Customer Churn dataset

Category	Causal Discovery
Description	https://github.com/raz1470/causal_ai/blob/main/notebooks/using causal graphs to answer causal questions.ipynb
Priority to test	☑
URL	https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset

Introduction

Customer churn refers to the phenomenon where customers discontinue their relationship or subscription with a company or service provider. It represents the rate at which customers stop using a company's products or services within a specific period. Churn is an important metric for businesses as it directly impacts revenue, growth, and customer retention.

In the context of the Churn dataset, the churn label indicates whether a customer has churned or not. A churned customer is one who has decided to discontinue their subscription or usage of the company's services. On the other hand, a non-churned customer is one who continues to remain engaged and retains their relationship with the company.

Here we use the platform named Karma 360 to help businesses to understanding customer churn for identify patterns, factors, and information that contribute to customer attrition. By analyzing churn behavior and its associated features, companies can develop strategies to retain existing customers, improve customer satisfaction, and reduce customer turnover. Through Causal Ai techniques. Companies can also be applied to forecast and proactively address potential churn, enabling companies to take proactive measures to retain at-risk customers.

Dataset (64,374 rows and 12 features)

CustomerI	Age	Gender	Tenure	Usage Fre	Support C	Payment I	Subscriptic	Contract L	Total Sper	Last Intera	Churn
1	22	Female	25	14	4	27	Basic	Monthly	598	9	1
2	41	Female	28	28	7	13	Standard	Monthly	584	20	0
3	47	Male	27	10	2	29	Premium	Annual	757	21	0
4	35	Male	9	12	5	17	Premium	Quarterly	232	18	0
5	53	Female	58	24	9	2	Standard	Annual	533	18	0
6	30	Male	41	14	10	10	Premium	Monthly	500	29	0
7	47	Female	37	15	9	28	Basic	Quarterly	574	14	1
8	54	Female	36	11	0	18	Standard	Monthly	323	16	0
9	36	Male	20	5	10	8	Basic	Monthly	687	8	0
10	65	Male	8	4	2	23	Basic	Annual	995	10	0
11	46	Female	42	27	9	21	Standard	Annual	526	3	1
12	56	Male	13	23	5	14	Basic	Quarterly	187	1	0
13	31	Male	2	7	0	25	Premium	Quarterly	758	24	0
14	42	Male	46	27	5	8	Premium	Quarterly	438	30	0
15	59	Male	21	17	2	14	Premium	Quarterly	663	15	0

This dataset employed in this analysis was consists of 64,374 customer records. And it's already been cleaned. Exist no missing value and outliers. Each record in the testing file corresponds to a customer.

The dataset comprises several features that are likely to influence customer churn. These factors include:

Age: Age of the customer
Gender: Gender of the customer
Tenure: The duration for which the customer has been using the service
Usage Frequency: How often the customer uses the service
Support Calls: The number of calls made to customer support
Payment Delay: The frequency of delayed payments
Subscription Type: The type of subscription the customer has
Contract Length: The length of the customer’s contract
Total Spend: The total amount spent by the customer

Problems we want to answer

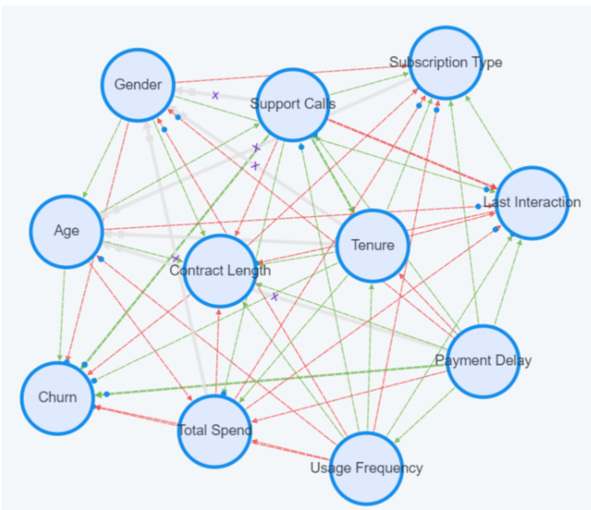
In our analysis, we aim to address the following key questions:

1. How do different factors correlate with the presence of customer churn?
2. What are the most significant predictors of customer churn in the dataset?
3. Is the factors considered by LLM will match our causal graph?

4. Can we predict the presence of customer churn based on these parameters?

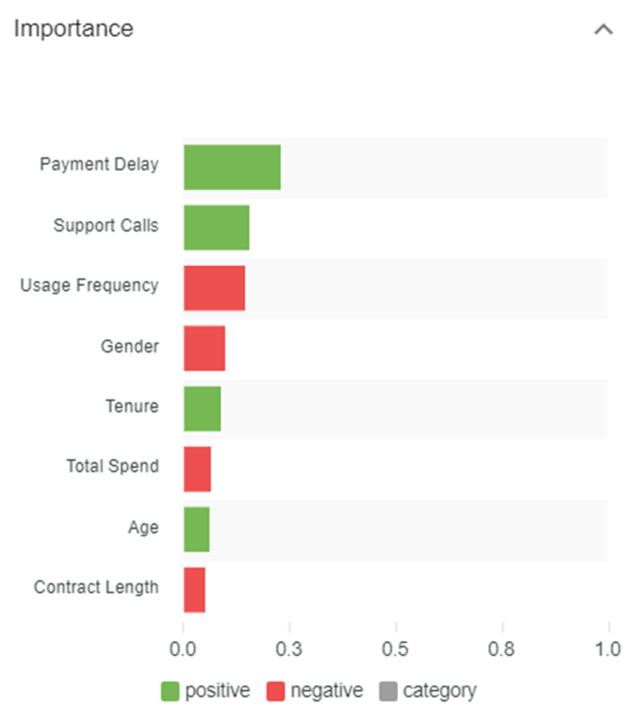
Analysis in Karma

After data cleaning(reduced to 600 row), We Used Karma 360 to visualize the relation of these factors. We could also see the information from every node.



Information_node

Features	Value
Name	Churn
Datatype	integer
total	600
empty	0
valid	600
mean	0.5
std	0.5
min	0
25%	0
50%	0.5
75%	1
max	1



- **Correlational Analysis**:

If you clicked the node in Karma 360, you can see all the causes of churn from karma.

Includes↓ (From most important to less important):

Factor Name: importance value

1. **Payment Delay:** 0.232
2. **Support Calls:** 0.159
3. **Usage Frequency:** -0.148
4. **Total Spend:** -0.101
5. **Tenure:** 0.091
6. **Total Spend:** -0.067
7. **Age:** 0.065
8. Contract Length: -0.054

If the factor is red, that means it is inversely proportional to the node, which is that the lower values can cause higher value of churn. If it's green, means that it is proportional to churn rate.

- **Feature Importance**:

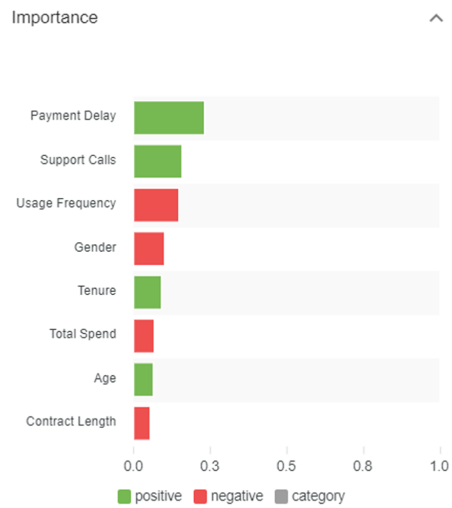
We found The most crucial indicator - "Payment Delay". And the higher payment delay frequency means the higher churn rate which is aligned with our understanding.

- **Operating with LLM (Sort by Importance)**

Here we use the newest LLM - Claude 3.5 Sonnet to interpreting our dataset and ask it to rank the factors of customer churn. We could found that the first three important factors(Payment Delay, Support Calls and Usage Frequency) which LLM output is same with our causal graph.

Based on the information provided about the churn dataset, I can offer an assessment of the top three factors that are likely to be most important for predicting customer churn:

1. Usage Frequency: This is likely the most critical factor in predicting churn. It directly reflects customer engagement and the value they're deriving from the service. Lower usage frequency often indicates that a customer is not finding the service useful or necessary, which significantly increases their likelihood of churning.
2. Payment Delay: The second most important factor is probably payment delay. Consistent late payments can be a strong indicator of potential churn. It may signal financial difficulties or dissatisfaction with the service, both of which are common precursors to a customer deciding to leave.
3. Support Calls: The number of support calls is likely the third most crucial factor. A high frequency of support calls could suggest that customers are experiencing ongoing issues or dissatisfaction with the service. Unresolved problems or frequent need for assistance might lead to frustration and ultimately, churn.



- **Causal Prediction(Wait till upload csv data function done)**:

Result

Based on the causal graphs and predictive models generated:

- We highlight significant factors that is directly affected to "Churn" such as **Payment Delay, Support Calls** and **Usage Frequency.** We also displayed the correlation from every node**.**
- We use LLM model to prove that our analyze is correct. More over, by using our platform, the company can see more information from every node and understanding their correlation.
- We discuss the predictive accuracy of our models to help company analyze whether customer will churn and develop strategies to reduce churn rate.

By understanding the factors contributing to customer churn and using Karma 360, businesses can proactively address churn and implement strategies to retain customers.

Resource link

Karma project link: http://192.168.50.3:28081/results?id=187&datasetGroupId=254 (ac: moo-fon.lee@vizuro.com, pw: 12345678)

Dataset: https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset

Code:
https://github.com/LeeMoofon0222/Vizuro_Intern/tree/main/cc_process