

Project Bootcamp AI Model for Customer Growth and Retention

Business Context

Bạn là **Data Scientist** tại một công ty **fintech / subscription / e-commerce**.

Doanh nghiệp đang gặp vấn đề:

- CAC tăng
- Churn cao ở nhóm khách hàng giá trị
- Campaign retention gửi đại trà → tốn chi phí, ROI thấp

Nên ưu tiên giữ khách hàng nào, bằng cách nào

- Giảm churn
- Tối đa hóa CLV
- Tránh over-treatment (spam khách hàng không cần giữ)

Dataset

Link:

<https://drive.google.com/drive/folders/1W13sZcd0cido1k5k1RuvHXjJzcvKQLic?usp=sharing>

- Transactions.csv
- Customers.csv
- Note:
 - Dataset này không có churn label sẵn.
 - Các bạn phải **định nghĩa churn** dựa trên inactivity window, và **giải thích vì sao định nghĩa đó hợp lý**.

Project Structure

1. Customer Value Foundations

Yêu cầu:

- Tính RFM
- Phân nhóm khách hàng:
 - High-value / At-risk

- New / Loyal / Hibernating
- Phân tích mối liên hệ:
 - RFM vs churn

Mục tiêu: **RFM có thể cung cấp phân tích ban đầu nhưng không đủ**

2. Churn Prediction as Classification

Churn = yes/no trong T tháng tới

Yêu cầu

- Define churn label (30 / 60 / 90 ngày)
- Feature engineering:
 - RFM
 - Frequency trend
- Train model (Logistic / Tree)

Deliverable

- AUC, Precision–Recall
- Confusion matrix
- Top features

3. Churn via BG-NBD

Yêu cầu

- Fit BG-NBD
- Estimate:
 - $P(\text{alive})$
 - Expected #transactions

So sánh bắt buộc

- Churn label vs $P(\text{alive})$

4. Churn with Survival Analysis

Time-to-churn modeling

Yêu cầu

- Define duration & event
- Fit CoxPH / Weibull
- Predict:

- Survival curve
- Expected remaining lifetime

5. CLV Modeling

Future value estimation based on customer lifetime

Approach 1 – BG-NBD + Gamma–Gamma

Yêu cầu

- Fit BG-NBD
- Predict:
 - Expected number of future transactions
 - Probability customer is alive
- Fit Gamma–Gamma
- Compute:
 - Expected monetary value
 - CLV over time horizon T

Approach 2 – Survival Analysis + Gamma–Gamma

Yêu cầu

- Use survival model output
- Predict:
 - Survival curve
 - Expected remaining lifetime
- Fit Gamma–Gamma
- Compute:
 - Time-dependent CLV

Final Business Question

Học viên phải trả lời:

Nếu ngân sách retention chỉ đủ giữ **20% customer base**, thì **chọn 20% này theo cách nào?**

3 chiến lược cần so sánh

1. High churn probability (classification)
2. Low P(alive) (BG-NBD)

3. High CLV × High churn risk (Survival-based)

Final Deliverables

Technical

- Github Code repo + README result presentation
- Clear pipeline: Data → Model → Decision → Outcome

Bonus (optional)

Mục tiêu: mô phỏng cách **churn & CLV model được dùng trong production**, không chỉ nằm trong notebook.

/score_customer

Unified customer scoring

Input

```
{  
  "customer_id": "C123"  
}
```

Output

```
{  
  "churn_probability": 0.42,  
  "p_alive": 0.63,  
  "expected_remaining_lifetime": 5.8,  
  "clv_bgnbd": 1250,  
  "clv_survival": 980  
}
```

Tổng hợp output của:

- Classification churn

- BG-NBD
- Survival analysis
- CLV modeling

/predict_churn

Churn prediction as classification

Input

```
{  
  "customer_id": "C123",  
  "horizon_days": 60  
}
```

Output

```
{  
  "churn_probability": 0.42,  
  "churn_label": "high_risk"  
}
```

/predict_survival

Time-to-churn prediction

Input

```
{  
  "customer_id": "C123"  
}
```

Output

```
{
```

```
"survival_curve": [  
  {"day": 30, "prob": 0.82},  
  {"day": 60, "prob": 0.65},  
  {"day": 90, "prob": 0.47}  
],  
"expected_remaining_lifetime": 5.8  
}
```

/estimate_clv

CLV estimation endpoint

Input

```
{  
  "customer_id": "C123",  
  "method": "bgnbd"  
}
```

or

```
{  
  "customer_id": "C123",  
  "method": "survival"  
}
```

Output

```
{  
  "method": "bgnbd",  
  "clv": 1250,  
  "horizon_months": 12  
}
```

/rank_customers_for_retention

Retention prioritization

Input

```
{  
  "top_k": 100,  
  "strategy": "high_clv_high_churn"  
}
```

Output

```
{  
  "strategy": "high_clv_high_churn",  
  "customers": [  
    {  
      "customer_id": "C123",  
      "churn_probability": 0.42,  
      "clv": 1250,  
      "priority_score": 0.78  
    }  
  ]  
}
```