Business Case Study – Amazon

PROBLEM –

How would you predict customer lifetime value (CLV) of Amazon shoppers? Provide a recommendation on how to increase the CLV to the VP of product.

1 CLARIFICATION

1.1 1. BUSINESS PROCESS

a. User Journey

1.2 2. BUSINESS REQUIREMENTS

a. Prediction Horizon - 1 Year

b. Prediction Point - Prediction based on the first month

c. Prediction Frequency - Realtime

1.3 3. DATA SOURCES

a. User (Name, Email, Address)

b. Device (Windows, Mac, iPhone, Android, IP Address)

c. Marketing (SEO, Email Notification, Push Notification)

d. Transaction (Payment Method, Product, Price, Quantity, Shipping, Address)

2 TECHNICAL SOLUTION:

2.1 1. EDA

a. Data Quality (Missingness, Rare Values, Outliers)

b. Univariate Analysis (Bar Charts, Histogram)

c. Correlation Analysis (Visuals like Line Charts and Scatter Plots, Pearson Statistics)

2.2 2. DATA PREPROCESSING

a. Data Cleaning

b. Scaling

2.3 3. Feature Engineering

a. **Aggregation** - Past X day total/avg/standard purchase $ / Quantity per Customer

b. **Numerical Encoding** - Past X day total/avg/standard purchase $/ Quantity per Product

c. **Address Fields** - Decomposition -> Zip Code, State, Country

2.4 4. Feature Selection

a. PCA
b. L1 regularisation regression

2.5 5. Training the Model

a. Model - Random Forest
b. Hyperparameter Tuning Depth, N Trees, Min Sample Split, Grid Search, Random Search, Hyper Opt

2.6 6. Model Evaluation

a. K-Fold Cross Validation
b. MAE

2.7 7. Product ionization

a. RĘST API - Model

i. Input -> Model -> Output
ii. Model caching

b. Storage

i. Model
ii. Inputs & Outputs

c. Monitoring

i. Model Performance
ii. Latency & Bugs

- *Model → Variable Importance → 5 to 7 signals → PDP*
- *First 4-week spending amount $ vs CLV*
- *Prime membership*
- *Ul Visibly / Email Campaign*