Assignment

Welcome to the assignment of customer prediction, in this section, you will work on a different Telecom Data, You will use logistic regression with different variations to build a churn model that will effectively predict customer churn rate.

Source of the Data:

[*https://www.kaggle.com/blastchar/telco-customer-churn*](https://www.kaggle.com/blastchar/telco-customer-churn)

**Columns :**

Customer ID

Gender Whether the customer is a male or a female

Senior 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

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

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

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

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

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

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

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

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

Streaming Movies. 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)

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

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

Monthly Charge The amount charged to the customer monthly

Total Charges The total amount charged to the customer

Churn. Whether the customer churned or not (Yes or No)

1. Clean and preprocess the data if needed
2. Shuffle the data
3. Split the data to training and testing
4. Use stats model to identify which factors are significant
5. Model all features using x\_train and x\_test
6. Use the model to predict on unseen data(y\_test)
7. How many false positives do you have?
8. What is the mean accuracy between actual and prediction?
9. What is the precision rate ?
10. What is the recall rate ?
11. Model the data again only with the significant variables?
12. What is the new mean accuracy between actual and prediction?