**Customer Lifetime Value Prediction Model**Project 2 — Internship Phase

**Introduction**

The objective of this project is to predict the lifetime value (LTV) of retail customers using historical transaction data. Accurate LTV estimates help businesses allocate marketing resources efficiently and improve customer retention strategies.

**Abstract**   
We analyzed 22,190 transactions made by 4,372 unique customers across 37 countries. After data cleaning, we engineered Recency, Frequency, Monetary (RFM) features and trained a Random Forest Regressor to predict customer LTV. Model evaluation achieved a Mean Absolute Error of 1763.98 and a Root Mean Squared Error of 10127.09. Customers were segmented into Low, Medium, and High value tiers for actionable insights.

**Tools Used**

Python 3.8, pandas, NumPy, scikit-learn, matplotlib, seaborn,

**Steps Taken**

1. Imported the dataset and resolved encoding issues.
2. Cleaned missing values and parsed invoice dates.
3. Created an Amount field and generated RFM metrics.
4. Trained a 200-tree Random Forest model to predict Monetary value.
5. Evaluated the model using MAE and RMSE.
6. Scored all customers, segmented them into Low, Medium, and High LTV tiers, and exported the results.

**Conclusion**

The Random Forest model effectively differentiates revenue potential among customers. High-value customers (top 33%) represent a substantial share of future revenue, indicating that retention campaigns should prioritize this segment. Future enhancements include incorporating additional behavioral features and experimenting with gradient boosting models to improve accuracy