

Royal University of Bhutan

**DAT 205 Data Science Project I**

Submitted By:

Cheki Dorji[07230004]

Gyembo Wangchuk[07230028]

Rinzin Dema[07230076]

Tshering Yangden[07230084]

**Module Tutor: Tenzin Wangchuk**

Table of Contents

[Introduction 1](#_Toc193658474)

[Dataset Description 1](#_Toc193658475)

[Problem Statement 2](#_Toc193658476)

[Aims 2](#_Toc193658477)

[Objective 2](#_Toc193658478)

[Goals 2](#_Toc193658479)

[Methodology 3](#_Toc193658480)

[Timeline 4](#_Toc193658481)

[Expected Outcome 5](#_Toc193658482)

[Conclusion 6](#_Toc193658483)

[Reference 7](#_Toc193658484)

**Title of Project:** Predicting Customer Churn in Banking

### Introduction

Customer churn is a globally rising issue faced by banking companies as it directly leads to loss of profit and affects long term growth of business. By understanding the hidden causes of customer churn, banks can develop strategies to enhance overall services quality and improve customer retention. Traditionally, banks have relied on basic transaction data and customer feedback to understand customer behavior. However, with the availability of large data sets, now banks can use advanced data analytics techniques to predict customer churn with higher accuracy. Moreover, machine learning algorithms have proven to be highly effective in identifying patterns within complex data compared to traditional analysis methods. These models can analyze a number of factors including customer demographics, account activity, transaction history, and customer services interaction to predict the probability of customer churn (Peng et al., 2023).

This project uses machine learning techniques to predict customer churn in the banking sector. By predicting potential customers, banks can develop targeted retention strategies, reduce financial losses and improve customer satisfaction. This data-driven approach offers knowledgeable insight for better decision making and enhanced customer retention. By integrating predictive models with interactive visualizations, banks can develop personalized retention strategies and competitive advantage in the marketplace.

### Dataset Description:

This dataset was uploaded by Saurabh Badole in Kaggle. It contains information about bank customers and their churn status. This dataset is suitable for exploring and analyzing factors affecting customer churn in banking industries and for building predictive models to identify customer churn. This dataset contain data of 10,000 customer and 14 attributes which include RowNumber, CustomerID, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary and Exited.

Data source: <https://www.kaggle.com/datasets/saurabhbadole/bank-customer-churn-prediction-dataset>

### Problem Statement: Customer Churn is one of the growing issues in the banking industry. Churn happens when customers stop using a bank’s services due to competitors or dissatisfaction. Understanding the reason behind customer churn is essential to formulate effective retention strategies. Currently, many banks lack a systematic approach to prevent and predict churn, resulting in financial losses. This project aims to fill this gap by developing predictive models to predict customer churn and provide actionable insight for retention strategies (Guliyev &Yerdelen Tatoğlu, 2021)

### Aims:

* To analyze key factor machine learning models capable of predicting customer churn in the banking system.
* To create data visualization that will help us understand churn trends.
* To analyze key factors leading to customer churn in the banking system.

### Objective:

* Collect, preprocess and analyze customer churn datasets.
* Use Exploratory Data Analysis to identify trends in churn behavior.
* Build a machine learning model to predict customer churn in the banking system.
* Evaluate models using performance metrics.
* Develop an interactive web-based visualization dashboard.

### Goals:

* To improve customer loyalty strategies through data-driven decision-making.
* Provide interpretable and scalable churn prediction systems for banking institutions.
* Help banking industries to reduce financial losses caused by customer churn.

### Methodology

1. Data Preprocessing

* Handling missing values and outliers.
* Normalization
* Encoding categorical data.

Technology to be used: Python with Panda Library using Jupiter Notebook.

1. Data Visualization

* Churn distribution
* Customer demographics analysis
* Customer behaviour
* Churn by region
* Customer services interaction analysis
* Correlation analysis
* Customer segmentation for churn analysis
* Interactive dashboard

Technology to be used: Python with visualization libraries like Matplotlib and seaborn.

1. Model Selection

* Logistic regression
* Decision tree
* Random forest analysis

Technology to be Used:

1. Scikit-learn(sklearn.linear\_model.LogisticRegression): For Logistic regression model implementation.
2. Scikit-learn(sklearn.tree\_DecisionTreeClassifier): For Building decision trees.
3. Scikit-learn (sklearn.ensemble.RandomForestClassifier): For building a random forest model
4. Model Evaluation

* Cross-validation and performance matrix

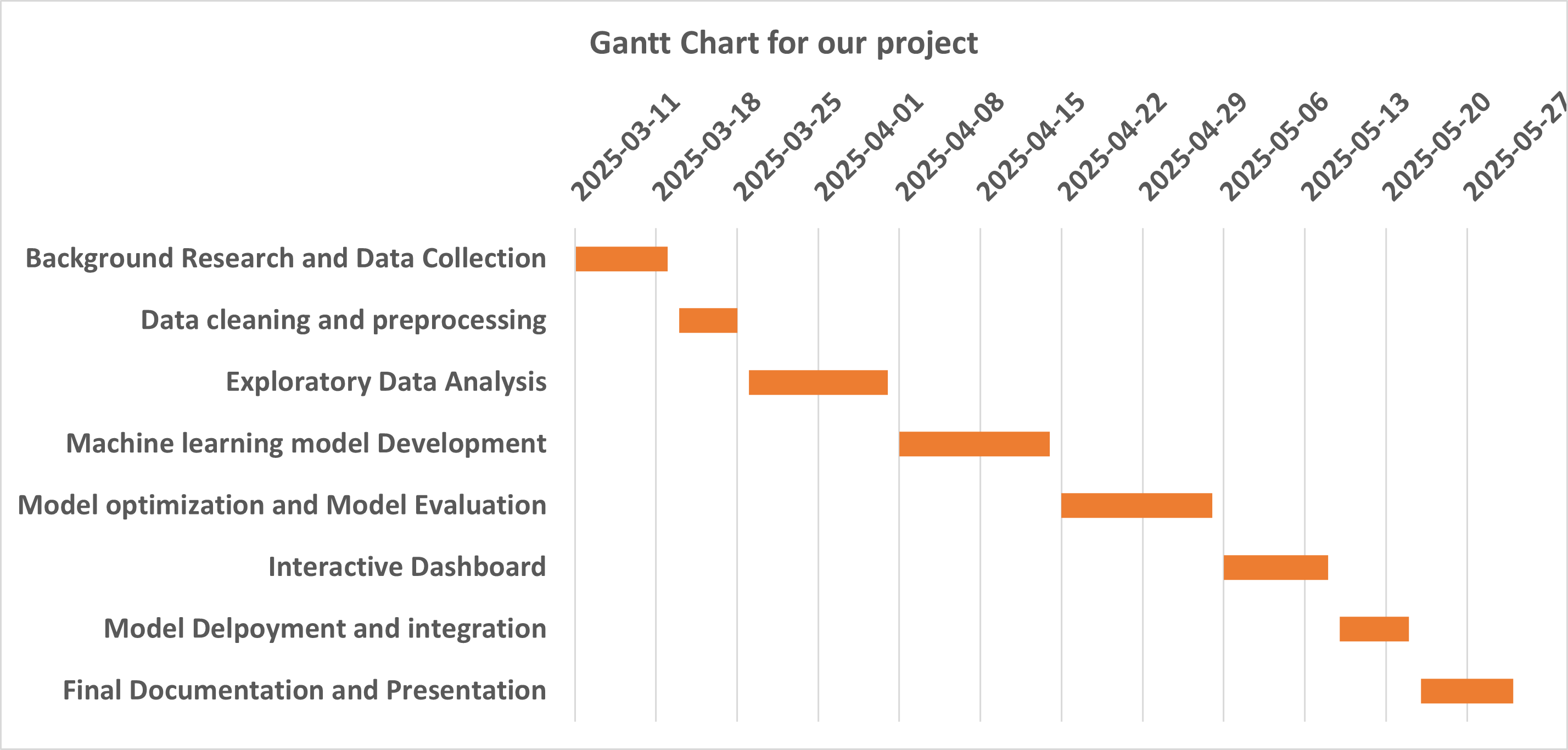
Technology to be used:

1. Scikit-learn(sklearn.model\_selection.cross\_val\_score, KFold): For cross-validation
2. Scikit-learn(sklearn.metrics): For classification reports and confusion matrix.
3. Dashboard development

* A web-based dashboard using a front-end framework.

Technology to be used: Streamlit: for interactive python-based dashboard

### Timeline



### Expected Outcome

* A machine learning model capable of predicting customer churn with high accuracy.
* Identification of key factors influencing churn.
* Recommendations for banks to reduce churn and improve customer retention.

### Conclusion

In the banking industry, losing customers is a major problem that results in monetary losses and reduced customer loyalty. In order to overcome the problem, this research uses machine learning approaches to anticipate possible client attrition based on important things like transaction records, account activity, and consumer demographics. Bank may implement proactive measures to improve their services and retain clients by spotting trends in customer behavior

The project follows a structured approach, starting with data collection, preprocessing and exploratory data analysis to identify trends, the project takes a methodical approach. Then in order to anticipate turnover with high accuracy, it creates machine learning models like logistic regression, decision trees, and random forests. The creation of an interactive web-based dashboard which gives banks clearer visual insights for improved decision- making, is the last step.

Banks may improve client retention tactics, lower churn rates, and boost profitability by putting this approach into place. In addition to reducing expenses, this predictive model strengthens ties with clients, guaranteeing stability and long-term growth in a market that is becoming more and more competitive.

### Reference

Guliyev, H., & Yerdelen Tatoğlu, F. (2021). Customer churn analysis in the banking sector: Evidence from an explainable machine learning model. *Journal of Applied Microeconometrics, 1*(2), 85–99.

Peng, K., Peng, Y., & Li, W. (2023). Research on customer churn prediction and model interpretability analysis. *PLoS ONE, 18*(12), e0289724.

<https://doi.org/10.1371/journal.pone.0289724>