Business Problem: Customer Churn

Customer churn, also known as customer attrition means the loss of clients or customers.

Telephone service companies, Internet service providers, pay TV companies, insurance firms, and alarm monitoring services, often use customer churn analysis and customer churn rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Companies from these sectors often have customer service branches which attempt to win back defecting clients, because recovered long-term customers can be worth much more to a company than newly recruited clients.

Companies usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another company or service provider, involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care facility, death, or the relocation to a distant location. In most applications, involuntary reasons for churn are excluded from the analytical models. Analysts tend to concentrate on voluntary churn, because it typically occurs due to factors of the company-customer relationship which companies control, such as how billing interactions are handled or how after-sales help is provided.

predictive analytics use churn prediction models that predict customer churn by assessing their propensity of risk to churn. Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn.

Data Description

Data is included in the file

Customer-Churn.csv

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:
- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Question 1 (20 points)
- Read the data from the data file
- Split data into training (70%) and test (30%) data set
- Based on training data, check the univariate distribution of each independent variables for churn and non-churn customers to get a basic idea of how customer churn is impacted by each independent variables.

Question 2 (20 points)

- Build logistic regression models with different independent variables and regularization parameters
- Evaluate and compare different model results

Question 3 (20 points)

- Build linear and nonlinear SVC models with different kernel functions and regularization parameters
- Evaluate and compare different model results

Question 4 (20 points)

- Build decision trees and ensemble of decision trees (Random forest, GBDT) models with different model parameters
- Evaluate and compare different model results

Question 5 (20 points)

- Evaluate and compare model results above
- Propose potential ways to further improve the model results

Appendix: Data Columns

customerID
Customer ID

gender
Whether the customer is a male or a female

SeniorCitizen
Whether the customer is a senior citizen or not (1, 0)

Partner
Whether the customer has a partner or not (Yes, No)

Dependents
Whether the customer has dependents or not (Yes, No)

tenure
Number of months the customer has stayed with the company

PhoneService
Whether the customer has a phone service or not (Yes, No)

MultipleLines

Whether the customer has multiple lines or not (Yes, No, No phone service)

InternetService

Customer's internet service provider (DSL, Fiber optic, No)

OnlineSecurity

Whether the customer has online security or not (Yes, No, No internet service)

OnlineBackup

Whether the customer has online backup or not (Yes, No, No internet service)

DeviceProtection

Whether the customer has device protection or not (Yes, No, No internet service)

TechSupport

Whether the customer has tech support or not (Yes, No, No internet service)

StreamingTV

Whether the customer has streaming TV or not (Yes, No, No internet service)

StreamingMovies

Whether the customer has streaming movies or not (Yes, No, No internet service)

Contract

The contract term of the customer (Month-to-month, One year, Two year)

PaperlessBilling

Whether the customer has paperless billing or not (Yes, No)

PaymentMethod

The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

MonthlyCharges

The amount charged to the customer monthly

TotalCharges

The total amount charged to the customer

Churn

Whether the customer churned or not (Yes or No)