Telecom Customer Churn

Group A:
Anusha Kokkinti
Shruti Jituri
Tejaswini Mummadi

Agenda

- Introduction
- Data
- Data Visualization
- Predictive Analysis
- Conclusion

Introduction

Customer Churn

Customer Analytics

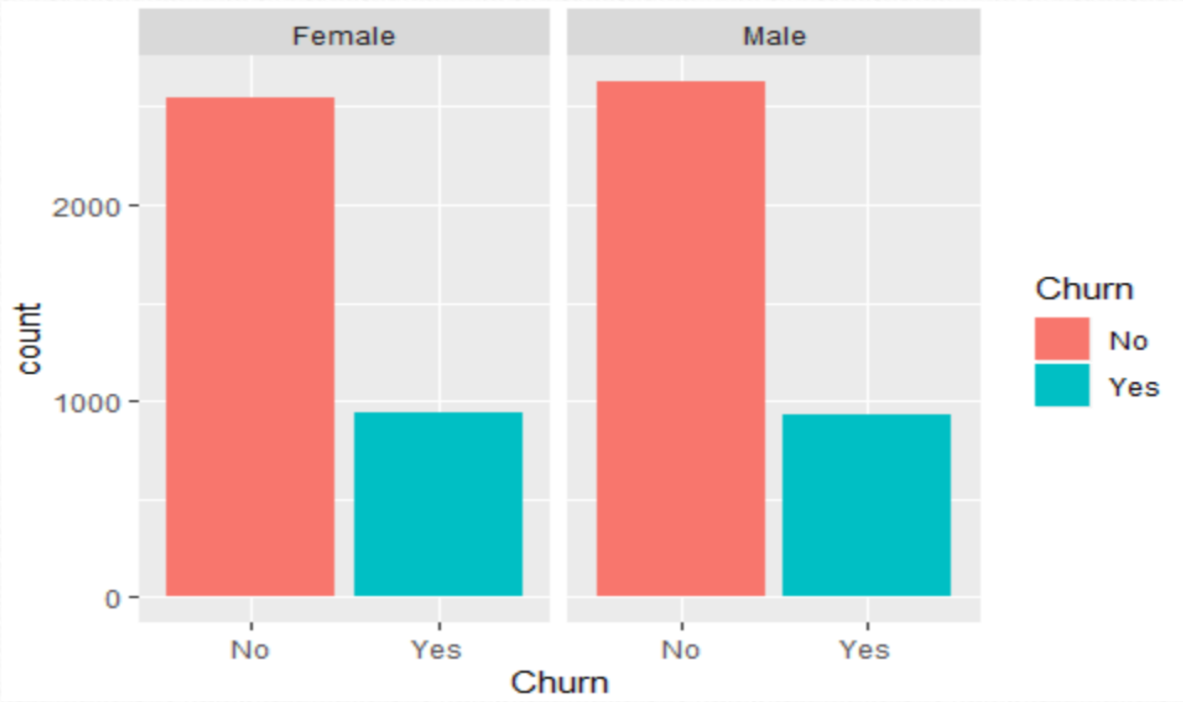
Data

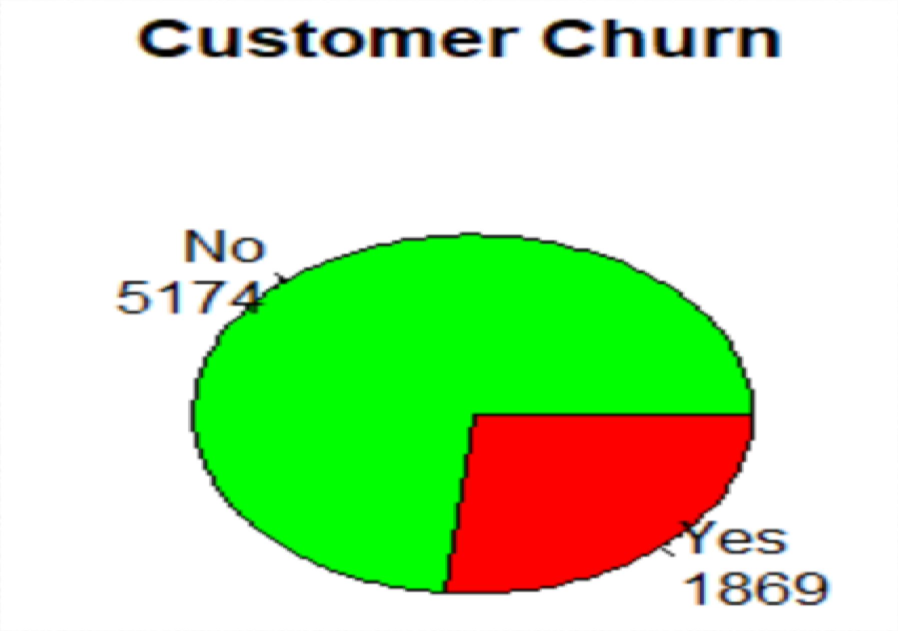
customerID	gender	SeniorCitizen	Partner	Dependents	tenure
PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection
TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod
MonthlyCharges	TotalCharges	Churn			

Descriptive Analytics

```
> summary(loadData$tenure)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00    9.00   29.00   32.37   55.00   72.00 
> summary(loadData$MonthlyCharges)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  18.25   35.50   70.35   64.76   89.85  118.75 
> x<-cor(loadData$tenure,loadData$MonthlyCharges)
> x
[1] 0.2478999
```

Data Visualization





Predictive Analysis

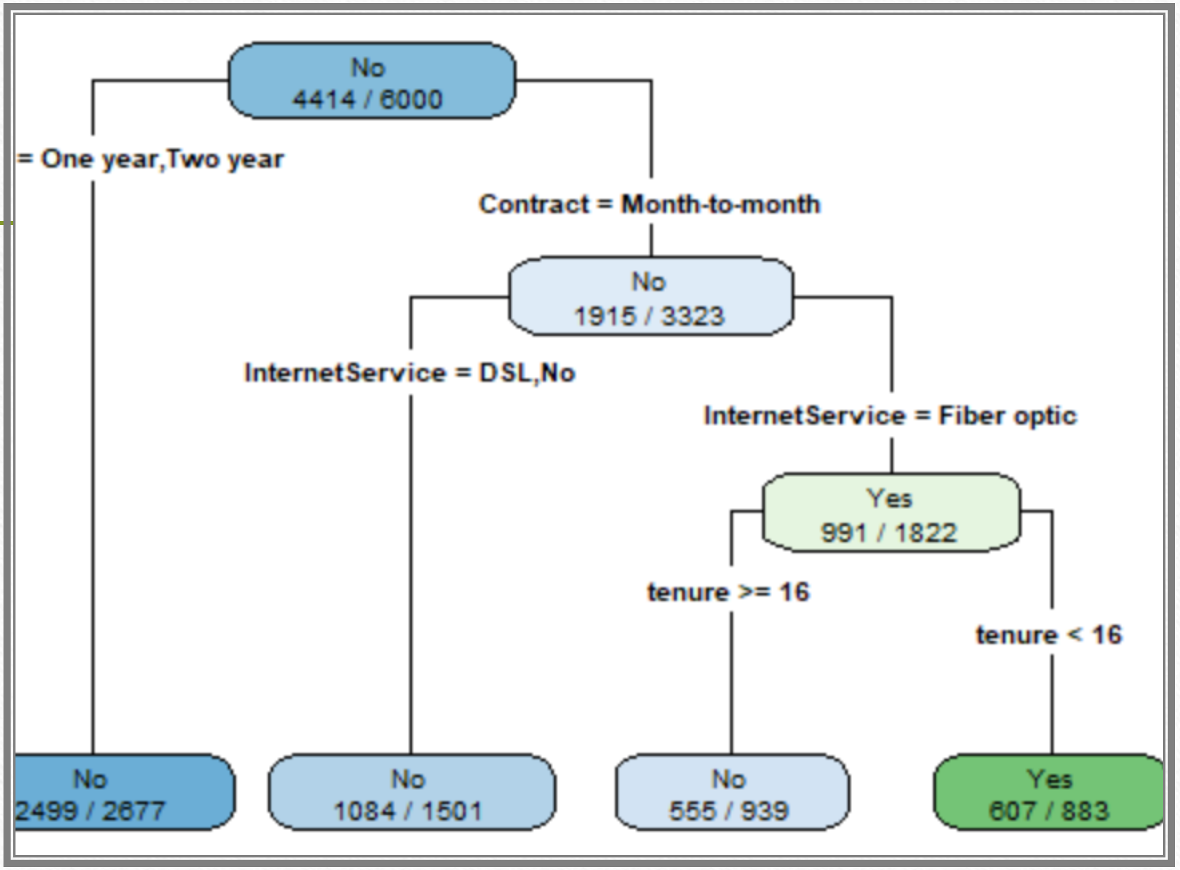
Classifiers

- Decision Tree
- Naïve Bayes
- Logistic Regression

Decision Tree

- **Decision Rules:**

- **If Contract=Month-to Month, InternetService = FiberOptic, AND tenure<16 then Churn is Yes**
- **If Contract=Month-to Month, InternetService = FiberOptic, AND tenure >=16 then Churn is No**
- **If Contract=Month-to Month AND InternetService = DSL,No, then Churn is No**
- **If Contract= One year, Two year then Churn is No**



Naïve Bayesian

Naive Bayes model is a generative model

It handles missing values and high dimensional data well when compared to other classifiers.

Input: Inserting all the variables into the model

Output: Calculating the posterior probability and providing the churn prediction model, accuracy and its overall performance

Implementation & Validation

Naïve Bayesian is a 3 – Step Process:

1- Prior probability for class

2-Likelihood of X given class C1:

3.Posterior probability of X1 being C1

1- Prior probability for Churn

```
A-priori probabilities:
Y
       No       Yes
0.7356667 0.2643333
```

2-Likelihood of X given class Churn

```
Conditional probabilities:
     SeniorCitizen
Y          [,1]      [,2]
  No  0.1293611 0.3356372
  Yes 0.2540984 0.4354905

     Dependents
Y            No       Yes
  No  0.6606253 0.3393747
  Yes 0.8253468 0.1746532

     tenure
Y         [,1]     [,2]
  No  37.43068 24.19653
  Yes 17.78310 19.19574

     PhoneService
Y             No        Yes
  No  0.09379248 0.90620752
  Yes 0.08953342 0.91046658

     MultipleLines
Y             No No phone service        Yes
  No  0.49388310       0.09379248 0.41232442
  Yes 0.45901639       0.08953342 0.45145019
```

3.Posterior probability of X being Churn

```
> testData1<-as.data.frame(loadData[6001,])
> results <- predict (model,testData1)
> results
[1] No
Levels: No Yes
> head(testData1)
     customerID gender SeniorCitizen Partner Dependents tenure PhoneService
6001 9503-XJUME   Male             0      No        Yes     52          Yes
     MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection
6001            No              No             No           No               No
     TechSupport StreamingTV StreamingMovies Contract PaperlessBilling
6001          No          No              No One year               No
               PaymentMethod MonthlyCharges TotalCharges Churn
6001 Credit card (automatic)          19.65        928.4    No
>
```

Performance

- Confusion Matrix

	Churn	
Prediction	No	Yes
No	632	95
Yes	129	188

- Accuracy

Accuracy
0.7854

Interpretations

632 correctly predicted as Non-churner customer from 727 which is (86.9%)

188 correctly predicted as Churner customer from 317 which is (59.3%).

Accuracy-78.4%

Limitations

- Naïve Bayes assumes all the features to be conditionally independent
- Here there exists some correlation between the variables {Internet services} and {OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies}.
- Naïve Bayes cannot handle numerical data well when compared to Decision Tree classifier.

Logistic Regression

DATA PREPARATION

SELECTION OF VARIABLES

FINAL MODEL

INTERPRETATIONS

TESTING

PERFORMANCE AS CLASSIFIER

Logistic Regression

- Data Preparation
  - Remove the colinearity.
- Selection of Variables
  - Log Likelihood Test
  - Backward Step wise method
  - Pchisq()

Logistic Regression- Final Model

```
Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)                           1.847697   0.343665   5.376 7.60e-08 ***
tenure                               -0.035454   0.002497 -14.197  < 2e-16 ***
MonthlyCharges                       -0.078703   0.011908  -6.609 3.86e-11 ***
as.factor(SeniorCitizen)1             0.252035   0.089566   2.814 0.004893 ** 
as.factor(PhoneService)Yes            0.998134   0.282241   3.536 0.000406 ***
as.factor(MultipleLines)Yes           0.686086   0.102517   6.692 2.20e-11 ***
as.factor(InternetService)Fiber optic 2.960670   0.298747   9.910  < 2e-16 ***
as.factor(InternetService)No         -2.790650   0.361137  -7.727 1.10e-14 ***
as.factor(OnlineBackup)Yes            0.257616   0.103364   2.492 0.012691 *  
as.factor(DeviceProtection)Yes        0.388734   0.105642   3.680 0.000233 ***
as.factor(StreamingTV)Yes             1.089116   0.148457   7.336 2.20e-13 ***
as.factor(StreamingMovies)Yes         1.069088   0.148330   7.207 5.70e-13 ***
as.factor(Contract)One year          -0.665529   0.115376  -5.768 8.00e-09 ***
as.factor(Contract)Two year          -1.488929   0.194464  -7.657 1.91e-14 ***
as.factor(PaperlessBilling)Yes        0.405710   0.080144   5.062 4.14e-07 ***
---
```

Logistic Regression - Interpretations

- The log odds of customers churning impacted by the variables

1. For a unit(month) increase in tenure, the log odds of the customer churning (leaving the company) decreases by 0.035.
2. For a unit (Dollar) increase in MonthlyCharges, the log odds of the customer churning (leaving the company) decreases by 0.0789.
3. The log odds of the customer churning are 0.252 more when the customer is a senior citizen than he is not.
4. The log odds of the customer churning are 0.99 more when the customer is using the phone service than he is not.
5. The log odds of the customer churning are 0.687 more when the customer has multiples lines of phone service than he is not given that he is having phone service.
6. a. The log odds of customer churning who are using the internet service with FiberOptics is 2.96 more when compared to the customer who are using the internet service with DSL option.

   b. The log odds of customer churning who are not using any InternetService is 2.79

   Less when compared to the customer who are using the internet service with DSL option.

Logistic Regression - Interpretations

The log odds of customers churning impacted by the variables.

7. The log odds of the customer churning are 0.257 more when he is using OnlineBackup service than he is not.
8. The log odds of the customer churning is 0.388 more w hen he is using Device Protection service than he is not.
9. The log odds of the customer churning are 1.089 more when he is using StreamingTV service than he is not.
10. The log odds of the customer churning are 1.069 more when he is using StreamingMovies service than he is not.
11. a. The log odds of customers churning who are on one-year contract is 0.665 less than those of the customers who are on month-month contract.

    b. The log odds of customers churning who are on Two-year contract is 1.488 less than those the customers who are on month-month contract.

12. The log odds of customers who have opted for paperless billing for churning is 0.405 more than those of the customer who have not opted for paperless billing.

Logistic Regression -Testing

Confusion Matrix

	ACTUAL	
PREDICTONS	No	Yes
No	694	133
Yes	66	150

Accuracy Rate

ACCURACY
80.90%

Decision Tree-Testing

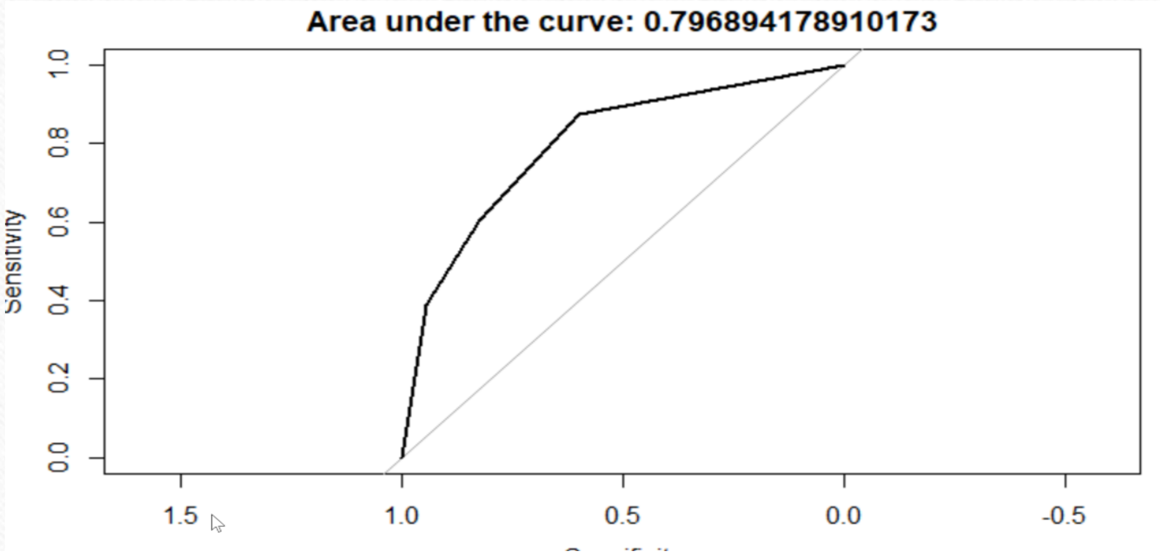
Confusion Matrix

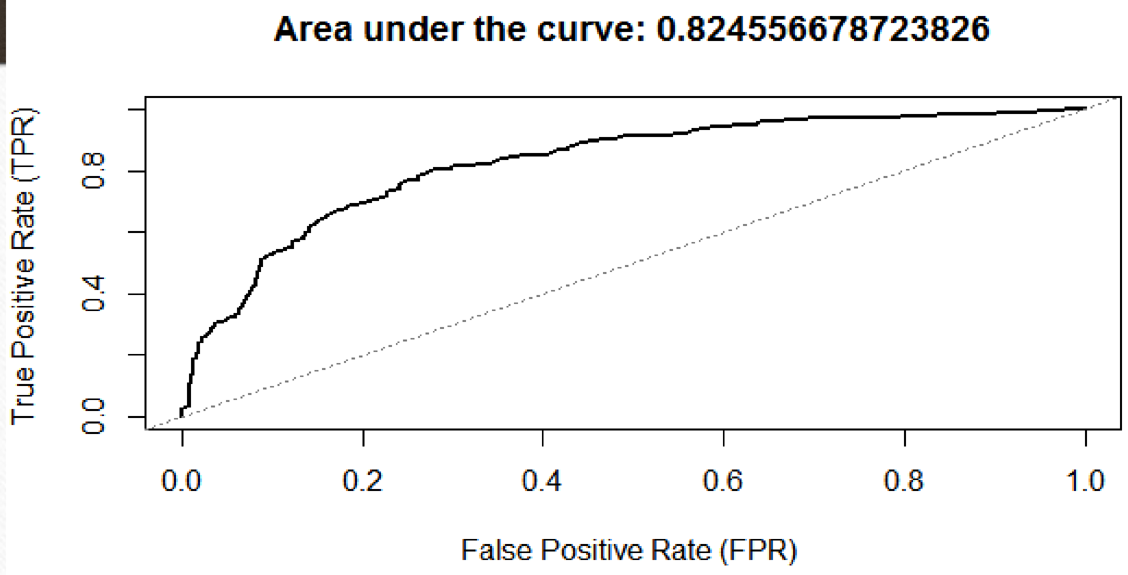
	ACTUAL	
PREDICTONS	No	Yes
No	717	173
Yes	43	110

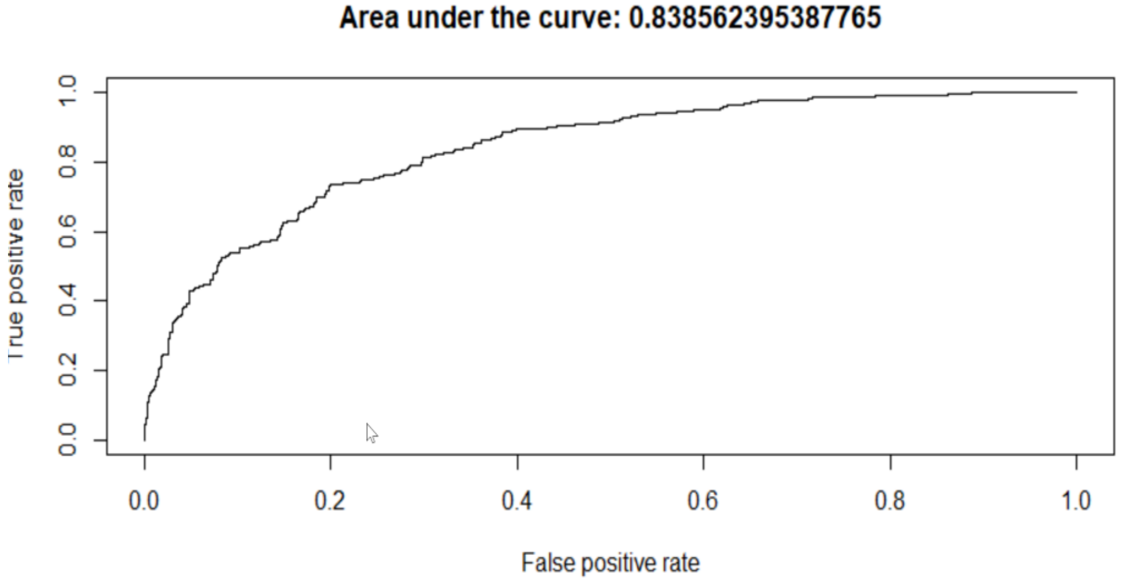
Accuracy Rate

ACCURACY
79.20%

Model Performance	
Model	**AUC**
Decision Tree	0.7969
Naïve Bayesian	0.8246
Logistic Regression	0.8386







Conclusion

- Based on the AUC values, Logistic Regression is giving the best prediction model with 83.86%
- From the models, we found
- Which Customers are likely to Churn
- Which features have the most impact on customer leaving

Conclusion

Feature	Impact
Tenure	Positive
Monthly Charges	Positive
Senior Citizen	Negative
Phone Service	Negative
Multiple Lines	Negative
Internet service- Fiber Optics	Negative

- Based on the prediction model, the features influencing the customers churn

Recommendations

- Approach the customers likely to churn with these features, and making appropriate changes will bind them to the company
- Building exclusive offers and promotions targeted towards seniors
- Adjust the relevant criteria
- On Boarding program
- Customer Feedback Loop
- Targeting Business users for Fiber Optics Internet Service
- Referral programs

Thank You