**A4. Telco Churn**

1. Describe the process by which you cleaned, processed, and partitioned data as necessary. (1 point)

* Dropped 11 rows that contained NA.
* Made Churn 0/1 dummy variable instead of chr or factor for use as dependent variable.
* Made yes/no factors: SeniorCitizen, InternetService, MultipleLines, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovie, contract .
  + Combined all “No” and “No internet service” to “No”.
  + Combined all “No” and “No phone service” to “No”.
  + Combined “one year” and “two year” to “Yes” for contract, considered “month-to-month” no contract.
* Changed all non-numeric attributes from character to factor data types.
* Sub-setted phone only, internet only, and both customers.
* Partitioned data to 75/25 train/test with seed(1024) for repeatable results.

1. What predictors do you think contributes to the churn of (i) only telephone customers, (ii) only Internet service customers, and (iii) customers who subscribe to both phone and Internet services? Explain the rationale for your answer. (2 points)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Telephone effect/rationale** | **Internet effect/rationale** | **Both effect/rationale** |
| DV: Churn | | | |
| gender | (+/-) One gender may have a higher propensity for churn than the other | | |
| SeniorCitizen | (-) May want to keep familiar service | (+) May be inclined to switch providers based on promotions | (-) Same reason as phone only |
| Partner | (+) Multiple decision makers could double churn likelihood  (2 customers instead of 1) | | |
| Dependents | (+/-) Presence of dependents may or may not increase churn | | |
| tenure | (-) The longer a customer stays with a company the less likely they are to churn | | |
| MultipleLines | (-) May be more of a hassle to switch providers | Excluded | (-) May be more of a hassle to switch providers |
| OnlineSecurity | Excluded | (-) Additional related services may decrease churn likelihood | |
| OnlineBackup | Excluded | (-) May not want to lose or transfer backed-up content | |
| DeviceProtection | Excluded | (-) Additional related services may decrease churn likelihood | |
| TechSupport | Excluded | (+/-) Depends on efficacy and outcome of support services used: Good may decrease, poor may increase churn | |
| StreamingTV | Excluded | (-) Additional related services may decrease churn likelihood | |
| StreamingMovies | Excluded | (-) Additional related services may decrease churn likelihood | |
| Contract | (-) May not want to pay to break contract | | |
| PaperlessBilling | (-) Paperless bills could be more automatic with customer paying less attention to charges | | |
| PaymentMethod | (+) Mailed and electronic check: more of a manual process, more attention to charges (may group into “manual payment”)  (-) Bank transfer and credit card are automatic and may mean customer pays less attention to what is being charged every month (may group into “Automatic payment”) | | |
| MonthlyCharges | (+) Higher monthly charges may increase churn | | |
| *Excluded*: customerID (individual customers not needed for model), PhoneService & InternetService (used as attributes to subset groups, not needed as predictor), TotalCharges (combination of monthly charge and tenure) | | | |

1. Create training and test data sets with a 75:25 split using a random seed of 1024. Use the training data to train three logit models with the variables you identified in Question 2. Combine the outputs of the three modes using stargazer. (3 points)

> p\_logit <- glm(Churn ~ gender + SeniorCitizen + Partner + Dependents + tenure + MultipleLines + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges, data=p\_train, family=binomial(link=logit))

> i\_logit <- glm(Churn ~ gender + SeniorCitizen + Partner + Dependents + OnlineBackup + OnlineSecurity + DeviceProtection + TechSupport + StreamingTV + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges, data=i\_train, family=binomial(link=logit))

> pi\_logit <- glm(Churn ~ gender + SeniorCitizen + Partner + Dependents + tenure + MultipleLines + OnlineBackup + OnlineSecurity + DeviceProtection + TechSupport + StreamingMovies + StreamingTV + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges, data=pi\_train, family=binomial(link=logit))  
> stargazer(p\_logit, i\_logit, pi\_logit, type="text", single.row=TRUE)

======================================================================================================

Dependent variable:

---------------------------------------------------------------

Churn

(1) (2) (3)

------------------------------------------------------------------------------------------------------

genderMale -0.204 (0.575) 83.277 (71,559.160) -0.092 (0.100)

SeniorCitizenYes -12.354 (1,342.667) 123.633 (119,038.600) 0.160 (0.123)

PartnerYes 0.832 (0.843) 4.185 (133,704.500) 0.038 (0.120)

DependentsYes -0.427 (0.844) -59.201 (209,128.100) -0.076 (0.139)

tenure -0.054\* (0.031) -2.829 (2,823.556) -0.037\*\*\* (0.004)

MultipleLinesYes 1.868 (2.633) 0.003 (0.118)

OnlineBackupYes 337.822 (497,861.200) -0.390\*\*\* (0.116)

OnlineSecurityYes 351.120 (507,978.200) -0.509\*\*\* (0.125)

DeviceProtectionYes 320.750 (500,731.200) -0.155 (0.119)

TechSupportYes 439.638 (452,703.300) -0.666\*\*\* (0.123)

StreamingMoviesYes 707.943 (965,923.300) -0.120 (0.131)

StreamingTVYes 730.843 (1,151,053.000) -0.093 (0.131)

ContractYes -1.525\* (0.831) -104.880 (143,358.600) -0.646\*\*\* (0.161)

PaperlessBillingYes -0.183 (0.610) -39.975 (109,101.300) 0.114 (0.115)

PaymentMethodBank transfer (automatic) 0.249 (0.922) 108.995 (199,676.400) -0.266 (0.186)

PaymentMethodCredit card (automatic) -0.662 (1.137) 14.353 (143,486.200) -0.140 (0.184)

PaymentMethodElectronic check -0.099 (1.158) 35.500 (145,673.300) 0.196 (0.154)

MonthlyCharges -0.496 (0.516) -68.539 (99,863.100) 0.041\*\*\* (0.005)

Constant 8.994 (10.366) 1,632.536 (2,529,895.000) -2.234\*\*\* (0.314)

------------------------------------------------------------------------------------------------------

Observations 243 64 2,521

Log Likelihood -46.363 -0.000 -1,209.780

Akaike Inf. Crit. 118.726 36.000 2,457.560

======================================================================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

1. What are the top three predictors of churn of (i) only telephone customers, (ii) only Internet service customers, and (iii) customers who subscribe to both phone and Internet services. Explain using marginal effects how much each predictor contributes to churn occurrence. (3 points)
   1. Phone only (p\_logit):
      1. Having **multiple lines** increases odd of churn by 6.48 times the odds of not having multiple lines (exp(MultipleLinesYes) = 6.476965e+00)
      2. Having a **partner** increases odds of churn by 2.3 times the odds of not having a partner (exp(PartnerYes) =2.297027e+00)
      3. Paying by **automatic bank transfer** increases odds of churn by 1.28 times the odds of paying by mailed check (exp(PaymentMethodBank transfer (automatic) ) = 1.283098e+00)
   2. Internet only (i\_logit): (this model seems unrealistic)
      1. Having **streaming TV** increases odds of churn by an infinite amount of times over not having streaming services (exp(StreamingTVYes )=Inf)
      2. Streaming movies increases the odds of churn by an extremely large amount of times over the odds of not streaming movies (exp(StreamingMoviesYes) = 2.855722e+307)
      3. Having **tech support** increases odds of churn by an extremely large amount of times the odds of not having tech support (exp(TechSupportYes) = 8.556024e+190)
   3. Phone & Internet (pi\_logit):
      1. Paying by **electronic check** increases odds of churn by 1.22 times the odds of paying by mailed check (exp(PaymentMethodElectronic check) = 1.2168401)
      2. Being a **senior citizen** increases odds of chuirn by 1.17 times the odds of not being a senior citizen (exp(SeniorCitizenYes) = 1.1734535)
      3. Using paperless billing increases odds of churn by 1.12 times the odds of not using paperless billing (exp(PaperlessBillingYes)= 1.1204706)
2. Fit your models using test data, and compute recall, precision, F1-score, and AUC values for each of your three models. Create a table with these values. (2 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Recall** | **Precision** | **F1** | **AUC** |
| p\_logit | 0.99831 | 0.92543 | 0.96049 | 0.5043616 |
| I\_logit | 0.7987 | 0.7953 | 0.7970 | 0.5909091 |
| pi\_logit | 0.8514 | 0.8189 | 0.8348 | 0.7271424 |