

Week 1 – Identify the problem and dataset.

The project is banking customer churn prediction. Telecommunications companies usually aren't the most popular among customers. It's pretty common to hear people express frustration with some aspect of their communications provider, such as unwanted marketing emails, hard-to-navigate customer service and high plan prices. As a result, it may be unsurprising to learn telecommunication companies experience higher rates of customer churn – also known as the rate at which customers cease doing business with a company. Based on a churn rate just under two percent for top companies, one source estimates carriers lose \$65 million per month from churn.

Also, it is long been known retention of existing customers is less expensive than acquisition of new ones. Going to the bottom of why telecom customers are so apt to churn can help companies make targeted decisions aimed at keeping them. It is also important to identify which customers are most at risk of churning based on their behaviors and account history. This way, companies can act to keep them before they bail out with a compelling offer or loyalty reward.

The dataset for this project is from Kaggle. There are 14 attributes in the dataset, such as geography, gender, age, tenure, balance and etc. The models will be built based on these variables. It is a binary classification problem. It is supervised machine learning problem because all the data is labeled. I will develop algorithms to predict whether a customer is going to churn or not and analyze feature importance to identify top factors that influence the results. I will use three models, which are logistic regression, KNN and random forest. The evaluation metrics will be accuracy, precision and recall. Based on model performance, I will choose a model that works the best.

Below is the 13-week project plan

Project Plan:

Week	Weekly Focus
1	Identify the problem statement and dataset
2	Ingest and explore the dataset
3	Perform exploratory data analysis
4	Make data model ready
5	Engineer Features
6	Develop 1 st modeling approach (simple, the baseline)
7	Develop 2 nd modeling approach (more complex)
8	Develop 3 rd modeling approach (even more complex)
9	Select the winning model
10	Data Centric AI
11	Explain the model, analyze risk, bias and ethical considerations
12	Save and package model for deployment
13	Bring it all together