Predicting **Customer Churn**

ISOM 599 Final Project
Group 2: Yifei Wang, Laurie Ye, Qing Ying

Business Problem

Customer churn is one of the biggest expenditures in many organizations, including telecommunication companies, where customer churn means the loss of customers who stop using their services.

Given customer-level data from a telecommunications company, our goal is to:

1) Understand what kind of customers are churning
2) Predict customers who will churn

Data Understanding

Variables

- Customer ID
- Demographic information (gender, senior citizen, partner, dependents)
- Tenure (how long they have been a customer)
- Services the customer signed up for (phone service, multiple lines, internet service, streaming TV, etc)
- Customer account information (contract, payment method, paperless billing, etc.)
- Churn

Multicollinearity

- Phone Service and Multiple Lines (Yes/No/No phone service)
- Internet Service and Online Security, Online Backup, Device Protection, Tech Support, Streaming TV and Streaming Movies
- Paperless Billing and Payment Method
- Tenure and Total Charge

These variables are highly correlated with one another and may reduce precision of the model.

Data Processing

Remove unnecessary columns:

- Drop Customer ID column
- Drop collinear columns
  - Online Security, Online Backup, Device Protection, Tech Support, Streaming Movies, Streaming TV, Multiple Lines, Payment Method, Total Charge

Prepare the data:

- Ensure categorical columns are factored (e.g., the Senior Citizen column)
- Standardize numerical columns
- Impute missing

Train-test split on a 70/30 ratio

Predictive Models

Logistic Regression

Feature Selection -> Stepwise approach: with direction ‘both’, we can add and remove data features to find the best combination for accuracy.

Boosting

Feature Selection - > Random Forest: Calculated the importance of each data feature to see which ones affect predictions the most.

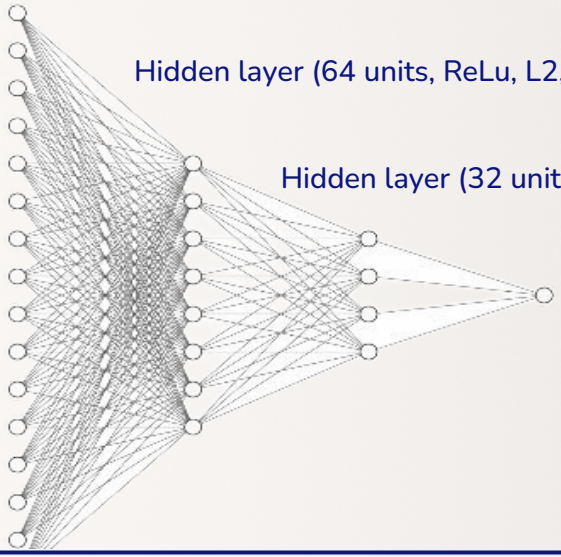
Algorithms:

1) XGBoost
2) AdaBoost

Deep Learning Model

Neural Network: Sequential neural network **designed for binary classification.**

Input layer(128 units, ReLu, L2, Dropout rate = 0.4)



Hidden layer (64 units, ReLu, L2, Dropout rate = 0.4)

Hidden layer (32 units, ReLu, L2, Dropout rate = 0.4)

Output layer (sigmoid activation)

- **Fine-turning**: Adam Optimizer
- **Measure Performance:** binary cross-entropy
- **Training**: batch size: 256, epochs: 200, 30% of data to validate learning
- **Prediction and Evaluation:** Predict on new data and access accuracy

Model Results

Model Names	Accuracy
Logistic Regression	0.81
XGboost	0.793
Adaboost	0.802
Neural Network Model	0.797

Inference: Logistic Regression

Variable	P-Value	Odds Ratio	Variable	P-Value	Odds Ratio
Intercept	0.0000	0.3642	SeniorCitizenYes	0.0064	1.3024
DependentsYes	0.0146	0.7906	tenure	0.0000	0.4111
PhoneServiceYes	0.0001	0.5277	InternetServiceFiber optic	0.0000	2.2989
InternetServiceNo	0.0686	0.6757			
ContractTwo year	0.0000	0.2052	ContractOne year	0.0000	0.4397
MonthlyCharges	0.0035	1.4381	PaperlessBillingYes	0.0000	1.5282

Interpretation: Odds Ratio

- **SeniorCitizenYes**: Being a senior citizen increases the odds of churn by about 30% compared to those who are not senior citizens
- **DependentsYes**: Customers with dependents decrease the odds of churn by about 21% compared to those without dependents
- **Tenure**: For every additional year in tenure, the odds of churn decreases by about 59%
- **PhoneServiceYes**: Having phone service decreases the odds of churn by about 47% compared to without phone service
- **InternetServiceFiber optic**: Choosing fiber optic internet service increases the odds of churn by about 1.3 times compared to not choosing DSL

Interpretation: Odds Ratio

- **InternetServiceNo**: Not having internet service decreases the odds of churn by about 32% compared to choosing DSL
- **ContractOne year**: Being a one-year contract decreases the odds of churn by about 56% compared to a month-to-month contract
- **ContractTwo year**: Being a two-year contract decreases the odds of churn by about 79% compared to a month-to-month contract
- **PaperlessBillingYes**: Choosing paperless billing increases the odds of churn by about 53% compared to not choosing paperless billing
- **MonthlyCharges**: For every additional dollar increase in monthly charges, the odds of churn increases by about 44%

Findings

1. Who are more likely to churn?

Senior Citizens	Without Dependents	Shorter Tenure	No Phone Service	Fiber Optic Internet
Internet Service	Monthly Contracts	Paperless Billing	Higher Monthly Charges	

2. Why customers are more likely to churn with fiber optic?

We often hear about fiber optics being a faster and more reliable internet service option, yet we observe higher churn in these customers. We believe this could be correlated with the fact that senior citizens are more likely to churn, who are more likely to have older devices that require DSL support.

Recommendations

Encourage subscription to Longer Contract

Provide discounts or upgraded service features for customers who are willing to sign longer contract durations. With a longer commitment period, customers are less likely to switch to a competitor during the contracted time

Promote Bundling Options

Currently, we observe that customers with no phone service and no internet service are more likely to churn that those who do. It may be beneficial to bundle services together and cross-sell them, as the commitment may increase customer loyalty.

Enhance Support and Services for Senior Citizen Customers

Offer more accessible and assistive services for seniors citizen customers, providing assistance for billing and plans

Thank you!

Q&A