

# Telecom Customer Churn



Kornkanok Somkul Shashi Bala



# **Project Introduction**



#### Data and Data Source

- Data taken from Kaggle adopted from IBM
- No. of Categorical Column 18 columns
- No. of Continuous Column 3 columns
- Final outcome categorical -> Churn (Yes or No)



#### Modeling and Validation Techniques

- Logistic Regression
- Anova Table
- Confusion Matrix etc.

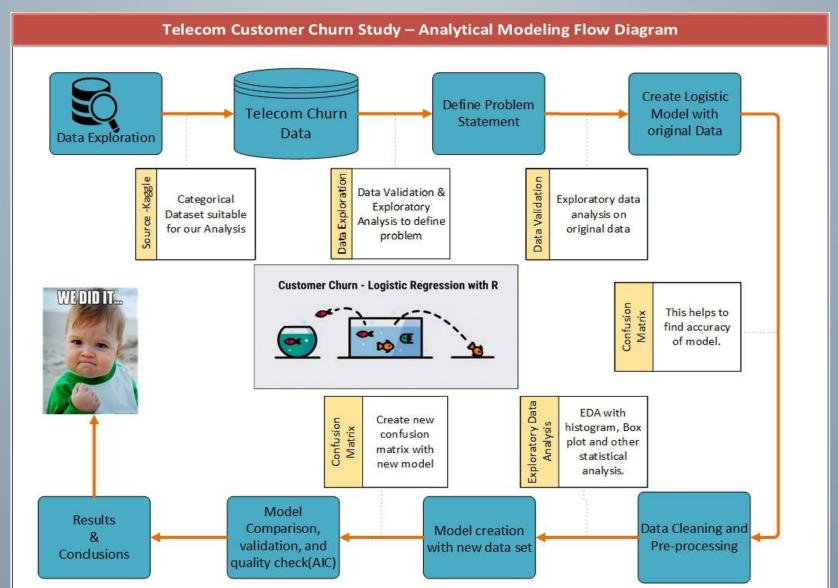


### Terms and Business Significance Discussion

- What is Churn?
- Is Churn a real threat?



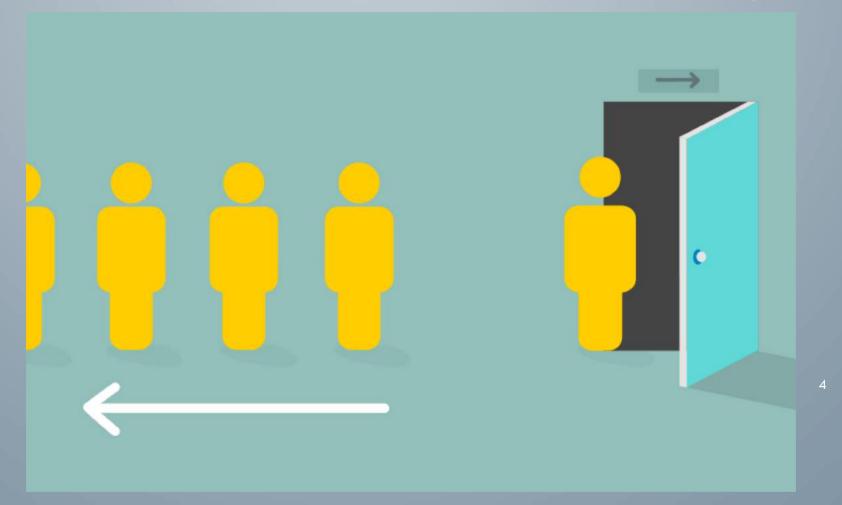
# **Project Overview and Flow Diagram**



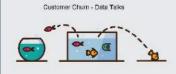


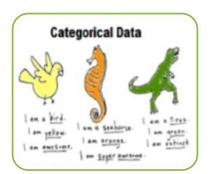
# **Problem Statement**

to understand and predict customer churn behavior based on customer's demographic, services, and the length of relationship



## Let's Talk Data





#### Service:

- phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.

#### **Customer information:**

- contract, payment method, paperless billing.

#### Demographic info about customers:

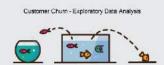
- gender, age range(Senior Citizen), and if they have partners and dependents.



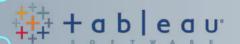


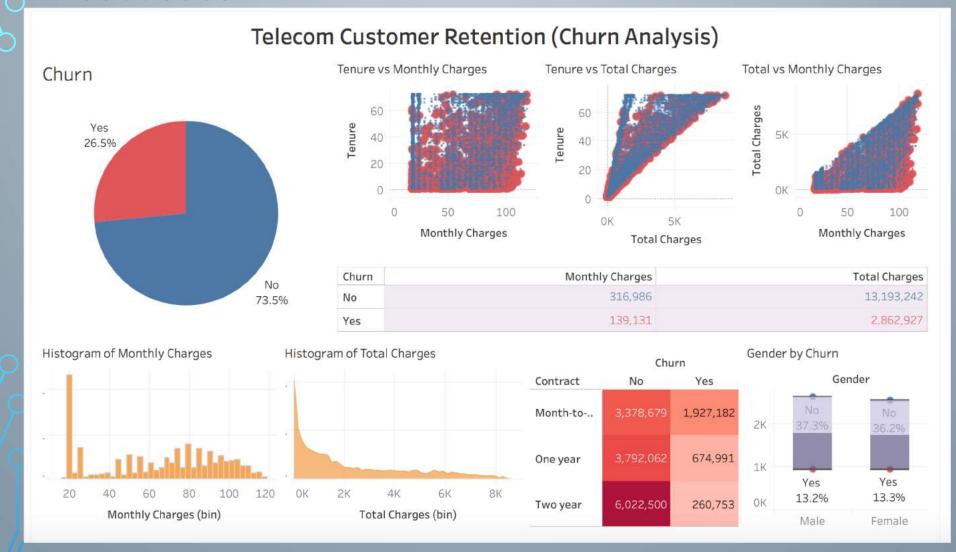
#### **Customer account information:**

- how long they've been a customer (tenure).
- monthly charges, and total charges



# **Exploratory Data Analysis** (on original data)





# **Data Pre-Processing**





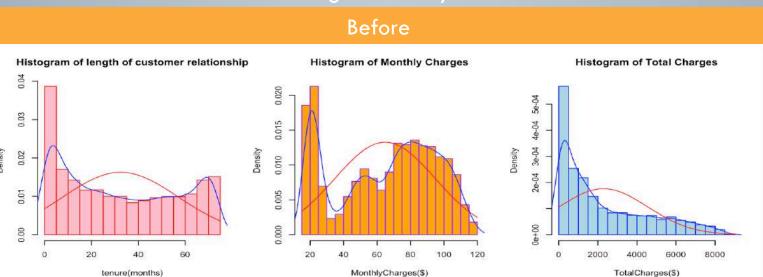
```
> str(Telcom2)
              7032 obs. of 20 variables:
'data.frame':
$ gender
                  : Factor w/ 2 levels "Female", "Male": 1 2 2 2 1 1 2 1 1 2 ...
$ SeniorCitizen : int 00000000000...
$ Partner
                  : Factor w/ 2 levels "No", "Yes": 2 1 1
$ Dependents
                  : Factor w/ 2 levels "No", "Yes": 1 1 1
$ tenure
                  : int 1 34 2 45 2 8 22 10 28 62 ...
$ PhoneService
                 : Factor w/ 2 levels "No", "Yes": 1 2 2
$ MultipleLines
                 : Factor w/ 2 levels "No", "Yes": 1 1 1
$ InternetService : Factor w/ 3 levels "DSL", "Fiber option
$ OnlineSecurity : Factor w/ 2 levels "No". "Yes": 1 2 2
$ OnlineBackup
                  : Factor w/ 2 levels "No", "Yes": 2 1 2
$ DeviceProtection: Factor w/ 2 levels "No". "Yes": 1 2 1
$ TechSupport
                  : Factor w/ 2 levels "No", "Yes": 1 1 1
$ StreamingTV
                  : Factor w/ 2 levels "No", "Yes": 1 1 1
$ StreamingMovies : Factor w/ 2 levels "No". "Yes": 1 1 1
                  : Factor w/ 3 levels "Month-to-month",
$ Contract
$ PaperlessBilling: Factor w/ 2 levels "No", "Yes": 2 1 2
                                                                              MultipleLines
                                                                                           OnlineSecurity
$ PaymentMethod : Factor w/ 4 levels "Bank transfer (au
$ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...
$ TotalCharges : num 29.9 1889.5 108.2 1840.8 151.7
$ Churn
                  : Factor w/ 2 levels "No", "Yes": 1 1 2
```

```
> str(Telcom)
'data.frame':
                7043 obs. of 21 variables:
                   : Factor w/ 7043 levels "0002-ORFBO", "0003-MKNFE", ...: 5376 3963 2565
$ customerID
2 1003 4771 5605 4535 ...
$ gender
                   : Factor w/ 2 levels "Female", "Male": 1 2 2 2 1 1 2 1 1 2 ...
$ SeniorCitizen
                  : int 0000000000 ...
$ Partner
                   : Factor w/ 2 levels "No", "Yes": 2 1 1 1 1 1 1 1 2 1 ...
                   : Factor w/ 2 levels "No", "Yes": 1 1 1 1 1 1 2 1 1 2 ...
$ Dependents
 $ tenure
                   : int 1 34 2 45 2 8 22 10 28 62 ...
                   : Factor w/ 2 levels "No", "Yes": 1 2 2 1 2 2 2 1 2 2 ...
 $ PhoneService
$ MultipleLines : Factor w/ 3 levels "No". "No phone service"...: 2 1 1 2 1 3 3 2 3 1
$ InternetService : Factor w/ 3 levels "DSL", "Fiber optic", ..: 1 1 1 1 2 2 2 1 2 1 ...
 $ OnlineSecurity : Factor w/ 3 levels "No", "No internet service",..: 1 3 3 3 1 1 1 3
                  : Factor w/ 3 levels "No", "No internet service", ..: 3 1 3 1 1 1 3 1
$ OnlineBackup
$ DeviceProtection: Factor w/ 3 levels "No","No internet service",..: 1 3 1 3 1 3 1 1
 3 TechSupport
                   : Factor w/ 3 levels "No", "No internet service", ...: 1 1 1 3 1 1 1 1
$ StreamingTV
                  : Factor w/ 3 levels "No", "No internet service", ..: 1 1 1 1 3 3 1
$ StreamingMovies: Factor w/ 3 levels "No", "No internet service"...: 1 1 1 1 1 3 1 1
                   : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1 1 1 2 ...
$ Contract
$ PaperlessBilling: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1 ...
$ PaymentMethod
                  : Factor w/ 4 levels "Bank transfer (automatic)"...: 3 4 4 1 3 3 2 4
$ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...
$ TotalCharges
                  : num 29.9 1889.5 108.2 1840.8 151.7 ...
                   : Factor w/ 2 levels "No", "Yes": 1 1 2 1 2 2 1 1 2 1 ...
$ Churn
```

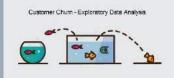


Customer Churn - Exploretory Data Analysis

Histogram Analysis

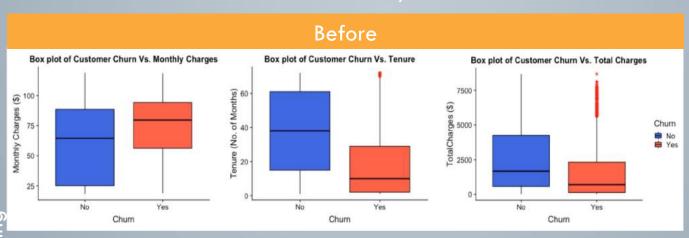


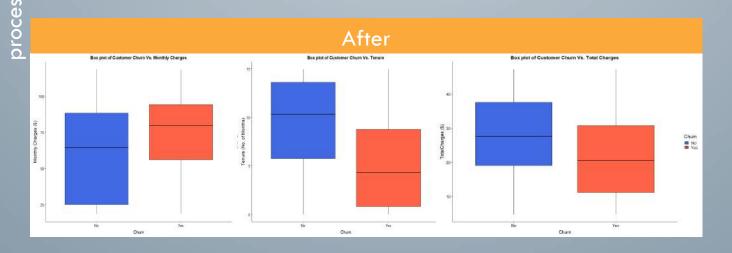




# **Exploratory Data Analysis**

**Box Plot Study** 





# Modeling



#### **ANOVA Table**

Analysis of Deviance Table

Model: binomial, link: logit

Response: Churn

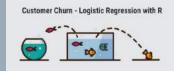
Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			4921	5699.5	
gender	1	0.79	4920	5698.7	0.3754116
SeniorCitizen	1	123.31	4919	5575.4	< 2.2e-16 ***
Partner	1	116.99	4918	5458.4	< 2.2e-16 ***
Dependents	1	32.02	4917	5426.4	1.522e-08 ***
PhoneService	1	2.11	4916	5424.3	0.1459938
MultipleLines	1	2.79	4915	5421.5	0.0947446 .
InternetService	2	474.34	4913	4947.1	< 2.2e-16 ***
OnlineSecurity	1	191.36	4912	4755.8	< 2.2e-16 ***
OnlineBackup	1	73.14	4911	4682.6	< 2.2e-16 ***
DeviceProtection	1	48.63	4910	4634.0	3.095e-12 ***
TechSupport	1	76.28	4909	4557.7	< 2.2e-16 ***
StreamingTV	1	0.27	4908	4557.5	0.6041299
StreamingMovies	1	0.10	4907	4557.4	0.7509795
Contract	2	299.16	4905	4258.2	< 2.2e-16 ***
PaperlessBilling	1	14.98	4904	4243.2	0.0001084 ***
PaymentMethod	3	29.12	4901	4214.1	2.112e-06 ***
MonthlyCharges	1	1.49	4900	4212.6	0.2221242
TotalCharges	1	230.40	4899	3982.2	< 2.2e-16 ***
Signif. codes: (	) ';	*** 0.001	L '**' 0.0	1 '*' 0.05	'.' 0.1' ' 1

#### Model

Log (P/1-P) = 0.8 (Gender) + 123.3 (Seniorcitizen) + 117.0 (Partner) + 32.0 (Dependents) + 2.1 (PhoneService) + 2.8 (MultipleLines) + 474.3 (InternetService) + 191.4 (OnlineSecurity) + 73.1 (OnlineBackup) + 48.6 (DeviceProtection) + 76.3 (TechSupport) + 0.3 (StreamingTV) + 0.1 (StreamingMovies) + 299.2 (Contract) + 15.0 (PaperlessBilling) + 29.1 (PaymentMethod) + 1.5 (MonthlyCharges) + 230.4 (TotalCharges)

# Modeling



#### **Binomial Logistic Regression**

```
Deviance Residuals:
              10
                  Median
                                3Q
                                        Max
-2.1388
         -0.6522 -0.2873
                            0.5835
                                     3.1278
Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)
                                      1.415771
                                                 0.981095
                                                            1.443 0.149006
genderMale
                                     -0.001872
                                                 0.079068
                                                            -0.024 0.981112
                                      0.356821
                                                 0.102329
                                                             3.487 0.000488 ***
SeniorCitizen1
                                      0.068337
PartnerYes
                                                 0.095452
                                                            0.716 0.474037
DependentsYes
                                     -0.158106
                                                 0.110564
                                                            -1.430 0.152716
                                      0.015158
                                                 0.070789
                                                            0.214 0.830445
tenure
                                                 0.790792
PhoneServiceYes
                                      0.604902
                                                            0.765 0.444312
MultipleLinesYes
                                      0.576043
                                                 0.216412
                                                            2.662 0.007773 **
                                      2.034555
                                                 0.968412
                                                            2.101 0.035648 *
InternetServiceFiber optic
InternetServiceNo
                                     -2.227301
                                                 0.982361
                                                            -2.267 0.023372 *
                                     -0.157673
                                                 0.217305
OnlineSecurityYes
                                                            -0.726 0.468094
OnlineBackupYes
                                      0.181373
                                                 0.212262
                                                             0.854 0.392841
DeviceProtectionYes
                                      0.286144
                                                 0.215435
                                                            1.328 0.184107
TechSupportYes
                                      0.015514
                                                 0.217194
                                                            0.071 0.943058
StreamingTVYes
                                      0.775448
                                                 0.396676
                                                            1.955 0.050599 .
StreamingMoviesYes
                                      0.728808
                                                 0.397728
                                                            1.832 0.066888 .
ContractOne year
                                     -0.628833
                                                 0.128487
                                                            -4.894 9.87e-07 ***
ContractTwo year
                                     -1.718207
                                                 0.226503
                                                           -7.586 3.30e-14 ***
PaperlessBillingYes
                                      0.365702
                                                 0.090826
                                                            4.026 5.66e-05
PaymentMethodCredit card (automatic) -0.045420
                                                 0.136989
                                                            -0.332 0.740225
PaymentMethodElectronic check
                                      0.229943
                                                 0.114007
                                                            2.017 0.043704
                                     -0.061775
PaymentMethodMailed check
                                                 0.137819
                                                            -0.448 0.653985
MonthlyCharges
                                     -0.035839
                                                 0.038675
                                                           -0.927 0.354099
TotalCharges
                                     -0.097058
                                                 0.033966
                                                           -2.858 0.004269 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 5699.5 on 4921 degrees of freedom
Residual deviance: 3982.2 on 4898 degrees of freedom
AIC: 4030.2
Number of Fisher Scoring iterations: 6
```

#### Model

```
Churn = 1.41 - 0.002 (GenderMale) + 0.36
(Seniorcitizen 1) + 0.07 (Partner Yes) - 0.16
(DependentsYes) + 0.02 (Tenure) + 0.06
(PhoneServiceYes) + 0.58 (MultipleLinesYes) + 2.03
(InternetServiceFiber optic) - 2.23 (InternetServiceNo)
- 0.16 (OnlineSecurityYes) + 0.18 (OnlineBackupYes)
+ 0.29 (DeviceProtectionYes) + 0.02 (TechSupportYes)
+ 0.78 (StreamingTVYes) + 0.73
(StreamingMoviesYes) – 0.63 (ContractOne year) -
1.72 (ContractTwo year) + 0.37 (PaperlessBillingYes) -
0.05 (PaymentMethodCredit card (automatic)) + 0.23
(PaymentMethodElectronic check) - 0.06
(PaymentMethodMailed check) -0.04
(MonthlyCharges) – 0.10 (TotalCharges)
```

# Customer Chum - Model Comparison

# **Model Comparison**

#### Initial Model

Coefficients:

Estimate Std. Error z value Pr(>|z|) (Intercept) 1.165e+00 8.151e-01 1.430 0.15284 genderMale -2.183e-02 6.480e-02 -0.337 0.73619 SeniorCitizenl 2.168e-01 8.453e-02 2.564 0.01033 \* -3.840e-04 7.783e-02 -0.0050.99606 PartnerYes DependentsYes -1.485e-01 8.973e-02 -1.655 0.09796 . tenure -6.059e-02 6.236e-03 -9.716 < 2e-16 \*\*\* PhoneServiceYes 1.715e-01 6.487e-01 0.264 0.79153 MultipleLinesYes 4.484e-01 1.773e-01 2.530 0.01142 \* 1.747e+00 7.981e-01 2.190 0.02855 \* InternetServiceFiber optic InternetServiceNo -1.786e+00 8.073e-01 -2.213 0.02691 \* OnlineSecurityYes -2.054e-01 1.787e-01 -1.150 0.25031 OnlineBackupYes 2.604e-02 1.754e-01 0.148 0.88197 DeviceProtectionYes 1.474e-01 1.764e-01 0.836 0.40339 TechSupportYes -1.805e-01 1.806e-01 -0.999 0.31759 StreamingTVYes 5.905e-01 3.263e-01 1.810 0.07035 . StreamingMoviesYes 5.993e-01 3.267e-01 1.834 0.06658 . ContractOne year -6.608e-01 1.076e-01 -6.142 8.15e-10 \*\*\* -1.357e+00 1.764e-01 -7.691 1.46e-14 \*\*\* ContractTwo year PaperlessBillingYes 3.424e-01 7.450e-02 4.596 4.31e-06 \*\*\* PaymentMethodCredit card (automatic) -8.779e-02 1.141e-01 -0.770 0.44156 PaymentMethodElectronic check 3.045e-01 9.450e-02 3.222 0.00127 \*\* PaymentMethodMailed check -5.759e-02 1.149e-01 -0.501 0.61627 MonthlyCharges -4.034e-02 3.176e-02 -1.270 0.20392 TotalCharges 3.289e-04 7.063e-05 4.657 3.20e-06 \*\*\*

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8143.4 on 7031 degrees of freedom coldual deviance: 5826.3 on 7008 degrees of freedom

AIC: 5874.3

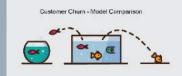
Number of Fisher Scoring iterations: 6

Pre-Processing

#### Final Model

```
Deviance Residuals:
   Min
             10 Median
                                    3.1278
-2.1388 -0.6522 -0.2873
                           0.5835
Coefficients:
                                     Estimate Std. Error z value Pr(>|z|)
                                                0.981095
(Intercept)
                                     1.415771
                                                           1.443 0.149006
genderMale
                                     -0.001872
                                                0.079068
                                                          -0.024 0.981112
SeniorCitizen1
                                     0.356821
                                                0.102329
                                                           3.487 0.000488 ***
PartnerYes
                                     0.068337
                                                0.095452
                                                           0.716 0.474037
DependentsYes
                                     -0.158106
                                                0.110564
                                                          -1.430 0.152716
                                     0.015158
                                                0.070789
                                                           0.214 0.830445
tenure
                                     0.604902
                                                0.790792
PhoneServiceYes
                                                           0.765 0.444312
                                     0.576043
                                                0.216412
                                                           2.662 0.007773 **
MultipleLinesYes
InternetServiceFiber optic
                                     2.034555
                                                0.968412
                                                           2.101 0.035648 *
InternetServiceNo
                                     -2.227301
                                                0.982361
                                                          -2.267 0.023372 *
                                     -0.157673
                                                0.217305
                                                          -0.726 0.468094
OnlineSecurityYes
OnlineBackupYes
                                     0.181373
                                                0.212262
                                                           0.854 0.392841
DeviceProtectionYes
                                     0.286144
                                                0.215435
                                                           1.328 0.184107
TechSupportYes
                                     0.015514
                                                0.217194
                                                           0.071 0.943058
StreamingTVYes
                                     0.775448
                                                0.396676
                                                           1.955 0.050599 .
StreamingMoviesYes
                                     0.728808
                                                0.397728
                                                           1.832 0.066888 .
ContractOne year
                                     -0.628833
                                                0.128487
                                                          -4.894 9.87e-07 ***
ContractTwo vear
                                     -1.718207
                                                0.226503
                                                          -7.586 3.30e-14 ***
PaperlessBillingYes
                                     0.365702
                                                0.090826
                                                           4.026 5.66e-05 ***
PaymentMethodCredit card (automatic) -0.045420
                                                0.136989
                                                          -0.332 0.740225
PaymentMethodElectronic check
                                     0.229943
                                                0.114007
                                                           2.017 0.043704 *
PaymentMethodMailed check
                                     -0.061775
                                                0.137819
                                                          -0.448 0.653985
MonthlyCharges
                                     -0.035839
                                                0.038675
                                                          -0.927 0.354099
TotalCharges
                                     -0.097058
                                                0.033966 -2.858 0.004269 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
   Null deviance: 5699.5 on 4921 degrees of freedom
  sidual deviance: 3982.2 on 4898 degrees of freedom
AIC: 4030.2
Number of Fisher Scoring iterations: 6
```

# **Model Comparison**



#### Confusion Matrix

Confusion Matrix and Statistics

Reference Prediction No Yes No 4639 833 Yes 524 1036

Accuracy: 0.807

95% CI: (0.7976, 0.8162)

No Information Rate: 0.7342 P-Value [Acc > NIR]: < 2.2e-16

Kappa : 0.478

Mcnemar's Test P-Value : < 2.2e-16

Sensitivity: 0.8985 Specificity: 0.5543 Pos Pred Value: 0.8478 Neg Pred Value: 0.6641 Prevalence: 0.7342 Detection Rate: 0.6597

Detection Rate: 0.6597
Detection Prevalence: 0.7782
Balanced Accuracy: 0.7264

'Positive' Class: No

Pre-Processing

#### Confusion Matrix

Confusion Matrix and Statistics

actual predicted No Yes No 3271 596 Yes 343 712

Accuracy : 0.8092

95% CI: (0.798, 0.8201)

No Information Rate: 0.7343 P-Value [Acc > NIR]: < 2.2e-16

Kappa : 0.479

Mcnemar's Test P-Value : < 2.2e-16

Sensitivity: 0.9051 Specificity: 0.5443

Pos Pred Value: 0.8459 Neg Pred Value: 0.6749

Prevalence: 0.7343

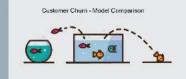
Detection Rate: 0.6646

Detection Prevalence: 0.7857

Balanced Accuracy : 0.7247

'Positive' Class: No

## **Random Forest**



OOB estimate of error rate: 19.99%

Confusion matrix:

No Yes class.error

No 3249 365 0.1009961

Yes 619 689 0.4732416

Low Error rate – "No"

High Error Rate — "Yes"

# Customer Chum - Results & Conclusion

## **Results and Conclusion**

#### Model

 $\begin{array}{l} \textbf{Churn} = 1.41 - 0.002 \ (\text{GenderMale}) + 0.36 \ (\text{Seniorcitizen1}) + 0.07 \ (\text{PartnerYes}) - 0.16 \ (\text{DependentsYes}) + 0.02 \ (\text{Tenure}) + 0.06 \ (\text{PhoneServiceYes}) + 0.58 \ (\text{MultipleLinesYes}) + 2.03 \ (\text{InternetServiceFiber optic}) - 2.23 \ (\text{InternetServiceNo}) - 0.16 \ (\text{OnlineSecurityYes}) + 0.18 \ (\text{OnlineBackupYes}) + 0.29 \ (\text{DeviceProtectionYes}) + 0.02 \ (\text{TechSupportYes}) + 0.78 \ (\text{StreamingTVYes}) + 0.73 \ (\text{StreamingMoviesYes}) - 0.63 \ (\text{ContractOne year}) - 1.72 \ (\text{ContractTwo year}) + 0.37 \ (\text{PaperlessBillingYes}) - 0.05 \ (\text{PaymentMethodCredit card (automatic)}) + 0.23 \ (\text{PaymentMethodElectronic check}) - 0.06 \ (\text{PaymentMethodMailed check}) - 0.04 \ (\text{MonthlyCharges}) - 0.10 \ (\text{TotalCharges}) \end{array}$ 

We can predict this model with accuracy of more than 80%, which can help company to know their future churn customers.

Company can use this model to know their churn customer in advance and can take appropriate action to retain their customer by resolving their concern and improving customer satisfaction.





https://github.com/KSomkul/finalproject



Analytics #ana625project

## References



https://www.kaggle.com/blastchar/telco-customer-churn

#### Image Sources: -

- https://www.optimove.com/wp-content/uploads/2014/02/Customer-Churn-Prediction-Prevention.png
- https://www.livechatinc.com/wp-content/uploads/2016/04/customer-churn@2x.jpg
- https://i.kym-cdn.com/photos/images/newsfeed/000/988/454/aa4.jpg
- https://hompal
  - stats.wabarr.com/presentations/categorical\_data/categorical\_data.html#3
- https://lc.gcumedia.com/hlt362v/the-visual-learner/images/continousdata.png





# Thank You Project Team



Shashi Bala MS — Data Analytics 2018-19



Kornkanok Somkul MS – Data Analytics 2018-19