Telecom customer churn prediction

-- Mihir Padsumbiya

Project Motivation

⇥ Increasing competition and rising need to protect the customers for a telecom company.

⇥ Continuous watch on customer behavior to predict whether a customer is more likely to leave the company in the near future.

⇥ Develop offer and discount programs catered specially according to the needs of the customers likely to churn to prevent the customers from leaving the company.

Image Beware of Subcontractors Trying to Steal Your Customers - PM Tips

Problem:

Telecom customers being poached by other companies or customers leaving for miscellaneous reasons. Company can take precautionary measures to prevent the customers from leaving.

Solution:

Business Intelligence models can be deployed to predict whether the customers will leave the company in the near future. Company can now take steps to lure the customers by providing special benefits etc.

About: Data

Source: IBM

Number of Rows: 7043

Number of Columns : 33

Number of Columns used : 33

Exploratory Data Analysis: Finding interesting patterns.

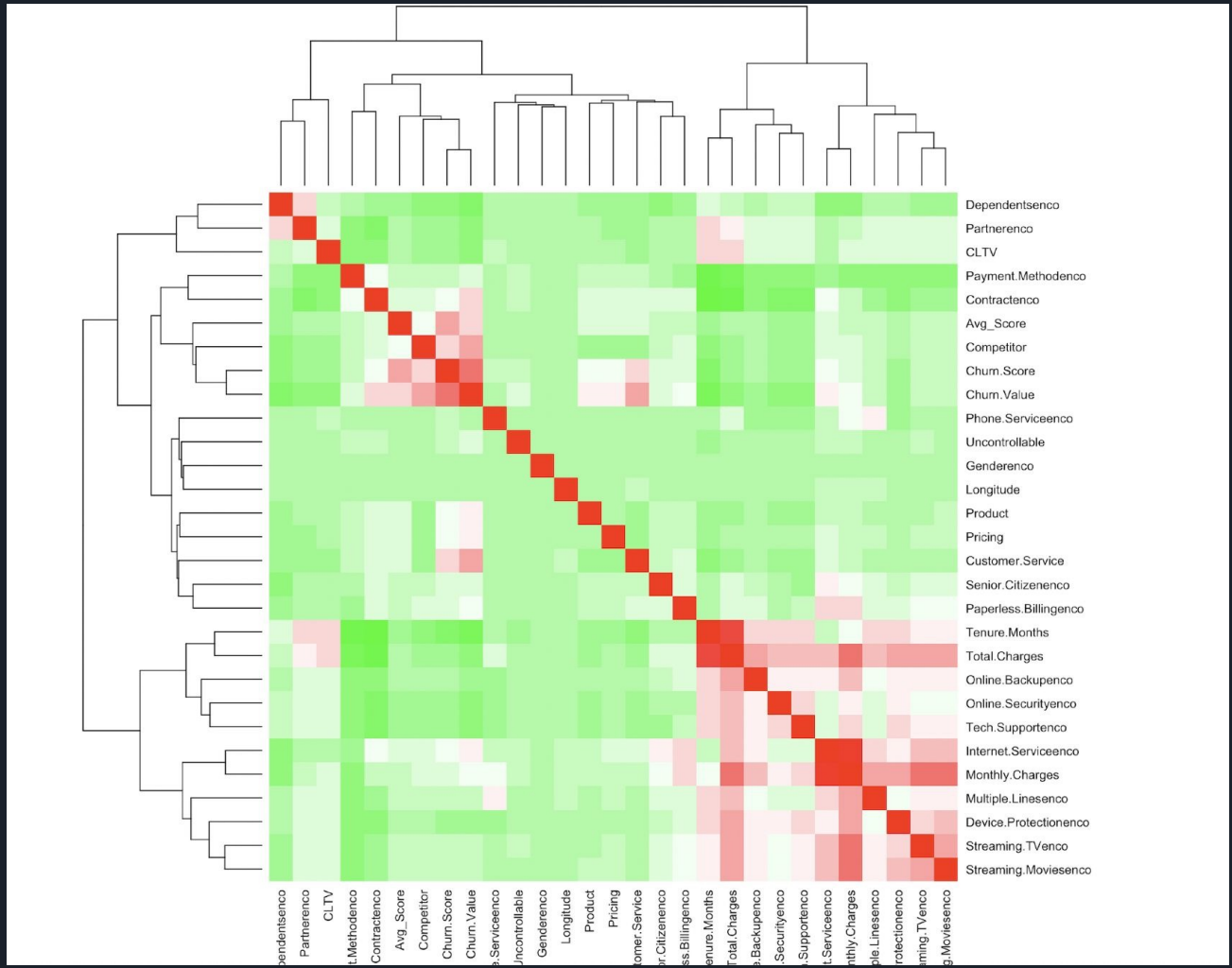
⇥ An interesting correlation between product variables. Shows that a customer is more likely to be using more than one products.

⇥ As expected, Tenure months and Total charges are highly correlated

⇥ A new moderate correlation b/w Contract encoded and Tenure Months

⇥ Monthly charges is moderately correlated with the product features

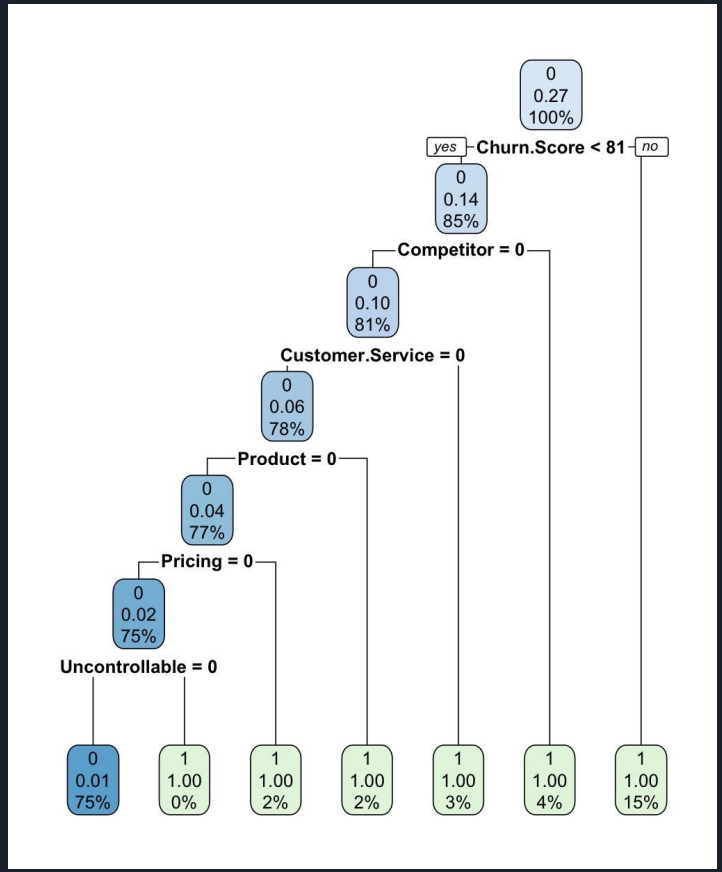
⇥ Zip code to be replaced with the mean of the churn scores for customers from that zipcode.


Dependentsenco
Partnerenco
CLTV
Payment.Methodenco
Contractenco
Avg_Score
Competitor
Churn.Score
Churn.Value
Phone.Serviceenco
Uncontrollable
Genderenco
Longitude
Product
Pricing
Customer.Service
Senior.Citizenenco
Paperless.Billingenco
Tenure.Months
Total.Charges
Online.Backupenco
Online.Securityenco
Tech.Supportenco
Internet.Serviceenco
Monthly.Charges
Multiple.Linesenco
Device.Protectionenco
Streaming.TVenco
Streaming.Moviesenco

BI Modelling

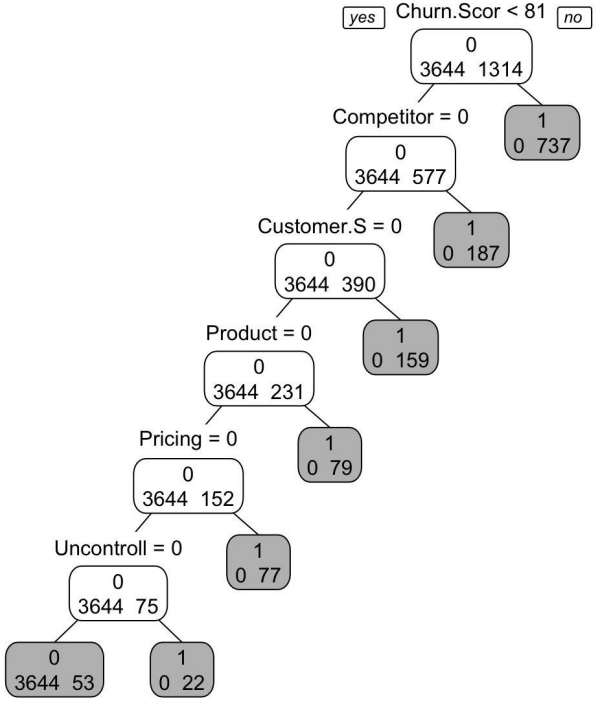
Decision Tree:

Predicted \ Reference		
	742	8
	0	280



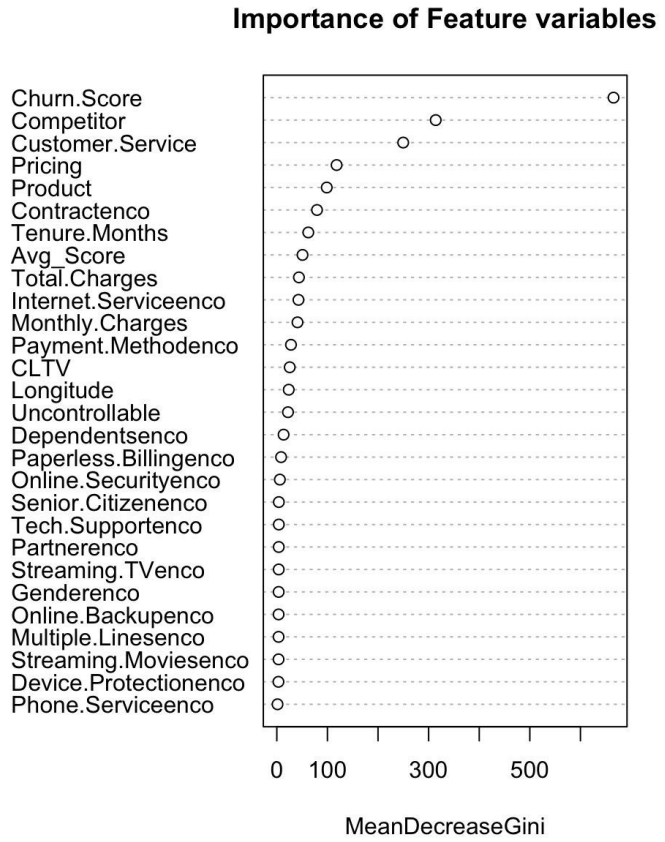
:Pruned Tree:

Predicted \ Reference		
	788	10
	0	257



Random Forests:

Predicted \ Reference		
	740	8
	2	280



Logistic Regression:

```
Coefficients:
                          Estimate    Std. Error z value Pr(>|z|)
(Intercept)            -24.10833954    9.02292251  -2.672  0.00754 **
Longitude               -0.00594947    0.07095547  -0.084  0.93318
Tenure.Months           -0.10418892    0.03542810  -2.941  0.00327 **
Monthly.Charges          0.15309021    0.14259081   1.074  0.28299
Total.Charges            0.00077539    0.00038750   2.001  0.04539 *
Churn.Score              0.24235571    0.02488269   9.740  < 2e-16 ***
CLTV                     0.00008177    0.00013071   0.626  0.53156
Customer.Service1       24.73868728 1165.78814154   0.021  0.98307
Competitor1             25.51356521 1012.74690042   0.025  0.97990
Product1                24.40061080 1897.03703769   0.013  0.98974
Pricing1                26.95780276 1414.76613154   0.019  0.98480
Uncontrollable1         25.23761372 3753.82567958   0.007  0.99464
Genderenco1             -0.50995526    0.30379573  -1.679  0.09323 .
Senior.Citizenenco1     -0.07532392    0.38415736  -0.196  0.84455
Partnerenco1             0.51459031    0.33153141   1.552  0.12062
Dependentsenco1         -2.32225173    0.67845608  -3.423  0.00062 ***
Phone.Serviceenco1      -3.49846165    2.94277943  -1.189  0.23451
Multiple.Linesenco1     -0.45555565    0.78304037  -0.582  0.56072
Online.Securityenco1    -1.64167243    0.80502257  -2.039  0.04142 *
Online.Backupenco1      -1.33725755    0.79001726  -1.693  0.09051 .
Device.Protectionenco1  -1.26462679    0.78728525  -1.606  0.10821
Tech.Supportenco1       -0.97599567    0.81592151  -1.196  0.23162
Streaming.TVenco1       -0.91420226    1.46190144  -0.625  0.53174
Streaming.Moviesenco1   -1.89667103    1.50552412  -1.260  0.20774
Paperless.Billingenco1   0.26062715    0.33916328   0.768  0.44222
Internet.Serviceenco1   -1.29489679    3.63244807  -0.356  0.72148
Internet.Serviceenco2   -5.29714732    7.15381772  -0.740  0.45902
Contractenco1           -1.05172060    1.15164821  -0.913  0.36112
Contractenco2            1.11297944    0.56538584   1.969  0.04901 *
Payment.Methodenco1      0.75925028    0.51300696   1.480  0.13887
Payment.Methodenco2      0.58708791    0.44276255   1.326  0.18485
Payment.Methodenco3     -0.35735324    0.57800846  -0.618  0.53641
Avg_Score                0.01668016    0.01550830   1.076  0.28212
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5733.92  on 4957  degrees of freedom
Residual deviance:  339.89  on 4925  degrees of freedom
AIC: 405.89
```

Our recommendation

Random Forest

Thank you!