BANA 273 Machine Learning

Analysis on Telecom Customer Churn Rate

Team 19
Harinishri Srikanth, Pei-Ling, Wu, Praveen Aravindar, Xinfa Pan

What question is it answering?

Analyze the customer data to predict whether the customer will churn or not. The ultimate question we are trying to find out is what attributes play an important role in determining the Churn rate. By implementing this analysis, we will be able to identify the key attributes and try to find out the major reason for the customers to leave the company.

Brief Description about the project

This is a pre-crawled dataset, taken as a subset of a bigger dataset that was created by extracting data from IBM, which ranks among the world's largest information technology companies. We would like to investigate the relationship between the Churn rate and other attributes in the dataset such as Tenure, Phone service etc.. This is important for the company because our analysis should indicate the most significant and priority attributes that contribute to the Churn rate. Thus, it can help decide what kind of service to provide as a priority to attract more customers and finally achieve the target of reducing the Churn rate on the application.

Data and Attributes

We found "WA_Fn-UseC_-Telco-Customer-Churn.csv" on Kaggle, which contains 7043 rows (customers) and 21 columns (features).

Columns description in below:

	Column	**Description**	**Sample value**
1	customerID		7590-VHVEG
2	gender	Whether the customer is a male or a female	male or a female
3	SeniorCitizen	Whether the customer is a senior citizen or not	(1, 0)
4	Partner	Whether the customer has a partner or not	(Yes, No)
5	Dependents	Whether customer has dependents or not	(Yes, No)
6	tenure	Number of months the customer has stayed with the company	34
7	PhoneService	Whether the customer has a phone service or not	(Yes, No)
8	MultipleLines	Whether the customer has multiple lines or	(Yes, No, No phone

		not	service)
9	InternetService	Customer's internet service provider	(DSL, Fiber optic, No)
10	OnlineSecurity	Whether the customer has online security or not	(Yes, No, No internet service)
11	OnlineBackup	Whether the customer has online backup or not	(Yes, No)
12	DeviceProtection	Whether the customer has device protection or not	(Yes, No)
13	TechSupport	Whether the customer has tech support or not	(Yes, No)
14	StreamingTV	Whether the customer has streaming TV or not	(Yes, No)
15	StreamingMovies	Whether the customer has streaming movies or not	(Yes, No)
16	Contract	The contract term of the customer	(Month-to-month, One year, Two year)
17	PaperlessBilling	Whether the customer has paperless billing or not	(Yes, No)
18	PaymentMethod	The customer's payment method	(Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)
19	MonthlyCharges	The amount charged to the customer monthly	
20	TotalCharges	The total amount charged to the customer	
21	Churn	Whether the customer churned or not	(Yes, No)

Output

The "Churn" column is our target. We are using Classification , regression algorithm, regression table, Decision tree,random forest, Heatmap and cluster to show attribute correlation.