OPIM 5604 – Predictive Modeling

Telecom Customer Churn Analysis

Presented By
Team 4

Yoonhye Sun
Vinay Reddy
Reza Rafa
Tarun Nattala

April 17th, 2023

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Agenda

- Business problem and objective
- Introduction to the dataset
- Overview of the data - preprocessing
- Application of disparate models
- Performance Comparison & choosing optimal model
- Business insights & Recommendations
- Conclusion

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Business Problem

Revenue

Increase in Churn rate directly effects the telecom company's revenue and profitability.

Dissatisfaction

Increase in churn rate means that customers are dissatisfied with the service.

Sustainability

Increase in churn rate impacts on business sustainability and business operations.

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Introduction to Dataset

- The dataset is churn data of a Telecommunications company that provides phone and internet services to 7043 customers in California and includes details about customer demographics, location, services, and current churn status.
- The dataset was acquired from IBM Cognos Analytics playground.

ⓒSaebyeol Yu. Saebyeol's PowerPoint

SEMMA - Sample

- Target Variable: Churn Label (categorical variable)
- 33 columns and 7043 rows

33/0 Cols

	CustomerID	Count	Country	State	City	Zip Code	Latitude	Longitude	Gender	Senior Citizen	Partner	Dependents	Tenure Months
1	3668-QPYBK	1	United States	California	Los Angeles	90003	33.964131	-118.272783	Male	No	No	No	2
2	9237-HQITU	1	United States	California	Los Angeles	90005	34.059281	-118.307420	Female	No	No	Yes	2
3	9305-CDSKC	1	United States	California	Los Angeles	90006	34.048013	-118.293953	Female	No	No	Yes	8
4	7892-POOKP	1	United States	California	Los Angeles	90010	34.062125	-118.315709	Female	No	Yes	Yes	28
5	0280-XJGEX	1	United States	California	Los Angeles	90015	34.039224	-118.266293	Male	No	No	Yes	49

Phone Service	Multiple Lines	Internet Service	Online Security	Online Backup	Device Protection	Tech Support	Streaming TV	Streaming Movies
Yes	No	DSL	Yes	Yes	No	No	No	No
Yes	No	Fiber optic	No	No	No	No	No	No
Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes
Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes
Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes

Contract	Paperless Billing	Payment Method	Monthly Charges	Total Charges	Churn Label	Churn Value	Churn Score	CLTV	Churn Reason	Churn Reason 2
Month-to-month	Yes	Mailed check	53.85	108.15	Yes	1	86	3239	Competitor mad...	Competitor made better offer
Month-to-month	Yes	Electronic check	70.7	151.65	Yes	1	67	2701	Moved	Moved
Month-to-month	Yes	Electronic check	99.65	820.5	Yes	1	86	5372	Moved	Moved
Month-to-month	Yes	Electronic check	104.8	3046.05	Yes	1	84	5003	Moved	Moved
Month-to-month	Yes	Bank transfer (au...	103.7	5036.3	Yes	1	89	5340	Competitor had ...	Competitor had better devices

erPoint

SEMMA - Sample

Specify Data Proportions

Training Set: 0.6

Validation Set: 0.2

Test Set: 0.2

- ○ Don't Alter Group Proportions
- ○ Alter Proportions in Training
- ◉ Alter Proportions in Both Training and Validation

Select Focal Group

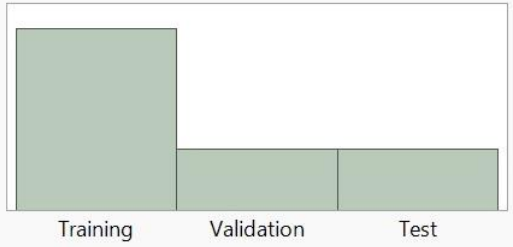
- No
- Yes

Focal Group Proportion: 0.5

- ☐ Balance Remaining Groups (Default is to maintain present ratios)
- ☐ Balance All Groups (Focal group proportion will be ignored)
- ☐ Bootstrap Augmentation (Leave unchecked for random trimming)

Distributions Churn Label=No

Validation Stratified by Churn Label



Frequencies

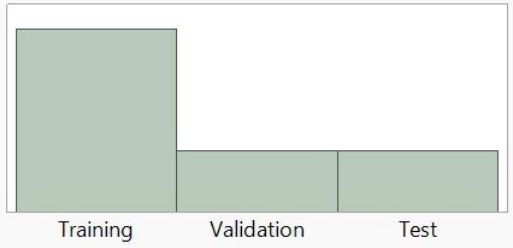
Level	Count	Prob
Training	3097	0.59985
Validation	1035	0.20046
Test	1031	0.19969
Total	5163	1.00000

N Missing 0

3 Levels

Distributions Churn Label=Yes

Validation Stratified by Churn Label



Frequencies

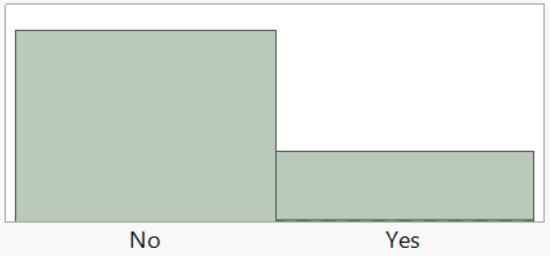
Level	Count	Prob
Training	1121	0.59979
Validation	374	0.20011
Test	374	0.20011
Total	1869	1.00000

N Missing 0

3 Levels

Distributions Validation Stratified by Churn Label=Training

Churn Label



Frequencies

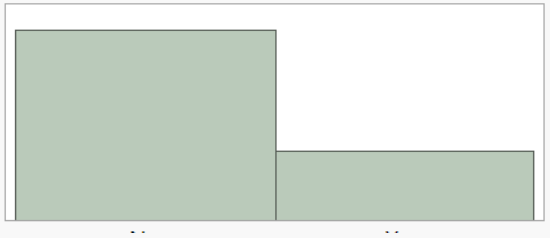
Level	Count	Prob
No	3098	0.73430
Yes	1121	0.26570
Total	4219	1.00000

N Missing 0

2 Levels

Distributions Validation Stratified by Churn Label=Validation

Churn Label



Frequencies

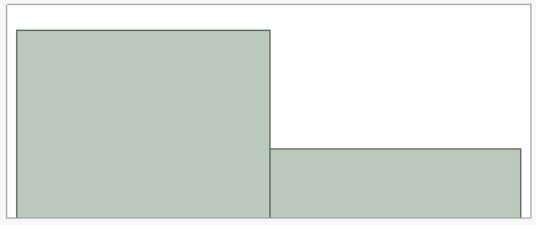
Level	Count	Prob
No	1032	0.73400
Yes	374	0.26600
Total	1406	1.00000

N Missing 0

2 Levels

Distributions Validation Stratified by Churn Label=Test

Churn Label



Frequencies

Level	Count	Prob
No	1033	0.73419
Yes	374	0.26581
Total	1407	1.00000

N Missing 0

2 Levels

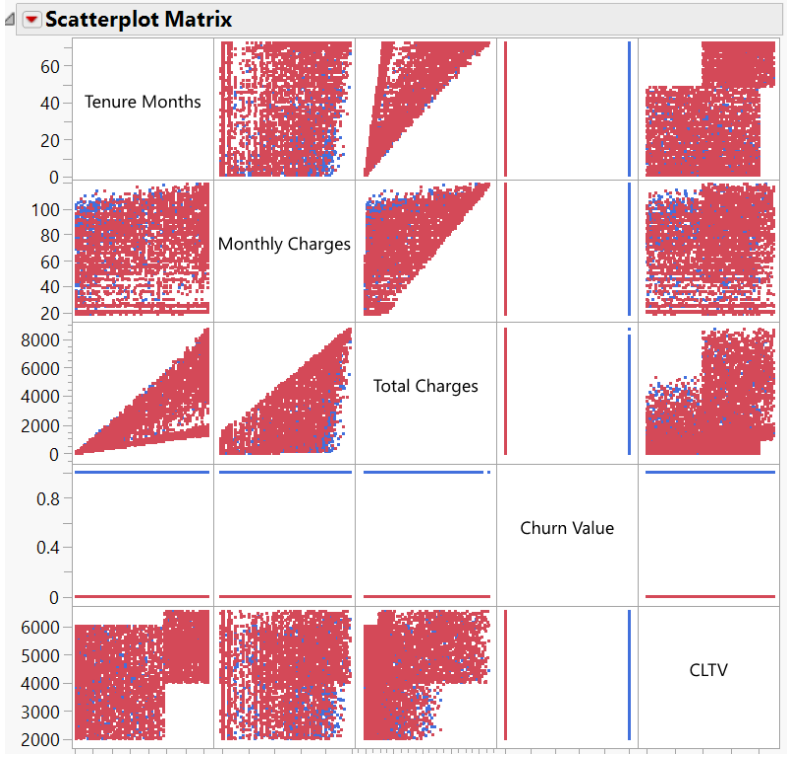
SEMMA - Explore

- Utilized Multivariate to find correlations between a target variable(categorical variable) and Predictor variables (continuous variables)

Multivariate

Correlations

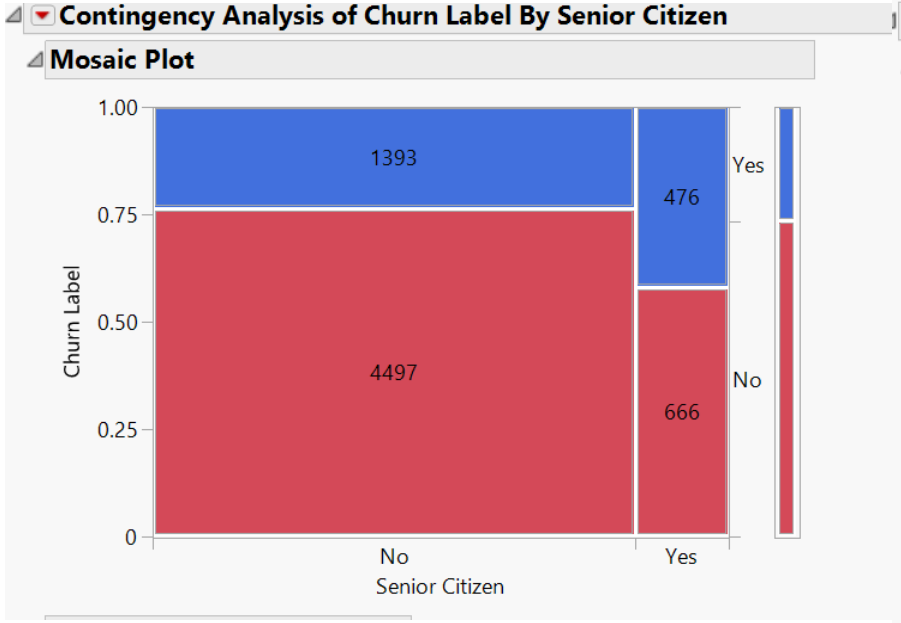
	Tenure Months	Monthly Charges	Total Charges	Churn Value	CLTV
Tenure Months	1.0000	0.2469	0.8259	-0.3540	0.3962
Monthly Charges	0.2469	1.0000	0.6511	0.1929	0.0986
Total Charges	0.8259	0.6511	1.0000	-0.1995	0.3418
Churn Value	-0.3540	0.1929	-0.1995	1.0000	-0.1283
CLTV	0.3962	0.0986	0.3418	-0.1283	1.0000

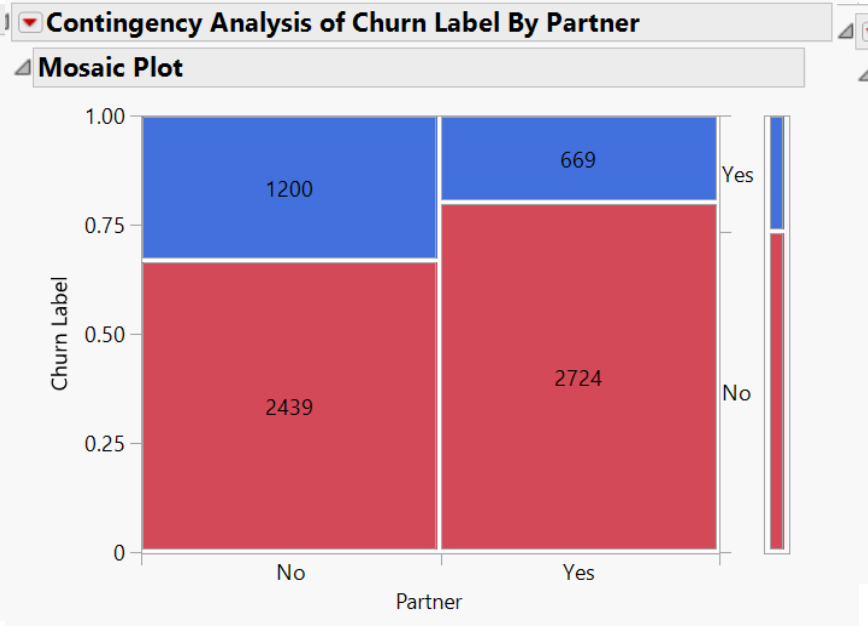


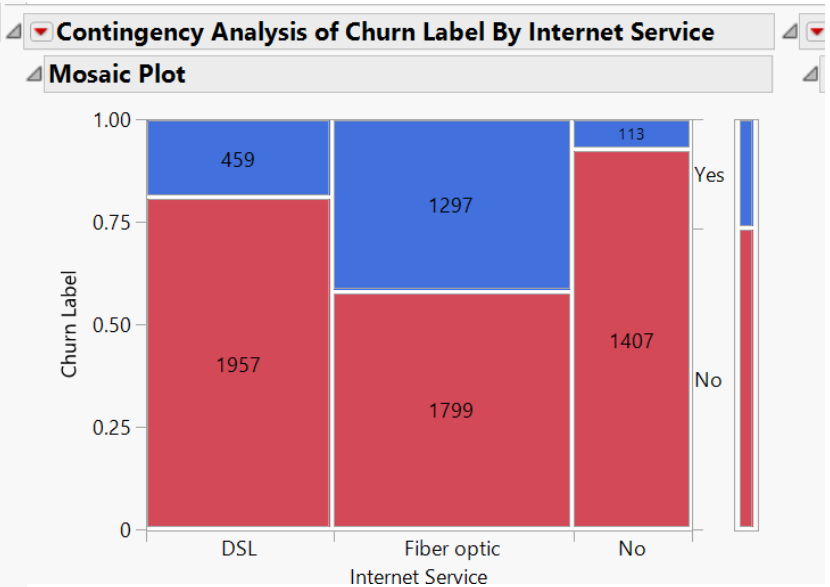
ⓒSaebyeol Yu. Saebyeol's PowerPoint

sEMMA - Explore

- Utilized Y by X to find correlations between target variables(categorical variable) and Predictor variables (categorical variables)
- Find Mosaic Plot that 'Yes' and 'No' are distinct



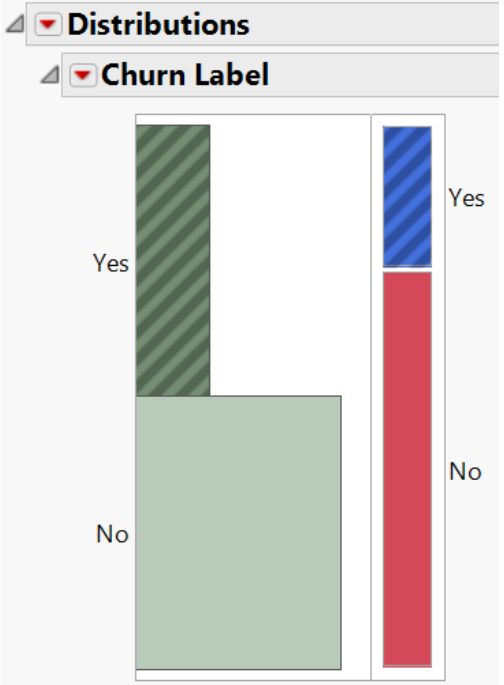




ⓒSaebyeol Yu. Saebyeol's PowerPoint

sEMMA - Explore

- Distribution of Target Variable: (Churn label)



Distributions

Churn Label

Frequencies

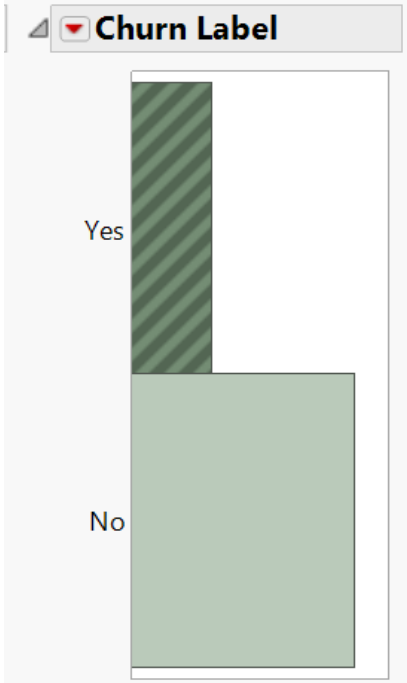
Level	Count	Prob
No	5163	0.73422
Yes	1869	0.26578
Total	7032	1.00000
N Missing	0	

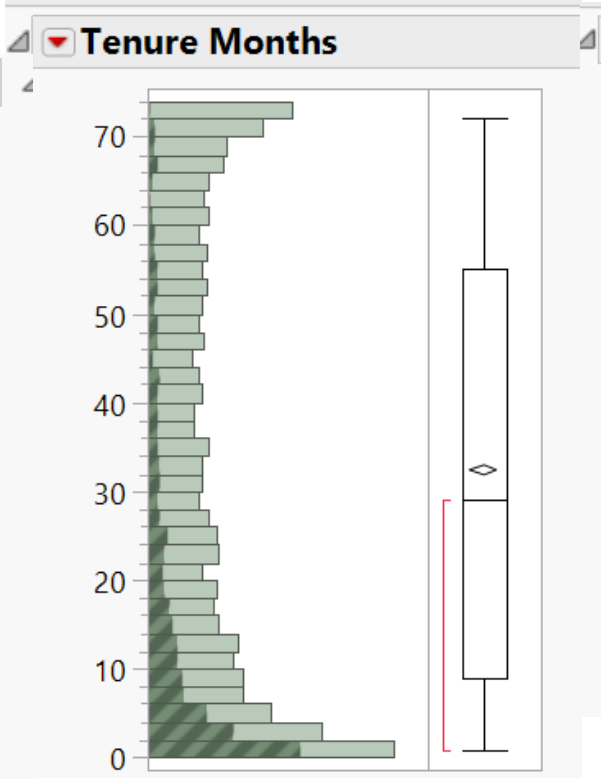
Yes-> 27%
No-> 73%

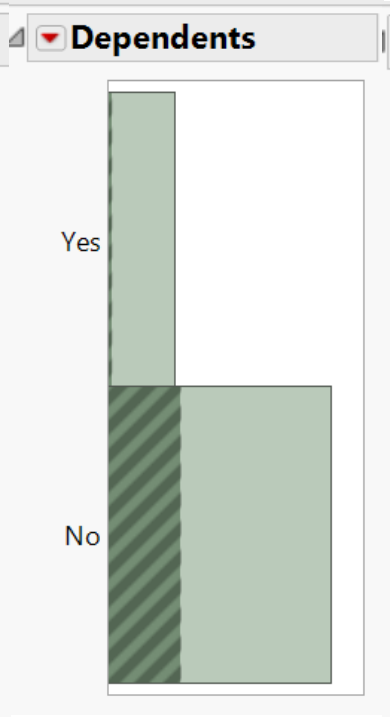
ⓒSaebyeol Yu. Saebyeol's PowerPoint

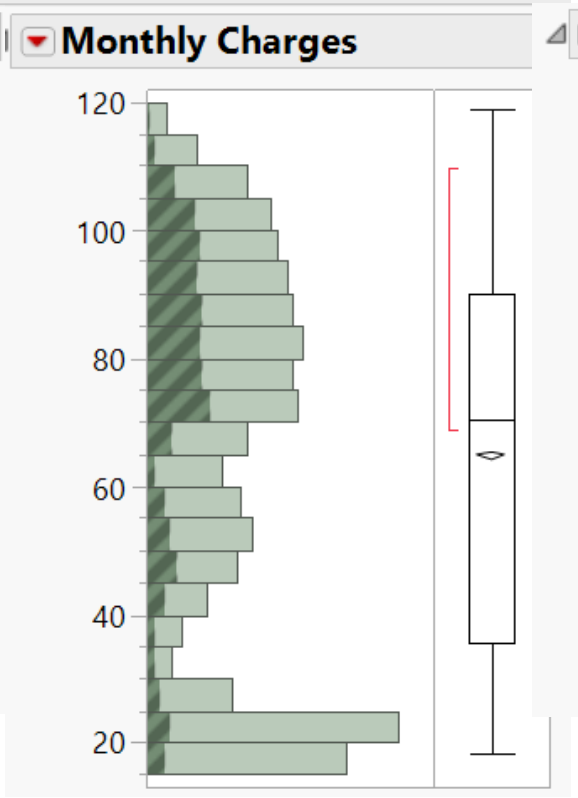
SEMMA - Explore

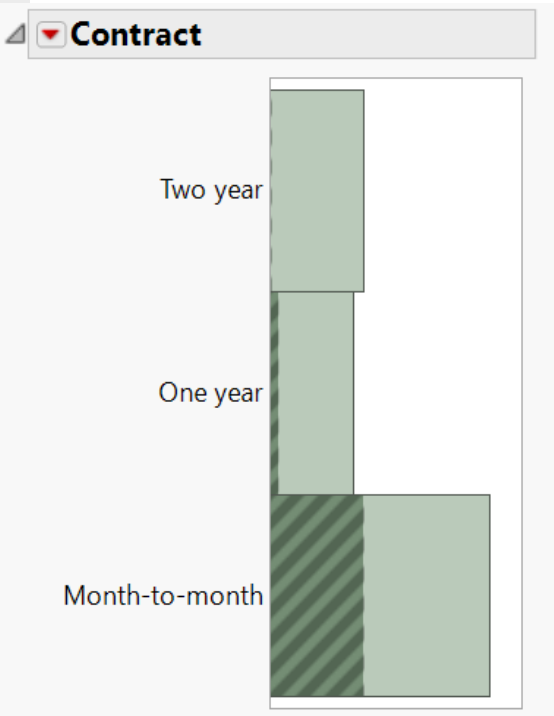
- Interesting findings:











Frequencies

Level	Count	Prob
No	5163	0.73422
Yes	1869	0.26578
Total	7032	1.00000

N Missing 0

2 Levels

Summary Statistics

Mean	32.421786
Std Dev	24.54526
Std Err Mean	0.2927037
Upper 95% Mean	32.995574
Lower 95% Mean	31.847999
N	7032

ⓒSaebyeol Yu. Saebyeol's PowerPoint

sEMMA - Explore

Predictor Screening

Predictor	Churn Label Contribution	Churn Label Portion	Rank
Contract	667.437	0.3203	1
Tenure Months	305.993	0.1468	2
Monthly Charges	271.927	0.1305	3
Dependents	193.787	0.0930	4
Total Charges	145.725	0.0699	5
Internet Service	134.301	0.0644	6
CLTV	67.402	0.0323	7
Payment Method	60.711	0.0291	8
Partner	28.545	0.0137	9
Paperless Billing	26.969	0.0129	10
Streaming Movies 2	26.859	0.0129	11
Online Security 2	25.599	0.0123	12
Tech Support 2	24.556	0.0118	13
Streaming TV 2	20.444	0.0098	14
Multiple Lines 2	20.336	0.0098	15
Phone Service	18.484	0.0089	16
Senior Citizen	18.217	0.0087	17
Online Backup 2	16.078	0.0077	18
Device Protection 2	10.496	0.0050	19

Copy

1. Contract
2. Tenure Months
3. Monthly Charges
4. Dependents
5. Total Charges
6. Internet service

contributes > 82%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

SEMMA - Modify (Missing values, Outliers)

- There are 11 missing values in the total charges ,It's better to hide and exclude them than imputing.
- By using the inter quantile range outlier we explored the outliers.

Missing Columns

☐ Show only columns with missing

Close

Select columns and choose an action.

Select Rows | Color Cells

Exclude Rows | Color Rows

Column	Number Missing
Senior Citizen	0
Partner	0
Dependents	0
Tenure Months	0
Phone Service	0
Multiple Lines 2	0
Internet Service	0
Online Security 2	0
Online Backup 2	0
Device Protection 2	0
Tech Support 2	0
Streaming TV 2	0
Streaming Movies 2	0
Contract	0
Paperless Billing	0
Payment Method	0
Monthly Charges	0
Total Charges	11
Churn Label	0
CLTV	0
Validation Stratified by Churn Label	0

Rows	
All rows	7,043
Selected	1,869
Excluded	11
Hidden	11
Labeled	0

Column	10% Quantile	90% Quantile	Low Threshold	High Threshold	Number of Outliers	Outliers (Count)
Tenure Months	2	69	-199	270	0	
Monthly Charges	20.05	102.65	-227.75	350.45	0	
Total Charges	84.53	5978.86	-17598	23661.9	0	
CLTV	2613	5865	-7143	15621	0	

Recode - Churn Reason - JMP Pro

New Column | Name: Churn Reason 2

Count	Old Values (21)	New Values (21)
5174		N/A
135	Attitude of service provider	Attitude of service provider
192	Attitude of support person	Attitude of support person
130	Competitor had better devices	Competitor had better devices
140	Competitor made better offer	Competitor made better offer
189	Competitor offered higher download speeds	Competitor offered higher download speeds
162	Competitor offered more data	Competitor offered more data
6	Deceased	Deceased
154	Don't know	Don't know
57	Extra data charges	Extra data charges
44	Lack of affordable download/upload speed	Lack of affordable download/upload speed
88	Lack of self-service on Website	Lack of self-service on Website
44	Limited range of services	Limited range of services
44	Long distance charges	Long distance charges
53	Moved	Moved
103	Network reliability	Network reliability
19	Poor expertise of online support	Poor expertise of online support
20	Poor expertise of phone support	Poor expertise of phone support
98	Price too high	Price too high
102	Product dissatisfaction	Product dissatisfaction
89	Service dissatisfaction	Service dissatisfaction

ⓒSaebyeol Yu. Saebyeol's PowerPoint

SEMMA-Modify (Recoded data)

No-> No
No internet/phone service-> No
Yes-> Yes
We have modified the old values to new values

New Column Name: Multiple Lines 2

Count	Old Values (3)	New Values (3)
3390	No	No
682	No phone service	No
2971	Yes	Yes

New Column Name: Online Security 3

Count	Old Values (3)	New Values (3)
3498	No	No
1526	No internet service	No
2019	Yes	Yes

New Column Name: Device Protection 2

Count	Old Values (3)	New Values (3)
3095	No	No
1526	No internet service	No
2422	Yes	Yes

New Column Name: Tech Support 2

Count	Old Values (3)	New Values (3)
3473	No	No
1526	No internet service	No
2044	Yes	Yes

New Column Name: Streaming TV 2

Count	Old Values (3)	New Values (3)
2810	No	No
1526	No internet service	No
2707	Yes	Yes

New Column Name: Online Backup 2

Count	Old Values (3)	New Values (3)
3088	No	No
1526	No internet service	No
2429	Yes	Yes

New Column Name: Streaming Movies 2

Count	Old Values (3)	New Values (3)
2785	No	No
1526	No internet service	No
2732	Yes	Yes

SEMMA-Modify (Hide and Exclude columns)

- RowID
- CustomerID
- Count
- Country
- State
- City
- Zip Code
- Lat Long
- Latitude
- Longitude
- Gender
- Senior Citizen
- Partner
- Dependents
- Tenure Months
- Phone Service
- Multiple Lines
- Multiple Lines 2
- Internet Service
- Online Security
- Online Security 2
- Online Backup
- Online Backup 2
- Device Protection
- Device Protection 2
- Tech Support
- Tech Support 2
- Streaming TV
- Streaming TV 2
- Streaming Movies
- Streaming Movies 2
- Contract
- Paperless Billing
- Payment Method
- Monthly Charges
- Total Charges
- Churn Label
- Churn Value
- Churn Score
- CLTV
- Churn Reason

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Predictor Variables

- Senior citizen
- Partner
- Dependents
- Tenure Months
- Phone Service
- Multiple Lines 2
- Internet Service
- Online Security 2
- Online Backup2
- Device Protection 2
- Tech Support 2
- Streaming Tv2
- Contract
- Paperless Billing
- Payment Method
- Monthly charges
- CLTV

ⓒSaebyeol Yu. Saebyeol's PowerPoint

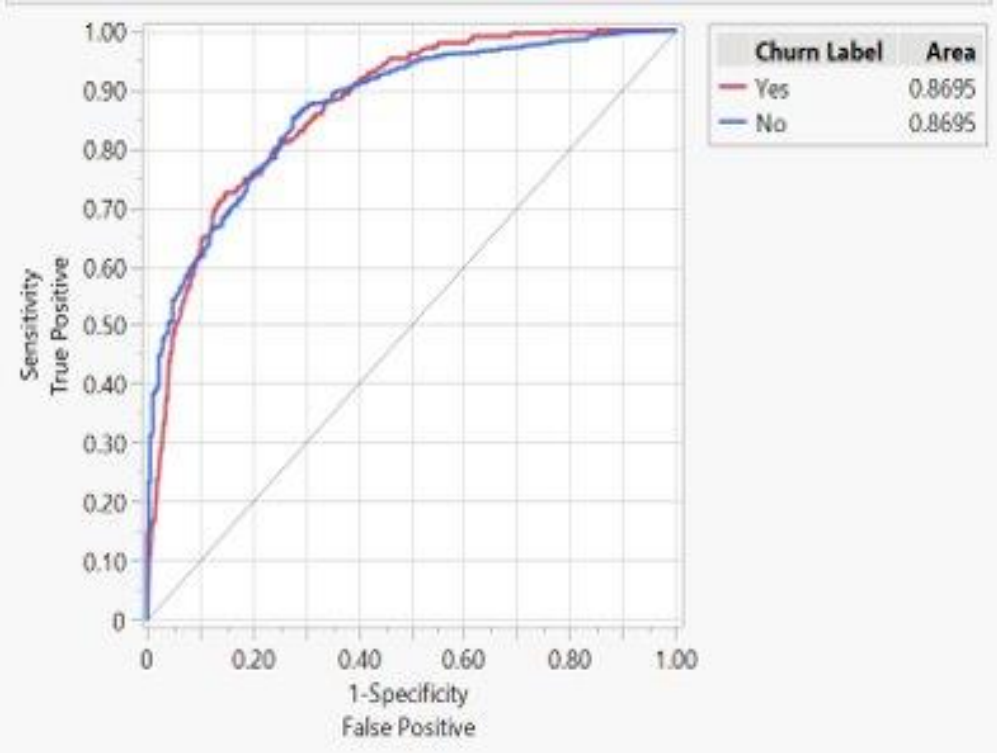
Variables	Status (Yes/No)	Reason
Customer ID	No	The mosaic plot doesn't show any significant difference to consider it as a predictor
Count	No	All variables are '1'
Country	No	There are many categorical variables
State	No	All variables are 'California'
City	No	There are many categorical variables
Zip Code	No	There are many continuous variables
Lat Long	No	There is no variable
Latitude	No	There are many continuous variables
Longitude	No	There are many continuous variables
Gender	No	The likelihood ration is low : 0.513
Senior Citizen	Yes	The likelihood ration: 148.131
Partner	Yes	The likelihood ration: 160.093
Dependents	Yes	The likelihood ration: 528.99
Tenure Months	Yes	The correlation rate: -0.354
Phone Service	Yes	The likelihood ration: 0.972
Multiple Lines 2	Yes	The likelihood ration: 11.226
Internet Service	Yes	The likelihood ration: 779.057
Online Security 2	Yes	The likelihood ration: 223.493
Online Backup 2	Yes	The likelihood ration: 48.722
Device Protection 2	Yes	The likelihood ration: 31.358
Tech Support 2	Yes	The likelihood ration: 205.538
Streaming TV 2	Yes	The likelihood ration: 27.88
Streaming Movies 2	Yes	The likelihood ration: 25.829
Contract	Yes	The likelihood ration: 1380.834
Paperless Billing	Yes	The likelihood ration: 268.893
Payment Method	Yes	The likelihood ration: 312
Monthly Charges	Yes	The correlation rate: 0.1929
Total Charges	Yes	The correlation rate: -0.1995
Churn Label	Yes	It is our target variable
Churn Value	No	It is same variable as our target variable
Churn Score	No	There is no information in data dictionary
CLTV	Yes	The correlation rate: -0.1283
Churn Reason	No	It doesn't impact to make model

ⓒSaebyeol Yu. Saebyeol's PowerPoint

SEMMA - (Modeling)

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-1.4211759	0.1723863	67.97	<.0001*
Partner[No]	-0.1484991	0.0488394	9.25	0.0024*
Dependents[No]	0.69759965	0.0728555	91.68	<.0001*
Tenure Months	-0.0533932	0.0080656	43.82	<.0001*
Phone Service[No]	0.31050902	0.08694	12.76	0.0004*
Multiple Lines 2[No]	-0.1217223	0.0524867	5.38	0.0204*
Internet Service[DSL]	0.0436266	0.0842328	0.27	0.6045
Internet Service[Fiber optic]	0.82918797	0.0831345	99.48	<.0001*
Online Security 2[No]	0.17898717	0.0560596	10.19	0.0014*
Tech Support 2[No]	0.2579766	0.0569363	20.53	<.0001*
Streaming TV 2[No]	-0.1607655	0.051684	9.68	0.0019*
Contract[Month-to-month]	0.88525087	0.1087284	66.29	<.0001*
Contract[One year]	0.00251809	0.1116302	0.00	0.9820
Paperless Billing[No]	-0.153834	0.0488329	9.92	0.0016*
Total Charges	0.00022857	8.9722e-5	6.49	0.0108*



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Logistic Regression Model:

Source	LogWorth	PValue
Dependents	24.719	0.00000
Internet Service	24.267	0.00000
Contract	16.817	0.00000
Tenure Months	11.945	0.00000
Tech Support 2	5.311	0.00000
Phone Service	3.435	0.00037
Online Security 2	2.874	0.00134
Paperless Billing	2.795	0.00160
Streaming TV 2	2.742	0.00181
Partner	2.645	0.00226
Total Charges	2.022	0.00950
Multiple Lines 2	1.694	0.02025

Significant variables contributing to the model:

- Dependents
- Internet Service
- Contract
- Tenure Months
- Tech Support 2
- Phone Service
- Online Security 2
- Paperless Billing

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.3103	0.2560	0.3248	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.4401	0.3740	0.4570	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3993	0.4310	0.3910	∑ -Log(ρ[j])/n
RASE	0.3600	0.3763	0.3549	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2605	0.2733	0.2523	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1906	0.2055	0.1727	∑ (ρ[j]≠ρMax)/n
N	4219	1406	1407	n

Training

Actual Churn Label	Predicted Count Yes	Predicted Count No
Yes	658	463
No	341	2757

Actual Churn Label	Predicted Rate Yes	Predicted Rate No
Yes	0.587	0.413
No	0.110	0.890

Validation

Actual Churn Label	Predicted Count Yes	Predicted Count No
Yes	203	171
No	118	914

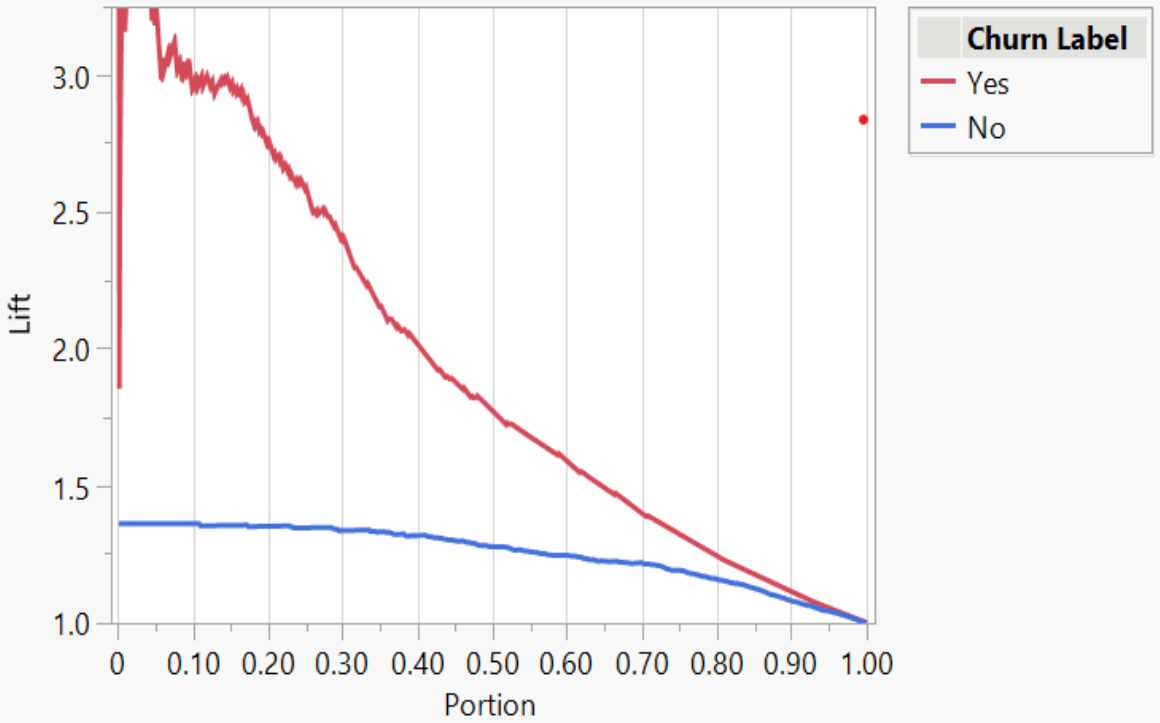
Actual Churn Label	Predicted Rate Yes	Predicted Rate No
Yes	0.543	0.457
No	0.114	0.886

Test

Actual Churn Label	Predicted Count Yes	Predicted Count No
Yes	225	149
No	94	939

Actual Churn Label	Predicted Rate Yes	Predicted Rate No
Yes	0.602	0.398
No	0.091	0.909

Lift Curve on Test Data



Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Logistic Regression	82.73%	70.53%	2.6532	86.95%	17.27%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Boosted Tree Model:

Overall Statistics

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.4052	0.2480	0.3208	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.5460	0.3640	0.4524	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3444	0.4356	0.3933	∑ -Log(ρ[j])/n
RASE	0.3330	0.3772	0.3575	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2376	0.2684	0.2521	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1590	0.2112	0.1777	∑ (ρ[j]≠ρMax)/n
N	4219	1406	1407	n

Confusion Matrix

Training

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	2859	239
Yes	432	689

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.923	0.077
Yes	0.385	0.615

Validation

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	921	111
Yes	186	188

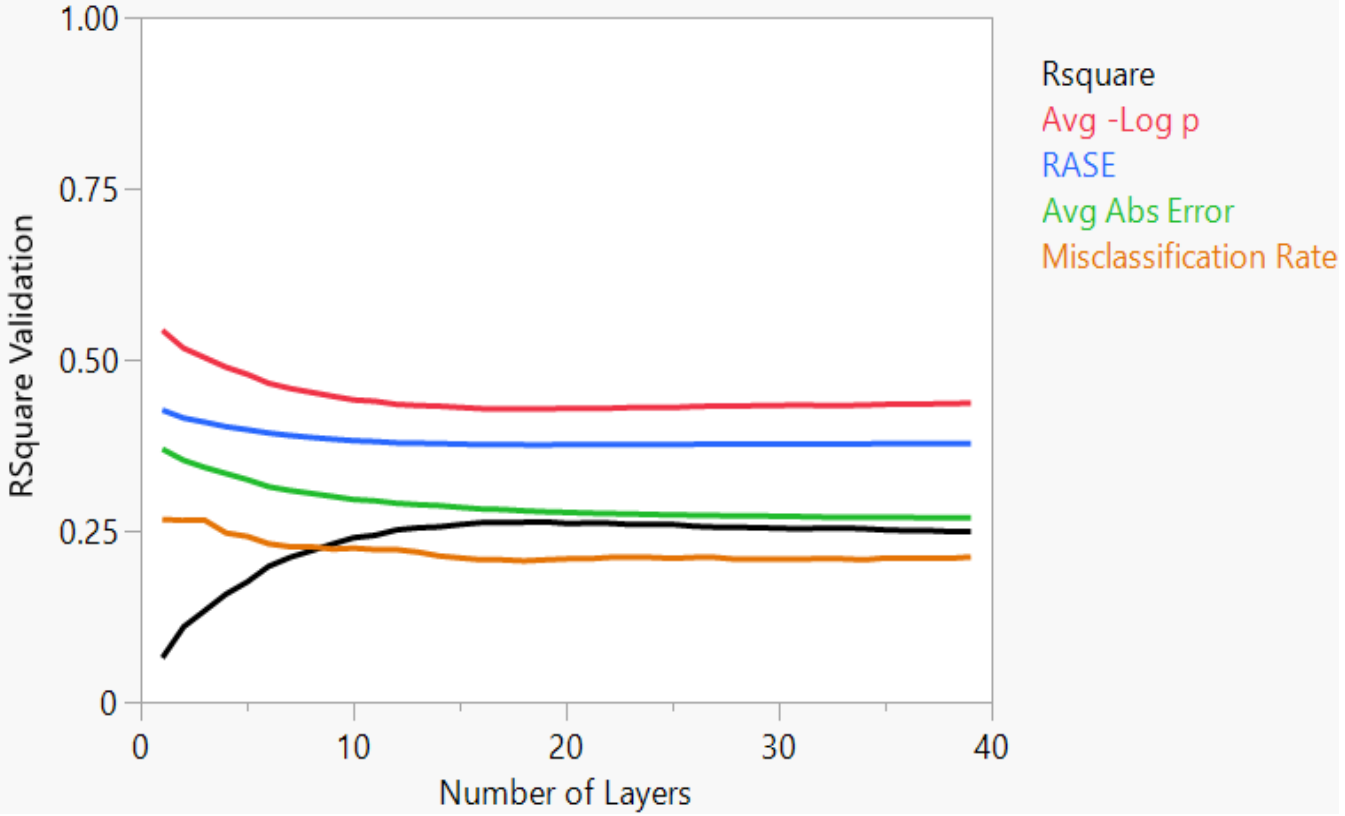
Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.892	0.108
Yes	0.497	0.503

Test

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	954	79
Yes	171	203

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.924	0.076
Yes	0.457	0.543

Cumulative Validation



ⓒSaebyeol Yu. Saebyeol's PowerPoint

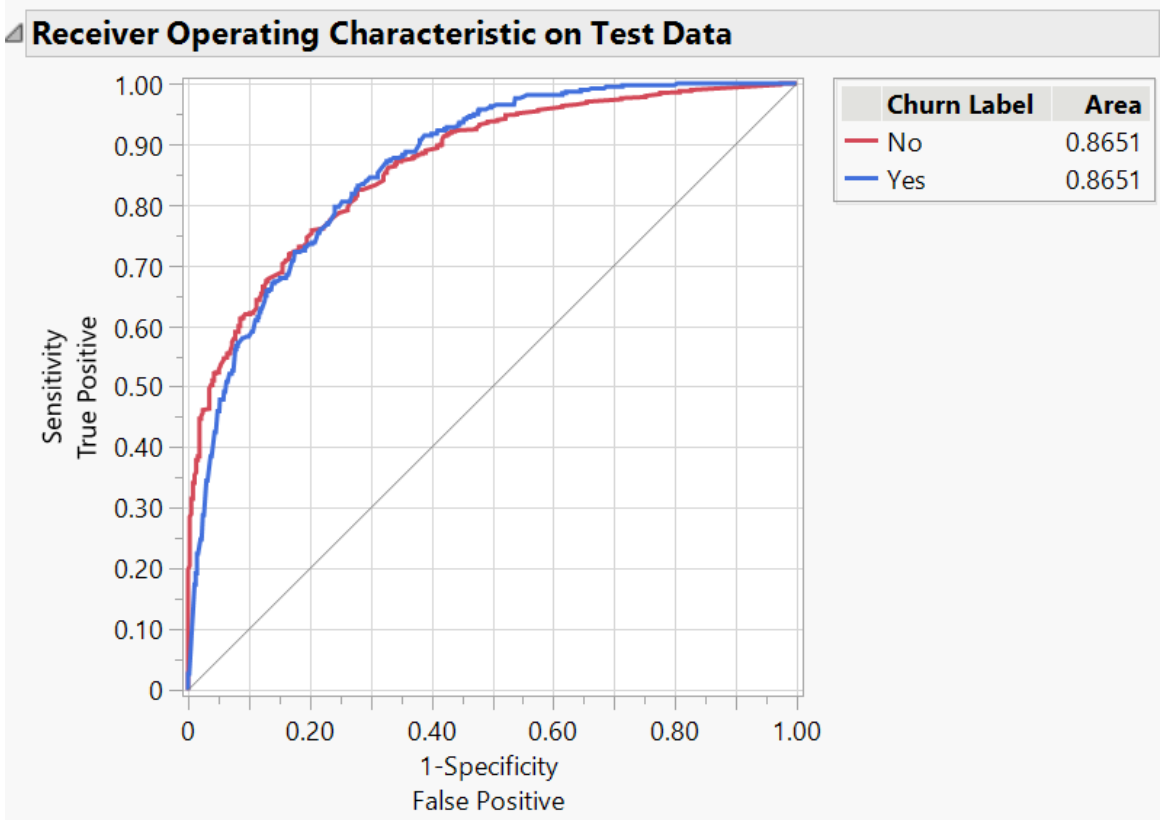
Column Contributions

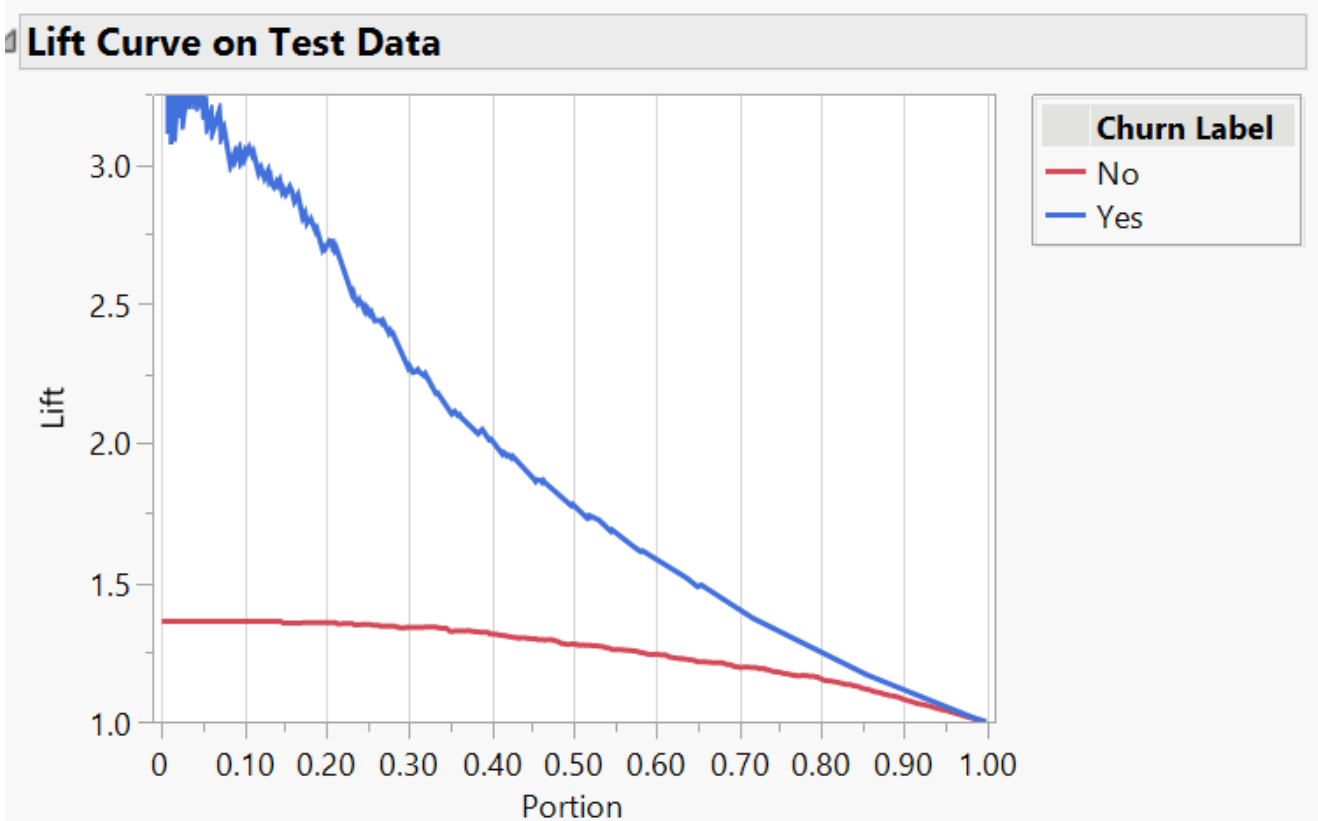
Term	Number of Splits	G^2	Portion
Partner	45	49839.565	0.1594
Monthly Charges	90	39619.8454	0.1267
Paperless Billing	39	26194.2853	0.0838
Contract	36	26033.5444	0.0833
Internet Service	30	24611.8761	0.0787
Tenure Months	45	23471.1467	0.0751
Streaming Movies 2	46	18088.5565	0.0579
Multiple Lines 2	23	16912.251	0.0541
Total Charges	66	16716.9434	0.0535
Phone Service	25	16693.7988	0.0534
Dependents	41	14945.3739	0.0478
Payment Method	19	14158.3191	0.0453
Tech Support 2	34	12686.5355	0.0406
Online Security 2	46	12629.0465	0.0404

Significant variables contributing to the model:

- Partner
- Monthly charges
- Paperless Billing
- Contract
- Internet Service
- Tenure months
- Streaming Movies2
- Multiple Lines2
- Dependents

ⓒSaebyeol Yu. Saebyeol's PowerPoint

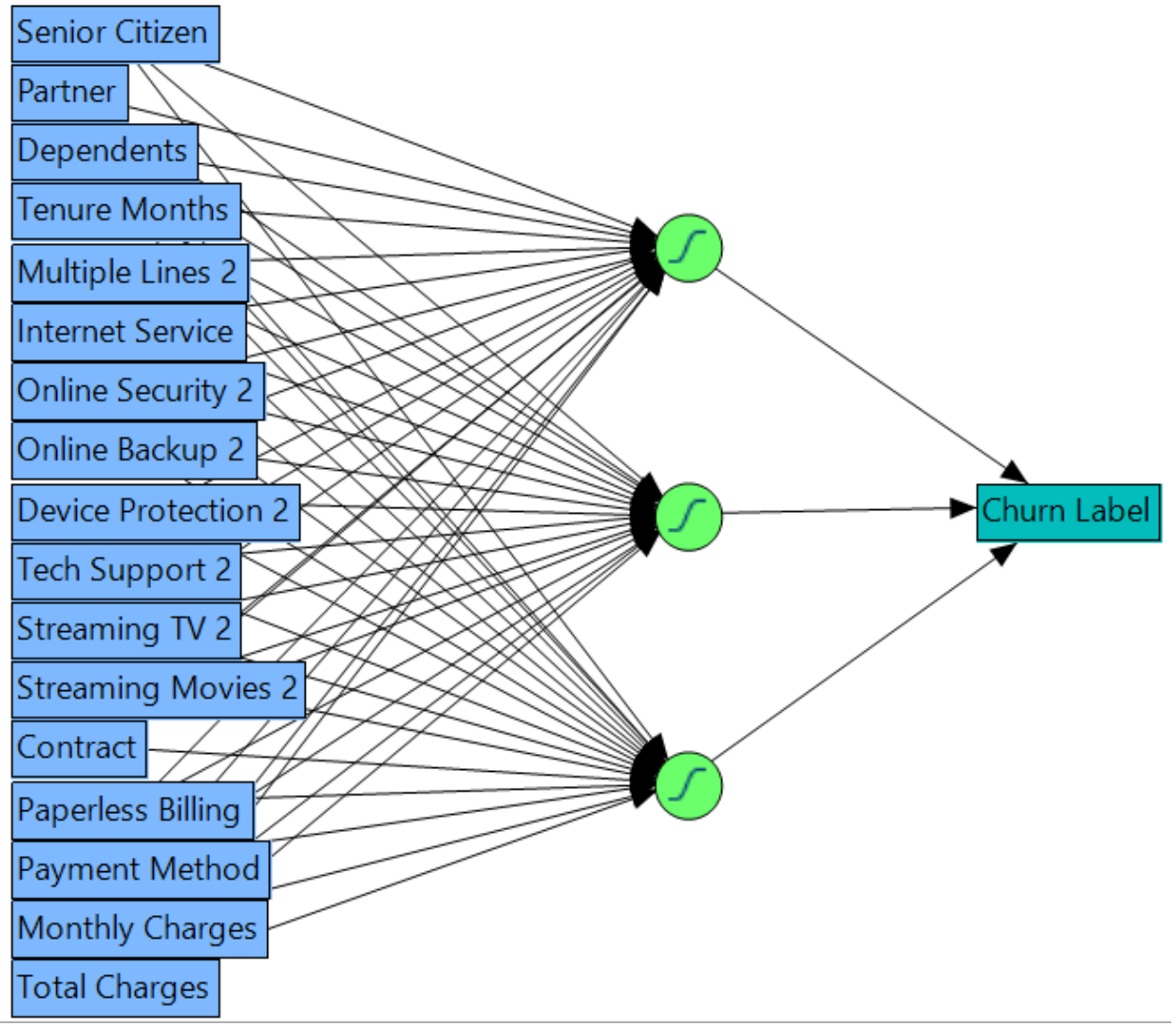


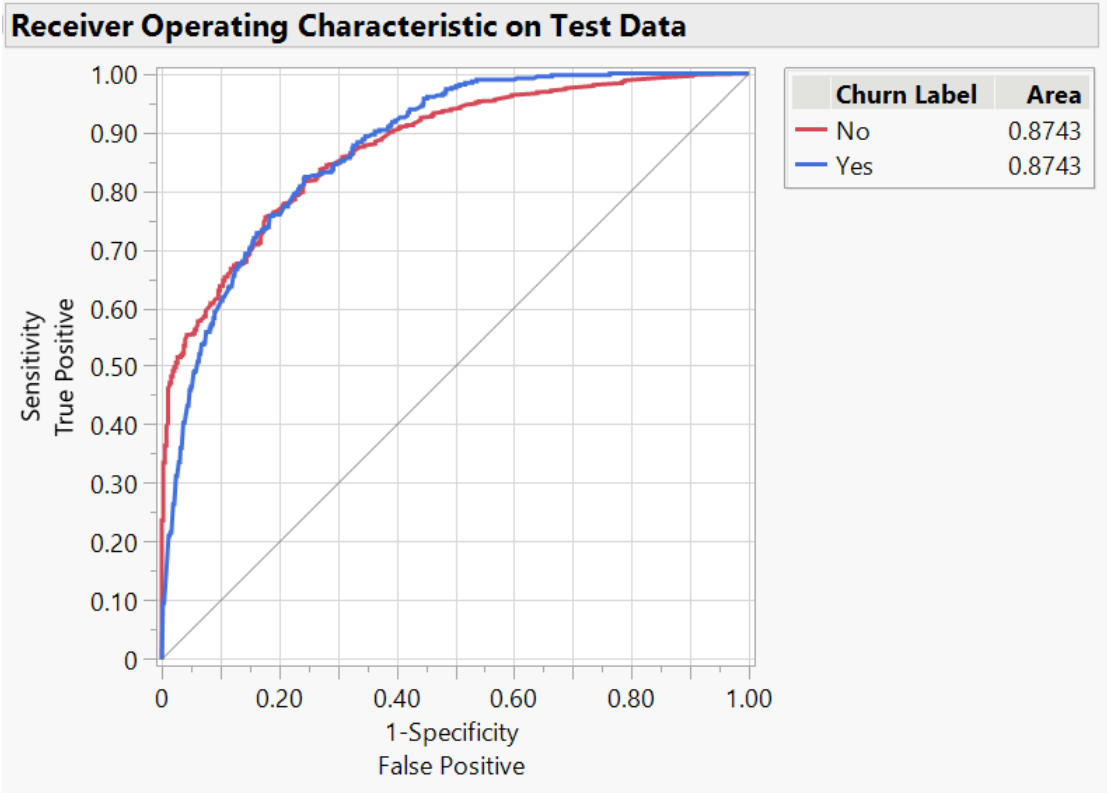


Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Boosted	82.23%	71.00%	2.671	86.51%	17.77%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Neural Network Model:





ⓒSaebyeol Yu. Saebyeol's PowerPoint

Model NTanH(3)

Training

Churn Label

Measures	Value
Generalized RSquare	0.4693918
Entropy RSquare	0.3355337
RASE	0.3542684
Mean Abs Dev	0.2544277
Misclassification Rate	0.1822707
-LogLikelihood	1622.9844
Sum Freq	4219

Confusion Matrix

Actual	Predicted Count	
Churn Label	No	Yes
No	2787	311
Yes	458	663

Confusion Rates

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.900	0.100
Yes	0.409	0.591

Validation

Churn Label

Measures	Value
Generalized RSquare	0.3882266
Entropy RSquare	0.2673383
RASE	0.3729554
Mean Abs Dev	0.2686076
Misclassification Rate	0.2083926
-LogLikelihood	596.69066
Sum Freq	1406

Confusion Matrix

Actual	Predicted Count	
Churn Label	No	Yes
No	911	121
Yes	172	202

Confusion Rates

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.883	0.117
Yes	0.460	0.540

Test

Churn Label

Measures	Value
Generalized RSquare	0.4749808
Entropy RSquare	0.3404066
RASE	0.3527779
Mean Abs Dev	0.249513
Misclassification Rate	0.1762615
-LogLikelihood	537.38662
Sum Freq	1407

Confusion Matrix

Actual	Predicted Count	
Churn Label	No	Yes
No	947	86
Yes	162	212

Confusion Rates

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.917	0.083
Yes	0.433	0.567

Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Neural 3TanH	83.10%	73.91%	2.788	87.11%	16.90%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Ensemble Model:

Model Comparison Validation Stratified by Churn Label=Training

Predictors

Measures of Fit for Churn Label

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.3103	0.4401	0.3993	0.3600	0.2605	0.1906	4219
Neural Model NTanH(3) - Save Formulas		0.3313	0.4645	0.3872	0.3554	0.2578	0.1846	4219
Boosted Tree		0.4052	0.5460	0.3444	0.3330	0.2376	0.1590	4219
Model Averaged		0.3582	0.4950	0.3716	0.3471	0.2520	0.1763	4219

Model Comparison Validation Stratified by Churn Label=Validation

Predictors

Measures of Fit for Churn Label

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.2560	0.3740	0.431	0.3763	0.2733	0.2055	1406
Neural Model NTanH(3) - Save Formulas		0.2692	0.3906	0.4233	0.3724	0.2707	0.2020	1406
Boosted Tree		0.2480	0.3640	0.4356	0.3772	0.2684	0.2112	1406
Model Averaged		0.2682	0.3893	0.4239	0.3730	0.2708	0.2048	1406

Model Comparison Validation Stratified by Churn Label=Test

Predictors

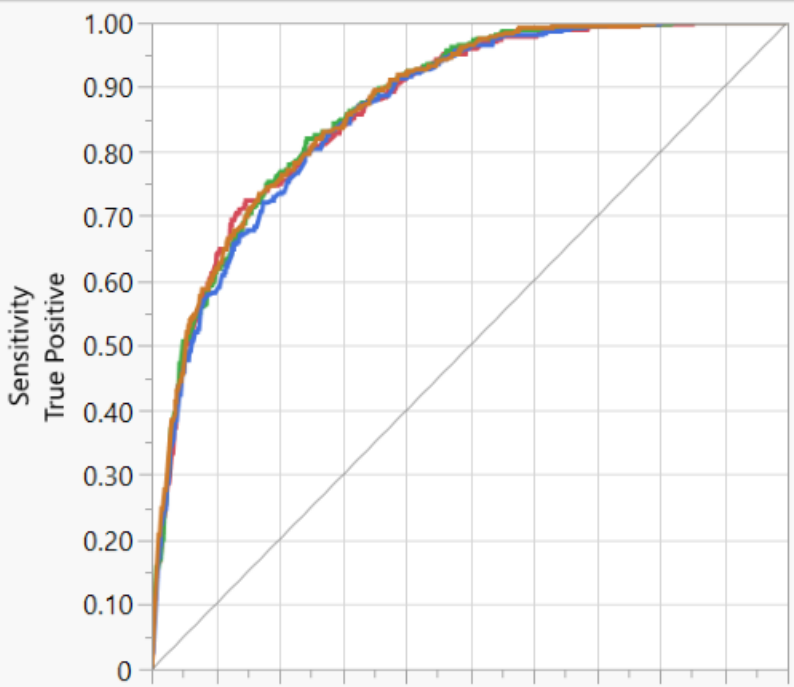
Measures of Fit for Churn Label

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.3248	0.4570	0.391	0.3549	0.2523	0.1727	1407
Neural Model NTanH(3) - Save Formulas		0.3361	0.4701	0.3844	0.3530	0.2521	0.1770	1407
Boosted Tree		0.3208	0.4524	0.3933	0.3575	0.2521	0.1777	1407
Model Averaged		0.3362	0.4702	0.3844	0.3527	0.2521	0.1677	1407

Confusion Matrix

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Ensemble model ROC:



Predictor	Creator	AUC
Prob[Yes]	Fit Nominal Logistic	0.8695
Probability(Churn Label=Yes)	Neural Model NTanH(3) - Save Formulas	0.8728
Prob(Churn Label==Yes)	Boosted Tree	0.8651
Churn Label Yes Avg Predictor By Validation Stratified by Churn Label	Model Averaged	0.8733

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Best Model : Ensemble Model

Predictor Model Averaged

Actual	Predicted Count	
Churn Label	No	Yes
No	952	81
Yes	155	219

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.922	0.078
Yes	0.414	0.586

Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Logistic Regression	82.73%	70.53%	2.6532	86.95%	17.27%
Boosted	82.23%	71.00%	2.671	86.51%	17.77%
Neural 3TanH	83.10%	73.91%	2.788	87.11%	16.90%
Ensemble(LR+Neural3T+Boosted)	83.22%	73.00%	2.7464	0.8733	16.78%

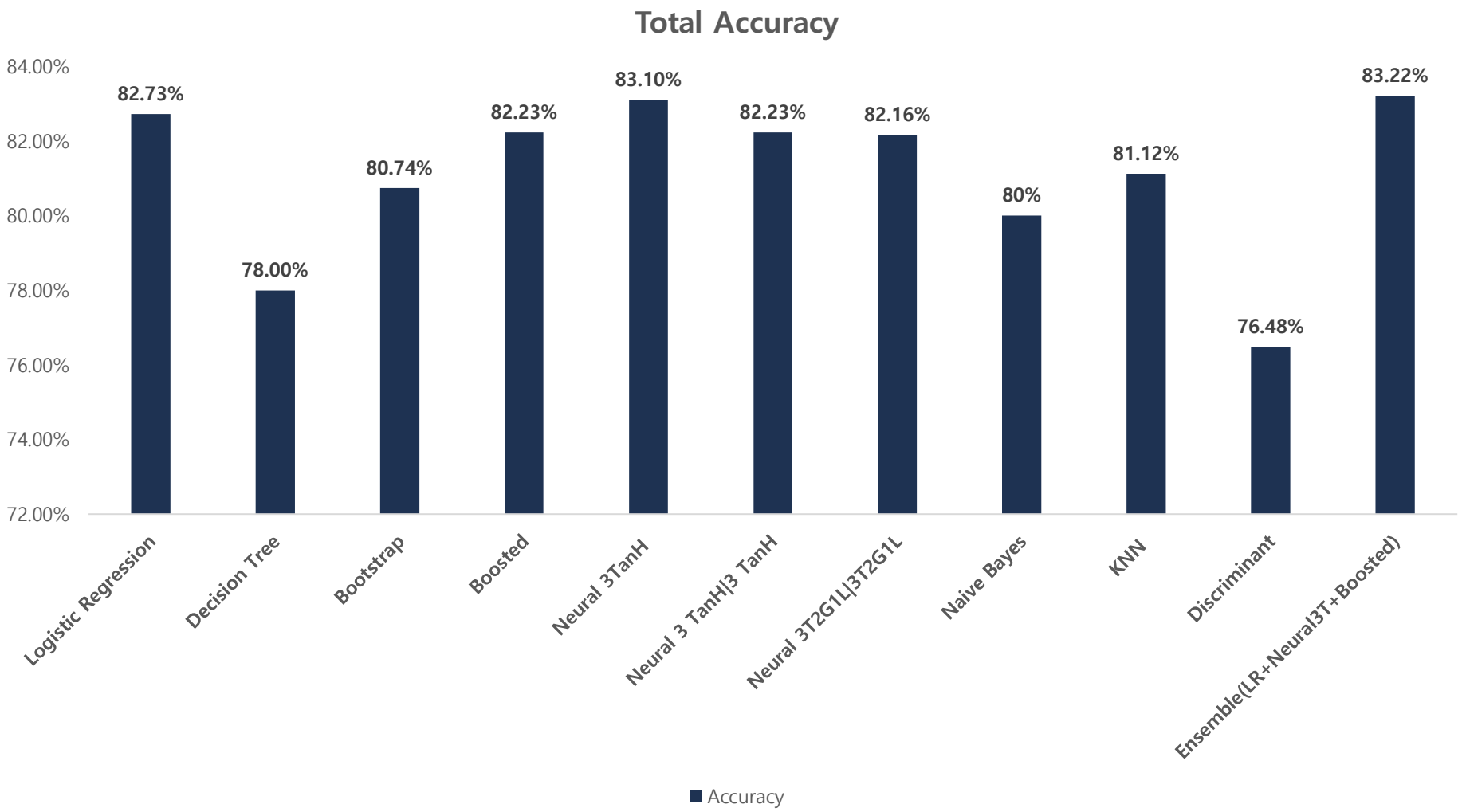
ⓒSaebyeol Yu. Saebyeol's PowerPoint

SEMMA - (Assess)

	Models Performance Metrics				
Models	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Logistic Regression	82.73%	70.53%	2.6532	0.8695	17.27%
Decision Tree	78.00%	68%	2.55	0.8368	22.00%
Bootstrap	80.74%	68.10%	2.562	0.8624	19.26%
Boosted	82.23%	71.00%	2.671	0.8651	17.77%
Neural 3TanH	83.10%	73.91%	2.788	0.8711	16.90%
Neural 3 TanH|3 TanH	82.23%	70.26%	2.6433	0.876	17.77%
Neural 3T2G1L|3T2G1L	82.16%	70.70%	2.659	0.8613	17.84%
Naive Bayes	80%	59%	2.26	0.8488	20.00%
KNN	81.12%	65.90%	2.482		18.88%
Discriminant	76.48%	53.91%	2.0282	0.8476	23.52%
Ensemble(LR+Neural3T+Boosted)	83.22%	73.00%	2.7464	0.8733	16.78%

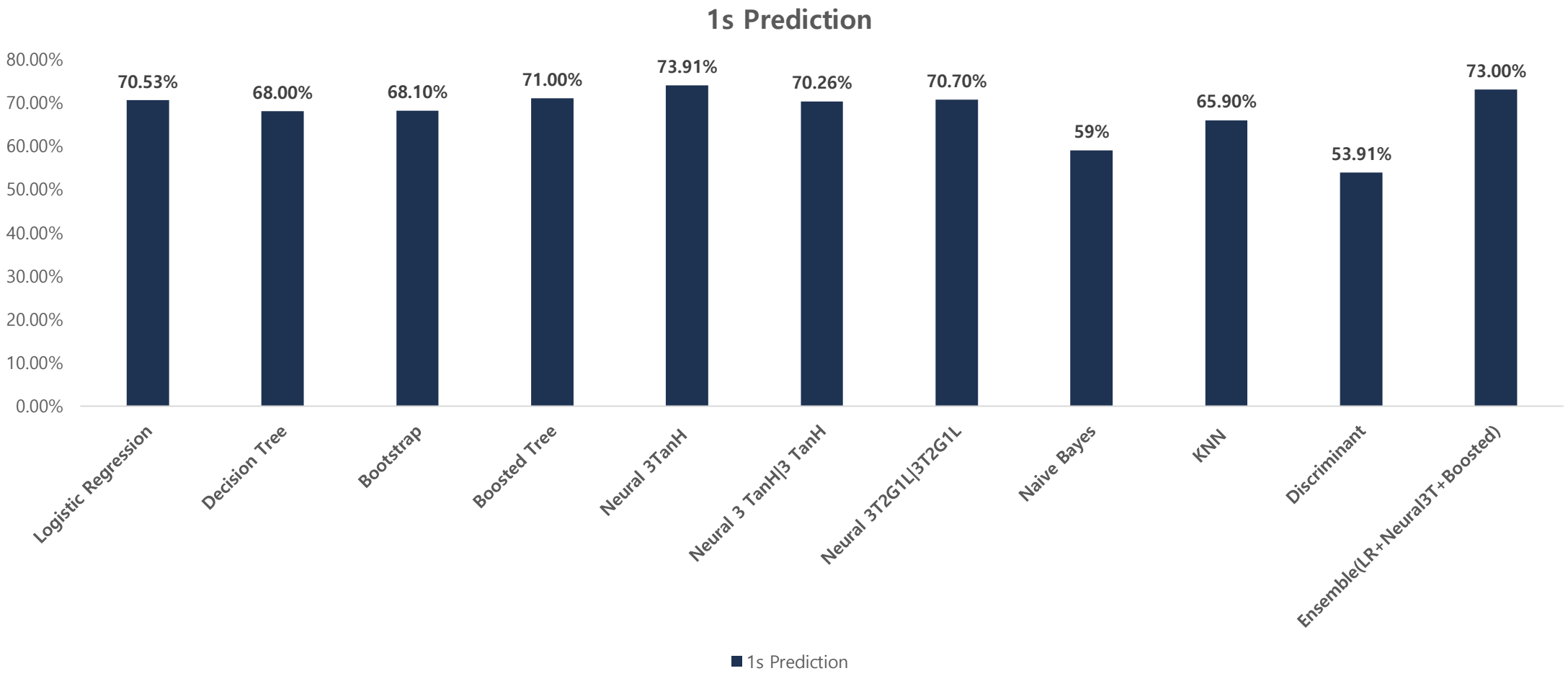
ⓒSaebyeol Yu. Saebyeol's PowerPoint

Models Total Accuracy Comparison:



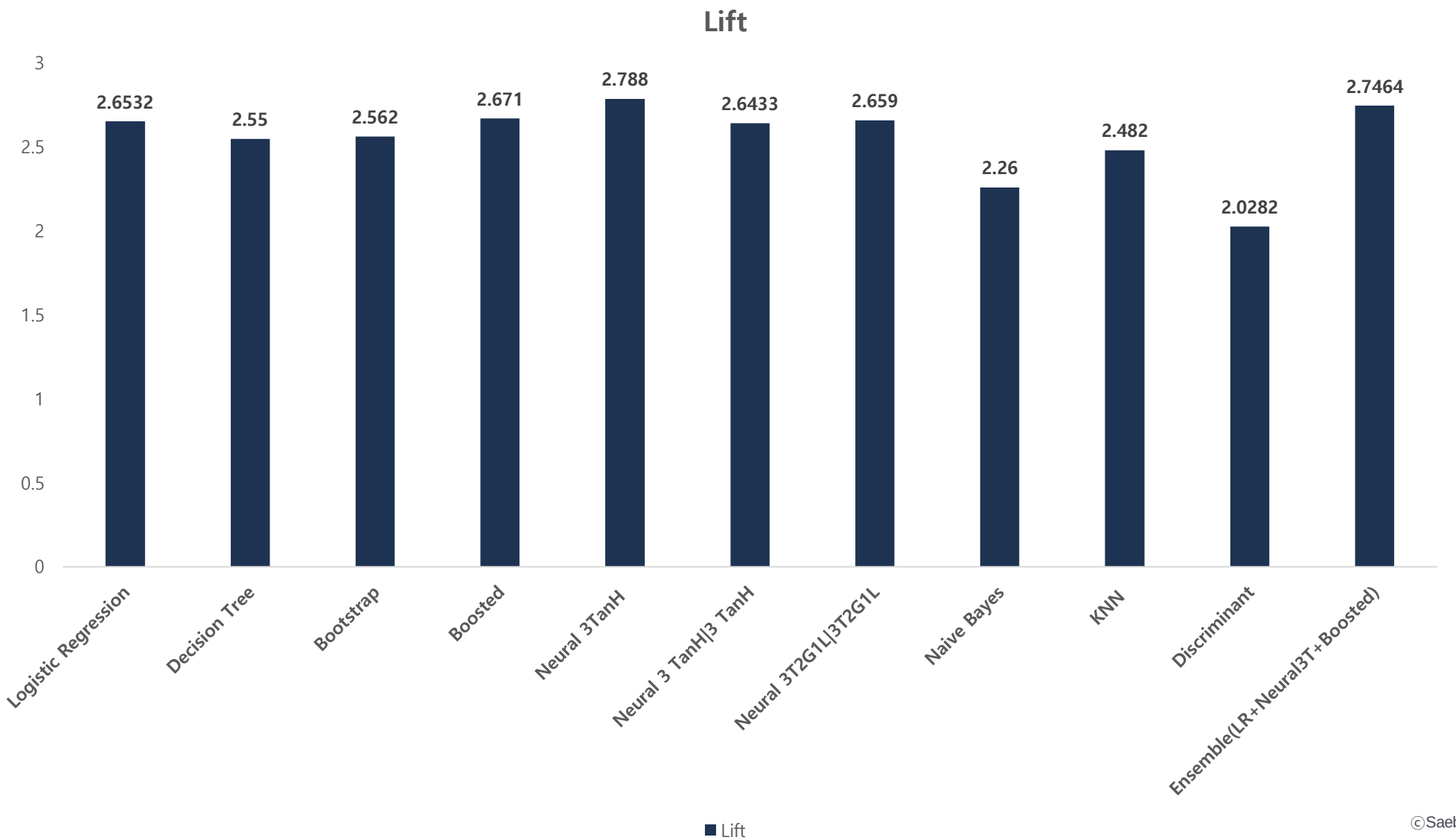
ⓒSaebyeol Yu. Saebyeol's PowerPoint

Models 1's Prediction Comparison:



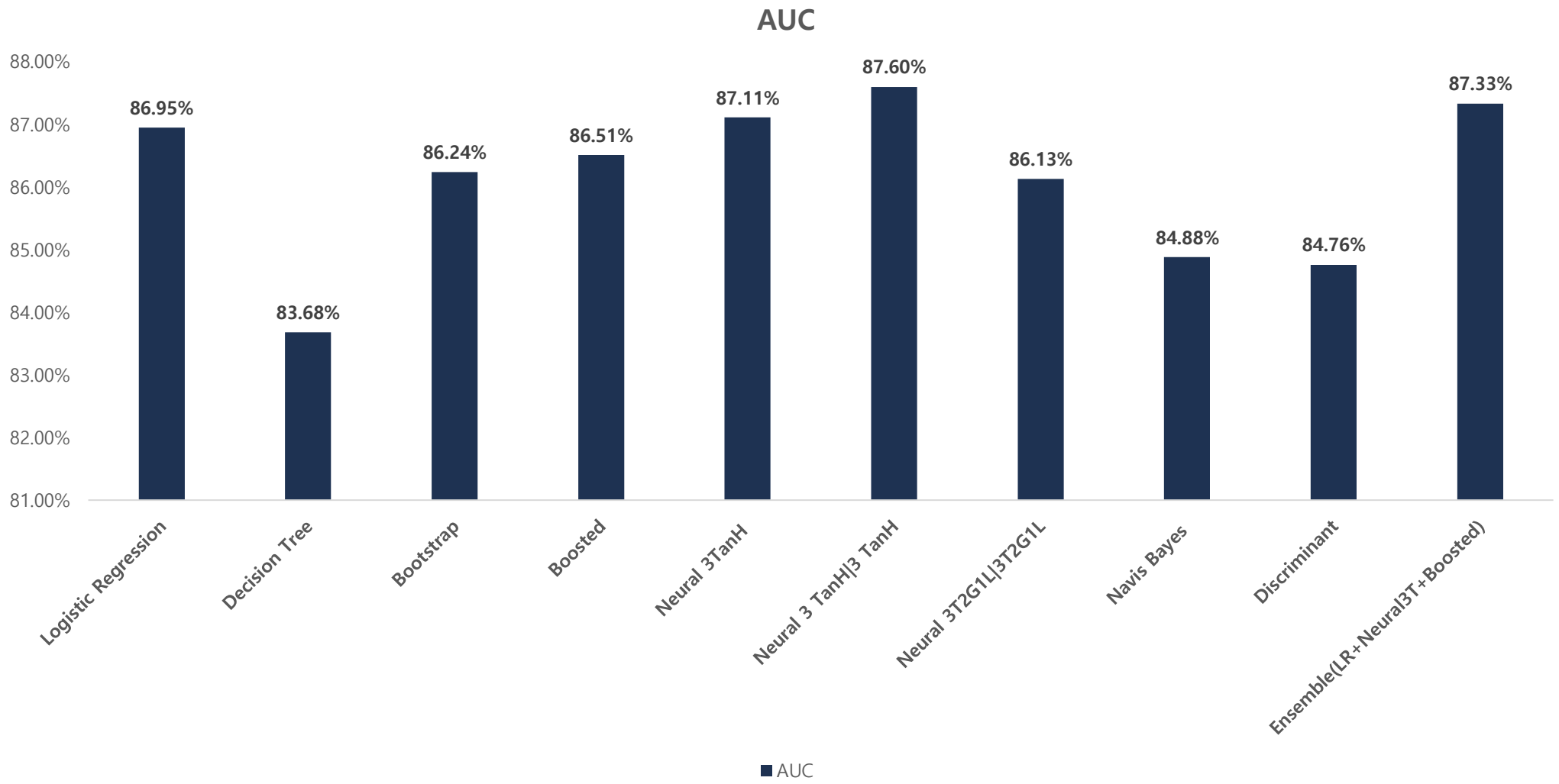
ⓒSaebyeol Yu. Saebyeol's PowerPoint

Models Lift Comparison:



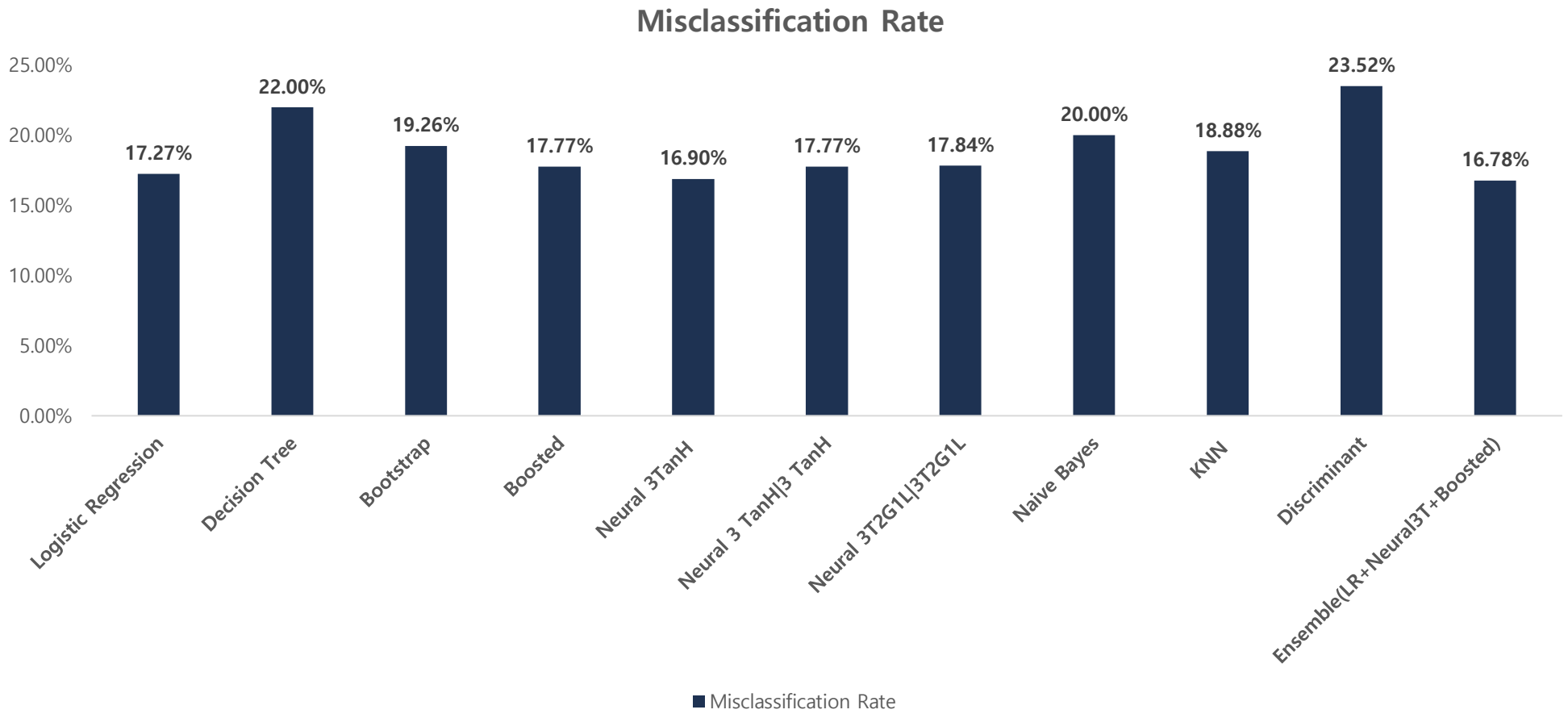
ⓒSaebyeol Yu. Saebyeol's PowerPoint

Model AUC Comparison:



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Models Misclassification Rate:



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Business Analysis & Recommendations:

Tenure months: Our average mean of 32.2 months has low churn rate, if we can retain the customer for 30 months the churn rate drops, so the telecom company can offer initial promotional offers/discounts to retain customers to experience the service.

Monthly Charges: The higher the monthly charges the more likely the customers are to churn. Firm can cut down the prices for services that are expensive.

Dependents: The customers having dependents are likely to stay with the service in our data, telecom company can offer referral discount and coupons to retain the customer.

Contract: month-on-month contract customers are churning out the most by 55.1%, Company can provide annual plan offers to their service, so that the churning rate drops.

Conclusion :

By leveraging customer data, advanced analytics, and predictive modeling, we can identify key churn drivers, forecast and predict customer churn.

Telecom Company can implement targeted retention strategies, improve customer satisfaction, optimize marketing and sales efforts, and enhance profitability.

We found Ensemble to be the best model for predicting customer churn with 83.3% accuracy.

By investing in robust churn analysis capabilities, we can gain valuable insights into customer behavior, preferences, and pain points, which can drive strategic decision-making and action.

By Using Ensemble model this will help the company to identify the potential users who are about to exit the company.

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Thank You !

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Questions ?

ⓒSaebyeol Yu. Saebyeol's PowerPoint

References:

Dataset:

- Telco customer churn (11.1.3+) (ibm.com)

Background research & articles

- https://www.mavenanalytics.io/data-playground?page=2&pageSize=5
- Curious how customer churn data can improve customer retention?
- use cases of Data Mining and SEMMA

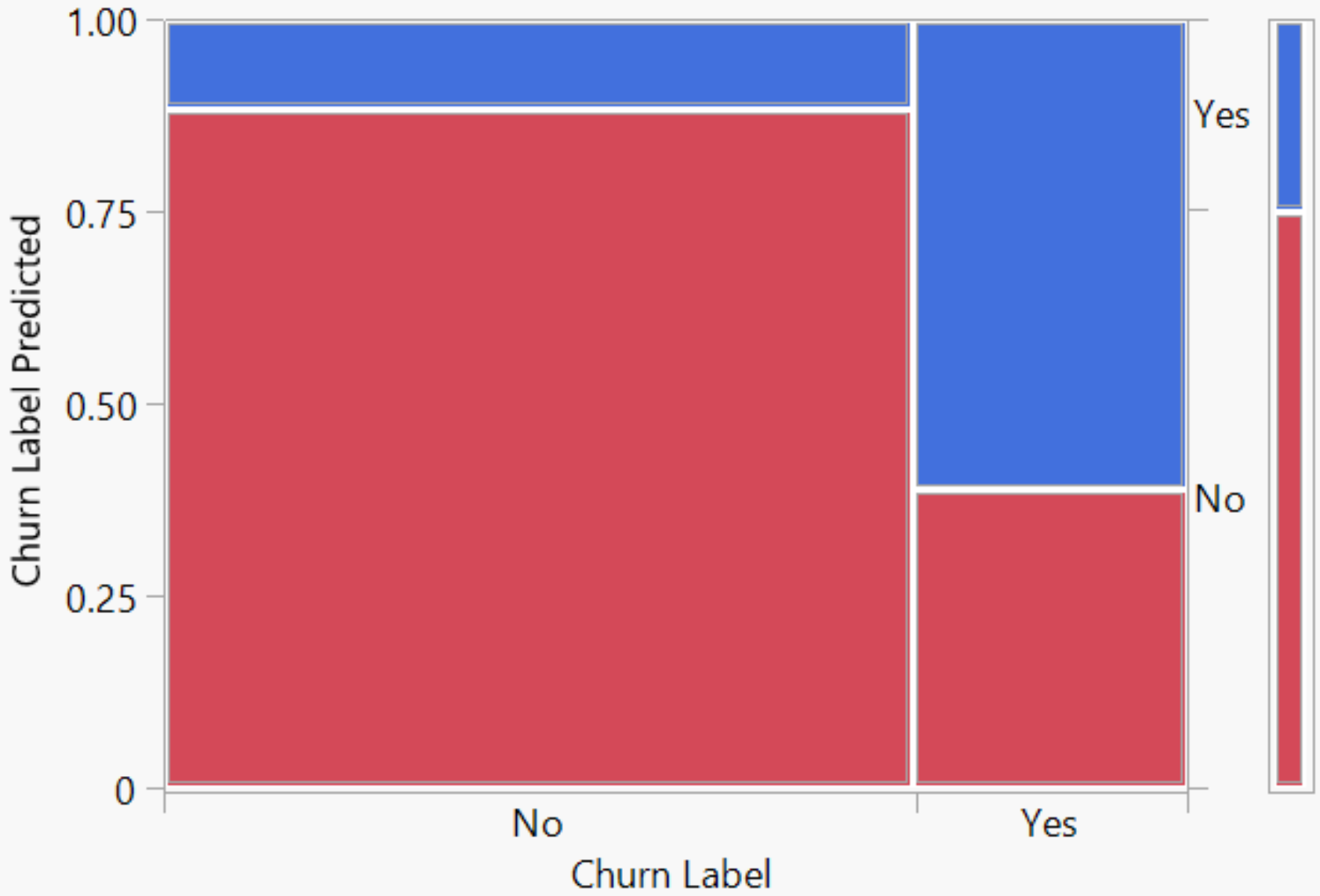
ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix

- Mosaic plot of Target variable vs Categorical
- Decision Tree
- Bootstrap Forest
- Estimates of Neural Network(3TanH)
- Neural Network (3T | 3T)
- Neural Network(3T2L1G | 3T2L1G)
- KNN
- Naive Bayes
- Discriminant Analysis

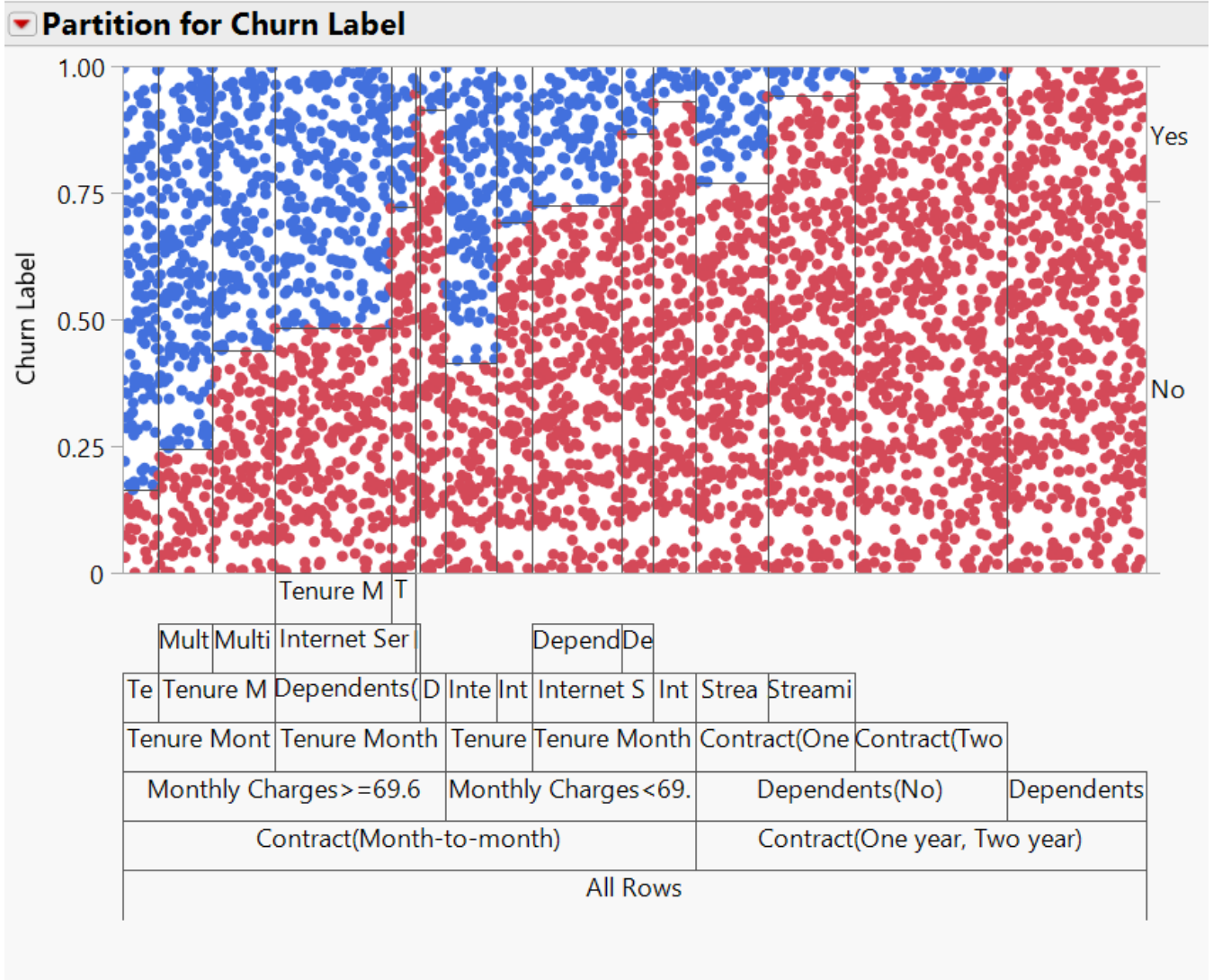
ⓒSaebyeol Yu. Saebyeol's PowerPoint

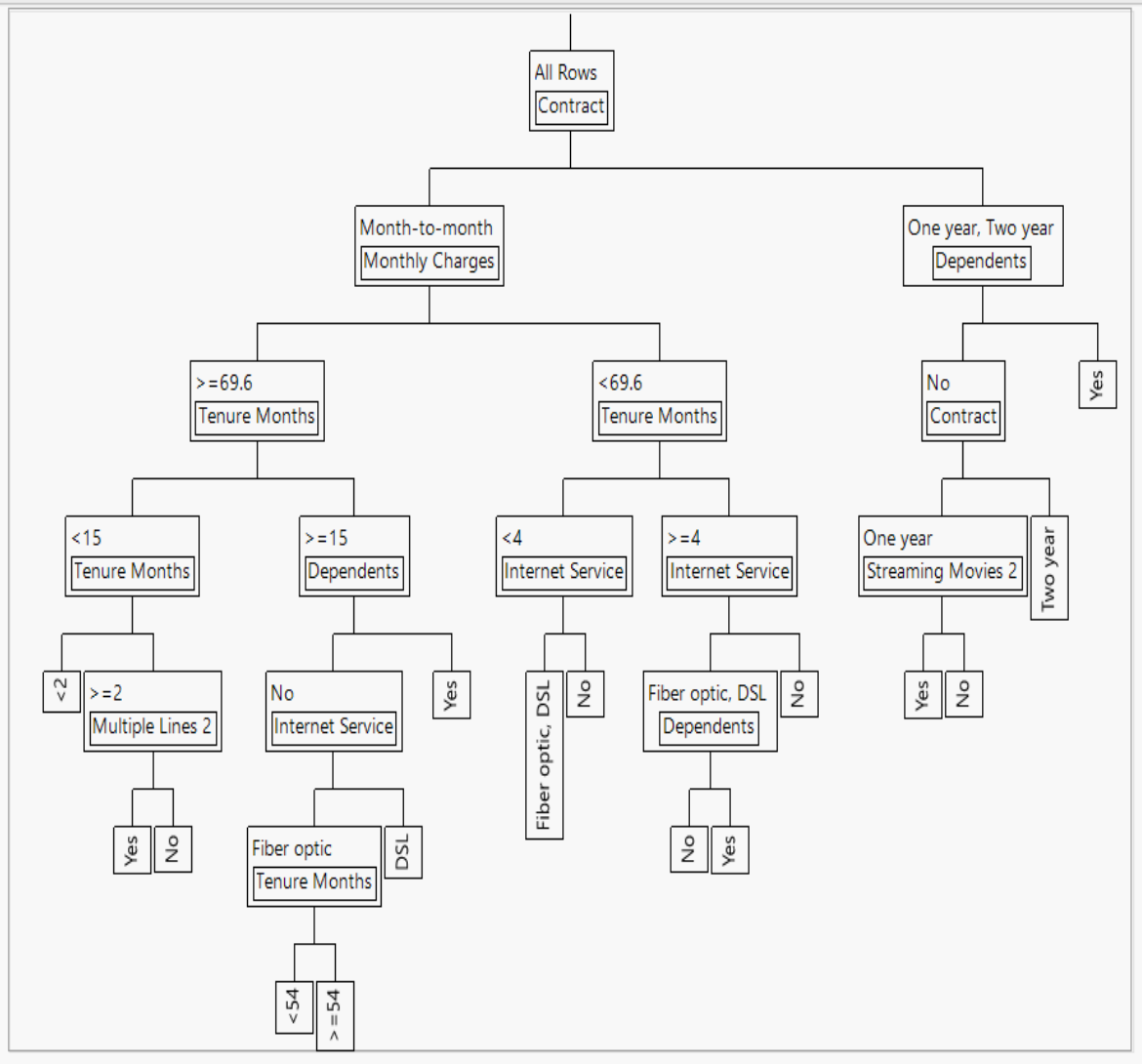
Test



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : Decision Tree:

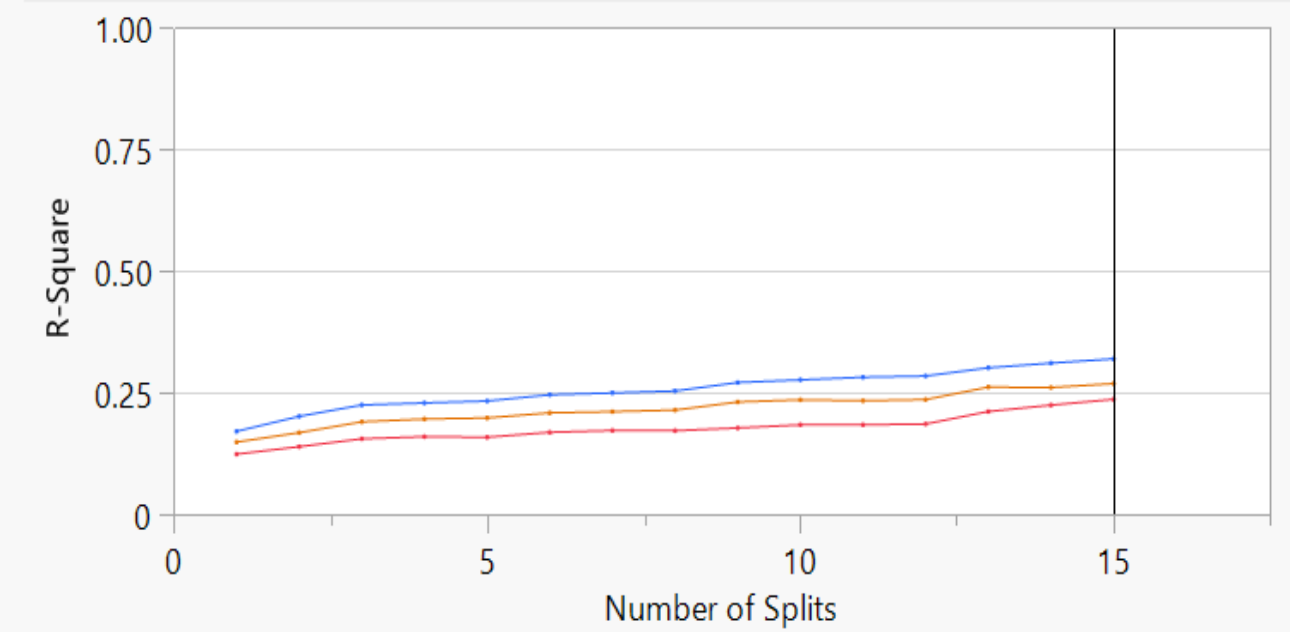




ⓒSaebyeol Yu. Saebyeol's PowerPoint

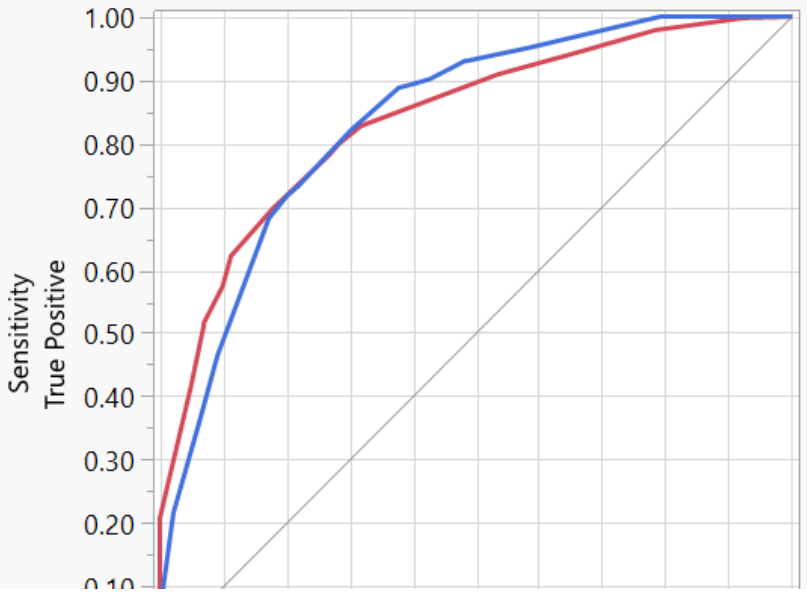
Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.3204	0.2371	0.2694	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.4519	0.3501	0.3907	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3935	0.4419	0.4231	∑ -Log(ρ[j])/n
RASE	0.3596	0.3839	0.3713	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2590	0.2791	0.2646	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1972	0.2326	0.2111	∑ (ρ[j]≠ρMax)/n
N	4219	1406	1407	n



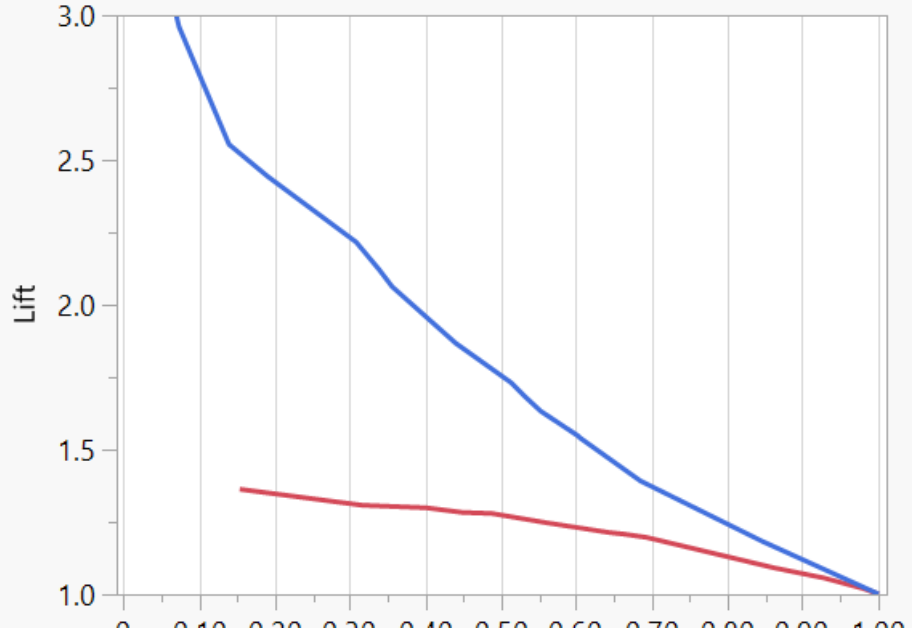
Leaf Label	No	Yes
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months<15&Tenure Months<2	0.1677	0.8323
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months<15&Tenure Months>=2&Multiple Lines 2(Yes)	0.2483	0.7517
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months<15&Tenure Months>=2&Multiple Lines 2(No)	0.4409	0.5591
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months>=15&Dependents(No)&Internet Service(Fiber optic)&Tenure Months<54	0.4857	0.5143
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months>=15&Dependents(No)&Internet Service(Fiber optic)&Tenure Months>=54	0.7221	0.2779
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months>=15&Dependents(No)&Internet Service(DSL)	0.9366	0.0634
Contract(Month-to-month)&Monthly Charges>=69.6&Tenure Months>=15&Dependents(Yes)	0.9137	0.0863
Contract(Month-to-month)&Monthly Charges<69.6&Tenure Months<4&Internet Service(Fiber optic, DSL)	0.4176	0.5824
Contract(Month-to-month)&Monthly Charges<69.6&Tenure Months<4&Internet Service(No)	0.6918	0.3082
Contract(Month-to-month)&Monthly Charges<69.6&Tenure Months>=4&Internet Service(Fiber optic, DSL)&Dependents(No)	0.7278	0.2722
Contract(Month-to-month)&Monthly Charges<69.6&Tenure Months>=4&Internet Service(Fiber optic, DSL)&Dependents(Yes)	0.8672	0.1328
Contract(Month-to-month)&Monthly Charges<69.6&Tenure Months>=4&Internet Service(No)	0.9299	0.0701
^&Dependents(No)&Contract(One year)&Streaming Movies 2(Yes)	0.7711	0.2289
^&Dependents(No)&Contract(One year)&Streaming Movies 2(No)	0.9438	0.0562
^&Dependents(No)&Contract(Two year)	0.9678	0.0322
Contract(One year, Two year)&Dependents(Yes)	0.9996	0.0004

Receiver Operating Characteristic on Test Data



Churn Label	Area
No	0.8368
Yes	0.8368

Lift Curve on Test Data



Term	Number of Splits	G^2	Portion
Contract	2	884.559084	0.5650
Tenure Months	4	238.365123	0.1522
Dependents	3	157.865418	0.1008
Monthly Charges	1	150.469199	0.0961
Internet Service	3	71.379782	0.0456
Streaming Movies 2	1	43.234468	0.0276
Multiple Lines 2	1	19.8491001	0.0127
Partner	0	0	0.0000
Phone Service	0	0	0.0000
Online Security 2	0	0	0.0000
Tech Support 2	0	0	0.0000
Paperless Billing	0	0	0.0000

Confusion Matrix

Training

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	2587	511
Yes	321	800

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.835	0.165
Yes	0.286	0.714

Validation

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	837	195
Yes	132	242

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.811	0.189
Yes	0.353	0.647

Test

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	855	178
Yes	119	255

Actual Churn Label	Predicted Rate No	Predicted Rate Y
No	0.828	0.1
Yes	0.318	0.6

Appendix Bootstrap Forest:

Overall Statistics

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.5302	0.2271	0.3067	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6689	0.3372	0.4359	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2720	0.4477	0.4014	∑ -Log(ρ[j])/n
RASE	0.2878	0.3818	0.3626	√ ∑(y[j]-ρ[j])^2/n
Mean Abs Dev	0.2034	0.2724	0.2533	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1045	0.2134	0.1912	∑ (ρ[j]≠ρMax)/n
N	4219	1406	1407	n

Training

Actual	Predicted Count	
Churn Label	No	Yes
No	2977	121
Yes	320	801

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.961	0.039
Yes	0.285	0.715

Validation

Actual	Predicted Count	
Churn Label	No	Yes
No	926	106
Yes	194	180

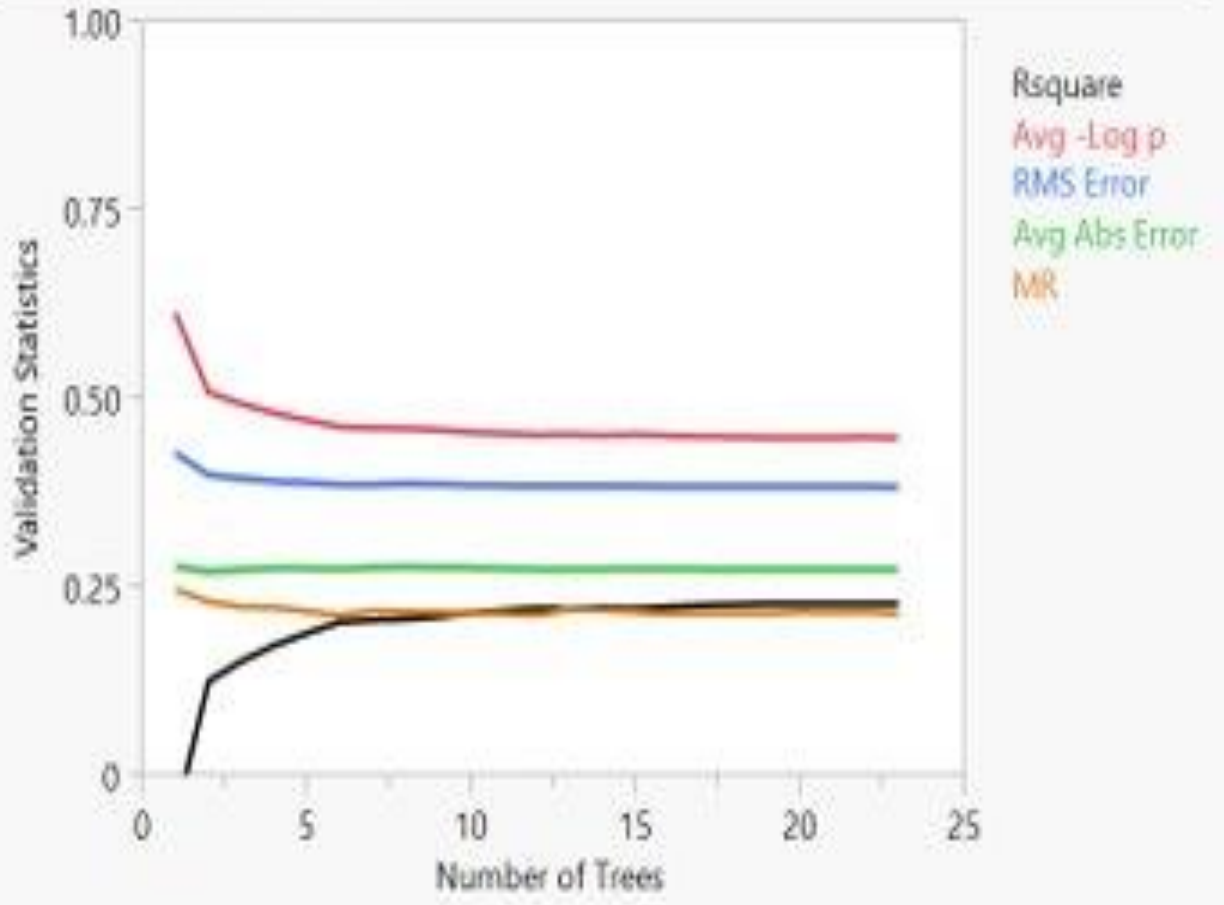
Actual	Predicted Rate	
Churn Label	No	Yes
No	0.897	0.103
Yes	0.519	0.481

Test

Actual	Predicted Count	
Churn Label	No	Yes
No	940	93
Yes	176	198

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.910	0.090
Yes	0.471	0.529

Cumulative Validation



ⓒSaebyeol Yu. Saebyeol's PowerPoint

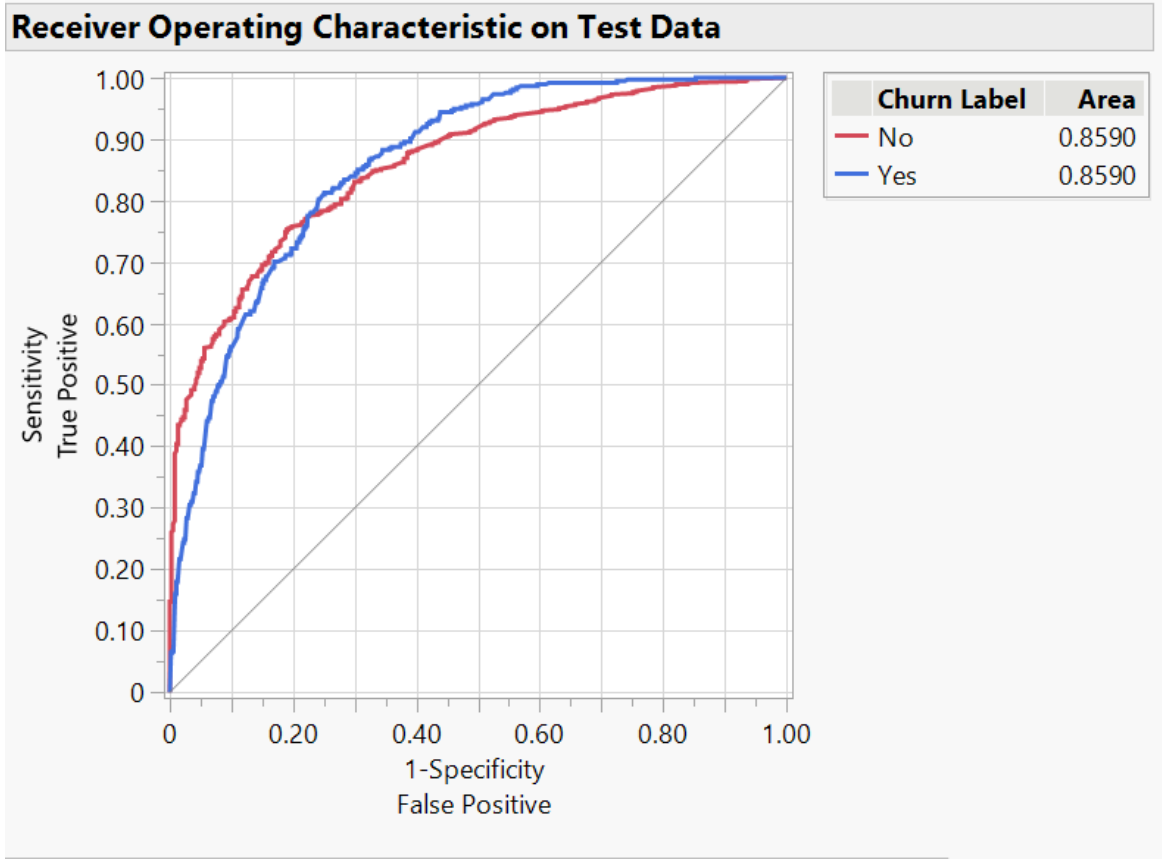
Column Contributions

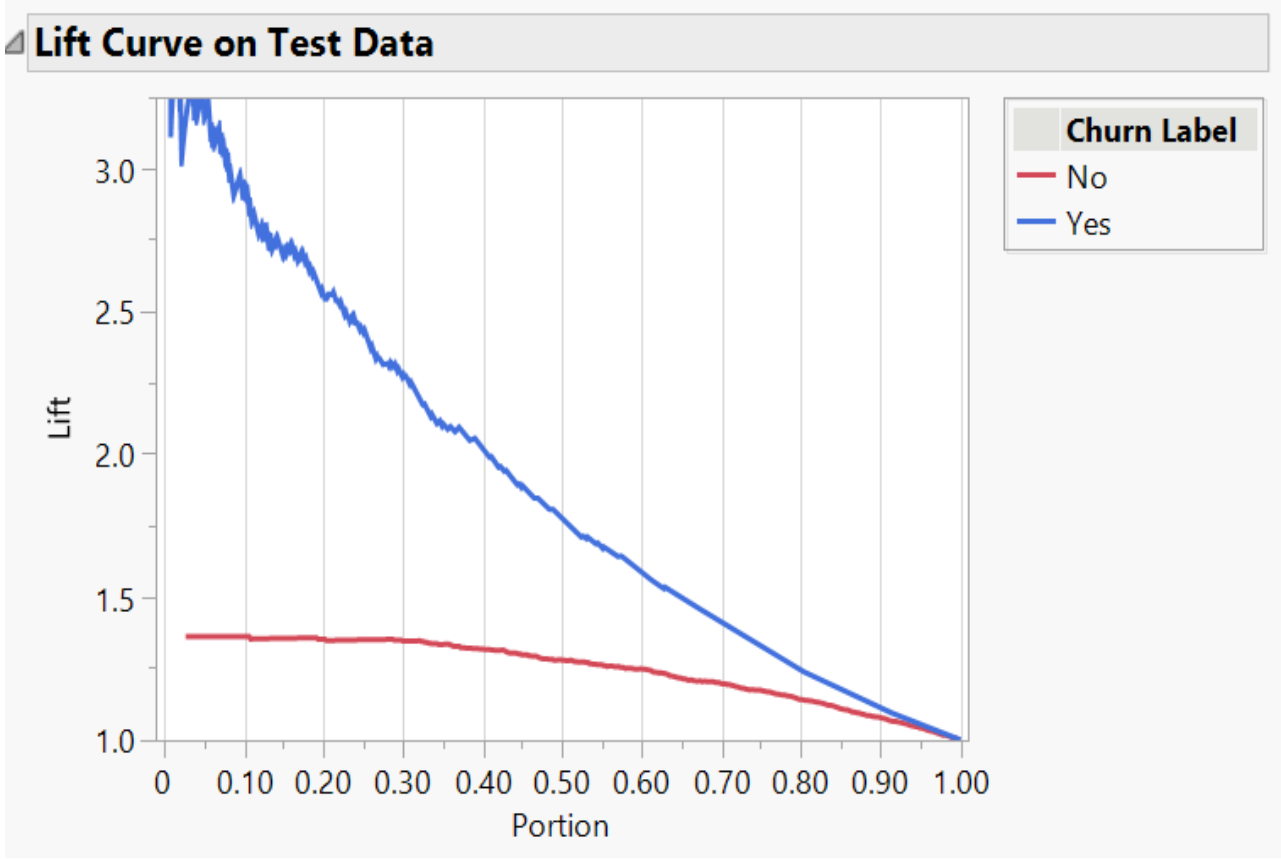
Term	Number of Splits	G^2	Portion
Contract	105	439.252009	0.2355
Monthly Charges	968	271.574349	0.1456
Tenure Months	739	247.329511	0.1326
Total Charges	769	227.01202	0.1217
Dependents	182	110.326203	0.0592
Internet Service	127	107.910625	0.0579
Payment Method	495	83.0465905	0.0445
Partner	410	56.4273595	0.0303
Paperless Billing	396	42.2483749	0.0227
Streaming Movies 2	256	40.0350715	0.0215
Multiple Lines 2	290	33.6173203	0.0180
Online Security 2	246	33.4081192	0.0179
Senior Citizen	303	33.2432682	0.0178
Tech Support 2	256	32.5311118	0.0174
Streaming TV 2	238	32.129732	0.0172
Online Backup 2	305	32.0399852	0.0172
Device Protection 2	247	26.9162278	0.0144
Phone Service	87	15.9968678	0.0086

Per-Tree Summaries

Tree	Splits	Rank	OOB Loss	OOB Loss/N
1	543	14	874.71124	0.5578516
2	575	6	824.98637	0.5406202
3	559	1	790.99617	0.5133006
4	583	12	869.85874	0.5551109
5	565	7	827.50914	0.5433415
6	569	19	909.81876	0.5722131
7	585	5	838.45292	0.5367816
8	549	3	810.49237	0.5232359
9	559	20	915.27341	0.5781891
10	573	22	902.82846	0.58247
11	539	21	891.39602	0.580714
12	557	13	872.65127	0.5572486
13	549	8	851.67529	0.5452467
14	563	17	868.58924	0.5647524
15	559	16	864.82061	0.5630343
16	549	11	873.16159	0.5529839
17	569	4	834.18774	0.5263014
18	573	2	815.0865	0.5198256
19	555	23	940.96379	0.6059007
20	533	9	857.12107	0.5501419
21	535	15	867.68407	0.5601576
22	563	10	847.85481	0.5519888
23	557	18	904.19926	0.5651245

erPoint





Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Bootstrap	80.74%	68.10%	2.562	86.24%	19.26%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : Neural 3T estimates

Estimates

Parameter	Estimate
H1_1:Senior Citizen:No	0.107646
H1_1:Partner:No	-0.26571
H1_1:Dependents:No	0.667329
H1_1:Tenure Months	0.011814
H1_1:Multiple Lines 2:No	-0.01121
H1_1:Internet Service:DSL	-0.14167
H1_1:Internet Service:Fiber optic	0.15052
H1_1:Online Security 2:No	0.116291
H1_1:Online Backup 2:No	-0.09359
H1_1:Device Protection 2:No	0.028881
H1_1:Tech Support 2:No	-0.02124
H1_1:Streaming TV 2:No	-0.10695
H1_1:Streaming Movies 2:No	-0.11716
H1_1:Contract:Month-to-month	-0.3984
H1_1:Contract:One year	0.158163
H1_1:Paperless Billing:No	-0.05141
H1_1:Payment Method:Bank transfer (automatic)	-0.05756
H1_1:Payment Method:Credit card (automatic)	0.144458
H1_1:Payment Method:Electronic check	-0.08302
H1_1:Monthly Charges	9.416e-5
H1_1:Total Charges	1.786e-5
H1_1:Intercept	-1.32658
H1_2:Senior Citizen:No	-0.28414
H1_2:Partner:No	-0.27472
H1_2:Dependents:No	-0.09423

Parameter	Estimate
H1_2:Dependents:No	-0.09423
H1_2:Tenure Months	0.00545
H1_2:Multiple Lines 2:No	0.056595
H1_2:Internet Service:DSL	-0.1657
H1_2:Internet Service:Fiber optic	-0.22809
H1_2:Online Security 2:No	0.207327
H1_2:Online Backup 2:No	-0.05929
H1_2:Device Protection 2:No	0.058716
H1_2:Tech Support 2:No	0.029994
H1_2:Streaming TV 2:No	-0.18457
H1_2:Streaming Movies 2:No	0.048521
H1_2:Contract:Month-to-month	-0.12044
H1_2:Contract:One year	-0.54192
H1_2:Paperless Billing:No	-0.06754
H1_2:Payment Method:Bank transfer (automatic)	-0.05313
H1_2:Payment Method:Credit card (automatic)	0.082107
H1_2:Payment Method:Electronic check	-0.16343
H1_2:Monthly Charges	0.000915
H1_2:Total Charges	7.721e-6
H1_2:Intercept	-0.00862
H1_3:Senior Citizen:No	-0.20842
H1_3:Partner:No	-0.02932
H1_3:Dependents:No	-0.20026
H1_3:Tenure Months	-0.01061
H1_3:Multiple Lines 2:No	-0.00618
H1_3:Internet Service:DSL	-0.01491
H1_3:Internet Service:Fiber optic	0.087308
H1_3:Online Security 2:No	0.071256

Parameter	Estimate
H1_3:Online Security 2:No	0.071256
H1_3:Online Backup 2:No	0.024373
H1_3:Device Protection 2:No	0.004735
H1_3:Tech Support 2:No	0.065221
H1_3:Streaming TV 2:No	-0.09771
H1_3:Streaming Movies 2:No	0.048216
H1_3:Contract:Month-to-month	0.393773
H1_3:Contract:One year	-0.45185
H1_3:Paperless Billing:No	-0.04276
H1_3:Payment Method:Bank transfer (automatic)	0.003188
H1_3:Payment Method:Credit card (automatic)	-0.05381
H1_3:Payment Method:Electronic check	-0.00595
H1_3:Monthly Charges	-0.00398
H1_3:Total Charges	-1.56e-5
H1_3:Intercept	0.228085
Churn Label(No):H1_1	-6.48563
Churn Label(No):H1_2	5.507905
Churn Label(No):H1_3	-10.0939
Churn Label(No):Intercept	-1.65503

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : Neural 3T | 3T:

Neural

Validation Column: Validation Stratified by Churn Label

Model Launch

Model NTanH(3)NTanH2(3)

Training

Churn Label

Measures	Value
Generalized RSquare	0.47618
Entropy RSquare	0.341484
RASE	0.3531345
Mean Abs Dev	0.2495633
Misclassification Rate	0.1817966
-LogLikelihood	1608.4504
Sum Freq	4219

Confusion Matrix

Actual Churn Label	Predicted Count No	Yes
No	2784	314
Yes	453	668

Confusion Rates

Actual Churn Label	Predicted Rate No	Yes
No	0.899	0.101
Yes	0.404	0.596

Validation

Churn Label

Measures	Value
Generalized RSquare	0.3790889
Entropy RSquare	0.259994
RASE	0.373507
Mean Abs Dev	0.2653642
Misclassification Rate	0.2027027
-LogLikelihood	602.67196
Sum Freq	1406

Confusion Matrix

Actual Churn Label	Predicted Count No	Yes
No	917	115
Yes	170	204

Confusion Rates

Actual Churn Label	Predicted Rate No	Yes
No	0.889	0.111
Yes	0.455	0.545

Test

Churn Label

Measures	Value
Generalized RSquare	0.4785625
Entropy RSquare	0.3435589
RASE	0.3501801
Mean Abs Dev	0.2441389
Misclassification Rate	0.177683
-LogLikelihood	534.81841
Sum Freq	1407

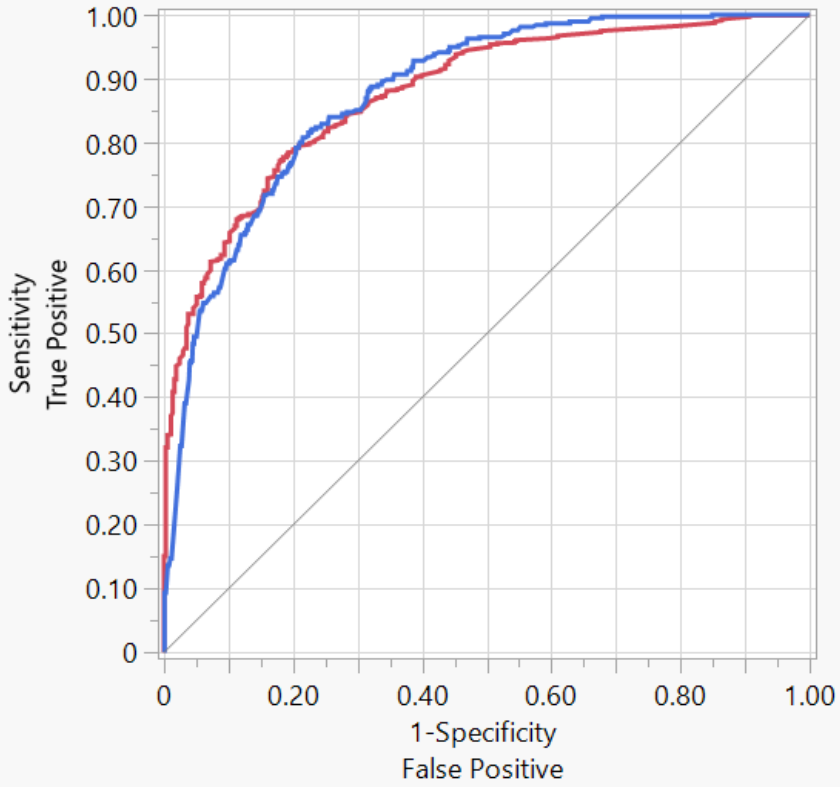
Confusion Matrix

Actual Churn Label	Predicted Count No	Yes
No	942	91
Yes	159	215

Confusion Rates

Actual Churn Label	Predicted Rate No	Yes
No	0.912	0.088
Yes	0.425	0.575

Receiver Operating Characteristic on Test Data



Churn Label	Area
No	0.8760
Yes	0.8760

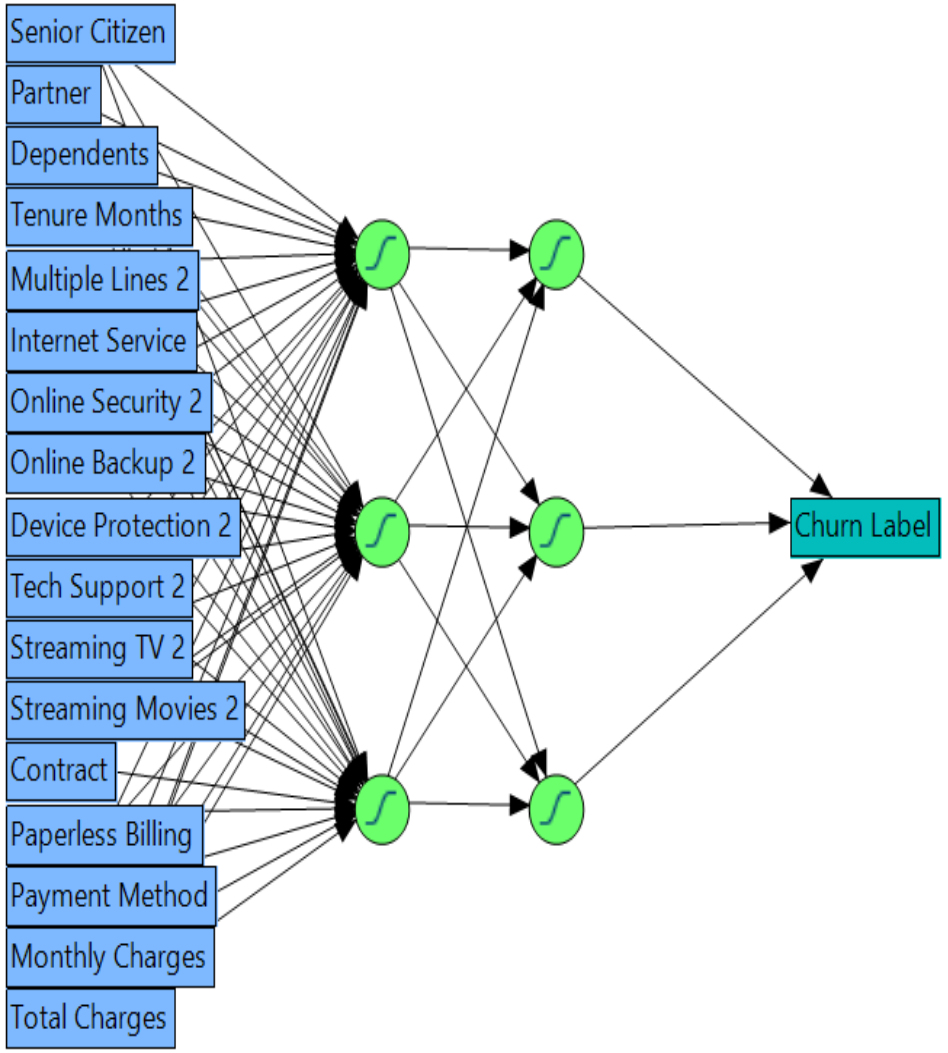
Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Neural 3 TanH|3 TanH	82.23%	70.26%	2.6433	87.60%	17.77%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

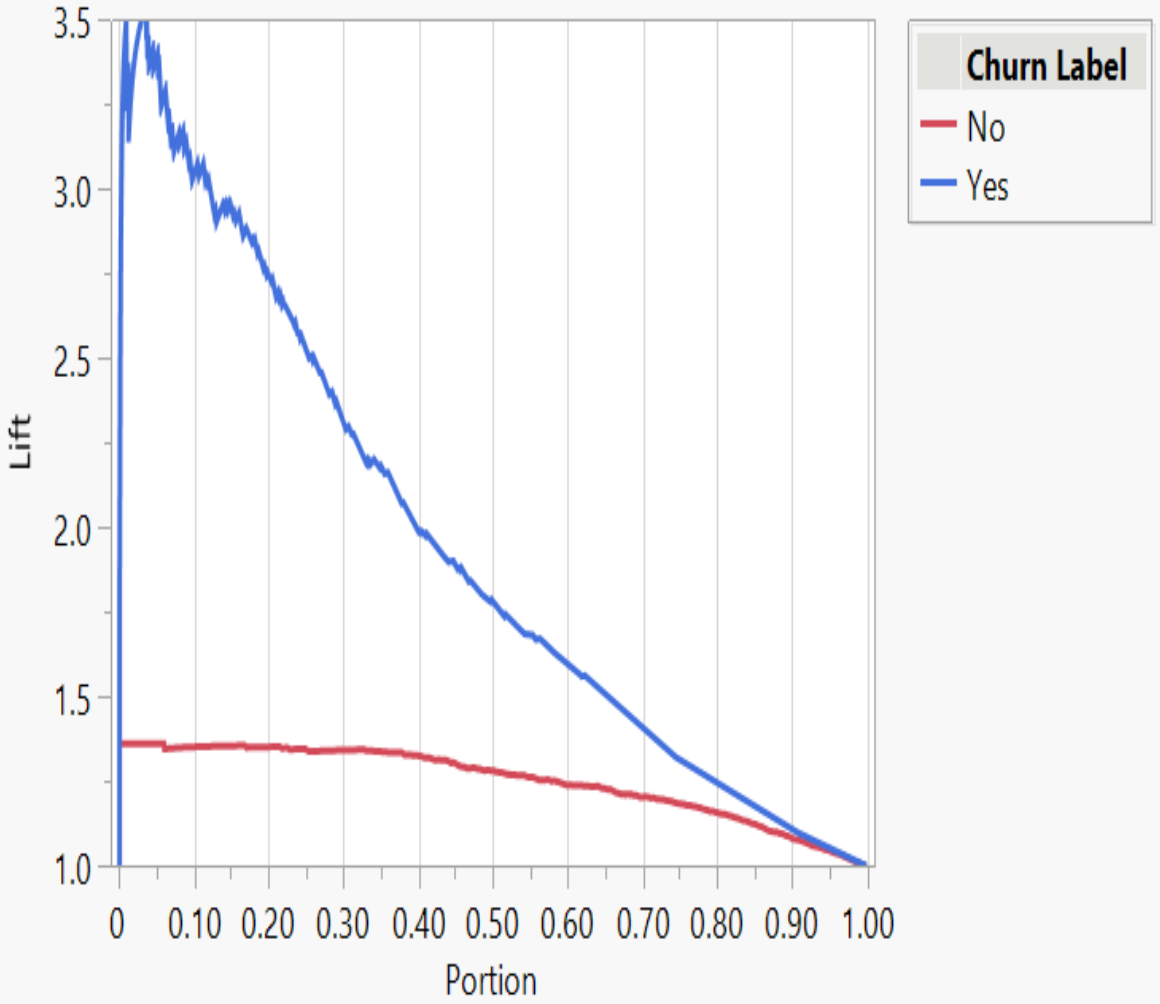
Parameter	Estimate
H2_1:Senior Citizen:No	-2.84267
H2_1:Partner:No	-0.83508
H2_1:Dependents:No	3.099481
H2_1:Tenure Months	-0.01106
H2_1:Multiple Lines 2:No	0.009629
H2_1:Internet Service:DSL	-0.85246
H2_1:Internet Service:Fiber optic	1.259404
H2_1:Online Security 2:No	0.485418
H2_1:Online Backup 2:No	0.459552
H2_1:Device Protection 2:No	0.182838
H2_1:Tech Support 2:No	-0.48402
H2_1:Streaming TV 2:No	-0.68736
H2_1:Streaming Movies 2:No	-0.77341
H2_1:Contract:Month-to-month	-1.20692
H2_1:Contract:One year	0.864591
H2_1:Paperless Billing:No	-0.30774
H2_1:Payment Method:Bank transfer (automatic)	-0.02088
H2_1:Payment Method:Credit card (automatic)	0.009743
H2_1:Payment Method:Electronic check	0.883815
H2_1:Monthly Charges	0.016699
H2_1:Total Charges	-3.83e-5
H2_1:Intercept	-1.85192
H2_2:Senior Citizen:No	2.64279
H2_2:Partner:No	0.323405
H2_2:Dependents:No	2.314346
H2_2:Tenure Months	-0.01615
H2_2:Internet Service:DSL	-2.02994
H2_2:Internet Service:Fiber optic	-0.14661
H2_2:Online Security 2:No	0.520018
H2_2:Online Backup 2:No	0.581783
H2_2:Device Protection 2:No	0.025716
H2_2:Tech Support 2:No	-0.31532
H2_2:Streaming TV 2:No	-0.09462
H2_2:Streaming Movies 2:No	-1.12283
H2_2:Contract:Month-to-month	1.401837
H2_2:Contract:One year	2.533576
H2_2:Paperless Billing:No	-0.52688
H2_2:Payment Method:Bank transfer (automatic)	-0.64112
H2_2:Payment Method:Credit card (automatic)	0.269951
H2_2:Payment Method:Electronic check	0.687891
H2_2:Monthly Charges	0.024267
H2_2:Total Charges	-0.00014
H2_2:Intercept	-4.0371
H2_3:Senior Citizen:No	0.024224
H2_3:Partner:No	-0.00238
H2_3:Dependents:No	-0.30023
H2_3:Tenure Months	-0.04407
H2_3:Multiple Lines 2:No	-0.13352
H2_3:Internet Service:DSL	0.524485
H2_3:Internet Service:Fiber optic	1.206421
H2_3:Online Security 2:No	-0.0838
H2_3:Online Backup 2:No	-0.20588
H2_3:Device Protection 2:No	-0.10344
H2_3:Tech Support 2:No	0.338598
H2_3:Streaming TV 2:No	-0.10717
H2_3:Contract:Month-to-month	1.32306
H2_3:Contract:One year	-0.4566
H2_3:Paperless Billing:No	-0.0101
H2_3:Payment Method:Bank transfer (automatic)	0.12500
H2_3:Payment Method:Credit card (automatic)	-0.137
H2_3:Payment Method:Electronic check	-0.1603
H2_3:Monthly Charges	-0.0355
H2_3:Total Charges	0.00023
H2_3:Intercept	1.27862
H1_1:H2_1	-0.7662
H1_1:H2_2	-0.4428
H1_1:H2_3	-0.9176
H1_1:Intercept	0.88592
H1_2:H2_1	-0.3850
H1_2:H2_2	0.18168
H1_2:H2_3	-1.1929
H1_2:Intercept	-0.8481
H1_3:H2_1	-0.0539
H1_3:H2_2	-0.9277
H1_3:H2_3	-0.1600
H1_3:Intercept	-0.8622
Churn Label(No):H1_1	2.1831
Churn Label(No):H1_2	3.24859
Churn Label(No):H1_3	2.40736

Appendix : Neural 31| 31.

Diagram



Lift Curve on Test Data



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : Neural 3T2L1G|3T2L1G:

Neural

Validation Column: Validation Stratified by Churn Label

Model Launch

Model NTanH(3)NLinear(2)NGaussian(1)NTanH2(3)NLinear2(2)NGaussian2(1)

Training

Churn Label

Measures	Value
Generalized RSquare	0.4499661
Entropy RSquare	0.3187287
RASE	0.3579529
Mean Abs Dev	0.2563714
Misclassification Rate	0.1900924
-LogLikelihood	1664.0311
Sum Freq	4219

Confusion Matrix

Actual	Predicted Count	
Churn Label	No	Yes
No	2770	328
Yes	474	647

Confusion Rates

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.894	0.106
Yes	0.423	0.577

Validation

Churn Label

Measures	Value
Generalized RSquare	0.3596621
Entropy RSquare	0.2445846
RASE	0.379856
Mean Abs Dev	0.2744193
Misclassification Rate	0.2183499
-LogLikelihood	615.22165
Sum Freq	1406

Confusion Matrix

Actual	Predicted Count	
Churn Label	No	Yes
No	907	125
Yes	182	192

Confusion Rates

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.879	0.121
Yes	0.487	0.513

Test

Churn Label

Measures	Value
Generalized RSquare	0.4402714
Entropy RSquare	0.3104409
RASE	0.3585638
Mean Abs Dev	0.2533267
Misclassification Rate	0.1783937
-LogLikelihood	561.80043
Sum Freq	1407

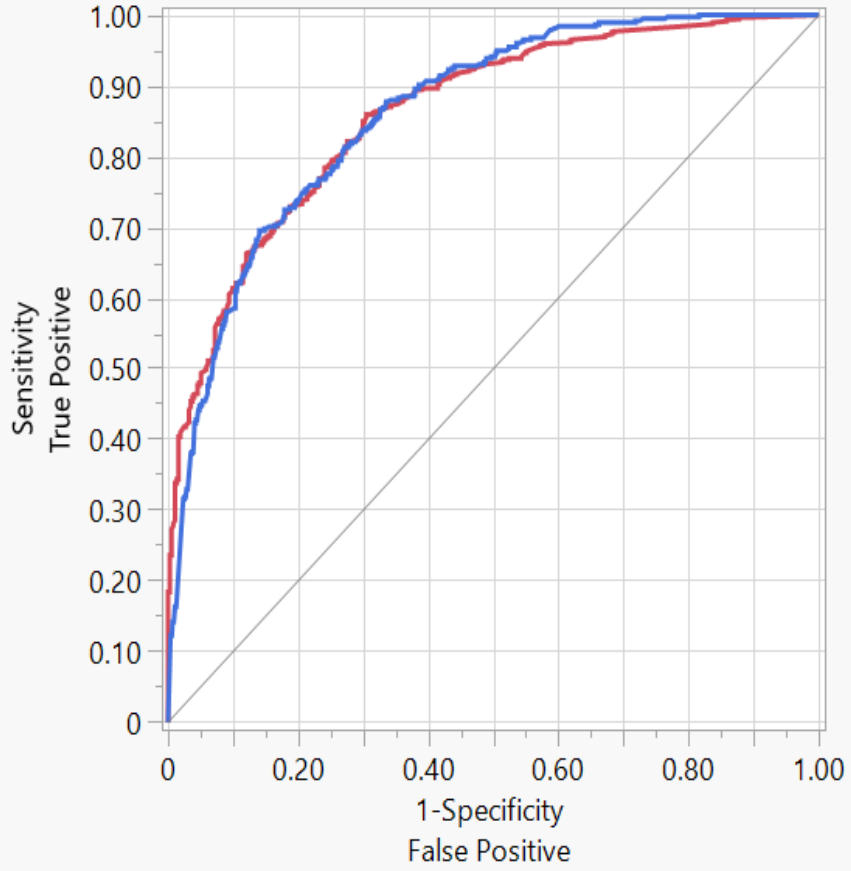
Confusion Matrix

Actual	Predicted Count	
Churn Label	No	Yes
No	946	87
Yes	164	210

Confusion Rates

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.916	0.084
Yes	0.439	0.561

Receiver Operating Characteristic on Test Data

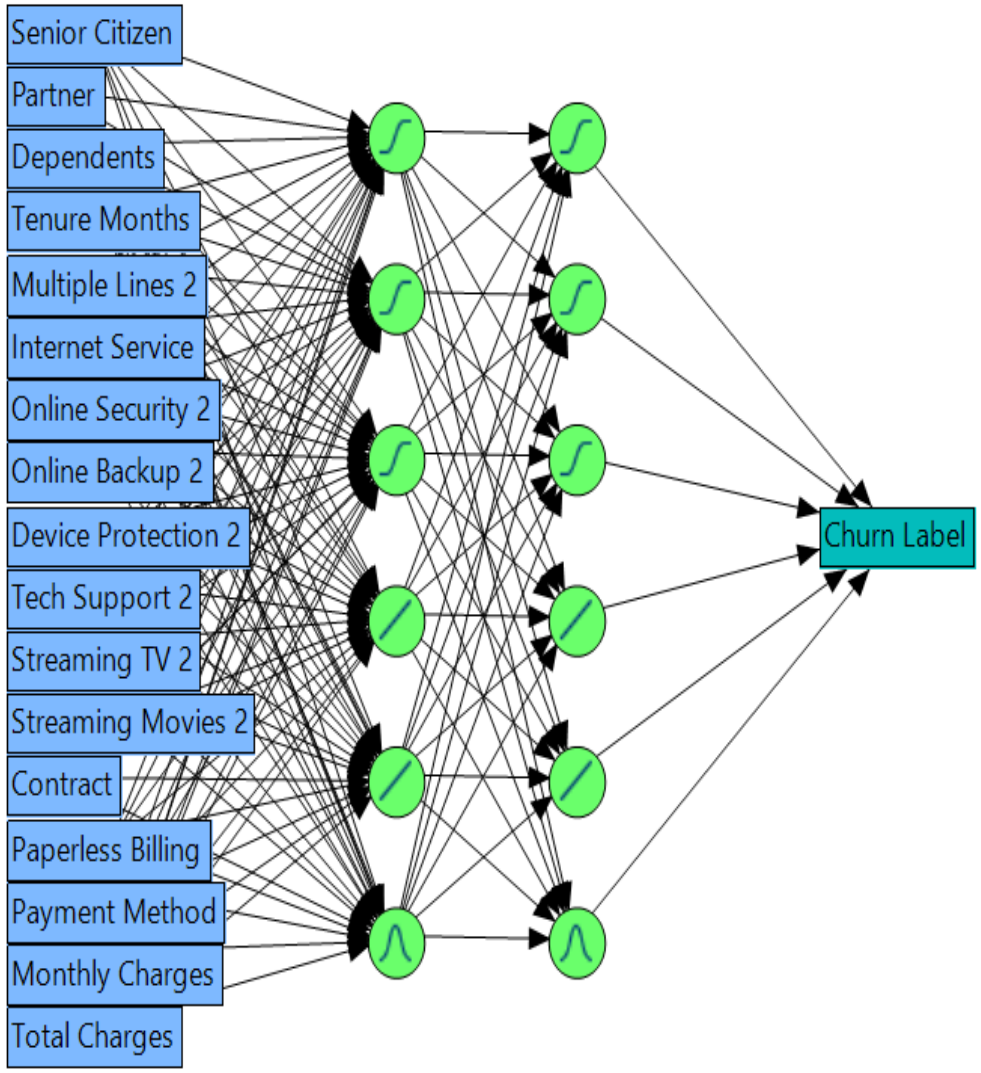


Churn Label	Area
No	0.8613
Yes	0.8613

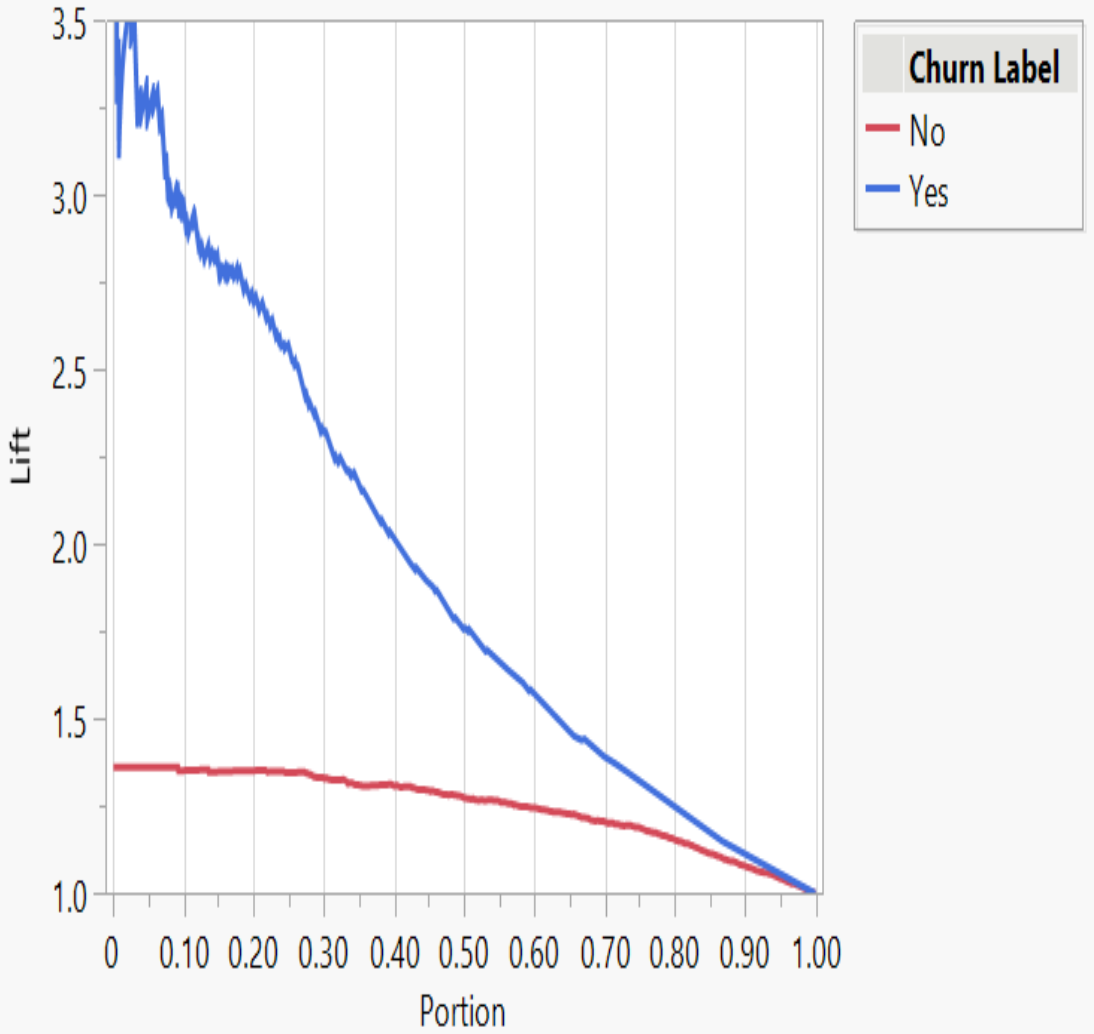
Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Neural 3T2G1L|3T2G1L	82.16%	70.70%	2.659	86.13%	17.84%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Diagram



Lift Curve on Test Data



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Estimates

Parameter	Estimate
H2_1:Senior Citizen:No	-30.9519
H2_1:Partner:No	0.368821
H2_1:Dependents:No	-0.4526
H2_1:Tenure Months	-0.01853
H2_1:Multiple Lines 2:No	-0.89322
H2_1:Internet Service:DSL	1.252298
H2_1:Internet Service:Fiber optic	0.705233
H2_1:Online Security 2:No	-0.1631
H2_1:Online Backup 2:No	2.040966
H2_1:Device Protection 2:No	1.974089
H2_1:Tech Support 2:No	-0.10164
H2_1:Streaming TV 2:No	-0.26111
H2_1:Streaming Movies 2:No	-1.48334
H2_1:Contract:Month-to-month	0.299192
H2_1:Contract:One year	-0.08058
H2_1:Paperless Billing:No	-2.22999
H2_1:Payment Method:Bank transfer (automatic)	0.246577
H2_1:Payment Method:Credit card (automatic)	-0.8637
H2_1:Payment Method:Electronic check	-2.2258
H2_1:Monthly Charges	0.021046
H2_1:Total Charges	-0.00026
H2_1:Intercept	19.88149
H2_2:Senior Citizen:No	-1.36653
H2_2:Partner:No	0.075692
H2_2:Dependents:No	-0.5717
H2_2:Tenure Months	-0.05587
H2_2:Multiple Lines 2:No	1.210862
H2_2:Internet Service:DSL	0.046118
H2_2:Internet Service:Fiber optic	0.339182
H2_2:Online Security 2:No	2.674172
H2_2:Online Backup 2:No	0.099904
H2_2:Device Protection 2:No	1.020439
H2_2:Tech Support 2:No	-0.98192
H2_2:Streaming TV 2:No	-0.14015
H2_2:Streaming Movies 2:No	0.598149
H2_2:Contract:Month-to-month	1.10147
H2_2:Contract:One year	1.362068
H2_2:Paperless Billing:No	0.123178
H2_2:Payment Method:Bank transfer (automatic)	0.985229
H2_2:Payment Method:Credit card (automatic)	-1.80577
H2_2:Payment Method:Electronic check	-0.21164
H2_2:Monthly Charges	-0.04677
H2_2:Total Charges	-0.00016
H2_2:Intercept	4.1872
H2_3:Senior Citizen:No	1.827127
H2_3:Partner:No	1.645003
H2_3:Dependents:No	0.986559
H2_3:Tenure Months	0.034832
H2_3:Multiple Lines 2:No	2.742521
H2_3:Internet Service:DSL	-1.23855
H2_3:Internet Service:Fiber optic	0.555784
H2_3:Online Security 2:No	-0.7752
H2_3:Online Backup 2:No	-0.63088
H2_3:Device Protection 2:No	-0.20286
H2_3:Tech Support 2:No	-0.26673
H2_3:Streaming TV 2:No	-0.89
H2_3:Streaming Movies 2:No	-0.58897
H2_3:Contract:Month-to-month	2.492478
H2_3:Contract:One year	0.452586
H2_3:Paperless Billing:No	-0.23481
H2_3:Payment Method:Bank transfer (automatic)	-1.74761
H2_3:Payment Method:Credit card (automatic)	-1.40637
H2_3:Payment Method:Electronic check	2.602057
H2_3:Monthly Charges	-0.00385
H2_3:Total Charges	-0.00012
H2_3:Intercept	-0.60178
H2_4:Senior Citizen:No	-0.14052
H2_4:Partner:No	1.295883
H2_4:Dependents:No	-4.13789
H2_4:Tenure Months	0.137904
H2_4:Multiple Lines 2:No	-0.14019
H2_4:Internet Service:DSL	-1.13799
H2_4:Internet Service:Fiber optic	-2.83704
H2_4:Online Security 2:No	-1.20907
H2_4:Online Backup 2:No	-0.77613
H2_4:Device Protection 2:No	0.099673
H2_4:Tech Support 2:No	-2.07688
H2_4:Streaming TV 2:No	0.081405
H2_4:Streaming Movies 2:No	0.726843
H2_4:Contract:Month-to-month	-5.50475
H2_4:Contract:One year	-0.3949
H2_4:Paperless Billing:No	0.289191
H2_4:Payment Method:Bank transfer (automatic)	-0.36213
H2_4:Payment Method:Credit card (automatic)	0.854796
H2_4:Payment Method:Electronic check	-0.95352
H2_4:Monthly Charges	-0.09127
H2_4:Total Charges	0.000957
H2_4:Intercept	7.1055
H2_5:Senior Citizen:No	1.583617
H2_5:Partner:No	0.677006
H2_5:Dependents:No	1.086698
H2_5:Tenure Months	-0.05726
H2_5:Multiple Lines 2:No	-2.81376
H2_5:Internet Service:DSL	-0.48467
H2_5:Internet Service:Fiber optic	0.571871
H2_5:Online Security 2:No	0.988464
H2_5:Online Backup 2:No	-0.27668
H2_5:Device Protection 2:No	2.38797
H2_5:Tech Support 2:No	-1.81293

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Estimates

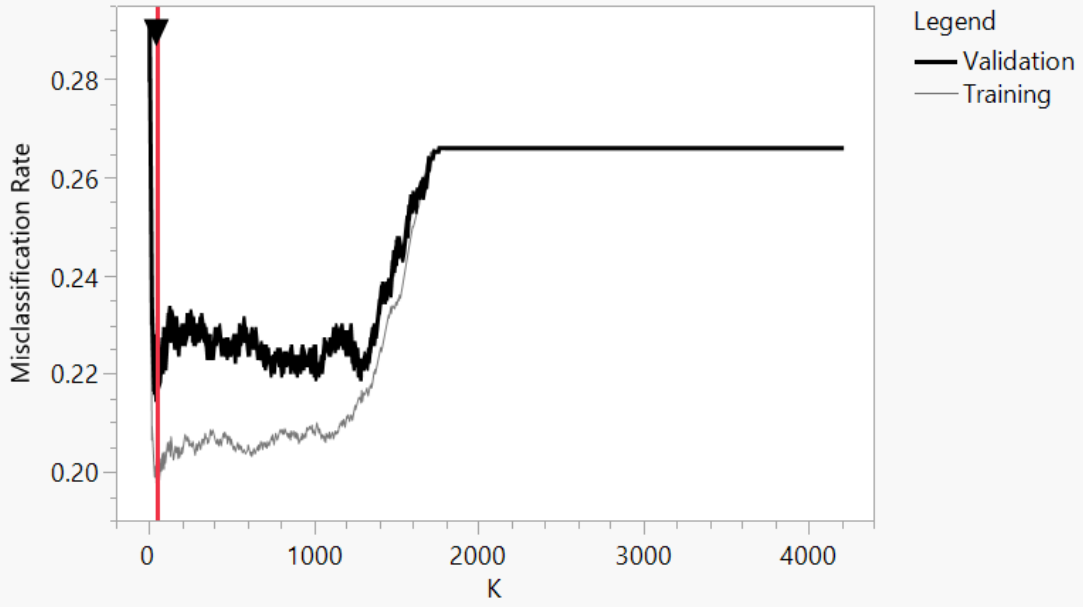
Parameter	Estimate
H2_1:Senior Citizen:No	-30.9519
H2_1:Partner:No	0.368821
H2_1:Dependents:No	-0.4526
H2_1:Tenure Months	-0.01853
H2_1:Multiple Lines 2:No	-0.89322
H2_1:Internet Service:DSL	1.252298
H2_1:Internet Service:Fiber optic	0.705233
H2_1:Online Security 2:No	-0.1631
H2_1:Online Backup 2:No	2.040966
H2_1:Device Protection 2:No	1.974089
H2_1:Tech Support 2:No	-0.10164
H2_1:Streaming TV 2:No	-0.26111
H2_1:Streaming Movies 2:No	-1.48334
H2_1:Contract:Month-to-month	0.299192
H2_1:Contract:One year	-0.08058
H2_1:Paperless Billing:No	-2.22999
H2_1:Payment Method:Bank transfer (automatic)	0.246577
H2_1:Payment Method:Credit card (automatic)	-0.8637
H2_1:Payment Method:Electronic check	-2.2258
H2_1:Monthly Charges	0.021046
H2_1:Total Charges	-0.00026
H2_1:Intercept	19.88149
H2_2:Senior Citizen:No	-1.36653
H2_2:Partner:No	0.075692
H2_2:Dependents:No	-0.5717
H2_2:Tenure Months	-0.05587
H2_2:Multiple Lines 2:No	1.210862
H2_2:Internet Service:DSL	0.046118
H2_2:Internet Service:Fiber optic	0.339182
H2_2:Online Security 2:No	2.674172
H2_2:Online Backup 2:No	0.099904
H2_2:Device Protection 2:No	1.020439
H2_2:Tech Support 2:No	-0.98192
H2_2:Streaming TV 2:No	-0.14015
H2_2:Streaming Movies 2:No	0.598149
H2_2:Contract:Month-to-month	1.10147
H2_2:Contract:One year	1.362068
H2_2:Paperless Billing:No	0.123178
H2_2:Payment Method:Bank transfer (automatic)	0.985229
H2_2:Payment Method:Credit card (automatic)	-1.80577
H2_2:Payment Method:Electronic check	-0.21164
H2_2:Monthly Charges	-0.04677
H2_2:Total Charges	-0.00016
H2_2:Intercept	4.1872
H2_3:Senior Citizen:No	1.827127
H2_3:Partner:No	1.645003
H2_3:Dependents:No	0.986559
H2_3:Tenure Months	0.034832
H2_3:Multiple Lines 2:No	2.742521
H2_3:Internet Service:DSL	-1.23855
H2_3:Internet Service:Fiber optic	0.555784
H2_3:Online Security 2:No	-0.7752
H2_3:Online Backup 2:No	-0.63088
H2_3:Device Protection 2:No	-0.20286
H2_3:Tech Support 2:No	-0.26673
H2_3:Streaming TV 2:No	-0.89
H2_3:Streaming Movies 2:No	-0.58897
H2_3:Contract:Month-to-month	2.492478
H2_3:Contract:One year	0.452586
H2_3:Paperless Billing:No	-0.23481
H2_3:Payment Method:Bank transfer (automatic)	-1.74761
H2_3:Payment Method:Credit card (automatic)	-1.40637
H2_3:Payment Method:Electronic check	2.602057
H2_3:Monthly Charges	-0.00385
H2_3:Total Charges	-0.00012
H2_3:Intercept	-0.60178
H2_4:Senior Citizen:No	-0.14052
H2_4:Partner:No	1.295883
H2_4:Dependents:No	-4.13789
H2_4:Tenure Months	0.137904
H2_4:Multiple Lines 2:No	-0.14019
H2_4:Internet Service:DSL	-1.13799
H2_4:Internet Service:Fiber optic	-2.83704
H2_4:Online Security 2:No	-1.20907
H2_4:Online Backup 2:No	-0.77613
H2_4:Device Protection 2:No	0.099673
H2_4:Tech Support 2:No	-2.07688
H2_4:Streaming TV 2:No	0.081405
H2_4:Streaming Movies 2:No	0.726843
H2_4:Contract:Month-to-month	-5.50475
H2_4:Contract:One year	-0.3949
H2_4:Paperless Billing:No	0.289191
H2_4:Payment Method:Bank transfer (automatic)	-0.36213
H2_4:Payment Method:Credit card (automatic)	0.854796
H2_4:Payment Method:Electronic check	-0.95352
H2_4:Monthly Charges	-0.09127
H2_4:Total Charges	0.000957
H2_4:Intercept	7.1055
H2_5:Senior Citizen:No	1.583617
H2_5:Partner:No	0.677006
H2_5:Dependents:No	1.086698
H2_5:Tenure Months	-0.05726
H2_5:Multiple Lines 2:No	-2.81376
H2_5:Internet Service:DSL	-0.48467
H2_5:Internet Service:Fiber optic	0.571871
H2_5:Online Security 2:No	0.988464
H2_5:Online Backup 2:No	-0.27668
H2_5:Device Protection 2:No	2.38797
H2_5:Tech Support 2:No	-1.81293

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : KNN:

Churn Label

Model Selection



Confusion Matrix for Best K=42

Training

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	2685	413
Yes	420	701

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.867	0.133
Yes	0.375	0.625

Validation

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	892	140
Yes	163	211

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.864	0.136
Yes	0.436	0.564

Test

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	914	119
Yes	145	229

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.885	0.115
Yes	0.388	0.612

Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
KNN	81.12%	65.90%	2.482		18.88%

ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : Naïve Bayes:

Confusion Matrix

Training

Actual	Predicted Count	
Churn Label	No	Yes
No	2518	580
Yes	306	815

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.813	0.187
Yes	0.273	0.727

Validation

Actual	Predicted Count	
Churn Label	No	Yes
No	807	225
Yes	119	255

Actual	Predicted Rate	
Churn Label	No	Yes
No	0.782	0.218
Yes	0.318	0.682

Test

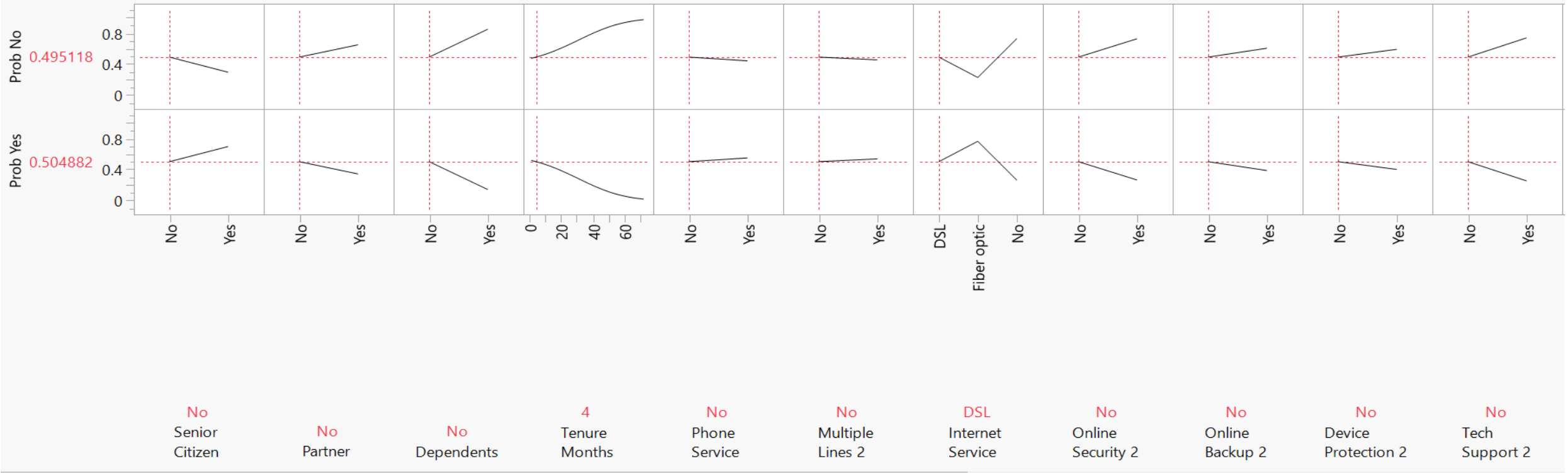
Actual	Predicted Count	
Churn Label	No	Yes
No	846	187
Yes	103	271

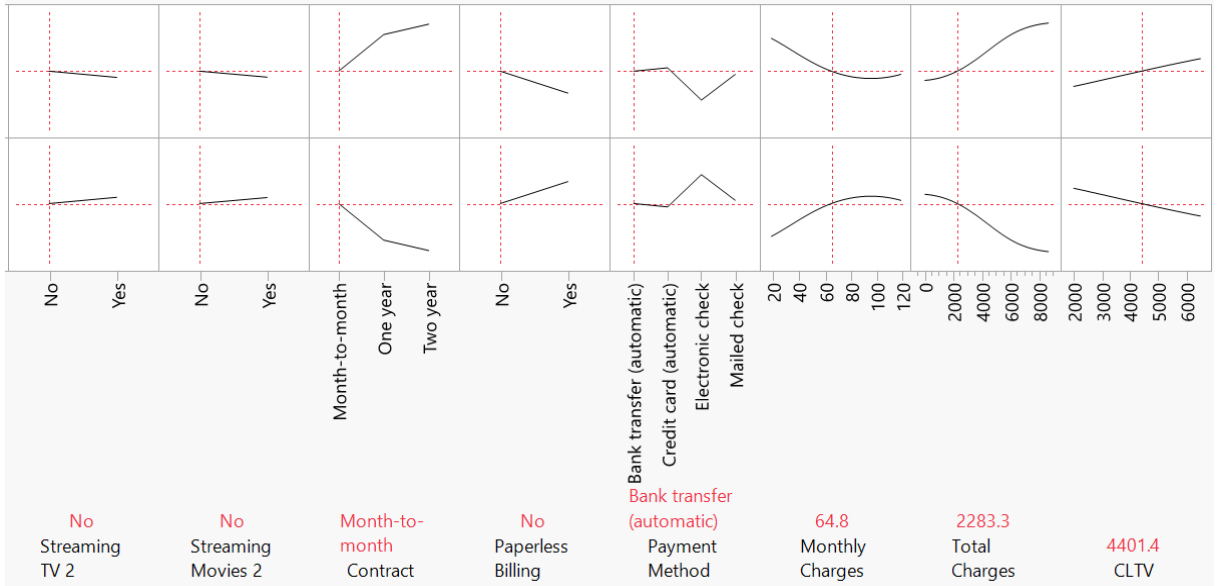
Actual	Predicted Rate	
Churn Label	No	Yes
No	0.819	0.181
Yes	0.275	0.725

Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Naive Bayes	80%	59%	2.26	84.88%	20.00%

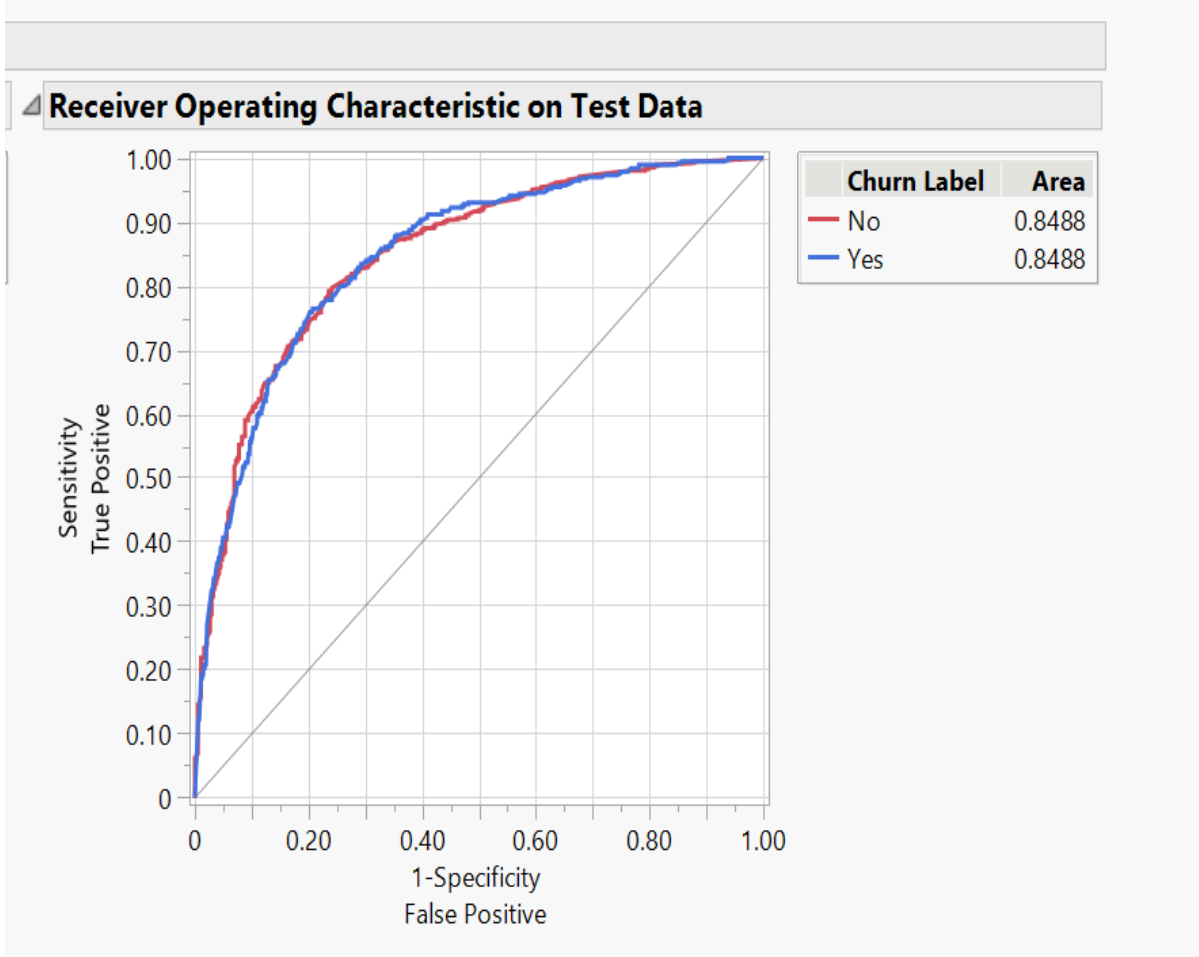
ⓒSaebyeol Yu. Saebyeol's PowerPoint

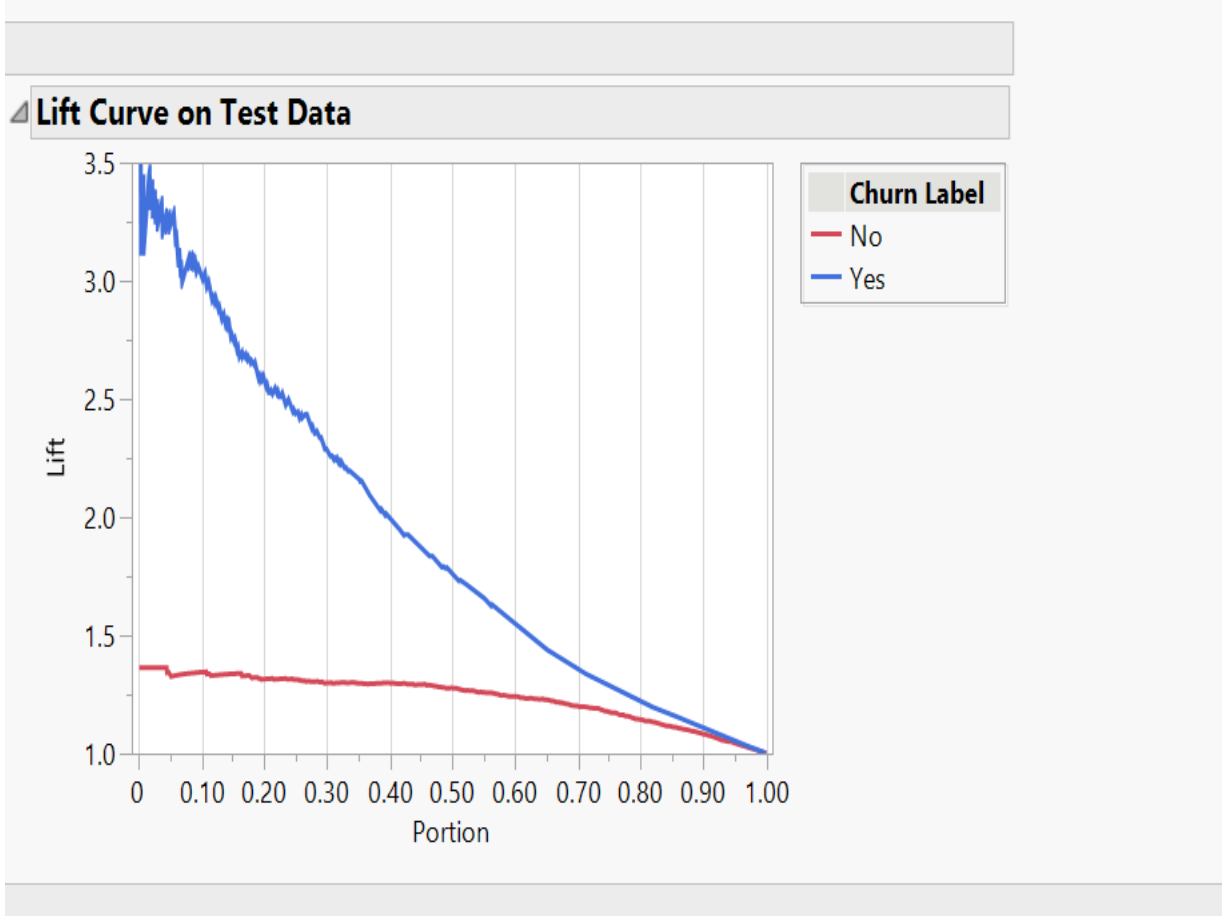
Prediction Profiler





ⓒSaebyeol Yu. Saebyeol's PowerPoint





ⓒSaebyeol Yu. Saebyeol's PowerPoint

Appendix : Discriminant Analysis:

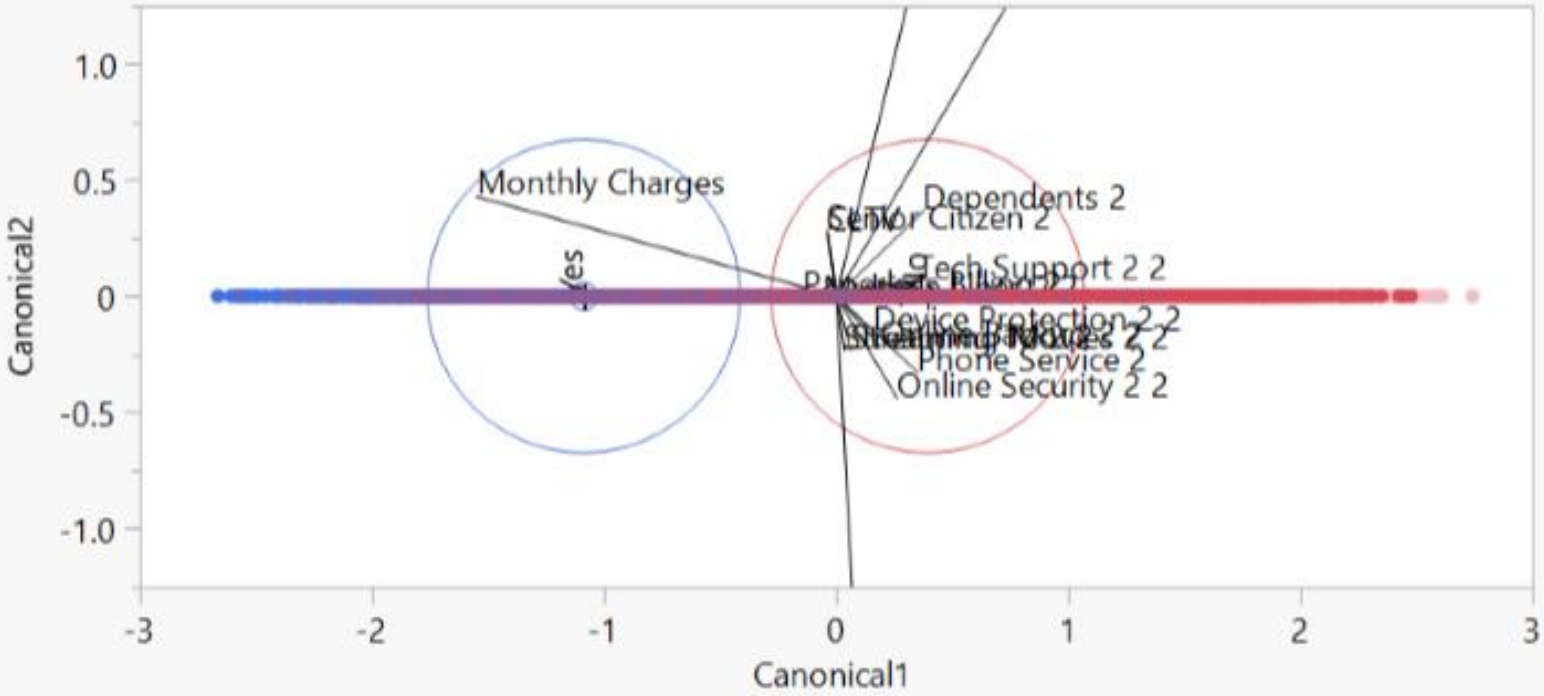
Discriminant Analysis

Discriminant Method: Linear

Classification: Churn Label

Validation: Validation Stratified by Churn Label

Canonical Plot



ⓒSaebyeol Yu. Saebyeol's PowerPoint

Discriminant Scores

Score Summaries

Source	Count	Number Misclassified	Percent Misclassified	Entropy RSquare	-2LogLikelihood
Training	4219	996	23.6075	0.15198	4142.65
Validation	1406	353	25.1067	0.10178	
Test	1407	331	23.5252	0.18833	

Training

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	2334	764
Yes	232	889

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.753	0.247
Yes	0.207	0.793

Validation

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	768	264
Yes	89	285

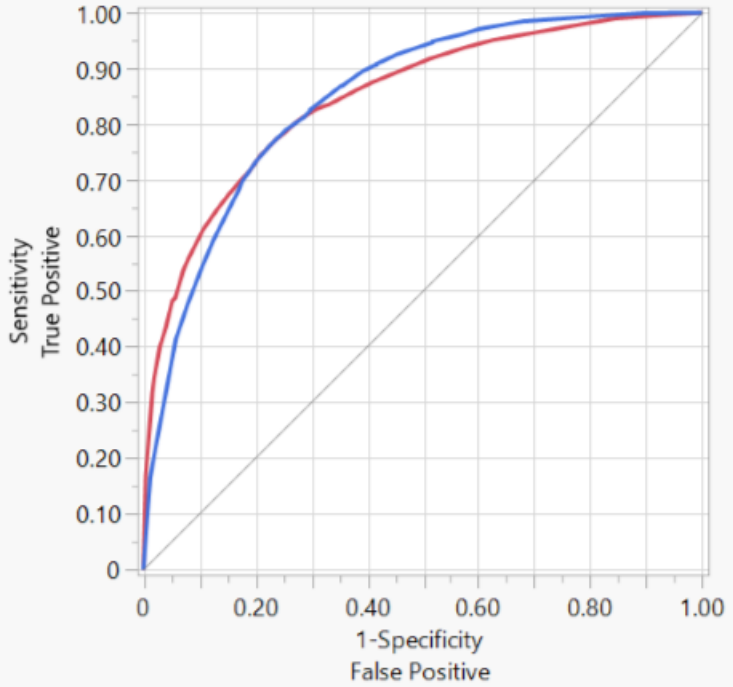
Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.744	0.256
Yes	0.238	0.762

Test

Actual Churn Label	Predicted Count No	Predicted Count Yes
No	785	248
Yes	83	291

Actual Churn Label	Predicted Rate No	Predicted Rate Yes
No	0.760	0.240
Yes	0.222	0.778

Receiver Operating Characteristic on Training Data



Churn Label	Area
No	0.8476
Yes	0.8476

Model	Accuracy	1's prediction	Lift	AUC	Misclassification Rate
Discriminant	76.48%	53.91%	2.0282	84.76%	23.52%

ⓒSaebyeol Yu. Saebyeol's PowerPoint