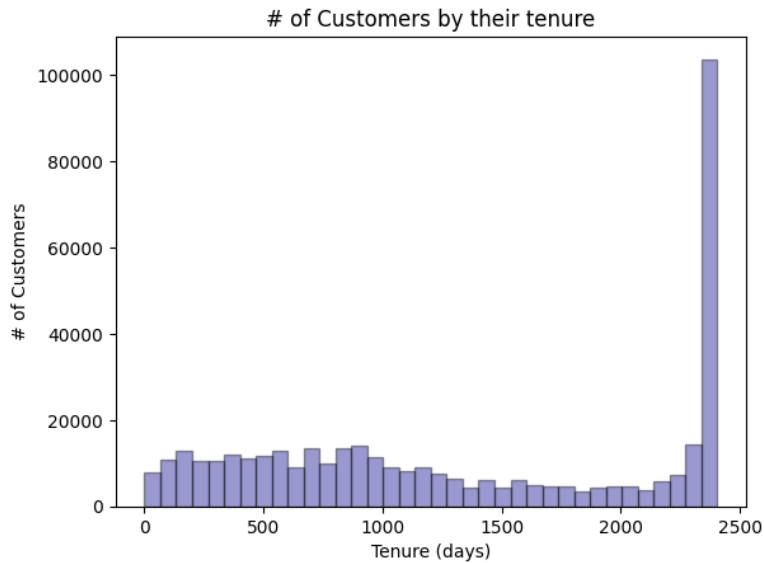
**c)**

**Figura 7 a)** el análisis de correlación para la variable ACTIVO **b)** el análisis de correlación para la variable COMPRAS\_FUTURAS **c)** el análisis de correlación para la variable MEAN

Con estas figuras sacamos varias conclusiones principales:

- Para la probabilidad de que este activo o no tiene mucho peso cuando se realizó la última compra, así como el número de compras que ha realizado cada cliente.
- Para la variable de COMPRAS\_FUTURAS podemos observar que todos los atributos tienen un mayor peso comparado con las otras figuras. Siendo muy característicos los atributos de Frequency, PCV y Recency, es decir, que existe una correlación con cuando tiempo ha pasado desde tu última compra y con cuánto valor tienes además del número de compras totales que tenga el cliente en ese periodo.
- Por último, en la variable objetivo MEAN no podemos determinar que exista una clara correlación siendo la más significativa el atributo PCV teniendo en cuenta que es el único atributo que tiene relación con el valor monetario al tratarse del beneficio generado por el cliente.

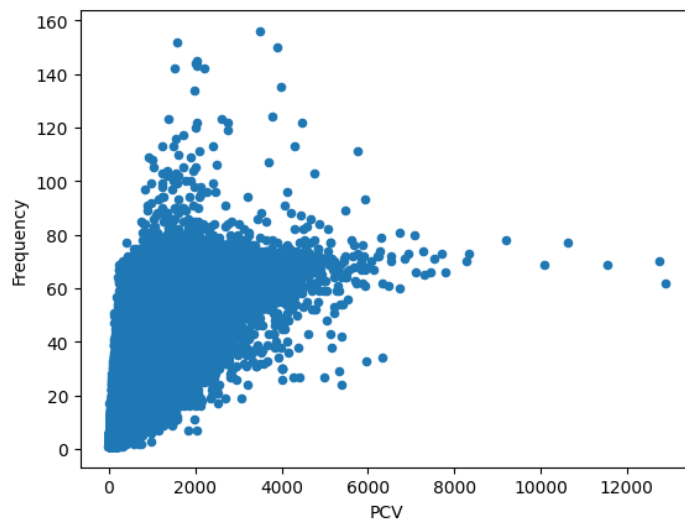
También vamos a analizar el tipo de clientes que tenemos en el dataset para el entrenamiento de predicción (**Figura 8**).



**Figura 8** Número clientes por su *Tenure*

Podemos observar que la mayoría de los clientes en el dataset son clientes antiguos, es decir, con mucha antigüedad. Esto significa que tenemos datos tanto de personas que llevan comprando desde el principio y siguen activas, como de aquellas que se han desactivado a lo largo del tiempo. Esta información es crucial para el modelo, ya que permite predecir si un cliente permanecerá activo o no.

En la siguiente figura, **Figura 9**, muestra la relación entre Frequency y PCV.



**Figura 9** Correlación de PCV-Frequency

La **Figura 9** muestra una gran variabilidad en el comportamiento de los clientes en términos de PCV y frecuencia de compras, con una concentración notable en valores bajos de PCV y una frecuencia de compras moderada.

los siguientes patrones:

- **Concentración de Datos en Valores Bajos de PCV:** La mayoría de los puntos se encuentran concentrados en la parte inferior izquierda de la gráfica, indicando que la mayoría de los clientes tienen un PCV bajo y una frecuencia de compras moderada.
- **Dispersión en Valores Altos de PCV:** Conforme el PCV aumenta, los datos se dispersan más, mostrando que hay clientes con valores de PCV más altos que realizan compras con menor frecuencia.
- **Algunos Clientes con Alta Frecuencia y PCV Moderado:** Hay algunos puntos que indican clientes con una frecuencia de compras muy alta, incluso por encima de 100, pero con un PCV moderado, alrededor de 2000 a 4000.
- **Outliers:** Existen algunos puntos atípicos con valores de PCV extremadamente altos (por encima de 10,000) y con diversas frecuencias de compras, lo cual podría indicar clientes excepcionales con un historial de compras muy elevado.

En términos de correlación, la gráfica sugiere una **relación positiva débil** entre el Past Customer Value (PCV) y la frecuencia de compras. Esto se observa de la siguiente manera:

- **Relación Positiva:** En general, a medida que el PCV aumenta, también tiende a aumentar la frecuencia de compras. Sin embargo, esta tendencia no es muy fuerte, ya que hay una gran dispersión de datos.
- **Dispersión Alta:** La alta dispersión de los puntos indica que, aunque hay una tendencia positiva, la correlación no es fuerte. Muchos clientes con un PCV bajo tienen una frecuencia de compras variada, y lo mismo ocurre con los clientes de alto PCV.
- **Outliers y Variabilidad:** Los outliers y la variabilidad en frecuencias de compra para diferentes niveles de PCV también sugieren que otros factores podrían estar influyendo en la frecuencia de compras además del PCV.

## 3.4 Modelos Anteriores

En esta sección se explican los diferentes acercamientos que se tomaron para abordar el problema antes de encontrar la solución planteada en el apartado posterior.

### 3.4.1 Modelo basado en Trabajos Previos

Al comenzar el desarrollo del trabajo se trató de darle un enfoque basado en el trabajo desarrollado por Wang Xiaojing, Liu Tianqi, y Miao Jingang (Wang et al., 2019) donde muestran una red neuronal densa para la predicción del Customer Lifetime Value con datos públicos extraídos de Kaggle.

Este trabajo parecía prometedor al observar los resultados obtenidos al utilizar la distribución ZILN (**zero-inflated lognormal distribution**). Sin embargo, al implementar las bibliotecas necesarias para verificar este tipo de modelo en los datos y desarrollar el proyecto, se descubrió que el código de la librería estaba incompleto y mal implementado, lo que impidió la creación de cualquier modelo.

### 3.4.2 Modelo Convolutacional

Tras descubrir que el modelo que empleaba la distribución **ZILN** no funcionaba se optó por crear un modelo neuronal desde cero basándonos en la investigación previamente realizada.

Al intentar implementar un modelo de redes neuronales convolucionales nos enfrentamos a varios desafíos que resultaron en un rendimiento subóptimo del modelo. En la siguiente **Figura 10** podemos observar la arquitectura que se siguió en este modelo.

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 9, 64)	192
max_pooling1d (MaxPooling1D)	(None, 4, 64)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 50)	12,850
dense_1 (Dense)	(None, 1)	51

**Figura 10** Resumen red CNN

La arquitectura de esta red es sencilla, ya que se trata de un prototipo para evaluar su funcionamiento. En esta etapa del proyecto, se optó por una red que devolviera directamente el Customer Lifetime Value (CLTV), por lo que solo tiene una salida. La estructura resultante es la siguiente:

1. **Capa Convolutacional (Conv1D):**

Esta es la primera capa del modelo, que aplica 64 filtros convolucionales a las secuencias de entrada con un tamaño de kernel de 2. Esta capa detecta patrones locales en los datos temporales de las transacciones del cliente, generando 64 mapas de características.

2. **Capa de Pooling (MaxPooling1D):**

La capa de pooling reduce la dimensionalidad de los mapas de características generados por la capa convolutacional anterior. En

este caso, utiliza una ventana de tamaño 2, manteniendo las características más relevantes y reduciendo la complejidad computacional.

**3. Capa de Aplanamiento (Flatten):**

Esta capa transforma los mapas de características multidimensionales en un vector unidimensional. El objetivo es preparar los datos para las capas completamente conectadas siguientes.

**4. Capa Densa (Dense):**

La primera capa densa recibe el vector aplanado y aplica una transformación lineal seguida por una función de activación ReLU. Esta capa tiene 50 neuronas y aprende combinaciones complejas de las características extraídas.

**5. Capa de Salida (Dense):**

La capa final es una capa densa con una sola neurona y una función de activación lineal. Esta capa genera la predicción final del CLTV, proporcionando un valor continuo que representa el valor futuro estimado del cliente.

A pesar de los esfuerzos dedicados a la creación del modelo no se logró obtener un funcionamiento correcto en el código de entrenamiento, siendo el principal problema la incapacidad de ingesta de datos del modelo. Debido al límite de tiempo se optó por desarrollar otro modelo, pero esta vez basado en redes densas.

**3.4.3 Modelo DNN**

Tras el fracaso de los dos modelos anteriores se optó por un modelo basado en redes densas, pero con el enfoque de que el output de la red sea el CLTV por lo que tendría una única salida dando como resultado la siguiente arquitectura representada en la siguiente **Figura 11**.

Layer (type)	Output Shape	Param #
dense_24 (Dense)	(None, 512)	4,608
dropout_6 (Dropout)	(None, 512)	0
dense_25 (Dense)	(None, 256)	131,328
dropout_7 (Dropout)	(None, 256)	0
dense_26 (Dense)	(None, 64)	16,448
dense_27 (Dense)	(None, 32)	2,080
dense_28 (Dense)	(None, 16)	528
dense_29 (Dense)	(None, 8)	136
dense_30 (Dense)	(None, 1)	9

**Figura 11** Resumen DNN fallido

### **1. Primera Capa Densa (Dense\_24)**

Esta es la primera capa densa del modelo que recibe la entrada de los datos iniciales. Aplica una transformación lineal con 512 unidades, seguida de una función de activación **ReLU**.

### **2. Primera Capa de Dropout (Dropout\_6)**

La primera capa de dropout se utiliza para prevenir el sobreajuste durante el entrenamiento al desactivar aleatoriamente un 50% de las neuronas de la capa anterior en cada actualización del proceso de entrenamiento.

### **3. Segunda Capa Densa (Dense\_25)**

Esta capa densa sigue a la capa de dropout y aplica una transformación lineal con 256 unidades. Permite aprender características más complejas a partir de la salida de la capa anterior.

### **4. Segunda Capa de Dropout (Dropout\_7)**

Otra capa de dropout para regularizar el modelo y prevenir el sobreajuste, desactivando aleatoriamente un 30% de las neuronas de la capa anterior.

### **5. Tercera Capa Densa (Dense\_26)**

Esta capa densa continúa el proceso de aprendizaje con 64 unidades, capturando patrones más específicos y complejos.

### **6. Cuarta Capa Densa (Dense\_27)**

Añade más capacidad de aprendizaje al modelo, con 32 unidades que ayudan a detectar relaciones aún más detalladas en los datos.

### **7. Quinta Capa Densa (Dense\_28)**

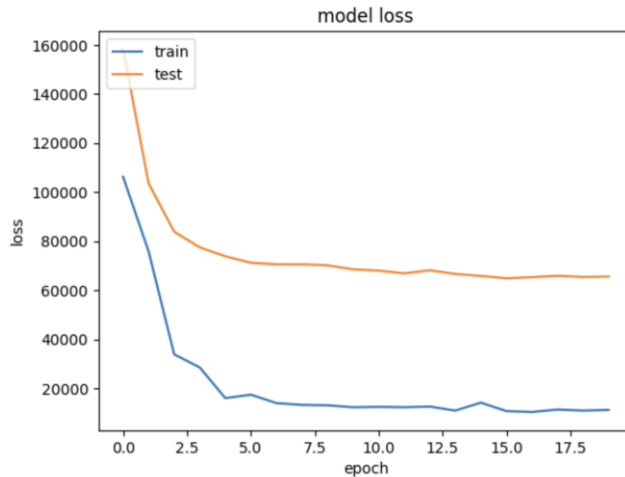
Esta capa con 16 unidades continúa fortaleciendo la capacidad de representación del modelo.

### **8. Sexta Capa Densa (Dense\_29)**

Con 8 unidades, esta capa reduce aún más la dimensionalidad, enfocándose en características muy específicas.

### **9. Séptima Capa Densa (Dense\_30)**

La capa final de la red, con una sola unidad, genera la predicción final del CLTV. Esta unidad probablemente utiliza una activación lineal para producir un valor continuo



**Figura 12** Entrenamiento DNN fallido

La figura 12 nos muestra como el modelo presenta claros signos de sobreajuste. A medida que avanza el entrenamiento, sigue mejorando su rendimiento en el conjunto de entrenamiento, pero no mejora en el conjunto de prueba, indicando que puede estar memorizando los datos de entrenamiento en lugar de aprender patrones generalizables. Este problema, junto con el hecho de que la salida del modelo era únicamente el CLTV, lo que lo hacía poco explicable, resultó en que el modelo no fuera válido para presentarlo a la empresa. Debido a esto, se exploraron otras formas de cálculo del CLTV y diferentes arquitecturas de red, lo que llevó al desarrollo del modelo siguiente.

### 3.5 Modelo DNN Implementado

En esta sección se detalla la construcción del modelo final que se implementa en la empresa que cedió los datos para la realización de este proyecto.

#### 3.5.1 Desarrollo del Modelo

Como ya se ha comentado el modelo elegido ha sido DNN y tras los modelos fallidos se cambió el método de cálculo de CLTV para ser más explicable para la empresa y ser una solución óptima para el uso que le van a implementar. Para poder realizar esta métrica es necesario tener como la salida las 3 variables para poder realizar el cálculo del Customer Lifetime Value. Las tres variables que predice el modelo corresponden a:

- **ACTIVO:** Si el cliente realiza una compra en los próximos 6 meses.
- **Compras futuras:** variable que determina el número de compras totales que tendrá el cliente dentro de 6 meses.
- **MEAN:** variable que indica la media de transacción por cliente.

En la siguiente figura, **Figura 13**, se representa la arquitectura de la red neuronal. Esta es la estructura representada en la figura:

1. **Capa de entrada (InputLayer):**
  - Esta capa recibe 18 características de entrada, los 18 parametros descritos en el apartado **3.3 Sábana de Predicción**
2. **Primera capa densa (Dense):**
  - Esta capa contiene 64 neuronas y utiliza la función de activación ReLU para capturar patrones no lineales en los datos de entrada.
3. **Segunda capa densa (Dense):**
  - Con 8 neuronas, esta capa actúa como un embotellamiento que ayuda a reducir la dimensionalidad y a extraer características más abstractas y significativas.
4. **Capas de salida (Dense):**
  - **Salidas específicas para cada variable objetivo:**
    - **Churn output:** predice la probabilidad de que este activo en 6 meses.
    - **Número de compras (num\_compras\_output):** predice la frecuencia de compra total en 6 meses.
    - **Valor del cliente (customer\_value\_output):** predice el valor monetario medio por transacción.
  - Cada una de estas capas está conectada a la segunda capa densa y tiene una sola neurona para proporcionar una salida específica para cada variable objetivo.

Layer (type)	Output Shape	Param #	Connected to
input_layer_1 (InputLayer)	(None, 18)	0	-
dense_2 (Dense)	(None, 64)	1,216	input_layer_1[0]...
dense_3 (Dense)	(None, 8)	520	dense_2[0][0]
churn_output (Dense)	(None, 1)	9	dense_3[0][0]
num_compras_output (Dense)	(None, 1)	9	dense_3[0][0]
customer_value_out... (Dense)	(None, 1)	9	dense_3[0][0]

**Figura 13** Arquitectura de la red neuronal densa

Esta arquitectura permite que el modelo haga predicciones más específicas y detalladas. En lugar de solo predecir el Customer Lifetime Value (CLTV), ahora también puede predecir la probabilidad de abandono (**churn\_output**), el número de compras (**num\_compras\_output**), y el valor del cliente (**customer\_value\_output**). Esta estructura mejora la capacidad explicativa del modelo y aborda el problema de sobreajuste observado anteriormente al proporcionar múltiples salidas predictivas.

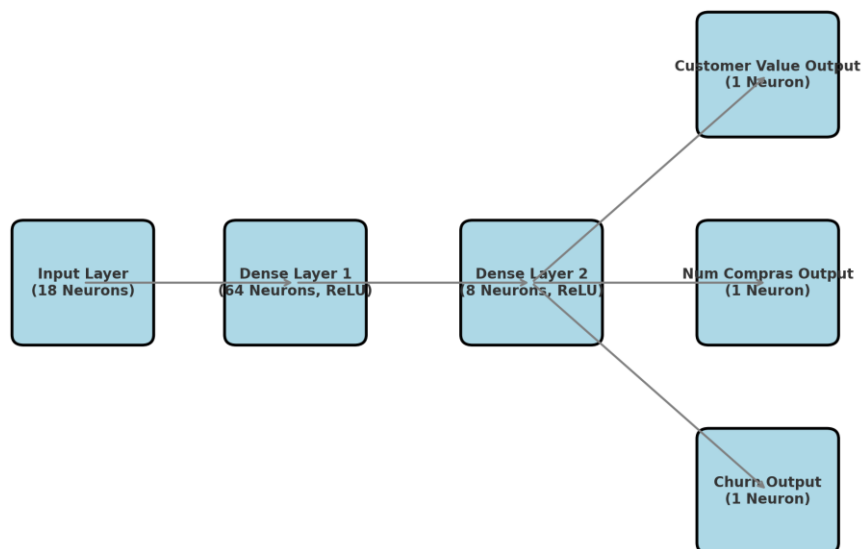
Durante el entrenamiento, el modelo emplea un enfoque de aprendizaje supervisado. Para las variables numéricas se emplea la métrica de MSE (error cuadrático medio) para la pérdida, mientras que para la variable ACTIVO se emplea el binary-crossentropy al tratarse de una probabilidad.

El algoritmo de optimización utilizado es Adam (Adaptive Moment Estimation), conocido por su eficiencia y capacidad para manejar grandes volúmenes de datos y modelos complejos (Chung et al., 2009). Adam ajusta dinámicamente las tasas de aprendizaje de los parámetros del modelo, mejorando la convergencia y estabilidad del entrenamiento. Además, se aplican técnicas de regularización como l2 dentro de las capas dense\_2 y dense\_3, que ayudan a prevenir el sobreajuste y mejorar la capacidad de generalización del modelo.

El desempeño del modelo se evalúa utilizando métricas específicas para cada variable objetivo, como el error absoluto medio (MAE) para el caso de MEAN y COMPRAS\_FUTURAS y el coeficiente de determinación (R cuadrado) para el caso de ACTIVO. Estas métricas proporcionan una visión clara de la precisión y fiabilidad del modelo en la predicción del valor monetario, la frecuencia de compra y el tiempo de retención de los clientes.

En esta figura, **Figura 14**, se muestra un esquema de como seria la red neuronal representada.

Esquema de la Red Neuronal para la Predicción del CLTV



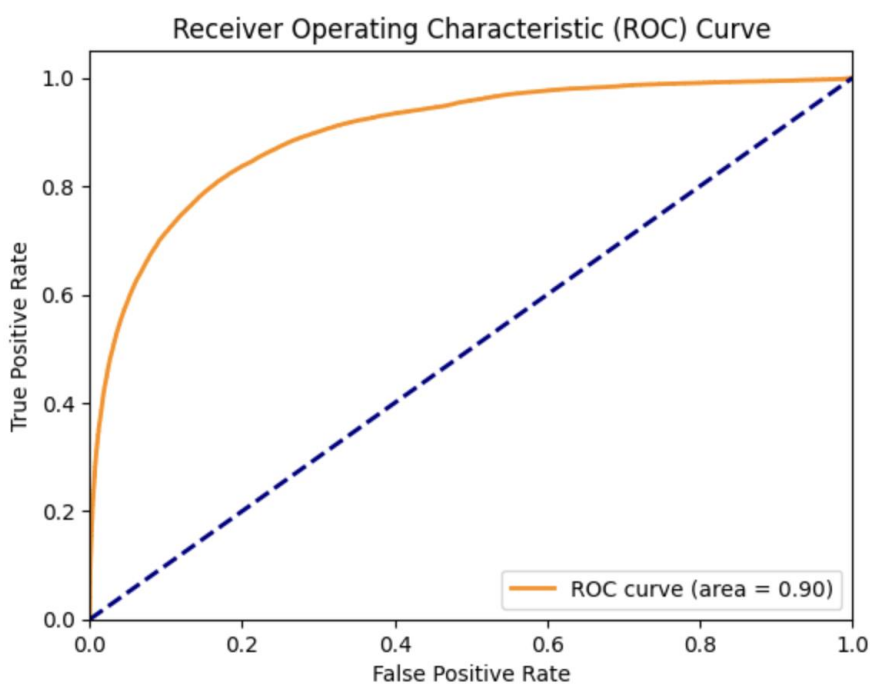
**Figura 14** Esquema red neuronal

### 3.6 Evaluación del Modelo

Para evaluar la efectividad del modelo de red neuronal, se reservó un periodo de 6 meses para llevar a cabo un exhaustivo proceso de evaluación. Durante este periodo, se recopilaron y analizaron datos de manera continua para garantizar la validez y la precisión de las predicciones realizadas por el modelo.

Para la evaluación de la variable ACTIVO:

Se emplearon varias métricas de rendimiento y visualizaciones para ilustrar los resultados de la evaluación. A continuación, se presentan dos figuras, **Figura 15** y **Tabla 2**, para la explicación del rendimiento del modelo para la variable ACTIVO.



**Figura 15** Curva ROC

**Curva ROC (Receiver Operating Characteristic):** Esta curva ilustra la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) para diferentes umbrales de clasificación. El área bajo la curva (AUC) es de 0.90, lo que indica un alto nivel de precisión en la discriminación entre las clases.

Neural Network:				
	precision	recall	f1-score	support
0.0	0.90	0.79	0.84	74849
1.0	0.71	0.86	0.78	44732
accuracy			0.82	119581
macro avg	0.81	0.82	0.81	119581
weighted avg	0.83	0.82	0.82	119581

	Predicted NO	Predicted YES
Actual NO	59268	15581
Actual YES	6417	38315

**Tabla 2** Tabla de métricas de evaluación

**Tabla de métricas de evaluación** La tabla muestra las métricas de evaluación para una red neuronal, enfocada en un problema de clasificación binaria. Aquí está la explicación de las métricas y una valoración del resultado:

### Métricas

- **Precisión**  
La precisión mide la proporción de verdaderos positivos sobre el total de predicciones positivas.  
**Para la clase 0 (NO ACTIVOS): 0.90**  
**Para la clase 1 (ACTIVOS): 0.71**
- **Recall**  
El recall mide la proporción de verdaderos positivos sobre el total de instancias reales de la clase.  
**Para la clase 0 (NO ACTIVOS): 0.79**  
**Para la clase 1 (ACTIVOS): 0.86**
- **F1-score**  
El F1-score es la media armónica de la precisión y el recall, proporcionando un balance entre ambos.  
**Para la clase 0 (NO ACTIVOS): 0.84**  
**Para la clase 1 (ACTIVOS): 0.78**
- **Support**  
El soporte indica el número reales de instancias reales de cada clase.  
**Para la clase 0 (NO ACTIVOS): 74849**  
**Para la clase 1 (ACTIVOS): 44732**
- **Accuracy**  
El Accuracy mide la proporción de todas las predicciones correctas sobre el total de instancias.  
**Accuracy: 0.82**
- **Macro Avg**  
**Precision: 0.81**  
**Recall: 0.82**  
**F1-score: 0.81**

- **Weighted Avg**  
El promedio ponderado considera el soporte de cada clase, proporcionando un promedio ponderado de las métricas.  
**Precision: 0.83**  
**Recall: 0.82**  
**F1-score: 0.82**

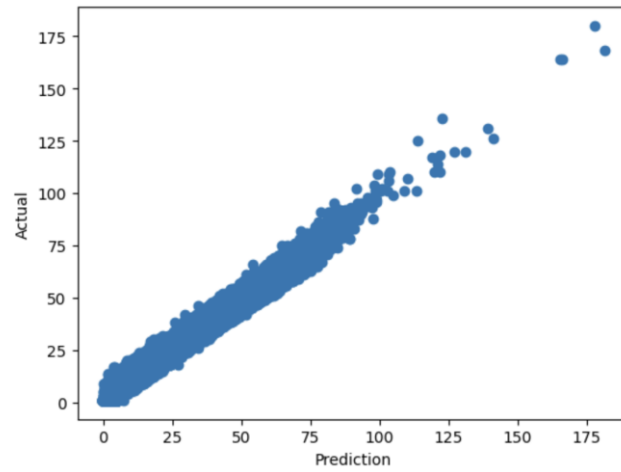
## Resultados

- **Precisión Alta para la Clase NO ACTIVOS:** La clase 0 (**NO ACTIVOS**) tiene una precisión alta (0.90), lo que significa que la mayoría de las predicciones de la clase NO ACTIVOS son correctas.
- **Recall Alto para la Clase ACTIVOS:** La clase 1 (**ACTIVOS**) tiene un recall alto (0.86), indicando que la mayoría de las instancias reales de la clase ACTIVOS son correctamente identificadas.
- **Balance de Métricas:** El F1-score para ambas clases está razonablemente balanceado, aunque ligeramente más alto para la clase **NO ACTIVOS** (0.84) en comparación con la clase **ACTIVOS** (0.78).
- **Exactitud General Buena:** Con una exactitud del 82%, el modelo está funcionando bien en general, identificando correctamente la mayoría de las instancias.
- **Desempeño en la Matriz de Confusión:** Hay un número notable de falsos positivos (15581) y falsos negativos (6417), pero la mayoría de las predicciones son correctas.

El modelo presenta un desempeño sólido con una exactitud del 82%. La precisión y el recall son razonablemente altos, especialmente para la clase ACTIVOS, lo que es importante si la clase ACTIVOS representa un evento crítico (como el abandono de clientes)

Para la variable COMPRAS\_FUTURAS, que indica el número de compras totales en un periodo de 6 meses se llevó a cabo un análisis exhaustivo de la siguiente **Figura 16**.

Total Sales Actual: 2168333  
Total Sales Predicted: 2190035.0  
Individual R2 score: 0.991046626454206  
Individual Mean Absolute Error: 1.2730047699701434



**Figura 16** COMPRAS\_FUTURAS evaluación de modelo

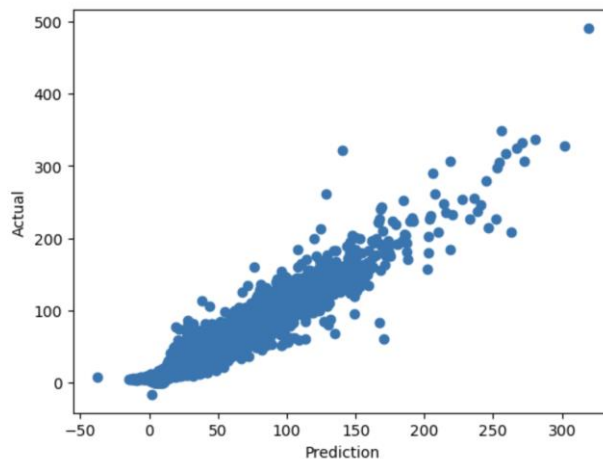
La **Figura 16** muestra una comparación entre los valores reales (Eje Y) con los predichos (Eje X). Los puntos de la gráfica están alineados cercanos a la diagonal lo que implica una fuerte correlación entre los valores reales con los predichos.

El valor de R cuadrado de 0.991 implica que el modelo puede explicar el 99,1% la variabilidad de los datos. Además, el valor de 1.27 de MAE indica que las predicciones del modelo son muy cercanas a las reales de media.

Con estas medidas podemos concluir que el modelo para la variable COMPRAS\_FUTURAS predice de manera precisa y de manera eficaz.

Para la variable MEAN, que indica la cantidad media de dinero gastado por el cliente por transacción en un periodo de 6 meses. Para su evaluación se emplea una figura similar a la variable COMPRAS\_FUTURAS, **Figura 16**.

Total Sales Actual: 3577200.0  
Total Sales Predicted: 3639134.0  
Individual R2 score: 0.9474927736435035  
Individual Mean Absolute Error: 2.1615326674270783



**Figura 17** MEAN evaluación de modelo

La **Figura 17** muestra una comparación entre los valores reales (Eje Y) con los predichos (Eje X). Los puntos de la gráfica están alineados cercanos a la diagonal lo que implica una fuerte correlación entre los valores reales con los predichos.

El valor de R cuadrado de 0.947 implica que el modelo puede explicar el 94,7% la variabilidad de los datos. Además, el valor de 2.16 de MAE indica que las predicciones del modelo son muy cercanas a las reales de media.

Con estas medidas podemos concluir que el modelo para la variable MEAN predice de manera precisa y de manera eficaz.

Para resumir, podemos concluir que el modelo predice de manera precisa cada una de las variables necesarias para el cálculo del CLTV. Mostrando para las variables numéricas una fuerte correlación entre los datos reales y predichos y para la probabilidad de cliente activo una curva ROC de 0.90 que nos indica una muy buena predicción de si un cliente se mantiene activo o si sin embargo se desactiva.

## 4 Comparativa de los Modelos Tradicionales

Para poder realizar la comparativa de resultados con el modelo basado en redes neuronales para la predicción de Customer Lifetime Value se ha desarrollado dos modelos tradicionales que funcionan como uno solo, BetaGeoFitter que predice la probabilidad de que el cliente se mantenga activo en el periodo de 6 meses y el número de compras que el cliente va a realizar en el mismo periodo, y el modelo Gamma-Gamma que es el encargado de predecir la variable monetaria por la que se calcula el CLTV.

Para la creación de estos modelos se tuvo que realizar una sábana diferente de datos para la predicción con algunas similitudes como el uso del Past Customer Value que es la métrica creada por el agregado del gasto bruto habiéndole sustraído su coste de adquisición de cliente.

La sabana de predicción tiene como índice el identificador del cliente y quedando como atributos los siguientes:

- **Frequency:** Número total de compras realizadas por el cliente.
- **Recency:** Días transcurridos desde la última compra.
- **T:** Días desde la primera compra del cliente.
- **Monetary\_value:** Variable que contiene el Past Customer Value

Tras la creación de la sabana se realizó un análisis de anómalos donde se encontraron un total de 1368 clientes sospechosos que fueron eliminados para mejorar el rendimiento de los modelos tradicionales.

## 4.1 Modelo BetaGeoFitter

Se comienza con la creación del modelo BetaGeoFitter, que será el encargado de calcular las variables de probabilidad de que los clientes se mantengan activos y el número de compras en el periodo de 6 meses.

A continuación, se proporcionará un resumen del modelo BetaGeoFitter, **Tabla 3**, que se desarrolla para la evaluación de resultados con coeficiente de penalización de 0.002

	coef	se(coef)	lower 95% bound	upper 95% bound
<b>r</b>	2.718222	0.006396	2.705685	2.730759
<b>alpha</b>	101.352804	0.260649	100.841933	101.863676
<b>a</b>	0.281785	0.000809	0.280199	0.283371
<b>b</b>	3.308155	0.012926	3.282820	3.333490

**Tabla 3** BetaGeoFitter resumen

El resumen del modelo BetaGeoFitter proporciona información crucial sobre los parámetros ajustados del modelo, además de sus intervalos de confianza, estos corresponden a las columnas *lower 95% bound* y *upper 95% bound*. Estos son los componentes principales del modelo creado:

- **r (coef = 2.718222)**

El parámetro r representa la tasa de repetición de las compras de los clientes. Un valor alto indica una mayor probabilidad de que un cliente realice compras repetidas.

El error estándar de r es de 0.006396 lo que indica que el modelo realiza una estimación precisa.

- **alpha (coef = 101.352804)**

El parámetro alpha está relacionado con la tasa de abandono de los clientes. Un valor más alto de alpha implica que los clientes se mantienen activos durante un mayor periodo de tiempo.

El error estándar del parámetro alpha es de 0.26 teniendo un intervalo de confianza de 100.84 a 101.86 lo que indica que los clientes se mantienen activos durante un periodo muy amplio.

- **a (coef = 0.281785)**

El parámetro a se relaciona con la variabilidad en la tasa de compra entre distintos clientes. Cuanto menor sea el parámetro a menos variabilidad y por lo tanto más homogeneidad en el comportamiento de los clientes. Al tener en cuenta que el negocio

vende cada quince días es normal que no haya mucha variabilidad de compra entre los clientes.

- **b (coef = 3.308155)**

El parámetro b representa la propensión a que los clientes realizan compras. A mayor sea el valor una mayor propensión a que el cliente compre.

Por lo tanto, indica que al menos el cliente compre unas 3 veces antes de desactivarse.

Con el análisis de estos parámetros podemos determinar que el modelo construido es de alta precisión al ser tan estrechos los intervalos de confianza.

La única medida que se puede sacar de este modelo son las compras futuras que al evaluarlo con el valor real quedan las siguientes medidas:

- **MAE: 0.5571**
- **R cuadrado: 0.8574**

## 4.2 Modelo Gamma-Gamma

Teniendo ya creado el modelo BetaGeoFitter comenzamos con la implementación del modelo Gamma-Gamma, este modelo es el encargado de predecir la variable monetaria para la predicción del CLTV.

Para poder implementar el modelo primero hay que realizar la comprobación del que el valor monetario no tenga una alta correlación con la frecuencia, como se puede ver en la siguiente tabla.

	frequency	monetary_value
frequency	1.000000	0.179761
monetary_value	0.179761	1.000000

**Tabla 4** Gamma-Gamma correlación de comprobación

Al observar que la correlación entre las dos variables es muy baja procedemos a montar el modelo y realizar un resumen que nos da información de los parámetros del modelo. El resultado del resumen es la siguiente tabla:

	coef	se(coef)	lower 95% bound	upper 95% bound
p	1.501276	0.009437	1.482780	1.519772
q	5.365878	0.015630	5.335243	5.396512
v	52.504438	0.443424	51.635326	53.373549

**Tabla 5** Gamma-Gamma resumen

El resumen del modelo indica información sobre los parámetros ajustados del modelo, como los intervalos de confianza que corresponden a las columnas *lower 95% bound* y *upper 95% bound*. A continuación, se procede a detallar cada uno de los parámetros del modelo:

- **p (coef = 1.501276)**

El parámetro p representa la variabilidad de la tasa de gasto de los clientes. Un valor más alto de p implica una menor variabilidad en los gastos individuales.

El error estándar de r es de 0.009437 lo que indica que el modelo realiza una estimación precisa.

- **q (coef = 5.365878)**

El parámetro q es otro componente de la distribución este está asociado a la variabilidad en la media de gasto de los clientes. A mayor sea el componente q sugiere una menor variabilidad de los gastos individuales.

- **v (coef = 52.504438)**

El parámetro v se refiere a la media de los valores monetarios. Ha mayor sea el valor del parámetro v mayor es la media en el gasto de los clientes.

Con los datos obtenidos del resumen podemos determinar que los valores monetarios tienen baja variabilidad y tienen una alta media de gasto.

Para la evaluación del modelo se ha comprobado los valores predichos con los reales obteniendo:

- **MAE: 32.72**
- **R cuadrado: 0.7325**

Con ambos modelos implementados podemos juntarlos para poder realizar la predicción del CLTV.

### **4.3 Evaluación con el Modelo DNN**

En esta sección vamos a comparar los resultados de la predicción de las variables, así como algunos de los resultados de los outputs de ambos modelos.

Podemos comparar dos medidas concretas que son equivalentes para los modelos que se han presentado en el documento. Estamos hablando del número de compras futuras y el valor monetario.

La probabilidad de que un cliente se mantenga activo en la red neuronal es una variable diferente en comparación con el modelo BetaGeoFitter. En el modelo BetaGeoFitter, se puede obtener una probabilidad de vida del cliente, pero esta probabilidad se refiere a la posibilidad de que varíe su frecuencia de compra, no a que permanezca activo. Por esta razón, no se compararán los resultados de estas probabilidades, aunque sí se realizarán comparaciones para las otras variables.

Para el número de compras futuras hay varias diferencias:

- La primera de ellas es que para el modelo de redes neuronales no calcula el número de ventas que va a realizar el cliente durante el periodo como sí lo hace el modelo BetaGeoFitter sino que calcula el agregado total de ventas que tendría el cliente en ese periodo. Pero en términos generales podemos decir que representan lo mismo.
- La segunda diferencia se encuentra en la variable monetaria. En términos de significado, no representan lo mismo porque el modelo Gamma-Gamma utiliza una variable monetaria más representativa en términos de beneficios, calculada incorporando costes y descuentos que no se han considerado en el modelo de red neuronal profunda (DNN). En el modelo DNN, se ha optado por utilizar el ingreso bruto para observar el dinero total que el cliente aporta a la empresa. Sin embargo, en términos generales, ambas variables representan lo mismo, aunque se miden de manera diferente.

Con las diferencias claras de las variables que se obtienen en los diferentes modelos se vamos a visualizar la comparativa de las métricas de evaluación para la variable referente al número de compras futuras.

	DNN	BetaGeoFitter
MAE	1.270	0.5571
R cuadrado	0.991	0.8574

**Tabla 6** Evaluación modelo Frequency

En esta tabla se comparan las métricas de evaluación de dos modelos: DNN y BetaGeoFitter. Observamos que, en términos del error absoluto medio (MAE), el modelo BetaGeoFitter funciona mejor, con un error de 0.5571. Este error, en términos de negocio, es prácticamente insignificante, representando menos de una compra de media. Por otro lado, el modelo de red neuronal (DNN) tiene un MAE de 1.270, lo que equivale a más de una compra de media. Aunque este error también es pequeño en términos de negocio, matemáticamente podemos afirmar que el modelo tradicional BetaGeoFitter tiene un mejor rendimiento en términos de minimizar el error absoluto de las predicciones individuales.

Sin embargo, cuando consideramos el coeficiente de determinación (R cuadrado), el modelo de red neuronal (DNN) muestra un mejor ajuste con un valor de 0.991, en comparación con 0.8574 del modelo BetaGeoFitter. Esto indica que el modelo DNN es más efectivo para capturar la variabilidad de los datos y puede proporcionar una mejor representación de los patrones subyacentes. En resumen, aunque el BetaGeoFitter minimiza mejor el error absoluto, el DNN ofrece un ajuste superior y una mejor capacidad para explicar la variabilidad en los datos.

Por otro lado, tenemos la medición de la variable monetaria que podemos ver las diferentes métricas aunadas en la siguiente tabla:

	DNN	Gamma-Gamma
MAE	2.1600	32.7200
R cuadrado	0.9474	0.7325

**Tabla 7** Evaluación modelo Monetary

En esta tabla se comparan las métricas de evaluación de los dos modelos: DNN y Gamma-Gamma para la variable Monetary. Aunque ambas variables representan lo mismo, no están en la misma escala, por lo que el error absoluto medio (MAE) no es directamente comparable.

1. **MAE:**

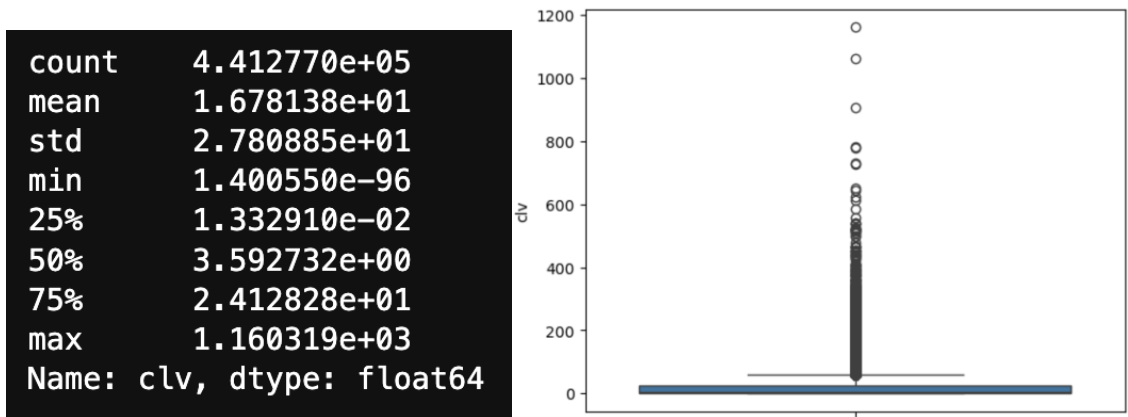
- El modelo Gamma-Gamma tiene un MAE de 32.72, lo cual no es un buen resultado para realizar un cálculo de CLTV de calidad en el contexto del negocio analizado.
- Por otro lado, el modelo de red neuronal (DNN) tiene un MAE de 2.16, que es considerablemente menor y más aceptable para el negocio.

2. **R<sup>2</sup>:**

- En términos de R<sup>2</sup>, el modelo DNN es claramente superior, con un valor de 0.9474, lo que indica que puede explicar casi la totalidad de la variabilidad en los datos.
- El modelo Gamma-Gamma, con un R<sup>2</sup> de 0.7325, no puede explicar un cuarto de los datos predichos, lo que indica un ajuste significativamente peor.

Aunque, como ya ha sido mencionado, no se pueden comparar directamente los MAE debido a las diferencias en las escalas de las variables, está claro que el modelo Gamma-Gamma tiene un rendimiento insatisfactorio con un MAE alto. Además, el modelo DNN demuestra ser mucho más efectivo para explicar la variabilidad en los datos, haciendo que sea el claro ganador en términos de R<sup>2</sup>.

Una ventaja que tienen los modelos tradicionales frente al desarrollado en este proyecto es que los modelos tradicionales te ofrecen un output de variable única de CLTV no teniendo que realizar el paso previo de buscar el cálculo de CLTV que encaje con tu objetivo de negocio. En las siguientes **Figura 18** queda una representación del CLTV que da como resultado el modelo tradicional.



**Figura 18** CLTV descriptivo

En esta figura se representa un boxplot del CLTV (Customer Lifetime Value), acompañado de un resumen estadístico de la distribución de los datos. El resumen estadístico muestra que hay 441,277 datos, con una media de 16.78 y una desviación estándar de 27.81. Los valores de CLTV varían desde un mínimo extremadamente bajo de  $1.40e-96$  hasta un máximo de 1160.32. Los percentiles revelan que el 25% de los valores son menores o iguales a 0.013, el 50% (mediana) es 3.59, y el 75% es 24.13.

El boxplot del CLTV muestra que el 75% de los valores de CLTV no superan los 24.13, pero la figura está distorsionada por la presencia de numerosos valores atípicos que alcanzan hasta 1200, indicando que hay clientes con un CLTV significativamente más alto que la mayoría. Esta gran cantidad de valores atípicos afecta la visualización y comprensión de la distribución real del CLTV.

En el contexto del negocio, la métrica calculada del CLTV parece no tener sentido. La mayoría de los clientes son fieles, compran cada 15 días y el gasto medio por compra es de 20-40 euros. En un periodo de 6 meses, estos clientes realizarían aproximadamente 12 compras, lo que supondría un gasto mínimo de 240 euros, superando ampliamente los 24.13 puntos de CLTV del percentil 75.

En conclusión, aunque los modelos tradicionales pueden ser útiles dependiendo del objetivo de la empresa y el trabajo realizado, en este caso parecen subestimar el CLTV. Esto subraya la necesidad de utilizar modelos más avanzados para capturar mejor las complejidades del comportamiento del cliente y proporcionar una estimación más precisa del Customer Lifetime Value.

## 5 Metodología de cálculo de CLTV

Uno de los objetivos de este trabajo de fin de grado es el diseño de una metodología de cálculo de Customer Lifetime Value, que pueda ser replicable en otro escenario de negocio o en algún caso práctico futuro que necesite la obtención de esta métrica.

A continuación, se detallan las diferentes fases por la que se deberá pasar para la obtención de la métrica CLTV.

## **5.1 Entendimiento de Negocio**

La primera fase corresponde a la comprensión del negocio al que se le va a realizar la métrica. Es muy necesario comprender el funcionamiento de la empresa, así como el comportamiento que tienen los clientes en ellas para poder determinar el valor que tiene un cliente.

En negocios como en el que está basado el proyecto tiene un peso equilibrado tanto las variables correlacionadas con la frecuencia de comportamiento de compra con la variable monetaria obteniendo un valor de cliente representativo de cada uno de sus clientes. Mientras que esté mismo tipo de cálculo para una empresa con frecuencias de compras anuales como puede ser el caso de una empresa dedicada al retail o a un modelo de suscripción puede no ser el óptimo.

Métricas tan importantes para la toma de decisiones de un negocio como es el Customer Lifetime Value deben de ser creadas a partir de dicho negocio, no al revés. Por ello hay que tener un pleno conocimiento y entendimiento de todos los factores que afecten en el valor de cliente y elegir el cálculo adecuado para su obtención.

## **5.2 Análisis, Limpieza y Creación de Sábana de Datos**

Una vez se haya logrado una comprensión profunda del negocio, los objetivos de la empresa y el comportamiento de los clientes para el Customer Lifetime Value que se desea realizar, el siguiente paso crucial es la preparación de los datos para el modelo a desarrollar.

Este proceso incluye un análisis de los datos que pueda proporcionar la empresa, la limpieza de estos y la creación de variables como las del modelo RFM que puedan aportar información dentro de la sábana de datos de la predicción.

La preparación de datos es una fase crítica en el cálculo del Customer Lifetime Value. La creación de una buena base de datos es la base para construir un modelo de predicción de CLTV robusto y de buena calidad.

## **5.3 Desarrollo del Modelo**

Teniendo la perspectiva del negocio y una base de datos por la que construir un modelo el siguiente paso es el desarrollo del modelo. Como se ha hablado en el documento existen ventajas e inconvenientes de cada tipo de modelo, tanto tradicionales como modelos más avanzados como el uso de redes neuronales.

Por esta razón es tener claro el objetivo de negocio para poder dedicar los recursos al modelo que genere un mejor cálculo.

Como resumen, los modelos tradicionales tienen un mejor rendimiento si la sabana de datos con la que se trabaja es de un tamaño menor y tiene un cálculo sencillo de Customer Lifetime Value que puede servir para los objetivos de negocio de la mayoría de las empresas, además de al llevar más tiempo en el sector es más fácil encontrar información de su implementación y su desarrollo. En términos de modelos avanzados, tienen un mejor rendimiento con bases de datos más grandes que los modelos tradicionales y tiene un mayor grado de personalización y sofisticación que puede servir para resolver algunos de los problemas más específicos que puedan tener las empresas.

## 5.4 Evaluación de resultados

Esta fase se realiza tras tener realizada alguna prueba con el modelo y se debe realizar mediante la comparativa de datos reales por lo que se ha tenido que reservar algún periodo de tiempo para poder realizar esta evaluación.

Se ha de evaluar todas de las variables que son empleadas en el cálculo de CLTV y el propio Customer Lifetime Value bajo una perspectiva de negocio.

Dentro de esta fase es donde se encuentra la comprensión de los datos del modelo y cómo se pueden aplicar en el cliente, por lo que pueden tener buenos resultados matemáticos pero la métrica que se está calculando no resuelve los problemas iniciales de la empresa.

## 5.5 Implementación

Teniendo un modelo funcional que devuelve resultados prometedores y con el visto bueno de la empresa el último paso es el de implementar dicho modelo dentro de la estructura de los datos de la empresa para que puedan ejecutarlo y obtener los insights necesarios para obtener sus objetivos o para resolver sus problemas que tuvieron inicialmente.

Es importante un seguimiento del modelo debido a que todos los cambios que realicen en la empresa con los insights que extraen de los resultados del modelo es posible que el comportamiento de los clientes cambie y haga falta un reentrenamiento del modelo.

## 5.6 Resumen de la Metodología

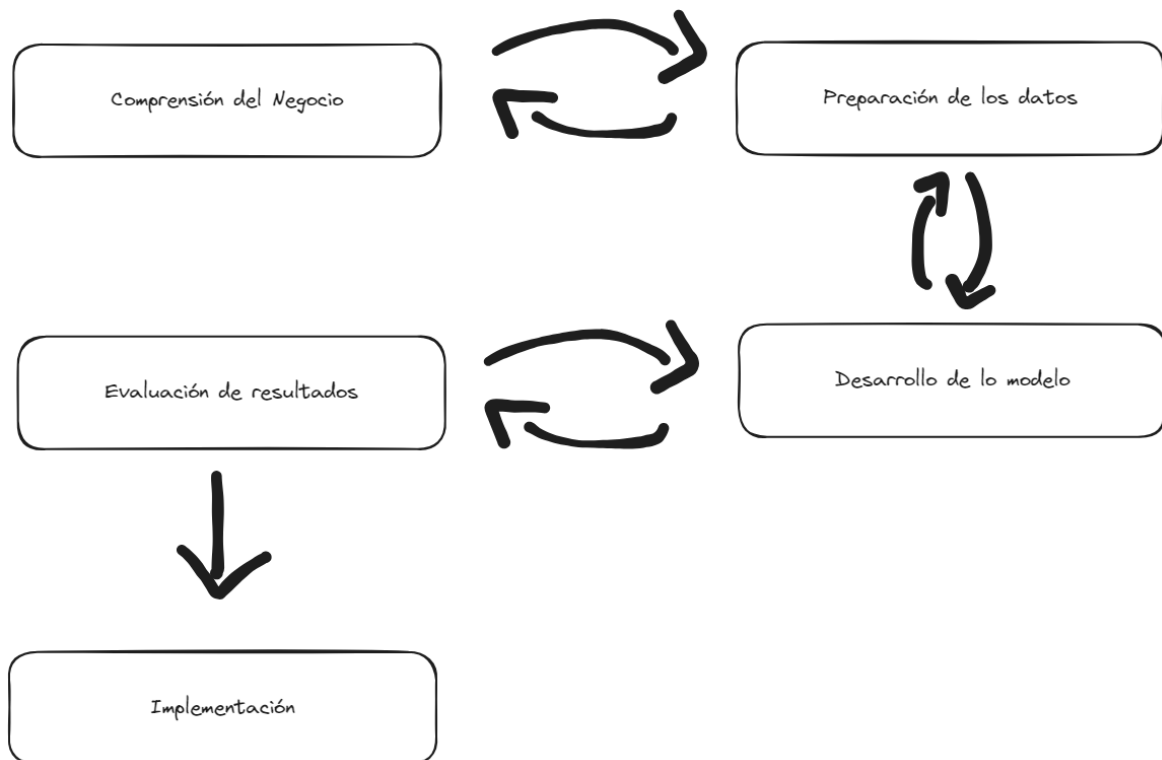
En resumen, esta metodología cuenta con 5 fases:

1. **Comprensión del Negocio:** Fase primordial donde se establece el cálculo del Customer Lifetime Value y donde se comprenden los objetivos o problemas de la empresa.
2. **Preparación de los datos:** Fase de análisis donde se comprenden los datos de los que se dispone para poder crear la sabana de datos de la

predicción teniendo en cuenta el modelo que se desarrollará en la fase siguiente.

3. **Desarrollo del modelo:** Fase donde se desarrolla el modelo capaz de predecir las variables necesarias para el cálculo del CLTV o incluso el propio CLTV.
4. **Evaluación de los resultados:** Fase donde se evalúan los resultados del modelo desde la perspectiva matemática y desde la perspectiva del negocio al que se le implementa la solución.
5. **Implementación:** Última fase donde se implementa el modelo dentro de la estructura de la empresa y se realiza un seguimiento de las decisiones que toma la empresa a través de los insights que proporciona el modelo.

En la imagen siguiente se hace una representación gráfica de las diferentes fases de la metodología.



**Figura 19** Esquema metodología de cálculo CLTV

## 6 Resultados

En este capítulo se resumirán los insights que se han conseguido con el desarrollo de los diferentes modelos para la resolución del problema de la empresa y con el diseño de la metodología anteriormente presentada, comenzando con los modelos fallidos, explicados en el apartado **3.4 Modelos Anteriores**, seguido con los modelos tradicionales desarrollados en los apartados **4.1 Modelo BetaGeoFitter** y **4.2 Modelo Gamma-Gamma**, finalmente el modelo final implementado detallado en el apartado **3.5 Modelo DNN Implementado**.

### 6.1 Modelo basado en ZILN

Al intentar implementar este modelo se concluyó que las librerías desarrolladas por los autores estaban incompletas y mal formadas por lo que no se pudo desarrollar y comprobar la efectividad de este modelo con los datos de la empresa.

### 6.2 Modelo CNN

Este modelo se consiguió desarrollar una primera arquitectura de red para poder realizar una primera prueba de funcionamiento de este tipo de red en el cálculo directo de la métrica Customer Lifetime Value. Sin embargo, en la fase de entrenamiento no se pudo conseguir la ingesta de los datos para la predicción imposibilitando la continuación de desarrollo de esta red CNN.

### 6.3 Modelo DNN

Tras el desarrollo de la red CNN se comenzó el desarrollo de una red densa que calculara la métrica objetivo de manera directa. Durante la fase de entrenamiento de esta red se observó un sobreajuste que se intentó paliar mediante diferentes regulaciones como l1 y l2 y la adición de capas ocultas de tipo dropout sin lograr mejorar los resultados.

Durante el desarrollo de esta red se determinó que el método de predicción directa para el Customer Lifetime Value no es una solución óptima para el negocio debido a la característica de las redes neuronales de caja negra y no poder explicar la solución que estaba proponiendo el modelo.

Esto provocó el cambio en el enfoque en el cálculo del CLTV y la necesidad de predicción de tres variables distintas:

- **Probabilidad de cliente activo**
- **Numero de compras futuras**
- **Coste medio de transacción del cliente**

## 6.4 Modelo BetaGeoFitter

El modelo BetaGeoFitter se desarrolló para la evaluación de desempeño del modelo final y es el encargado de predecir las variables relativas al comportamiento de cliente.

El rendimiento de este modelo fue muy eficiente en capturar el patrón de compras de los clientes obteniendo métricas de validación de:

- **MAE: 0.55**
- **R cuadrado: 0.85**

Estas métricas indican que el modelo se equivoca de promedio en media compra que en términos de negocio es casi imperceptible y tiene una explicabilidad de la varianza en el 85% de los datos.

Estas métricas nos indican que es un modelo apto para la predicción de esta variable y de ser empleado para el cálculo del CLTV.

## 6.5 Modelo Gamma-Gamma

El modelo Gamma-Gamma se desarrolló en conjunto al modelo BetaGeoFitter para poder realizar la evaluación del desempeño del modelo implementado. Este modelo es el encargado de predecir las variables monetarias que serán empleados en el cálculo del Customer Lifetime Value. Para evaluar su desempeño se calcularon medidas como el MAE y R cuadrado:

- **MAE: 32.72**
- **R cuadrado: 0.73**

Estas métricas indican que el modelo se equivoca de promedio en 32.72, este valor es demasiado alto para poder realizar una buena predicción del Customer Lifetime Value, y es capaz de explicar el 73% de la variabilidad de los datos dejando más de un cuarto de los datos sin poder ser captados provocando en parte el valor subóptimo del MAE.

Este resultado provoca la mala calidad del CLTV generado por el modelo que no corresponde con el valor real que tienen los clientes de la empresa de los datos con los que se ha realizado la predicción.

## 6.6 Modelo Implementado

Finalmente, el modelo implementado en la empresa para el cálculo de su Customer Lifetime Value es una red densa que predice tres valores:

- **Probabilidad de cliente activo**
- **Número de compras futuras**

### - **Coste medio de transacción del cliente**

Para medir el desempeño de la red se ha evaluado la eficacia en la predicción de cada una de las variables devueltas por el modelo.

#### **Probabilidad de cliente activo**

Para la evaluación de esta variable se crearon las siguientes métricas de medición:

- **Precisión: 0.82**
- **Recall: 0.82**
- **F1-score: 0.81**

Estos resultados implican una buena segmentación de clientes activos y no activos, en este caso la red predice en un 90% el grupo de los no activos frente a un 71% del grupo activo por lo que la red tiene mayor efectividad para encontrar aquellos clientes que no van a realizar una compra en un periodo de 6 meses.

#### **Número de compras futuras**

Para la evaluación de esta variable se forma el MAE y el R cuadrado obteniendo los siguientes resultados:

- MAE: 1.27
- R cuadrado: 0.99

Estos resultados indican que la variabilidad de los datos es explicada en casi un 100% de los datos y que el modelo falla de media en la predicción en 1.27 compras, que en términos de cálculo de CLTV no afecta en exceso al valor de clientes, aunque puede ser mejorable empleando otros modelos como el anterior comentado BetaGeoFitter.

#### **Coste medio de transacción del cliente**

Para la evaluación de esta variable se forma el MAE y el R cuadrado obteniendo los siguientes resultados:

- MAE: 2.16
- R cuadrado: 0.94

Estos resultados indican que la variabilidad de los datos es explicada en un 94% de los datos y que el modelo falla de media en la predicción en 2.16 euros por transacción, un muy buen resultado comparándolo con el devuelto por el modelo tradicional Gamma-Gamma.

Con estas tres variables se consiguió el cálculo de un Customer Lifetime Value que fue capaz de representar un valor de cliente identificativo para cada uno de los clientes de la empresa permitiendo a esta desarrollar diferentes estrategias de negocio centrados en la retención de los clientes con mayor CLTV y la captación de clientes con un potencial de CLTV alto.

## 7 Conclusión

El presente Trabajo de Fin de Grado ha abordado con éxito el objetivo principal de diseñar una metodología de cálculo de Customer Lifetime Value adaptada y aplicada un entorno empresarial real, dentro del sector alimenticio.

Este objetivo se ha logrado mediante el desarrollo de una red neuronal densa capaz de predecir con precisión y eficacia cada uno de los componentes necesarios para el cálculo del CLTV. A continuación, se detallan las conclusiones derivadas del cumplimiento de los objetivos específicos planteados en el apartado **1.1 Objetivos del TFG**:

- **Comprensión del Negocio**

El análisis de los datos proporcionados permitió una comprensión profunda de los objetivos y necesidades de la empresa en tener y aplicar una métrica de comportamiento del cliente como el Customer Lifetime Value.

- **Análisis Exploratorio de los Datos**

Se implementaron técnicas de análisis de datos que ayudaron a identificar tendencias, patrones y anomalías en los datos. Estas técnicas incluyeron visualización de datos, así como el análisis en el comportamiento de compra del cliente. Este análisis exploratorio fue crucial para la preparación y la limpieza de datos antes de la modelización del modelo predictivo.

- **Aplicación de Metodologías de Análisis de Datos**

Se investigaron y aplicaron dentro del proceso de realización de los modelos diversas técnicas de análisis de datos como la estandarización y la extracción de características relevantes que alimentaron el modelo para una mejor precisión y rendimiento.

- **Modelización del Modelo**

Se diseñó, desarrolló e implementó un modelo basado en redes neuronales para la predicción del CLTV. Este modelo aprovechó las características de las redes neuronales para capturar patrones complejos de los datos en las transacciones de los clientes y su capacidad de tener varios outputs sin necesidad de implementar diferentes modelos.

- **Comparativa con Modelos Tradicionales**

Se realizó una comparativa entre el modelo desarrollado con los modelos tradicionales como el BetaGeoFitter y el Gamma-Gamma

donde el modelo basado en redes ofreció una mejor precisión y un mejor CLTV bajo la perspectiva de negocio.

- **Descripción de la Metodología Diseñada**

Se elaboró un esquema detallado de la metodología diseñada, describiendo cada fase desde la comprensión del negocio hasta la implementación y seguimiento del modelo. Esta metodología puede ser aplicable a otros casos de negocio con características similares.

- **Documentación del Desarrollo del Trabajo**

Se recopiló y documentó toda la información y los procesos seguidos durante el desarrollo del proyecto. Desde las investigaciones previas necesarias para el correcto desarrollo del trabajo hasta el final desarrollo de este documento.

En conclusión, este proyecto ha logrado diseñar e implementar con éxito una metodología de predicción del Customer Lifetime Value empleando redes neuronales, demostrando su eficacia y comparando los resultados con modelos tradicionales. Este trabajo proporciona además una metodología sólida y bien documentada que puede ser replicada en otros contextos de negocio similares.

## **7.1 Trabajos Futuros**

A pesar del trabajo desarrollado con el modelo de redes neuronales para la predicción del Customer Lifetime Value, se ha probado a implementar el modelo para otro negocio dedicado al sector del *retail* dando resultados no óptimos. Para abordar estas limitaciones, futuros trabajos deberían enfocarse en varias áreas claves. Primero en explorar las diferentes características que capturen mejor el comportamiento de un cliente dentro del sector del *retail*. Segundo, se podría investigar diferentes arquitecturas de redes neuronales más sofisticadas como LSTM explicadas en este mismo documento que puedan manejar secuencias de datos temporales de manera más efectiva. Por último, sería útil realizar pruebas de validación cruzadas más exhaustivas y ajustar hiperparámetros del modelo presentado para que sea más específico del sector del *retail*. Estos enfoques no solo mejorarán la precisión y eficacia del modelo desarrollado, sino que también proporcionarían conocimientos más valiosos y aplicables para estrategias de negocio en términos de retención y captación de clientes además de *insights* dentro del mundo del marketing en los diferentes sectores de las empresas.

## 8 Análisis de Impacto

Este proyecto ha sido capaz de diseñar una metodología para el cálculo del Customer Lifetime Value, comparado con modelos tradicionales como el Gamma-Gamma y BetaGeoFitter. A continuación, se presentan las posibles contribuciones y los potenciales beneficios en diversos ámbitos.

### 8.1 Ámbito Académico

El proyecto ha contribuido al campo de investigación en análisis exploratorio y predictivo, además del uso de machine learning y Deep learning aplicados a la gestión de clientes. La comparativa del modelo desarrollado en este escrito con modelos tradicionales proporcionan un nuevo enfoque sobre las fortalezas y limitaciones de cada modelo dependiendo del negocio al que se le aplique.

Este trabajo puede servir como base para futuros estudios y desarrollos para la predicción del CLTV, promoviendo el uso de nuevas tecnologías como pueden ser las redes del tipo LSTM y nuevas técnicas en el ámbito académico.

### 8.2 Sector Empresarial

La implementación de un modelo basado en redes neuronales para la predicción de diferentes métricas, como son la probabilidad de que se mantenga activo el cliente, el número de compras que va a realizar y la cantidad media de dinero que va a realizar por transacción, así como el CLTV, tiene un impacto significativo dentro del sector empresarial, en esencia en campos relacionados con el CRM, gestión de relaciones con el cliente, como son el marketing o las ventas. Las empresas pueden emplear este tipo de modelo para diferentes objetivos:

- **Optimización de estrategias de marketing:** Al poder identificar clientes con un alto CLTV, las empresas pueden enfocar su inversión en la captación de estos segmentos mejorando la eficiencia y la efectividad de las campañas de marketing.
- **Retención de clientes:** Tener una métrica como el CLTV permite a las empresas poder crear estrategias para mejorar la retención de los clientes con un CLTV rentable reduciendo la tasa de abandono y la inversión al poder centrar las acciones de una manera más óptima.
- **Mejor toma de decisiones:** Con las métricas proporcionadas por el modelo se pueden extraer diferentes insights sobre el futuro comportamiento de los clientes y poder generar una imagen de la compañía en un futuro facilitando la toma de decisiones al tener información fiable de la imagen a futuro de la empresa.

### **8.3 Impacto en la Comunidad Científica y Tecnológica**

El desarrollo de un modelo DNN y realizar la comparativa con modelos tradicionales contribuye al avance de la ciencia y la tecnología. Además, este trabajo demuestra la aplicabilidad y los beneficios de las técnicas modernas de machine learning en un entorno empresarial real, incentivando a otros profesionales a adoptar enfoques similares en sus futuras investigaciones.

## 9 Bibliografía

- Kotler, P. (1974). Marketing during periods of shortage. *Journal of Marketing*, 38(3), 20-29.
- Fader, P., Hardie, B., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415-430.
- Fader, P. S., & Hardie, B. G. S. (2009). Customer Lifetime Value: Rescaling the Value of the Customer Base. *Journal of Interactive Marketing*, 23(2), 71-89. doi:10.1016/j.intmar.2008.11.001
- Borle, S., Sing, S., Jain, D. (2008). Customer Lifetime Value Measurement. *En Management Science*, Vol. 54, No. 1, pp. 100-112.
- Cui, D., Curry, D. 2005. Prediction in marketing using Support Vector Machine. *En Marketing Science*, Vol. 24, No. 4. <https://doi.org/10.1287/mksc.1050.0123>
- Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model." *Marketing Science*, 24(2), 275-284
- Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., & Sriram, S. (2006). "Modeling Customer Lifetime Value." *Journal of Service Research*, 9(2), 139-155.
- Gupta, S., Hanssens, D. M., Hardie, B. G. S., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., & Sriram, S. (2006). "Customer Lifetime Value Modeling: Review and Future Directions." *Journal of Interactive Marketing*, 20(2), 37-53.
- Zhang, Y., Luo, Y., Sun, L., Liu, J., & Zeng, D. (2018). "Customer Lifetime Value Prediction Using Deep Learning." *IEEE Access*, 6, 39756-39765.
- Zhang, S., Li, Z., & Ren, J. (2019). "Customer Lifetime Value Prediction Based on Convolutional Neural Networks: A Case Study in Retail Industry". *Journal of Artificial Intelligence Research*, 65, 987-1002.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*, 86(11), 2278-2324.
- Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing Customers. *Journal of Marketing Research*, 41(1), 7-18. doi:10.1509/jmkr.41.1.7.25084
- Farris, P. W., Bendle, N. T., Pfeifer, P. E., & Reibstein, D. J. (2010). *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance* (2nd ed.). Pearson Education.
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, 80(6), 69-96. <https://doi.org/10.1509/jm.15.0420>
- UC Berkeley School of Information. (2023). *Predicting Lifetime Value of Users Using Machine Learning*. Recuperado de <https://www.ischool.berkeley.edu/projects/2023/predicting-lifetime-value-users-using-machine-learning>

- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- Gers, F. A., Schraudolph, N. N., & Schmidhuber, J. (2002). "Learning Precise Timing with LSTM Recurrent Networks". *Journal of Machine Learning Research*, 3, 115-143.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". *arXiv preprint arXiv:1412.3555*.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Chamberlain, B. P., Cardoso, A., Liu, C. H., Pagliari, R., & Deisenroth, M. P. (2017). Customer lifetime value prediction using embeddings. En *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1753–1762). ACM. <https://arxiv.org/pdf/1703.02596.pdf>.
- Zhao, H., Wu, J., & Xu, X. (2019). A Deep Probabilistic Model for Customer Lifetime Value Prediction. *arXiv preprint arXiv:1912.07753*. Recuperado de <https://arxiv.org/abs/1912.07753>
- Zhao, X., Liu, X., & Wang, H. (2019). "Using RFM Model and Data Mining for Customer Value Analysis: A Case Study in Retail Industry". *Journal of Retail Analytics*.
- Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science*, 24(2), 275-284.
- Wang, Xiaojing, Liu, Tianqi, and Miao, Jingang. (2019). A Deep Probabilistic Model for Customer Lifetime Value Prediction : [arXiv:1912.07753](https://arxiv.org/abs/1912.07753)

## 10 Anexos

TFG_Memoria.pdf			
ORIGINALITY REPORT			
12%	10%	%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	Submitted to Universidad Internacional de la Rioja Student Paper		1%
2	Submitted to Infile Student Paper		1%
3	Submitted to Universidad Politécnica de Madrid Student Paper		1%
4	Submitted to Corporación Universitaria Minuto de Dios, UNIMINUTO Student Paper		<1%
5	Submitted to University of Leeds Student Paper		<1%
6	ri.itba.edu.ar Internet Source		<1%
7	Submitted to ISM Vadybos ir ekonomikos universitetas, UAB Student Paper		<1%
8	oa.upm.es Internet Source		<1%

9	<a href="http://repositorio.unal.edu.co">repositorio.unal.edu.co</a> Internet Source	<1 %
10	Submitted to Universidad Nacional Abierta y a Distancia, UNAD,UNAD Student Paper	<1 %
11	Submitted to University of the Sunshine Coast Student Paper	<1 %
12	<a href="http://eprints.ucm.es">eprints.ucm.es</a> Internet Source	<1 %
13	<a href="http://www.coursehero.com">www.coursehero.com</a> Internet Source	<1 %
14	<a href="http://dergipark.org.tr">dergipark.org.tr</a> Internet Source	<1 %
15	Submitted to Nyenrode Business Universiteit Student Paper	<1 %
16	<a href="http://digibuo.uniovi.es">digibuo.uniovi.es</a> Internet Source	<1 %
17	<a href="http://www.orpconference.org">www.orpconference.org</a> Internet Source	<1 %
18	<a href="http://hdl.handle.net">hdl.handle.net</a> Internet Source	<1 %
19	<a href="http://essay.utwente.nl">essay.utwente.nl</a> Internet Source	<1 %
20	<a href="http://www.springerprofessional.de">www.springerprofessional.de</a> Internet Source	<1 %

		<1 %
21	<a href="http://insis.vse.cz">insis.vse.cz</a> Internet Source	<1 %
22	<a href="http://www.vse.cz">www.vse.cz</a> Internet Source	<1 %
23	<a href="http://www.researchgate.net">www.researchgate.net</a> Internet Source	<1 %
24	<a href="http://www.tandfonline.com">www.tandfonline.com</a> Internet Source	<1 %
25	Submitted to Escuela Politecnica Nacional Student Paper	<1 %
26	<a href="http://repositorio.uam.es">repositorio.uam.es</a> Internet Source	<1 %
27	<a href="http://link.springer.com">link.springer.com</a> Internet Source	<1 %
28	<a href="http://upcommons.upc.edu">upcommons.upc.edu</a> Internet Source	<1 %
29	<a href="http://repository.unab.edu.co">repository.unab.edu.co</a> Internet Source	<1 %
30	<a href="http://www.aurelielemmens.com">www.aurelielemmens.com</a> Internet Source	<1 %
31	<a href="http://ela.kpi.ua">ela.kpi.ua</a> Internet Source	<1 %

32	<a href="http://qdoc.tips">qdoc.tips</a> Internet Source	<1%
33	<a href="http://www.biblio-sepi.esimez.ipn.mx">www.biblio-sepi.esimez.ipn.mx</a> Internet Source	<1%
34	Submitted to BPP College of Professional Studies Limited Student Paper	<1%
35	<a href="http://mecnica.uniandes.edu.co">mecnica.uniandes.edu.co</a> Internet Source	<1%
36	Submitted to University of Greenwich Student Paper	<1%
37	<a href="http://dspace.espol.edu.ec">dspace.espol.edu.ec</a> Internet Source	<1%
38	<a href="http://stats.stackexchange.com">stats.stackexchange.com</a> Internet Source	<1%
39	<a href="http://www.vnunet.es">www.vnunet.es</a> Internet Source	<1%
40	<a href="http://aprenderly.com">aprenderly.com</a> Internet Source	<1%
41	<a href="http://www.riuc.bc.uc.edu.ve">www.riuc.bc.uc.edu.ve</a> Internet Source	<1%
42	<a href="http://docobook.com">docobook.com</a> Internet Source	<1%
43	Submitted to Universidad TecMilenio	

	Student Paper	<1 %
44	<a href="http://rua.ua.es">rua.ua.es</a> Internet Source	<1 %
45	Submitted to Universidad Manuela Beltrán Student Paper	<1 %
46	<a href="http://www.usb.ac.za">www.usb.ac.za</a> Internet Source	<1 %
47	<a href="http://doku.pub">doku.pub</a> Internet Source	<1 %
48	<a href="http://uvadoc.uva.es">uvadoc.uva.es</a> Internet Source	<1 %
49	Submitted to Pontificia Universidad Javeriana Cali Student Paper	<1 %
50	Submitted to Universidad Católica de Santa María Student Paper	<1 %
51	<a href="http://businessbse.wordpress.com">businessbse.wordpress.com</a> Internet Source	<1 %
52	<a href="http://publikationen.bibliothek.kit.edu">publikationen.bibliothek.kit.edu</a> Internet Source	<1 %
53	<a href="http://discovery.researcher.life">discovery.researcher.life</a> Internet Source	<1 %

54	<a href="https://es.slideshare.net">es.slideshare.net</a> Internet Source	<1%
55	<a href="https://libros.catedu.es">libros.catedu.es</a> Internet Source	<1%
56	<a href="https://repositorio.cuc.edu.co">repositorio.cuc.edu.co</a> Internet Source	<1%
57	<a href="http://www.itson.mx">www.itson.mx</a> Internet Source	<1%
58	<a href="https://cl2.bumeran.com">cl2.bumeran.com</a> Internet Source	<1%
59	<a href="https://gacetasanitaria.org">gacetasanitaria.org</a> Internet Source	<1%
60	<a href="https://reunir.unir.net">reunir.unir.net</a> Internet Source	<1%
61	<a href="https://smartinezcardenas.wordpress.com">smartinezcardenas.wordpress.com</a> Internet Source	<1%
62	<a href="https://www.cambridge.org">www.cambridge.org</a> Internet Source	<1%
63	<a href="https://atos.net">atos.net</a> Internet Source	<1%
64	<a href="https://dspace.umh.es">dspace.umh.es</a> Internet Source	<1%
65	<a href="https://herbshop.es">herbshop.es</a> Internet Source	<1%

66	<a href="http://ideas.repec.org">ideas.repec.org</a> Internet Source	<1 %
67	<a href="http://journal.untar.ac.id">journal.untar.ac.id</a> Internet Source	<1 %
68	<a href="http://masterdisseny.com">masterdisseny.com</a> Internet Source	<1 %
69	<a href="http://prezi.com">prezi.com</a> Internet Source	<1 %
70	<a href="http://propolis-colmena.blogspot.com">propolis-colmena.blogspot.com</a> Internet Source	<1 %
71	<a href="http://repositorio.uchile.cl">repositorio.uchile.cl</a> Internet Source	<1 %
72	<a href="http://www.goalsystems.com">www.goalsystems.com</a> Internet Source	<1 %
73	<a href="http://www.meff.es">www.meff.es</a> Internet Source	<1 %
74	<a href="http://www.opus-bayern.de">www.opus-bayern.de</a> Internet Source	<1 %
75	<a href="http://www.unapec.edu.do">www.unapec.edu.do</a> Internet Source	<1 %
76	<a href="http://1library.co">1library.co</a> Internet Source	<1 %
77	Submitted to National College of Ireland Student Paper	<1 %

78	Submitted to Universidad Carlos III de Madrid Student Paper	<1 %
79	aquaticcommons.org Internet Source	<1 %
80	bibliotecadigital.econ.uba.ar Internet Source	<1 %
81	cabbib2.cnea.gov.ar Internet Source	<1 %
82	de.slideshare.net Internet Source	<1 %
83	ebin.pub Internet Source	<1 %
84	epdf.pub Internet Source	<1 %
85	mobbingopinion.bpweb.net Internet Source	<1 %
86	ndl.ethernet.edu.et Internet Source	<1 %
87	repositorio.roca.utfpr.edu.br Internet Source	<1 %
88	ricaxcan.uaz.edu.mx Internet Source	<1 %
89	www.e-radiocontrol.com.ar Internet Source	<1 %


90	<a href="http://www.grafiati.com">www.grafiati.com</a> Internet Source	<1%
91	<a href="http://www.ie.edu">www.ie.edu</a> Internet Source	<1%
92	<a href="http://www.intotheminds.com">www.intotheminds.com</a> Internet Source	<1%
93	<a href="http://www.martinez-ribes.com">www.martinez-ribes.com</a> Internet Source	<1%
94	<a href="http://www.poderjudicial.es">www.poderjudicial.es</a> Internet Source	<1%
95	<a href="http://www.tdx.cat">www.tdx.cat</a> Internet Source	<1%
96	<a href="http://www.upo.es">www.upo.es</a> Internet Source	<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Este documento esta firmado por



<b>Firmante</b>	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
<b>Fecha/Hora</b>	Mon Jun 03 14:18:08 CEST 2024
<b>Emisor del Certificado</b>	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
<b>Numero de Serie</b>	561
<b>Metodo</b>	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)