

# **Final Internship Project Report**

## **Celebal Technologies**

**Title: Customer Lifetime Value Prediction**

**Submitted by:**

Sudeshna Pendyala (Email: sudeshnapendyala11@gmail.com)  
CSI-Intern, Celebal Technologies.

---

### **1. Abstract**

This project aims to predict the Customer Lifetime Value (CLV) using historical online retail data. The CLV represents the total monetary value a business expects to earn from a customer throughout their relationship. This analysis helps businesses improve decision-making in marketing, sales, and customer retention strategies. Machine Learning techniques were used for prediction, and visual insights were presented using Tableau and Streamlit.

---

### **2. Problem Statement**

In the e-commerce domain, understanding customer behavior and predicting their future value is crucial. The project focuses on predicting the future value of a customer to a business based on their past purchase history and behavioral patterns.

**Goal:** Predict the monetary value (CLV) for each customer using recency and frequency metrics.

---

### **3. Dataset Description**

- **Dataset Source:** Kaggle - Online Retail II Dataset
  - **Files Used:**
    - online\_retail\_2009.xlsx
    - online\_retail\_2010.xlsx
  - **Key Features:**
    - Invoice, StockCode, Description, Quantity, InvoiceDate, UnitPrice, Customer ID, Country
-

## **4. Data Preprocessing & Feature Engineering**

- Combined 2009 and 2010 data.
  - Removed null and duplicate values.
  - Excluded cancelled invoices (starting with 'C').
  - Calculated RFM (Recency, Frequency, Monetary) metrics:
    - **Recency:** Days since last purchase
    - **Frequency:** Number of orders
    - **Monetary:** Total amount spent
- 

## **5. Modeling**

- **Features Used:** Recency, Frequency
- **Target:** Monetary Value (CLV)
- **Algorithms Tried:**
  - Linear Regression
  - Gradient Boosting
  - Random Forest Regressor

**Best Model:** Random Forest Regressor

- MSE: ~155,772
- $R^2$  Score: ~0.576

The model was saved using pickle as `clv_model.pkl` for deployment.

---

## **6. Deployment using Streamlit**

A Streamlit web app was developed to make predictions based on user input:

- Users input Recency and Frequency
- The model predicts the CLV instantly

**Key Features:**

- Clean UI with input form
- Visualization of customer trends
- Top customers by predicted CLV

- Recency vs CLV and Frequency vs CLV graphs

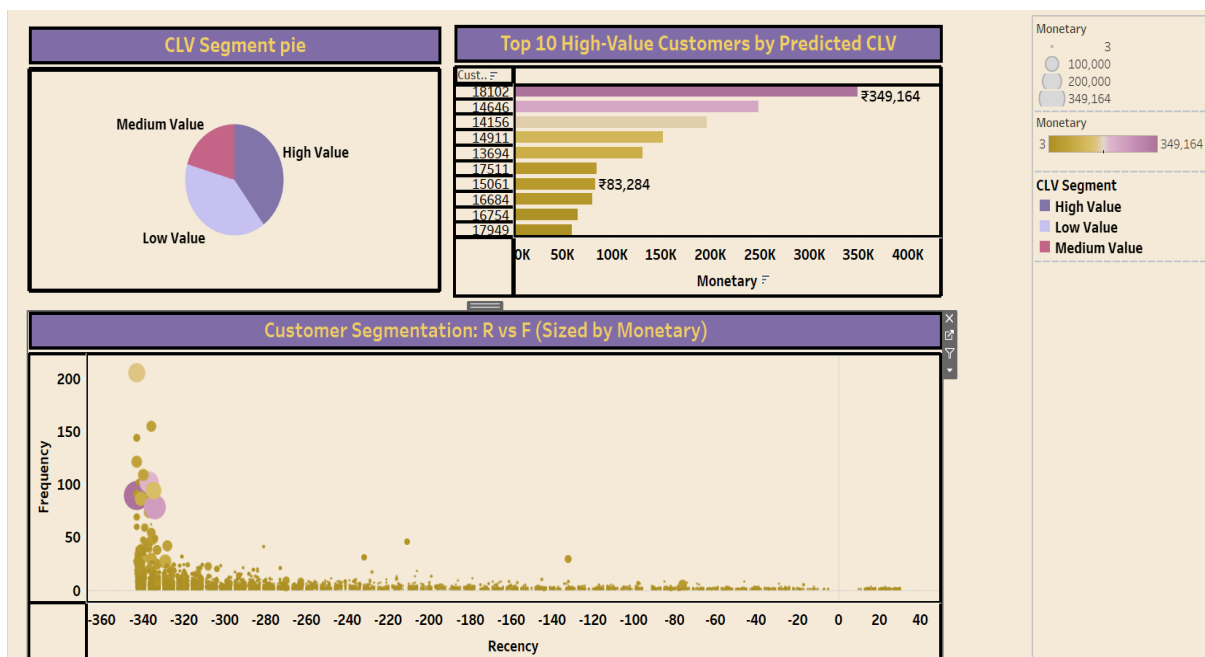
---

## 7. Tableau Dashboards

To support the analysis, Tableau dashboards were built on the same dataset:

### Customer Segmentation Dashboard

- RFM Score Distribution
- Customer categories: Loyal, At Risk, New, etc.



---

## 8. Conclusion & Future Work

This project successfully predicts Customer Lifetime Value using Recency and Frequency features. The deployed app and visual dashboards allow business users to make data-driven decisions.

### Future Enhancements:

- Include demographic or session data for improved predictions
  - Model customer churn and retention scores
  - Integrate with live transactional systems
-

## **9. References**

- Kaggle: <https://www.kaggle.com/datasets/lakshmi25npathi/online-retail-dataset>
- Scikit-learn Documentation
- Streamlit Docs
- Tableau Public: [Dashboard](#)