A logo for college computing

Description automatically generated

**Predicting Customer Churn and Enhancing Retention Strategies Through Machine Learning**

| *Swan Saung Lwin* |  |
| --- | --- |
| *2025282* |  |
| *Strategic Thinking* |  |
| *CA 1 – Capstone Project Proposal* |  |
| *Sunday, 30th March 2025 23:59* |  |
| *Saturday, 29th March 2025* |  |



Contents

[1. Introduction 1](#_heading=h.cersum6jhy0z)

[2. Problem Definition 1](#_heading=h.8x540d88x0yp)

[3. Objectives 1](#_heading=h.cxop9cih540k)

[4. Literature review 2](#_heading=h.6609l845gxfr)

[5. Scope and Methodology 2](#_heading=h.xa3gw8clg8cy)

[5.1. Project Scope 2](#_heading=h.aek9xqu1iobz)

[5.2. Models & Evaluation 3](#_heading=h.3dg72zou7o26)

[5.3. Timeline 4](#_heading=h.vi9ohtk2s3l2)

[6. Data Sources 4](#_heading=h.uwb5ld322ztz)

[7. Ethical Considerations 4](#_heading=h.hlv9o24igx5)

[References 5](#_heading=h.5akkuwhp012l)

# 1. Introduction

In recent years, digital businesses have experienced rapid growth, leading to fierce competition among companies. As a result, expanding and sustaining a loyal customer base becomes more important than ever. Customer churn is a critical challenge for businesses, directly impacting revenue and long-term sustainability. Customer churn is the number of existing customers lost, for any reason at all, over a given period of time. The rate at which customer churn happens is called churn rate and it provides companies with an understanding of customer satisfaction and customer loyalty, and also identify potential changes in a company’s bottom line. (O'Brien and Downie, 2024).

This project utilizes machine learning techniques to predict customer churn while analyzing the significance of each service that business offers, and segmenting customers into focus groups to work on targeted strategies for customer retention. For this project, I will focus on building the model for a telco company which provided home phone and Internet services to customers in California in Q3.

# 2. Problem Definition

Customer Churn is a critical challenge faced by businesses across industries, especially in the digital market. High churn rates may directly affect a company with revenue loss and increased operational costs as acquiring a new customer can cost 5 to 25 times more than keeping an existing customer (Gallo, 2014).

Qualities and services provided by the company play a crucial role in customer satisfaction and retention. Dewan et al. (2020) found that effective engagement strategies, such as personalized communication and tailored product offerings, significantly enhance customer satisfaction and retention rates. Understanding how different services impact customer satisfaction can help businesses focus their efforts on high-impact areas, thereby improving overall retention. Customer segmentation is also a key strategy for improving retention by tailoring services to different customer groups. It is also important to analyse and identify high value customers so that the company can spend resources for retention effectively.

Many companies struggle to predict customer churn accurately and have difficulties in carrying out effective retention strategies. Key challenges include ineffective traditional methods, lack of insights into impact of different services, generalized retention strategies, and the need to have cost-effective retention strategies.

# 3. Objectives

The key objectives of this project are defined as follow

* to develop an accurate churn prediction model and identify potential churn customers by analyzing customer behavior, customer lifetime value, service usage, etc.
* to implement customer segmentation to improve prediction of customer churn and ensure that retention strategies are more personalized and effective rather than being generalized.
* to identify the factors that influence customer churn by assessing the impact of different services on churn.
* to optimize retention strategies to be more cost-effective by identifying high lifetime value potential churn customers with optimal operation cost.

# 4. Literature review

According to research done by Park and Ahn (2022), customer churn can be divided into two categories; voluntary, where the customers choose to use another company’s service due to dissatisfaction, and involuntary, where the customers are not able to use services due to other circumstances. Therefore I will work on this project assuming that all the churn customers are voluntary churn.

On the impact of services offered on churn, Keaveney (1995) conducted an investigation of how certain critical incidents caused customers to switch from one service provider to another. She surveyed more than 500 service customers, producing a list of more than 800 critical behaviors on the part of the firm. These behaviors were classified into eight categories of critical incidents, including pricing, core service failures, service encounter failures, employee responses to service failures, and competition. Based on this research, we can see how the quality of service can lead to customer dissatisfaction and churn.

There are lots of different research methods carried out by researchers on churn prediction in the telecommunication industry. Lalwani et al. (2022) applied a gravitational search algorithm to perform feature selection and to reduce the dimensions of the data-set. Vafeiadis et al. (2015) perform Monte Carlo simulations using five of the most popular, state-of-the-art classification methods for the churn prediction problem. The two top performing methods in terms of corresponding testing error were the two-layer Back-Propagation Network with 15 hidden units and the Decision Tree classifier; both methods achieved accuracy 94% and F-measure 77%, approximately. Tsai and Lu (2009) took an approach using hybrid neural networks and achieved higher than 88% accuracy.

# 5. Scope and Methodology

## 5.1. Project Scope

This project focuses on predicting customer churn using machine learning and identifying the impact of key services offered by a telecommunication company. The study will analyze customer behavior, and service usage patterns to predict churn.

Scope for this project will be defined as follows in Figure 5.1.

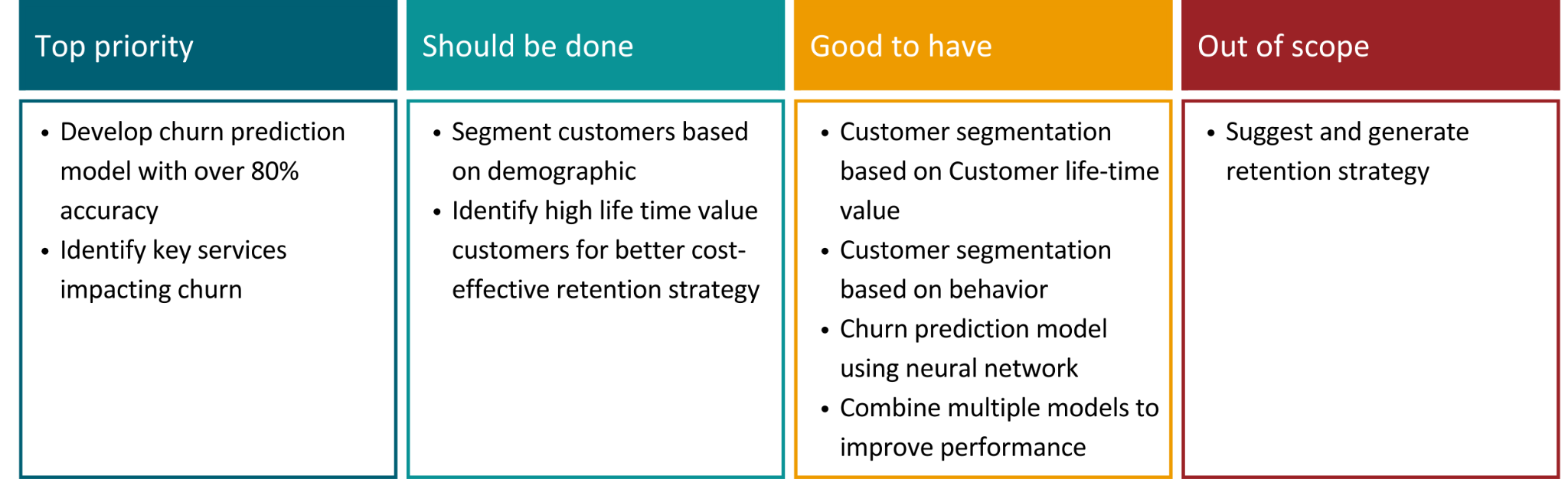


Figure 5.1.1. Project Scope & Task Prioritization

This project follows the CRISP-DM methodology, which provides a structured approach to data-mining.

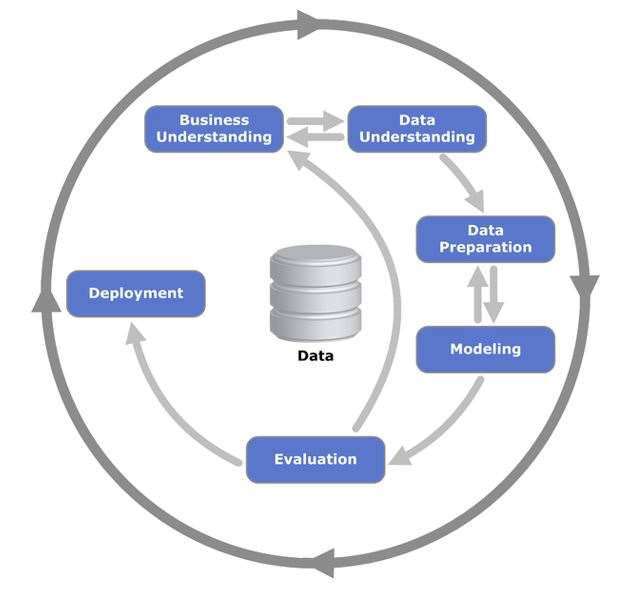


Figure 5.1.2. CRISP-DM process diagram *(Wikimedia Commons, 2012).*

## 5.2. Models & Evaluation

For predicting churn, I will be using supervised learning models and comparing them based on performance. Models include

* Logistic regression
* Random Forest
* Gradient Boosting Models

l will also be using neural networks (multi-layer perceptron) as an optional scope.

For customer segmentation, I will be using unsupervised learning models such as K-means and Hierarchical Clustering to segment customers based on demographics and behavior.

To identify most influential factors and services impacting churn, I will be using methods such as

* Gravitational search
* SHAP (SHapley Additive Explanations)
* Information Gain and Mutual Information

To evaluate the performance of each supervised model, I will be using F1-score and AUC-ROC metrics.

I will be using the following measures to evaluate customer segmentation:

* Silhouette Score
* Inertia
* Davies-Bouldin Index
* Dunn Index

## 5.3. Timeline

High level timeline for this project is described as follows in Figure 5.3.1.

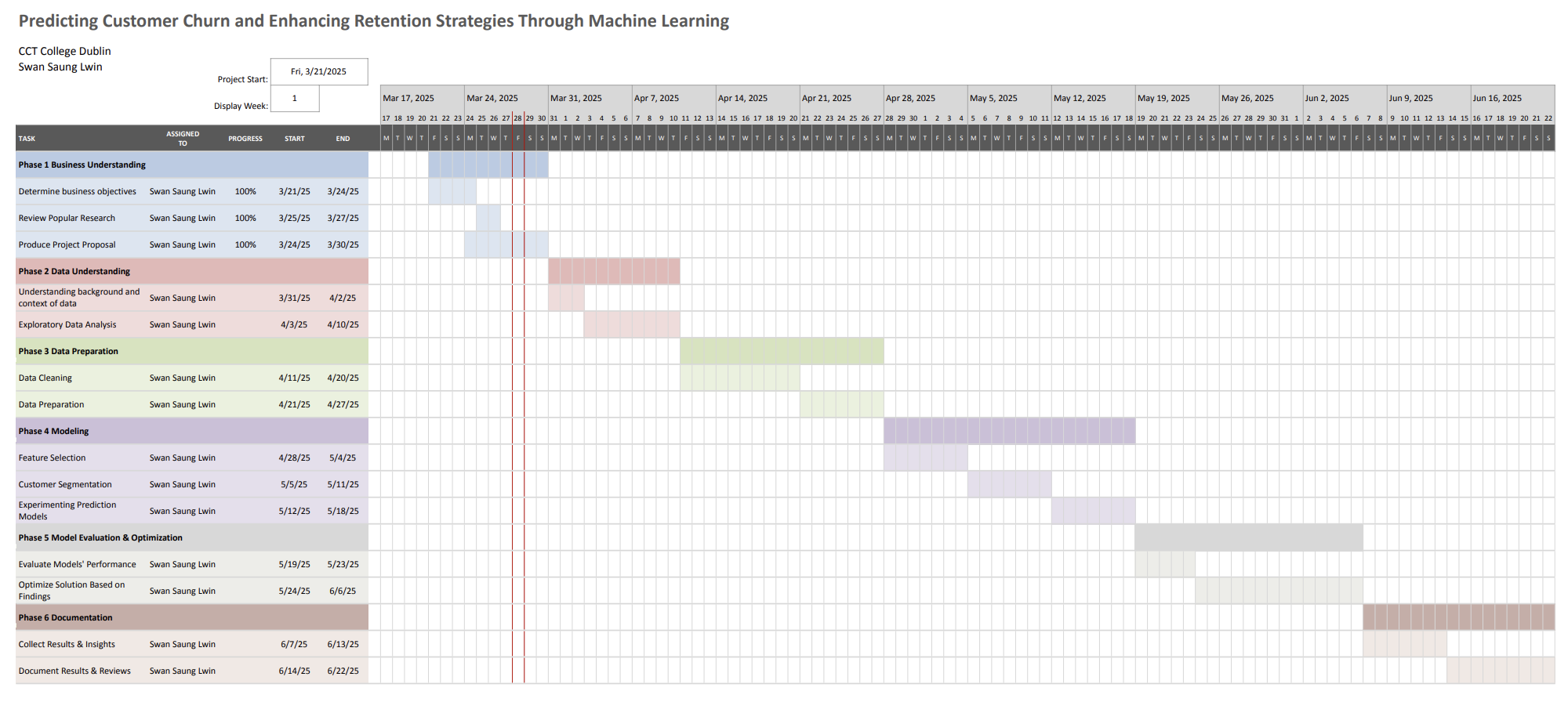


Figure 5.3.1. High level project timeline for Customer Churn Prediction

# 6. Data Sources

The Telco customer churn data contains information about a fictional telco company that provided home phone and Internet services to 7043 customers in California in Q3. It indicates which customers have left, stayed, or signed up for their service as well as the services being used by each customer. Multiple important demographics are included for each customer, as well as a Satisfaction Score, Churn Score, and Customer Lifetime Value (CLTV) index. (IBM

<https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113>

# 7. Ethical Considerations

The dataset contains sensitive customer information, including demographic attributes, service usage patterns, and satisfaction scores. Even though the dataset is fictional, real-world implementations must comply with data protection regulations such as General Data Protection Regulation (GDPR). To ensure privacy, transparency, and ethics , the following measures will be implemented.

* excluding personal identification
* ensuring that decisions are not disproportionately influenced by demographic variables
* implementing interpretable models alongside with complex models
* ensuring insights will be used on retention strategies rather than marketing tactics
* ensuring fair treatment for all customers

# References

O'Brien, A. and Downie, A., 2024. *What is customer churn?* IBM.

Gallo, A., 2014. *The Value of Keeping the Right Customers.* Harvard Business Review.

Park, W. and Ahn, H., 2022. *Not All Churn Customers Are the Same: Investigating the Effect of Customer Churn Heterogeneity on Customer Value in the Financial Sector.*

Keaveney, S. M., 1995. Customer switching behavior in service industries: An exploratory study. *Journal of Marketing, 59(2)*, 71–82.

Dewan, S., Wu, D. J., & Trivedi, R., 2020. Customer engagement in fintech: A study of factors affecting retention. Journal of Banking and Finance, 118, 105888.

Lalwani, P., Mishra, M.K., Chadha, J.S. and Sethi, P., 2022. Customer churn prediction system: a machine learning approach. *Computing, 104*(2), pp.271-294.

Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C., 2015. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, pp.1-9.

Tsai, C.F. and Lu, Y.H., 2009. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, *36*(10), pp.12547-12553.