AI Project 1

Description: Perform an Exploratory Data Analysis on customer churn using a Telco dataset. After reviewing several datasets, we decided to use the dataset from the following Grid Dynamics blog post. We chose this dataset because it had more columns of data than others we looked at. This will enable us to look at a greater number of possible factors when we perform our Exploratory Data Analysis.

<https://blog.griddynamics.com/churn-analytics-in-the-technology-and-telecom-industries-using-google-vertex-ai-a-reference-notebook/>

The project has its own GitHub repository at the following location.

<https://github.com/griddynamics/rnd-gcp-starter-kits/tree/main/churn-prevention-vertex-ai>

Project Team**:** David Gerhart, Cody Cushing, Doug Francis

# Project “Catechism”

1. What are we trying to do?
   1. Perform Exploratory Data analysis for a customer churn analysis.
   2. After Exploratory Data Analysis, if time permits, perform an analysis of churn to build a predictive model. This should enable us to probabilistically predict who is most likely to churn. In turn, this can be used to build a strategy of how to retain as many customers as possible.
2. How is it done today? What are the limits to the current practice?
   1. We speculate that medium and large-size businesses would already be using ML techniques as well as Exploratory Data Analysis. We further speculate that companies are probably using other non-AI data mining techniques.
   2. Currently the dataset we are looking at does not have any mitigation data in it. That is there is no indication that customer retention techniques are being used and how they impact the data provided.
3. What is new in our approach? Why do you think it will be successful?
   1. We hope to expand on our exploratory data analysis and use machine learning to find patterns in the data that we were not able to identify.
   2. We believe a combination of manual data analysis and Machine learning will reveal new patterns in the churn data. There are members of our team who have had a long-standing interest in customer churn and have created retention programs. We are very interested in the insights that AI will provide.
   3. We intend to test any predictions by withholding a subset of data from the model and comparing any predictions against that real data. This will allow us to evaluate the accuracy of the predictions.
4. “Who cares”? If you are successful, what difference will it make?
   1. Any business that depends on recurring revenue from its customers’ needs to minimize churn. Insights gained through churn analysis and the success of various retention techniques may apply to multiple industries.
   2. Being able to probabilistically know which customers have the highest probability of churning would allow a company to better target their limited customer retention spending.
5. What are the risks?
   1. If the data collected does not have the key variables, then the analysis would not yield a good predictive model.
   2. Since no company will in truth be using our analysis, there is little risk.
6. How much will it cost?
   1. There are no direct project costs since this is a student exercise. The cost will just be in terms of time. No other expenses will be incurred.
7. How long will it take?
   1. The project must be completed by February 13th.
8. What are the mid-term and final exams to check for success?
   1. Initial data cleaning and exploration
   2. Creation of 6 or 8 exploratory graphs
   3. Drafting of conclusions and follow-up analysis.
   4. Creation of

Source Control Process

We have created the following structure.

main

dev

Task Branches  
 . . .

Each ach task will be a branch from Dev. Once tasks are completed, we will perform a pull request and push our changes. Once approved those changes will be merged into Dev. Dev will be merged into the main at the end of the project and possibly after key milestones are met.

**Possible investigations**

Male vs. Female

Tenure

Phone Service (Yes/No & minutes)

Internet Service (Yes/No & minutes)

Combined Phone/Internet usage

Number of customer service calls

\*\* Ecommerce Customer Churn Analysis and Prediction

Grid Dynamics:

Paper: <https://blog.griddynamics.com/churn-analytics-in-the-technology-and-telecom-industries-using-google-vertex-ai-a-reference-notebook/>

Good number of columns

<https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data>

zzz

\*\* Telco Customer churns

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn> <- the data on kaggle.

Someone's analysis they have completed.

<https://365datascience.com/tutorials/python-tutorials/how-to-build-a-customer-churn-prediction-model-in-python/#5>

Bank Customer Churn Dataset

<https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset/data>

Repo: <https://github.com/griddynamics/rnd-gcp-starter-kits/tree/main/churn-prevention-vertex-ai>

Notebook: <https://github.com/griddynamics/rnd-gcp-starter-kits/blob/main/churn-prevention-vertex-ai/notebooks/Churn_Predictions_for_Telco_with_Vertex_AI.ipynb>

A paper with writing on it

Description automatically generated