CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Predicting Customer Lifetime Value Using Deep Learning on Aggregated Retail Data

Submitted by
Kartikeya Sinha (2448029)

Submitted to
Dr. Sathya P.

For the course
Course Code: MDS471
Course Name: Neural Networks and Deep Learning

July-2025

1. Dataset Selection

The dataset captures anonymised retail transactions, giving us a detailed look into how customers shop. Each entry represents a single purchase and contains important information such as:

- **User and demographic details**: User_ID, Gender, Age, Occupation, City Category, Duration of stay in their current city, and Marital Status
- **Purchase information:** Product_ID, Product Category, and the amount spent on the transaction

To better understand customer behaviour and make the most of the data, I have aggregated it at the user level. I have combined all the transactions for each individual into a single record. Here's what this consolidated view reveals:

- **Target Variable**

The total amount spent by the user (Customer Lifetime Value - CLV)

- **Features**

How many different product categories the user has explored, their average spending per transaction, the number of transactions made by each user, how much the user spends on different product categories, and I have also retained all the original user details for better context

This user-level dataset is a powerful tool for understanding Customer Lifetime Value (CLV). By compiling each customer's entire transaction and behavioral history into one comprehensive record, we can better grasp the long-term value each customer brings to our business. To model this, I will employ a tabular deep learning approach. This involves using a neural network that features embedding layers for high-cardinality categorical variables and dense layers to process numeric and aggregated behavioral data. With this method, we will be able to effectively identify individual customer patterns while also recognizing broader behavioral trends across our dataset.

2. Justification of Relevance

Customer Lifetime Value (CLV) modelling is an invaluable tool for any business focused on connecting with its customers. Here’s why it matters:

- **Smart Resource Allocation**: By predicting how much value each customer will bring in the long run, businesses can prioritise their marketing and retention efforts. This means focusing on those customers who are most likely to generate significant profits, rather than treating everyone the same.

- **Better Business Planning**: When businesses have accurate CLV predictions, they can gain insights into future revenue streams. This helps with budgeting and improves sales forecasts, which are crucial for managing inventory and overall finances.

- **Spotting Your Best Customers**: CLV modelling helps businesses understand what makes their most profitable customers tick. With this knowledge, companies can tailor their strategies to engage and nurture these valuable customers more effectively.

- **Reducing Churn**: By identifying customers who might churn, businesses can take early action with retention strategies. This proactive approach helps reduce churn and boosts overall profitability.

- **Building Long-Term Success**: Companies that embrace CLV modelling develop a deeper understanding of their customer base. This knowledge supports smarter, data-driven decisions that can create a lasting competitive edge in the market.

In a nutshell, CLV modelling turns raw sales data into insightful, forward-thinking strategies. It empowers businesses to be more efficient and customer-focused, making it an essential part of modern retail analytics.

3. Exploratory Data Analysis

- **Cardinality and Frequency Distribution of Categorical Variables**

The part investigates how many unique values each categorical feature contains and visualises the frequency distribution. This helps identify dominant categories and those with sparse representation.

- **Outlier and Distribution Check for Numerical Variables**

Distributions of numerical variables are assessed using visualisations such as histograms and box plots. This step is crucial for identifying anomalies and understanding if any transformation of variables is required.

- **Correlation Matrix**

A correlation heatmap is a useful tool for examining the linear relationship between numerical features in our data. We can see how features influence each other. This not only helps us understand the behaviour of the features better but also uncovers hidden patterns that might exist within our dataset.

- **Skewness of the Target Variable**

This part examines the distribution of the target variable, checking for skewness and transforming it if required.

- **Relationship of Features with Target Variable**

The final step of EDA explores how each feature correlates or associates with the target variable. This has been achieved using boxplots and scatterplots.