A logo of a company

Description automatically generated A close-up of a logo

Description automatically generated

**Joint Tech Internship Community Program**

**Assignment: Predicting Customer Lifetime Value (CLTV) in E-Commerce**

**Problem Statement:**

An e-commerce company wants to predict the Customer Lifetime Value (CLTV) of its users based on various factors such as purchase history, customer demographics, website engagement, and customer service interactions. Your task is to develop a machine learning model that predicts the CLTV for each customer given these attributes.

**Objective:**

Build a predictive model to estimate the Customer Lifetime Value (CLTV). Evaluate the model using appropriate metrics and provide insights into the factors that most influence CLTV.

**Dataset:**

You are provided with a dataset containing the following columns:

1. **CustomerID**: Unique identifier for each customer.
2. **Age**: Age of the customer.
3. **Gender**: Gender of the customer (Male/Female).
4. **Location**: The geographic location of the customer (e.g., Urban, Suburban, Rural).
5. **AnnualIncome**: The annual income of the customer (in USD).
6. **PurchaseFrequency**: The average number of purchases made by the customer per month.
7. **AverageOrderValue**: The average value of the orders placed by the customer (in USD).
8. **TotalPurchases**: The total number of purchases made by the customer.
9. **CustomerTenure**: The number of years the customer has been with the company.
10. **WebsiteEngagement**: The level of engagement the customer has with the website (e.g., Low, Medium, High).
11. **CustomerServiceInteractions**: The number of times the customer has interacted with customer service.
12. **MarketingSpend**: The total amount spent by the company on marketing targeted to the customer (in USD).
13. **CLTV**: The target variable representing the Customer Lifetime Value (in USD).

**Tasks:**

1. **Data Exploration and Preprocessing:**
   * Load the dataset and perform initial exploration to understand the data.
   * Identify and handle any missing values appropriately.
   * Convert categorical variables into numerical ones using techniques such as One-Hot Encoding or Label Encoding.
2. **Feature Engineering:**
   * Perform feature scaling (e.g., Standardization or Normalization) on continuous variables.
   * Create new features if relevant, such as average purchase value or engagement score.
   * Use dimensionality reduction techniques (e.g., PCA) to reduce the feature space if necessary.
3. **Model Building:**
   * Split the dataset into training and testing sets (e.g., 80-20 split).
   * Train different regression models (e.g., Linear Regression, Decision Trees, Random Forest, Gradient Boosting).
   * Perform hyperparameter tuning using techniques like Grid Search or Random Search to optimize model performance.
4. **Model Evaluation:**
   * Evaluate your models using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
   * Compare the performance of different models and select the best one.
   * Analyze feature importance to understand the most significant factors contributing to CLTV.
5. **Insights and Recommendations:**
   * Provide insights based on your model analysis regarding the factors that influence CLTV.
   * Suggest actionable strategies for the company to increase customer retention and maximize CLTV based on the findings.
6. **Documentation:**
   * Document your process, including data exploration, preprocessing steps, model selection, and evaluation.
   * Include visualizations where necessary to support your findings.

**Deliverables:**

* A Jupyter notebook (or Python script) with the entire workflow.
* A report summarizing your findings, including the model's performance and recommendations for optimizing CLTV strategies.

This assignment problem focuses on predicting Customer Lifetime Value (CLTV) in the e-commerce domain, a linear regression problem. The problem involves data preprocessing, feature engineering, model training, and evaluation, with a focus on understanding the factors that drive customer value and making data-driven decisions to optimize customer retention and profitability.

Top of Form

Bottom of Form