May 19, 2025

CUSTOMER CHURN PREDICTION

MCI Data Analyst Intern test assignment

Prepared and coded by Nguyễn Khánh Toàn

May 19, 2025

TABLE OF CONTENTS

Prepared and coded by Nguyễn Khánh Toàn

May 19, 2025

CHURN PREDICTION PROBLEM

Customer churn prediction is the task of identifying customers who are likely to stop using a company's services. It is crucial for businesses to take early action and improve customer retention strategies.
This problem is highly relevant in many industries, such as:

- Telecommunications
- Banking and Finance
- E-commerce
- Subscription services
- Insurance

By predicting churn, companies can proactively offer promotions, support, or engagement efforts to retain valuable customers.

Prepared and coded by Nguyễn Khánh Toàn

IMPACT OF CHURN RATE ON STAKEHOLDERS

- Customer:
  - A rising churn rate could highlight service issues or pricing misalignment, prompting the need for better retention strategies.
- MCI:
  - Churn reflects their experience and the availability of better alternatives in the market.
- Investors and decision-makers:
  - Churn rate trends offer valuable signals about business health and future revenue stability.

DATASET OVERVIEW

The dataset consists of customer data from a telecom company, aiming to predict churn behavior. It includes the following types of information:

- Geographic and account information:
    - State, Area code, Account length
- Service plans:
    - International plan, Voice mail plan, Number vmail messages
- Usage statistics:
    - Daily, evening, night, and international call minutes, number of calls, and corresponding charges:
        - e.g. Total day minutes, Total eve calls, Total intl charge, etc.
- Customer behavior:
    - Customer service calls (number of support interactions)
- Target label:
    - Churn (whether the customer has churned: True/False)

This dataset enables both behavioral and usage-based analysis to identify patterns associated with customer churn.

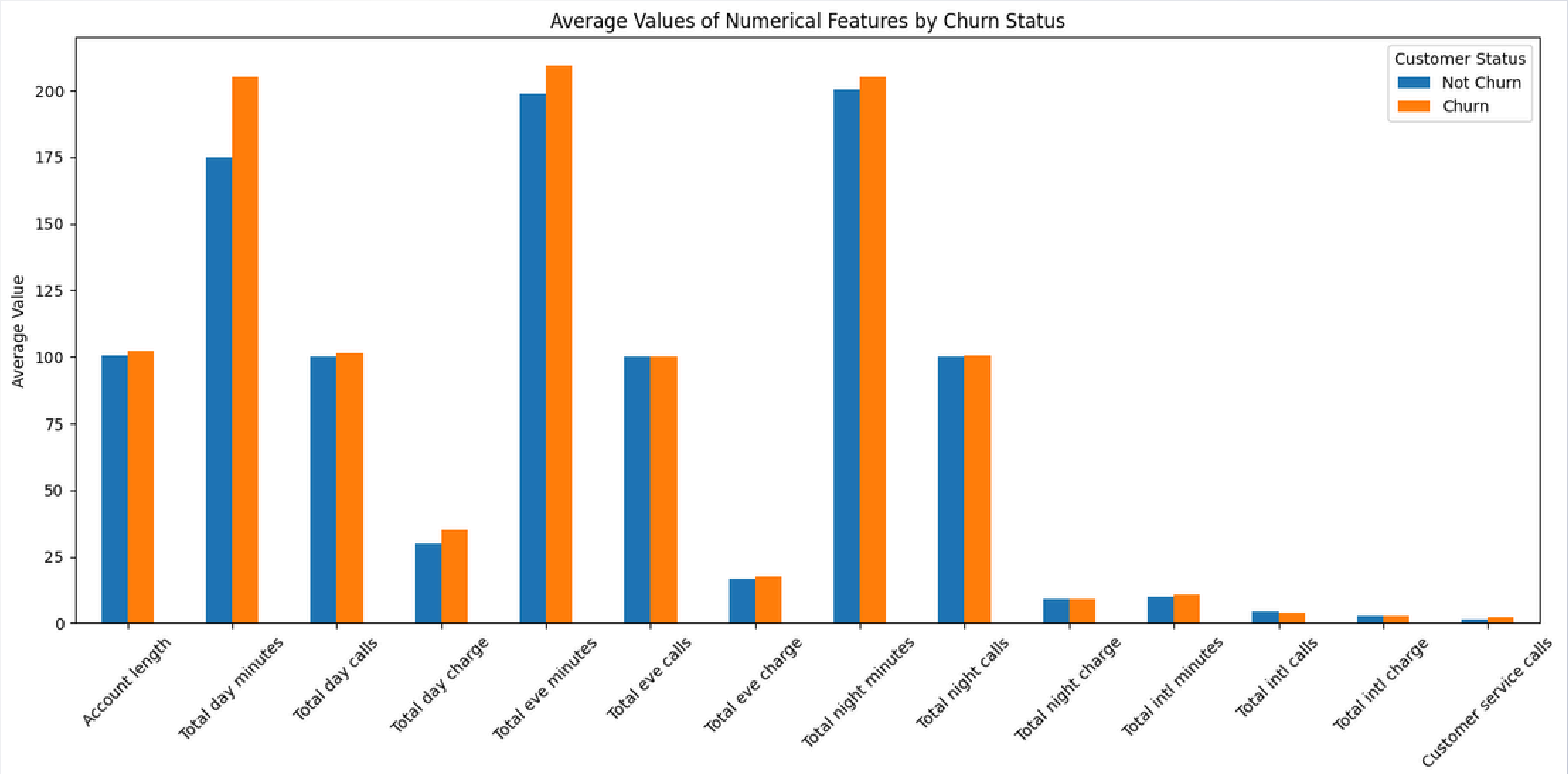
May 19, 2025

CHURN VS. NON-CHURN: CUSTOMER TRAITS

- **Churn**

They tend to have higher values for:

- Account length
- Total day minutes
- Total day charge
- Total evening minutes
- Total night minutes
- Most of them do not use the Voice Mail Plan.
- They make many customer service calls.



Prepared and coded by Nguyễn Khánh Toàn

CHURN VS. NON-CHURN: CUSTOMER TRAITS

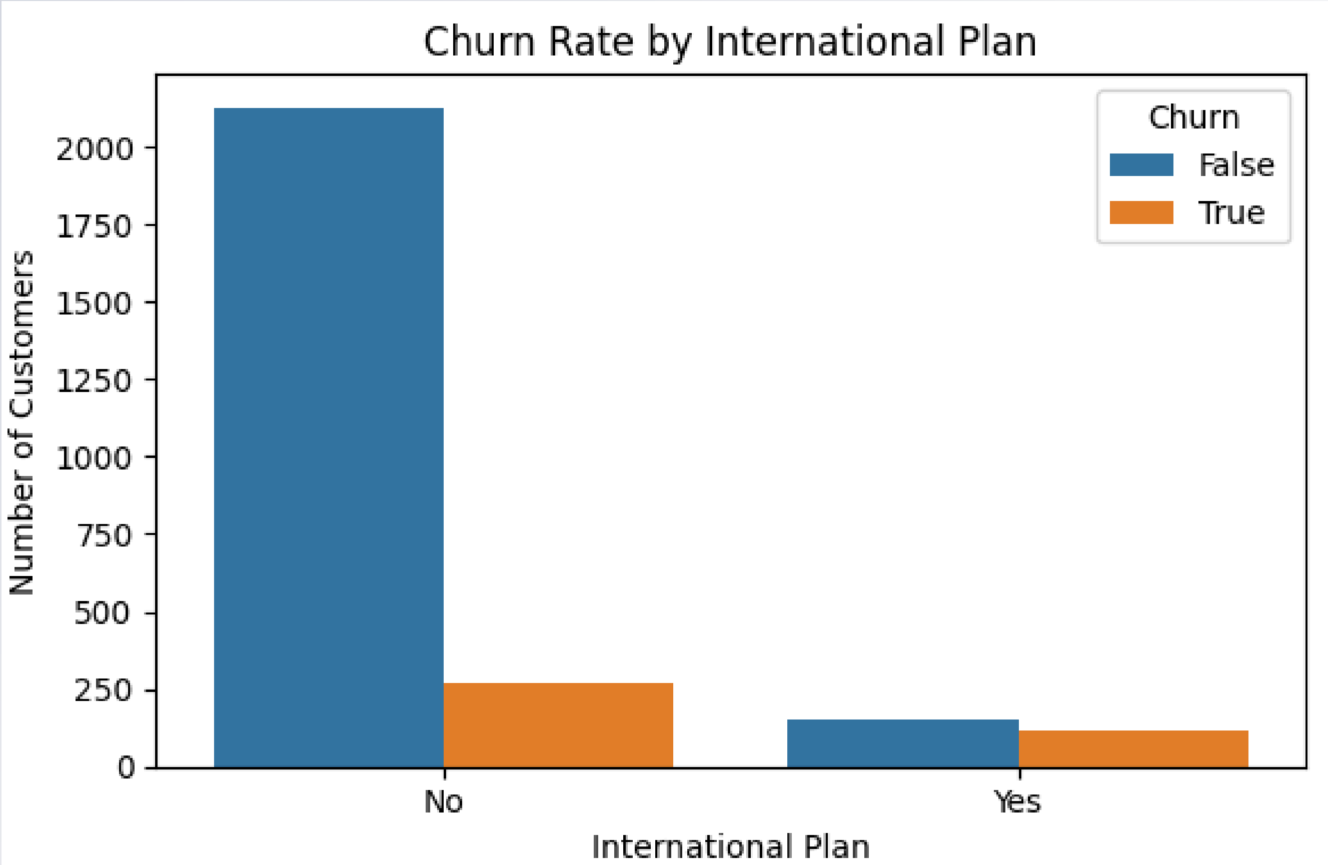
- **Churn**

Interpretation:

- These customers are heavy users, especially during the day, which results in high charges. If the perceived value doesn't match what they're paying, it may lead to dissatisfaction.
- Frequent customer service calls suggest problems or dissatisfaction. If those issues are not resolved effectively, customers are more likely to leave.

CHURN VS. NON-CHURN: CUSTOMER TRAITS

- **Non-churn**



- Most of them use don’t international plan.

CHURN VS. NON-CHURN: CUSTOMER TRAITS

- **Non-churn**

Interpretation:

- These customers likely incur lower overall charges, especially by avoiding international calls, which are typically more expensive.
- Their simpler service usage patterns may lead to fewer complications and fewer negative experiences, contributing to greater satisfaction and loyalty.

MODELING APPROACH AND RESULTS

- **Modeling Approach**

For binary classification churn prediction, several machine learning models can be implemented: Logistic Regression, Tree-based model likes Decision Trees, Random Forest, Gradient Boosting Machines, Support Vector Machine. In this project, I implemented Logistic Regression, Random Forest, LightGBM.

- **Features input**

Beside original provided features, there are synthetic features created in feature engineering

- Average call durations: avg_day_call_duration, avg_eve_call_duration, avg_night_call_duration, avg_intl_call_duration
- Proportion of call activity : day_ratio, eve_ratio, night_ratio, intl_ratio.
- Summarize overall call activity: total_minutes, total_calls, intl_calls_ratio.

MODELING APPROACH AND RESULTS

- **Logistic Regression results and features important.**

Model results:

- In 95 churn customer, model can correctly predict for 74 customer, with recall 77.9%. while in 572 no churn customer, 75.17% of them truly predicted.

Feature importance: The plot displays the top 5 most important features ,ranked by how much they contribute to improving the model's predictions of customer churn.

- Day_ratio: A higher proportion of daytime calls is associated with lower churn
- Night_ratio: In contrast, A higher proportion of night time calls is associated with higher churn
- Total day minutes: Higher usage during the day may lead to higher rate to churn.
- Total day charge: Higher charges (correlated with more minutes) are associated with higher churn risk.
- Intl_calls_ratio: A higher share of international calls slightly reduces churn.

MODELING APPROACH AND RESULTS

- **Random Forest results and features important.**

Model results:

- Model can precisely predict 78.95% churn customer and 99.3% no churn customer.

Feature importance:

- Total charge: Overall charges seem to be the most critical factor—high billing may increase churn risk.
- Customer service calls: Frequent contact with customer service can indicate issues or dissatisfaction, likely leading to churn.
- Total day minutes: High day-time usage may reflect higher bills or certain user behavior linked to churn.
- Total_minutes: Total call time (across all periods) reflects overall engagement; could correlate with cost or service usage habits.
- Total day charge: Like total day minutes, but directly related to cost—higher charges may drive churn.

MODELING APPROACH AND RESULTS

- **LightGBM results and features important.**

Model results:

- In 95 churn customer, model can predict true for 13 customer, with recall 0.86. while in 572 no churn customer, 98% of them truly predicted.

Feature importance:

- State_freq: Frequency or encoding of the customer's state.
- Avg_intl_call_duration: Average length of international calls. Longer calls may suggest business or high-value users with different churn behavior.
- Total_minutes: usage level in minute, reflect engagement with the service.
- Total charge: Total cost incurred by the customer. High charges might lead to dissatisfaction and churn
- Intl_calls_ratio: The proportion of international calls relative to total calls. A strong indicator—perhaps churners use more/less international calls.
-

May 19, 2025

RETENTION STRATEGIES BASED ON QUALITATIVE & QUANTITATIVE ANALYSIS

- **Quantitative Actions**

  These are actions based on numerical data and measurable outcomes.
  - Build churn prediction models using customer activity, complaint records, or billing history.
  - Segment users into groups (like heavy users or occasional users) to offer more suitable retention tactics.
  - Try A/B testing to compare different offers and see which one encourages customers to stay.
  - Use RFM analysis to spot those who recently reduced their activity and may churn.
  - Estimate customer lifetime value to focus on keeping the most valuable users.
- **Qualitative Actions**

  These are actions based on non-numerical insights such as opinions, experiences, or observations.
  - Ask customers why they leave through exit surveys and try to find common patterns.
  - Check social media or call center feedback for early signs of dissatisfaction.
  - Reach out personally to loyal or high-value users who may churn soon.
  - Look at competitors to understand what they’re offering that attracts your customers.
  - Create loyalty programs to reward long-term users and keep them engaged.

Prepared and coded by Nguyễn Khánh Toàn

May 19, 2025

THANK YOU !

There may still be some shortcomings in this work, and I sincerely welcome any feedback for improvement.

Prepared and coded by Nguyễn Khánh Toàn