Name – Khushi Goyal

Customer Lifetime Value (LTV) Prediction Model Project Report

Introduction

Customer Lifetime Value is a crucial metric in customer relationship management and marketing analytics. It represents the total revenue a business can reasonably expect from a single customer throughout their relationship. Predicting LTV allows businesses to:

- Focus marketing strategies on profitable customers.
- Personalize offers and communication.
- Optimize acquisition costs and retention efforts.

The **goal of this project** is to build a machine learning model that predicts the LTV of customers based on past purchase transactions. The demo_transaction dataset provided sufficient features such as transaction history, frequency, and monetary values to train the model.

Abstract

This project focuses on predicting the Customer Lifetime Value (LTV) using transactional data. LTV prediction helps businesses estimate the future value a customer will bring over a specific time period. By analyzing historical purchase behavior and customer characteristics, companies can identify high-value customers, improve customer retention strategies, and optimize marketing investments. The model was implemented in Google Colab using the demo_transaction dataset, ensuring a practical and scalable approach to predictive analytics.

Tools Used

- **Google Colab** – For coding, data preprocessing, model training, and evaluation.
- **Python** – The primary programming language.
- **Pandas & NumPy** – For data manipulation and cleaning.
- **Matplotlib & Seaborn** – For exploratory data analysis and visualization.
- **Scikit-learn** – For model building, training, and evaluation.

Steps Involved in Building the Project

1. **Dataset Loading** – Imported the demo_transaction dataset into Google Colab.

2. **Data Preprocessing** – Handled missing values, standardized column names, and converted dates into appropriate formats. New features such as Recency, Frequency, and Monetary value (RFM) were derived.
3. **Exploratory Data Analysis (EDA)** – Visualized purchase frequency, distribution of monetary values, and customer segmentation. This helped in understanding patterns in customer spending.
4. **Feature Engineering** – Extracted meaningful features (average transaction value, frequency of purchases, total transaction value, etc.) to strengthen model performance.
5. **Model Selection** – Applied regression-based algorithms (Linear Regression, Random Forest Regressor) to predict continuous LTV values.
6. **Model Training & Testing** – Split the dataset into training and testing sets to evaluate generalization. Hyperparameters were tuned for optimal accuracy.
7. **Model Evaluation** – Used metrics such as R^2 Score, Mean Squared Error (MSE), and Mean Absolute Error (MAE) to measure performance.
8. **Results** – The model successfully predicted LTV values for customers, allowing future transaction forecasting and customer segmentation.
9. **Final Deliverables**: A Google Colab notebook with preprocessing, EDA, model building, visualizations, and a CSV file of predicted Customer Lifetime Value.

Conclusion

The project demonstrated how **predictive analytics** can be applied to customer data to estimate future value. By using Google Colab and the demo_transaction dataset, we were able to preprocess data, engineer features, and build a robust regression model for LTV prediction. The results highlight the importance of customer segmentation and forecasting in business decision-making. This project can be extended further by incorporating advanced models like XGBoost, time-series forecasting, or deep learning for improved accuracy.