**Project Plan: Customer Lifetime Value (CLV) Prediction Model**

**1. Project Introduction**

This project involves the development of a predictive model aimed at calculating the **Customer Lifetime Value (CLV)**. CLV is a critical metric in understanding the long-term value of customers, helping businesses make informed decisions regarding customer acquisition, retention, and marketing strategies.

The project involves leveraging machine learning techniques to predict CLV based on customer data, focusing on four critical features: 'Lifespan (years)', 'Average Yearly Revenue', 'Churn Rate', and 'Discount Rate'.

**2. Project Objectives**

The primary objective of this project is to accurately predict the **Customer Lifetime Value** using the available customer data. Accurate predictions can help guide customer-oriented strategies, enabling the business to prioritize high-value customers, allocate resources efficiently, and improve overall customer retention and satisfaction.

**3. Data Preparation**

**Data Creation:** The project begins with the creation of a demo dataset using a formula for CLV. The CLV calculation typically uses parameters like frequency of purchase, gross margin, lifespan, and churn rate. It's essential to explain each term used in the formula:

* Frequency of Purchase: The number of repeat purchases a customer makes over a specified period.
* Gross Margin: The profit made from the customer after deducting the cost of goods sold.
* Lifespan: The projected length of the customer relationship.
* Churn Rate: The percentage of customers who end their relationship with a company in a given period.

**Feature Selection:** Implementing a strategy for selecting suitable columns from the dataset through feature engineering. The focus is on columns that directly or indirectly influence the CLV.

**4. Model Development**

**Model Development:** Utilizing the TF-IDF (Term Frequency-Inverse Document Frequency) Vectorization technique for feature extraction, followed by the application of a regression algorithm to predict the Customer Lifetime Value.

The selected regression algorithm will model the relationship between our chosen features ('Lifespan (years)', 'Average Yearly Revenue', 'Churn Rate', 'Discount Rate') and the target variable, the CLV.

**5. Output**

**Model Output:** Writing the CLV predictions into a Hadoop table named 'Customer Lifetime Values'. This table would act as a ready reference for various customer-centric strategic decisions and tactics.

**6. Evaluation and Testing**

**Model Evaluation:** Regularly assessing the model's accuracy and effectiveness using appropriate metrics for regression models, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared value.

**Model Improvement:** Continuously refining the model based on evaluation results and feedback. This iterative process ensures the model's adaptability to changing trends and customer behaviours.

**7. Deliverables**

**Model:** A predictive model capable of accurately calculating Customer Lifetime Value.

**Data:** A Hadoop table named 'Customer Lifetime Values', containing the output from the predictive model.

**Performance Reports:** Detailed reports outlining the performance and accuracy of the model.

**8. Benefits to the Business**

**Understanding CLV:** Accurate CLV prediction helps the business understand the potential value of its customers over the long term.

**Strategic Planning:** The CLV prediction can guide customer-centric strategies, focusing on customer acquisition, retention, and marketing efforts.

**Resource Allocation:** Helps in the efficient allocation of resources by prioritizing high-value customers.

**Revenue Forecasting:** Provides an estimate of the revenue to be generated by a customer over their relationship span with the company.

**Customer Retention:** Enables the creation of tailored strategies to enhance customer retention and satisfaction.