INTRODUCTION

1.1 Introduction

In the fast-paced growing environment, retail sales transform business model into online sales to speed up transaction procedure to keep up with the digital trend. According to Ikhlass Boukrouh (2025), this is known as electronic commerce (e-commerce) in which business transactions are conducted on the internet. This transformation brings general challenges such as high competition between e-commerce platforms, high return and refund rate from customers. This leads to customer churn and the trend keeps growing. Based on the statistics from Jack M. Germain (2023), the 42% if B2C companies are churning 3% or more, followed by another 16% that churns 4% or more. Most of the time, people decide to cease using a service after becoming increasingly unhappy with it over time. Customer churn happens when customer has decided to stop using the service. This reduces the brand loyalty, increasing future customer acquisition cost and revenue down, leading to business loss (Daniyal Asif, 2025). Churn prediction serves as a guidance for management team to refine marketing strategies, ensuring the enhanced approach fits the marketing objectives, leading to reduction in revenue (Sulim Kim, Heeseok Lee, 2022). However, few studies have implemented the churn analysis because it is difficult to define who the churners are in e-commerce. Research shows that it might cost five to twenty-five times more to acquire a new customer than to maintain an existing one, hence reducing churn is economically vital (Daniyal Asif et al, 2025). Churn evaluation is required to find out the root cause behind the customer churn. The churn evaluation is conducted in a monthly basis, observing the trends over month. E-commerce management able to adjust the marketing campaign plan, tailoring to the user needs, hence reducing the risk of customers getting churned (Nagaraj et al, 2025).

1.2 Problem Background

Churn prediction allows the e-commerce management to identify whether the customers are likely to stop using the e- commerce platform or not, allowing for preventive intervention. Without addressing customer churn, e-commerce has higher tendency to lose the

current customers, leading to lesser commission from transaction between customers and sellers. At the same time, it would be increasing future customer acquisition costs through marketing, advertisements, and promotions. Therefore, conducting research is necessary to reveal reasons that caused customers lose interest in future engagement on the platform. Current method would use manual check to identify whether the customer is churn or not. The moment to identify whether the customer is churn or not, the customer already churn. With the use of predictive model, marketing able to find out a list of customers that has sign to stop using the platform early. Marketing able to implement early prevention strategies such as free shipping/discount vouchers, lucky draw contest to keep these customers from leaving the platform. E-commerce revenue would be increasing in proportion to increased transaction.

There is a form used to collect data, due to lack of information. Demographic and conceptual data is provided at the beginning, but environmental and behavioral data don't have all. The steps to identify the churning customers for the case study is as follows. First, the determined churn customer is identified by group. Then, the determined groups are assigned with goal index. Next, the pattern is extracted using call tree method. The churn customers are revealed. They are officers and engineers. Their similarities are having Persia Insurance as their major business. Due to dissatisfaction towards the service, they decided to leave (Nagaraj P et al., 2023).

1.3 Problem Statement

E-commerce Customer daily active time has been gradually reducing. Reducing active time indicates that the customer might be leaving the e-commerce platform in the future. The number of customers would be reducing gradually, leading to lesser transactions to be made in the e-commerce platform. The seller would be directly influenced followed by the e- commerce platform. Seller unable to sell the products to the customer. E-commerce platform would gain lesser commission from overall transactions. By increasing e-commerce customer active time, customer would have higher tendency to make more transactions in e-commerce platform. This requires customer daily active period data, total time spent on the platform, purchase history data to train a model that predicts the tendency of customer churn. For the current solution, management predict the users that is frequently active in the platform would be purchasing more items. In the fact that customer is active but without purchasing items. This gives a

contradiction point that customer active time does not directly influence the number of customer transactions.

1.4 Research Question

The research question would be focusing on 3 different steps to conduct the churn analysis. The question starts from data preprocessing, followed by identification of churn attribute, at last visualizing the facts using statistical approach.

a) What are the steps to preprocess the customer churn dataset?
b) How does the known characteristics affect the customer churn rate?
c) How does the predicted result give insights on the customer churn?

1.5 Research Aim

To develop and enhance a predictive model that uses Random Forest algorithm, by labelling the customer as churn or not churn, provide actionable insights in improving the sales revenue and customer active time.

1.6 Research Objectives

The objectives of the research are:

(a) To preprocess the customer churn prediction data, leading to cleaned data for model training.

(b) To identify the key relevant attributes that affects the customer churn rate using Correlation coefficient matrix.

(c) To develop a machine learning model using Random Forest Algorithm that predicts the potential churning customers, visualizing the results in dashboard.

1.7 Research Scope

The scopes of the research are:

(a) The customer churn dataset will be collected from Kaggle Open Data Source. (https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data)

(b) Key Customer Lifetime Value (CLV) Variables such as Purchase Amount, Promo Code Used, Category, Previous Purchases, Purchase Date, Customer ID, and Churn will be used to calculate the impact towards customer churn pattern.

(c) The study would be focusing on e-commerce industry.

(d) The research direction would be predictive analysis.

(e) Python programming language would be used to preprocess the data.

(f) Random Forest algorithm would be used to predict the potential churning customers.

(g) Power BI would be used to visualize the data.

1.8 Research Significance

Random Forest algorithm implemented in the study improves the accuracy of churn prediction when the dataset is balanced. The Feature Importance able to identify the most important attribute that affects the customer churn rate. The identified attribute would be used as the metric to train the model recognizing the customer as churn. When the dataset is balanced, it aids in random sampling, which shuffles the training dataset. By each time the dataset is shuffled and trained, the model gets improved by time.

Random Forest require iterative training to get the model. Each iterative training with different sampling of dataset will give different result. Different results combine together forming a new model with higher accuracy. Each iterative training will give the model output. When the model output is then combined with the existing output, the overall accuracy improves. Iterative training consumes larger computational resource, as the training output count increases. More capital would be required investing in hardware to improve the efficiency of obtaining outputs. By using the current dataset, the longer the training count, the higher the accuracy of identifying customer with churn signals.

The research will flag up the customers that will start to churn. It alerts the ecommerce that the customers are going to churn, allowing them have time to make preventive measures to retain the customers. The diverse attributes would be helping managements to detect the signs the customer is going to churn. Hence, identifying the key attributes leading to customer churn is vital, as it indicates the hidden reason that the customer gives up on continuous support in the future.

1.9 Thesis Structure

The thesis consists of 5 chapters. Introduction is the first chapter to be illustrated, followed by second chapter, which is Literature Review. Then, Research Methodology would be fully addressed in Chapter 3. Random Forest algorithm-based model to be implemented on

Chapter 4. At the final chapter, a conclusion would be including the insights found across previous chapters.

1.10 Summary

This chapter introduces the problem background, problem statement related to customer churn, which is under e-commerce domain. With the support of problem background and problem statement, the research goals and objectives are focused on e-commerce domain. The research scope highlights the metrics to be used from data preparation to model building, followed by visualizing insights using dashboard. The research significance denotes the importance of conducting the research, expecting the insights found in the result aid in customer churn reduction.

The next chapter would be literature review, which would be comparing the existing methodological approaches, highlighting the strengths and limitations on each solution. The research methods mentioned in previous studies need to analyze deeper. At the same time, identifying the unanswered questions in previous research and highlight the areas that requires further research. The study would be helpful in determining the best model for my e-commerce use case.

References

Daniyal Asif, M. S. (2025). A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results in Engineering*, 104629. doi:https://doi.org/10.1016/j.rineng.2025.104629

Das, S. (2025). *Calculating Customer Lifetime Value and Churn using Beta Geometric Negative Binomial and Gamma-Gamma Distribution in a NFT based setting.* arXiv. Retrieved from https://arxiv.org/abs/2501.04719

Germain, J. M. (28 July, 2023). *Subscription Commerce Merchants Innovate Amid Rising Churn Rates*. Retrieved from E-commerce Times:

https://www.ecommercetimes.com/story/subscription-commerce-merchants-innovate-amid-rising-churn-rates-177730.html

Ikhlass Boukrouh, A. A. (2024). Explainable Machine Learning Models Applied to Predicting Customer Churn for E-Commerce. *IAES International Journal of Artificial Intelligence*, 286-297.

Kim, S., & Heeseok, L. (2022). Customer Churn Prediction in Influencer Commerce: An Application of Decision Trees. *Procedia Computer Science*, 1332-1339.

Nagaraj P, M. V. (2023). E-Commerce Customer Churn Prediction Scheme Based on Customer Behaviour Using Machine Learning. *International Conference on Computer Communication and Informatics (ICCCI)*, (pp. 1-6). Coimbatore, India.