Starcraft 2

Team 3

Neehar Namjoshi, Nishita Rao, Pooja Shah, Shivam Kumar

Table of Contents

1. Introduction

1. Exploration of Data

1. Model Comparison

1. Business Recommendation

1. Appendix

Introduction

Starcraft 2 is a Real-Time Strategy (RTS) game

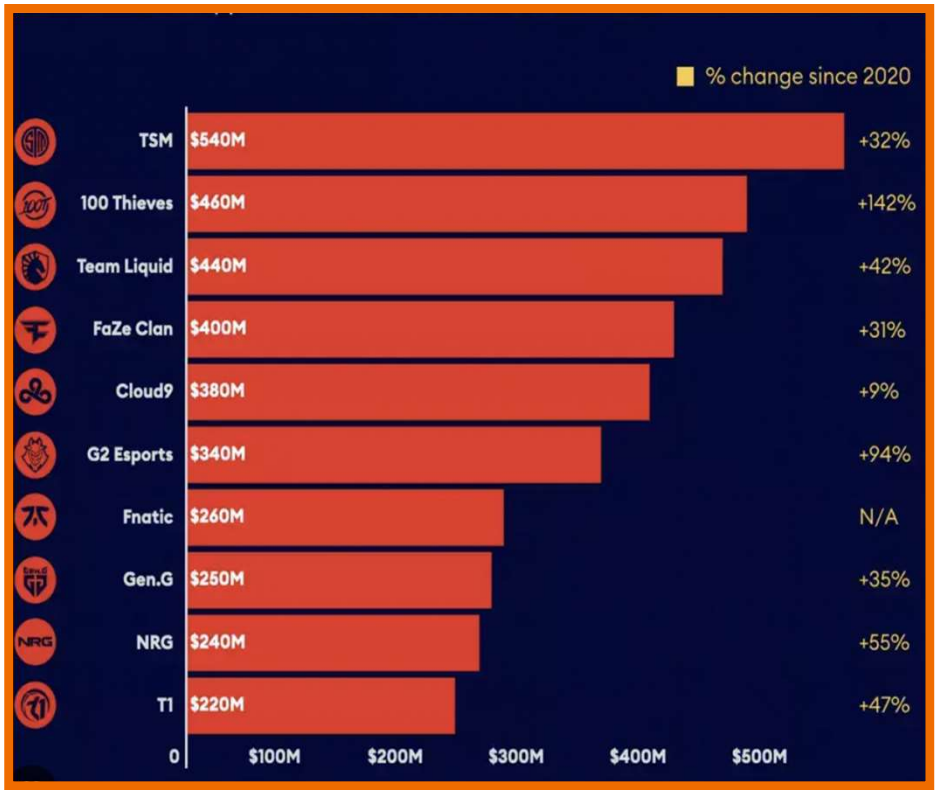
- 1v1 for professional tournaments
- Continuous not turn-based
- 20 x 3396
- Target Variable: Finding the highest skill level players

Business Case

Why do we want to find the best SC2 players for Electronic Sports Teams?

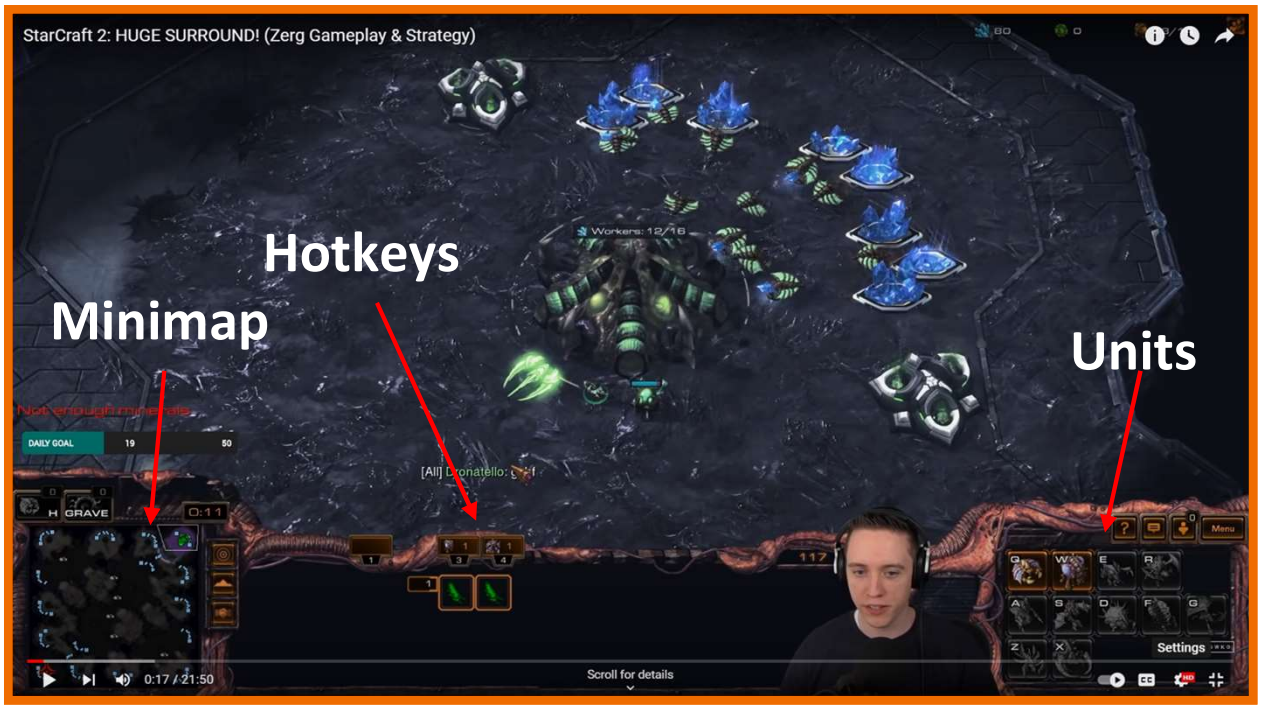
Esports Industry Valuation at $3.5billion (22% Growth)

- Prize Pool
- Sponsors
- Recognition
- Merchandise Sales



Breaking down the Dataset

- 1 timestamp = 1/88.5 seconds = 11.3 milliseconds
- Action : mouse-click or keystroke
- Hotkey : a key or combination of keys for quick access (eg. Ctrl + S)

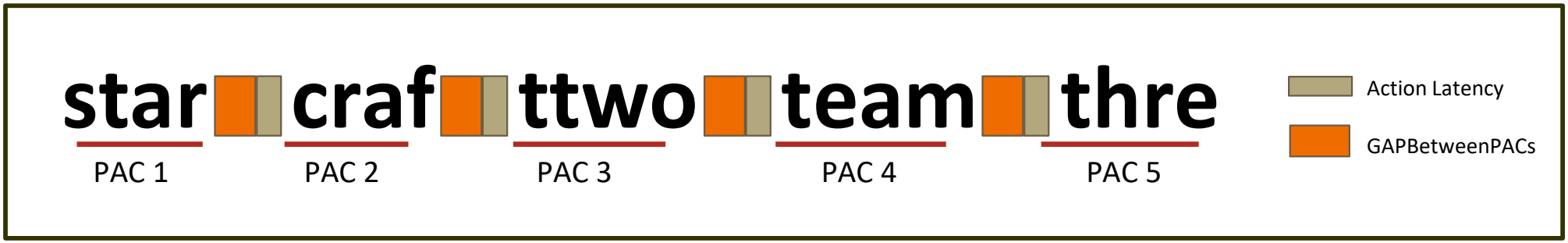


Variables
SelectByHotkeys
AssignToHotkeys
UniqueHotkeys
MinimapAttacks
No. of actions per time stamp
UniqueUnitsMade
ActionLatency
GapBetweenPACs

Breaking down the Dataset (Contd.)

What is Perception Action Cycle (PAC) ?
Example: starcraft two team three

Variables
✓ ActionLatency
✓ GapBetweenPACs



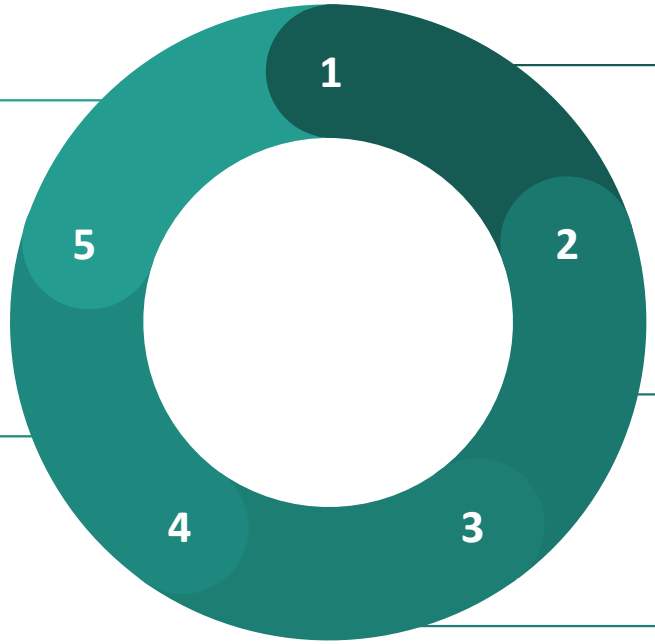
Exploration of Data

Significant Predictors

- ➔ Checking for statistical significance
- ➔ Assessing correlations between predictors

Reducing Dimensionality

- ➔ Interaction terms



Exploring Variables

- ➔ Verifying data type of variables in data set
- ➔ Changes to target variable
- ➔ Determining non-relevant variables

Missing Data

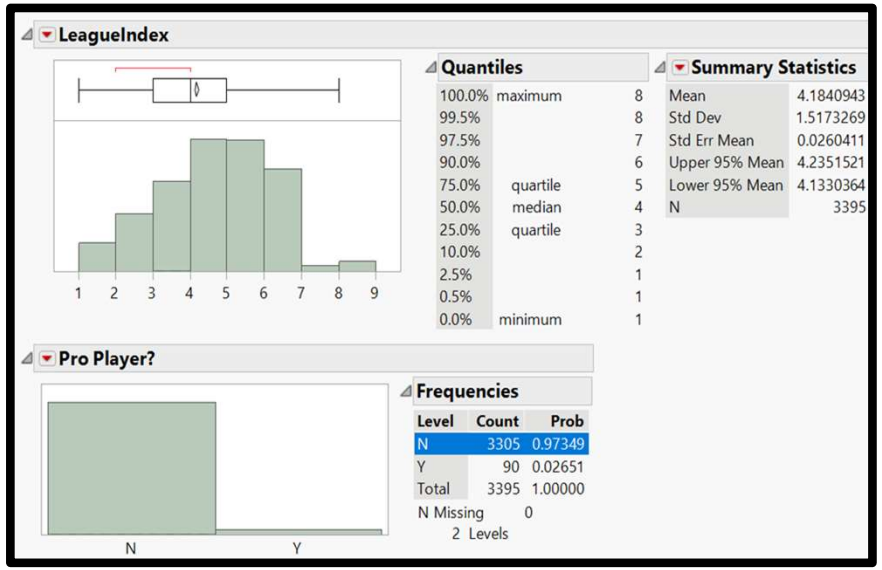
- ➔ Undertaking checks throughout dataset

Outliers

- ➔ Reviewing distribution of predictor variables
- ➔ Transformation of key variables

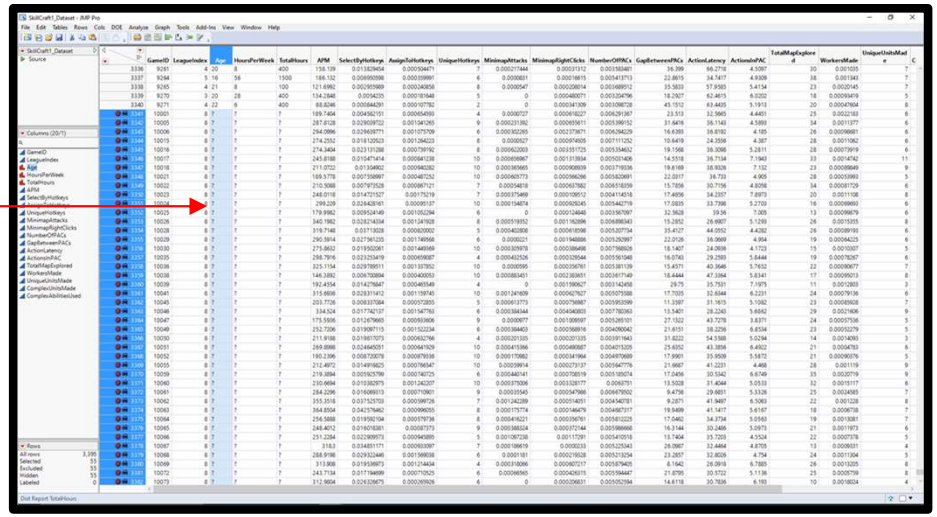
Exploring Variables

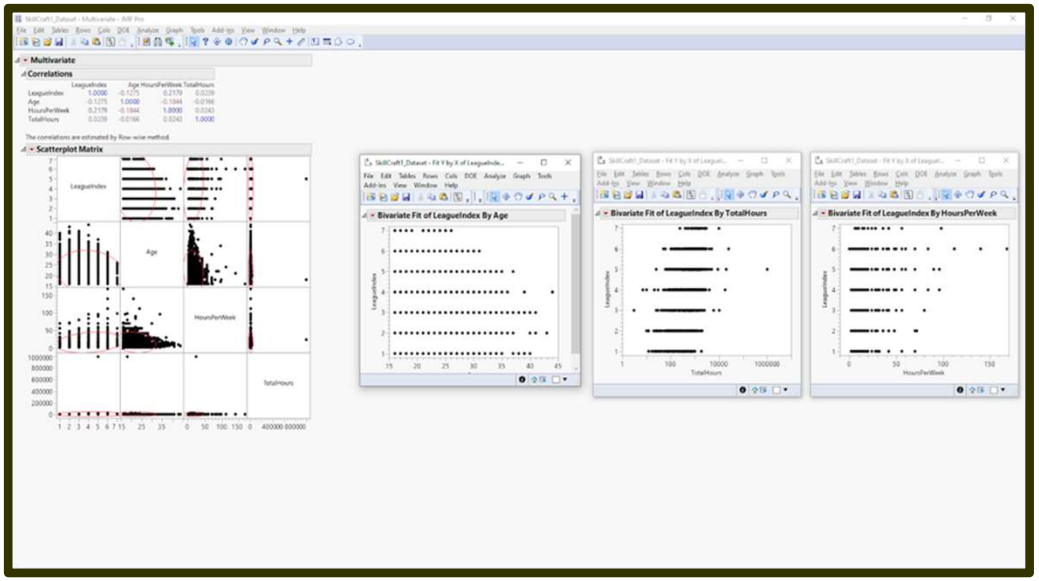
01	Properties of Predictor Variables	• Checks for data types of predictor variables
02	Recategorization of Target Variable	• Creation of Pro Player? as a categorical variable (Yes/No)
03	Determining non-relevant variables	• Exclusion of GameID



Missing Data

➔ Missing values found in predictor variables: Age, HoursPerWeek and Total Hours



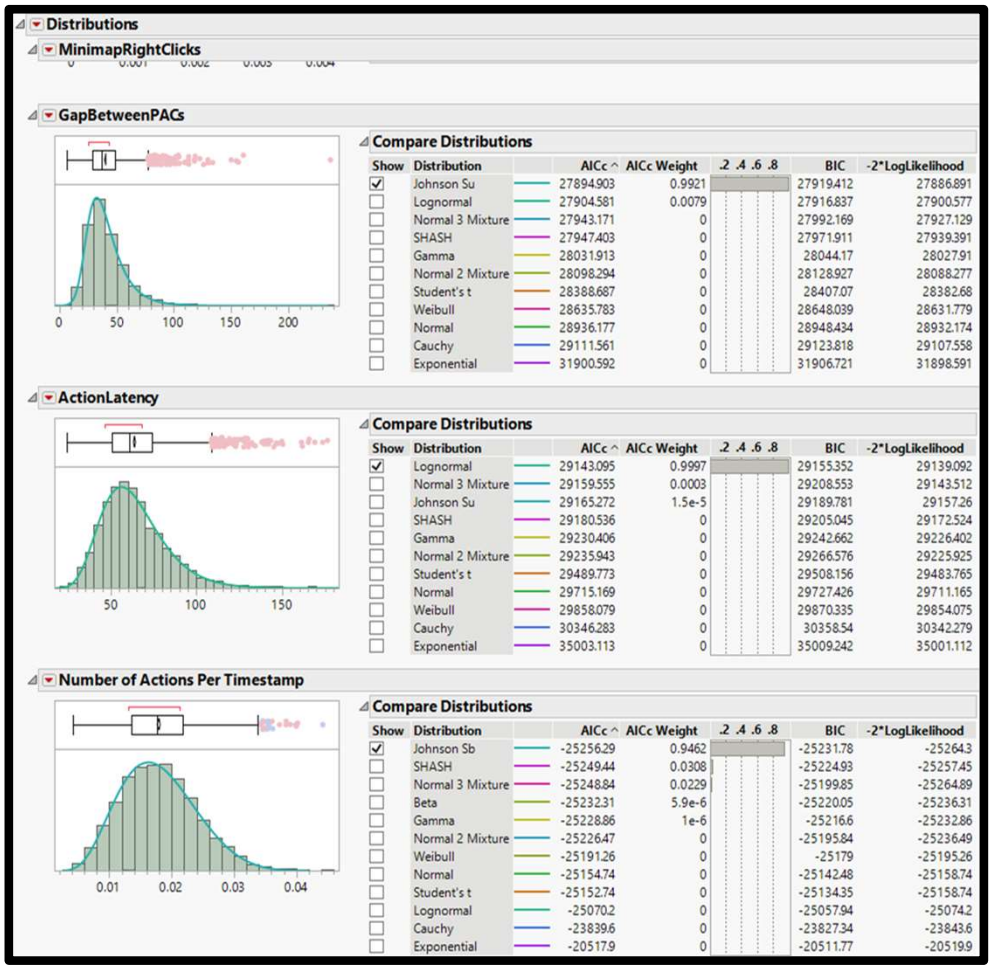


➔ Relationship between these predictor variables and target variable were explored

Decision

Hide and exclude Age, HoursPerWeek and TotalHours

Outliers

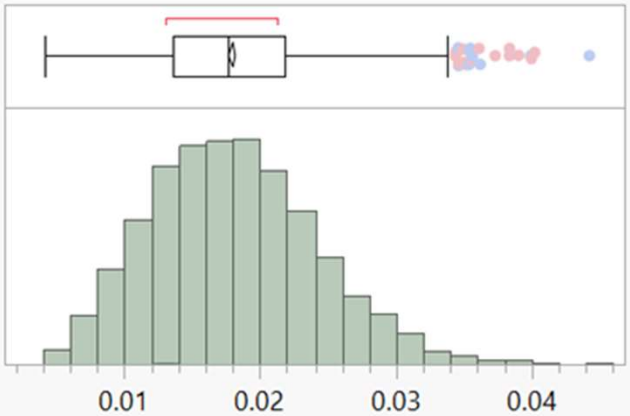


- ➔ Lot of outliers were identified across predictor variables
- ➔ Highly skewed distributions
- ➔ Exclusion of rows could lead to loss of information

<u>Decision</u>
Transformation of predictor variables
to scale data

Reducing Dimensionality

Number of Actions Per Timestamp



Quantiles

100.0%	maximum	0.0442407
99.5%		0.0350272
97.5%		0.0304564
90.0%		0.0258129
75.0%	quartile	0.0218188
50.0%	median	0.0176815
25.0%	quartile	0.0135288
10.0%		0.0103464
2.5%		0.0075392
0.5%		0.0056408
0.0%	minimum	0.0040994

Summary Statistics

Mean	0.0179157
Std Dev	0.0059518
Std Err Mean	0.0001021
Upper 95% Mean	0.018116
Lower 95% Mean	0.0177154
N	3395

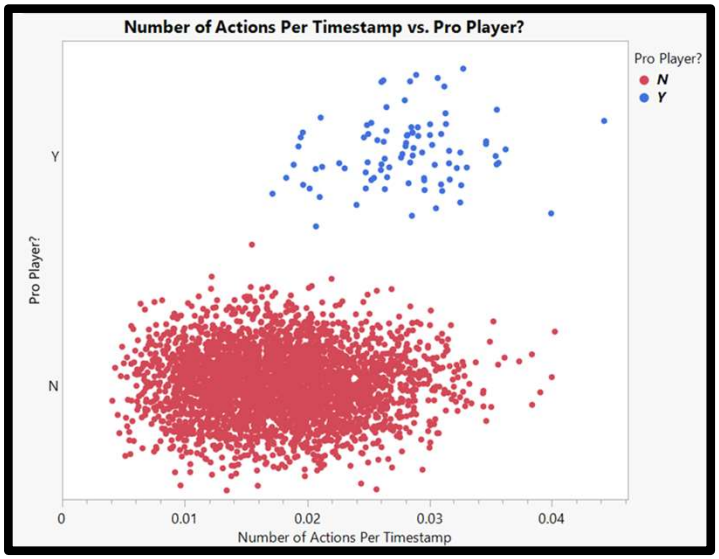
*Number of Actions Per Timestamp = Number of PACs **x** Actions in PACs*

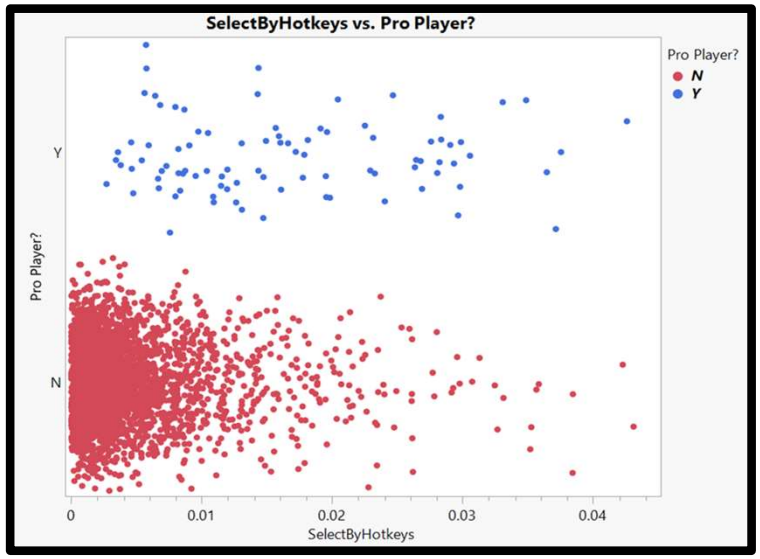
<u>Decision</u>

➔ Introduction of interaction term to increase interpretability

➔ Ensure predictor variables are measured per timestamp

Significant Predictors - Correlation





➔ Correlation between target and predictor variables were explored

<u>Decision</u>
Spread of data confirms some differentiation between target class

Significant Predictors

Selecting the most important variables

20 → 8

Nominal Logistic Fit for Pro Player?

Effect Summary

Source	LogWorth	PValue
ActionLatency	6.348	0.00000
MinimapAttacks	5.419	0.00000
AssignToHotkeys	4.116	0.00008
UniqueHotkeys	2.465	0.00343
SelectByHotkeys	1.949	0.01125
Number of Actions Per Timestamp	1.912	0.01225
UniqueUnitsMade	0.833	0.14699
GapBetweenPACs	0.788	0.16311

Predicting the Cream

Comparing the best sets of models to predict our RARE (2.7%) class!

Starcraft_2_SecondIterationjmp - Model Comparison - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

Model Comparison Validation=Training

Predictors

Measures of Fit for Pro Player?

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.6140	0.6423	0.0466	0.1100	0.0248	0.0137	2037
Neural Model NGaussian(3)NGaussian2(3) - Save Formulas		0.6003	0.6290	0.0482	0.1138	0.0275	0.0167	2037
Bootstrap Forest		0.8258	0.8427	0.021	0.0751	0.0165	0.0079	2037
Neural Model NGaussian(3) - Save Formulas		0.6126	0.6409	0.0467	0.1103	0.0243	0.0142	2037
Model Averaged		0.7118	0.7361	0.0348	0.1001	0.0233	0.0123	2037

Confusion Matrix

Model Comparison Validation=Validation

Predictors

Measures of Fit for Pro Player?

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.4508	0.4699	0.0421	0.1105	0.0250	0.0162	679
Neural Model NGaussian(3)NGaussian2(3) - Save Formulas		0.4709	0.4900	0.0406	0.1071	0.0266	0.0191	679
Bootstrap Forest		0.5397	0.5587	0.0353	0.1027	0.0225	0.0162	679
Neural Model NGaussian(3) - Save Formulas		0.5160	0.5352	0.0371	0.1046	0.0220	0.0162	679
Model Averaged		0.5280	0.5470	0.0362	0.1027	0.0240	0.0162	679

Model Comparison Validation=Test

Predictors

Measures of Fit for Pro Player?

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Fit Nominal Logistic		0.6508	0.6881	0.0584	0.1330	0.0328	0.0221	679
Neural Model NGaussian(3)NGaussian2(3) - Save Formulas		0.6298	0.6682	0.0619	0.1307	0.0352	0.0177	679
Bootstrap Forest		0.5625	0.6032	0.0732	0.1393	0.0351	0.0236	679
Neural Model NGaussian(3) - Save Formulas		0.6146	0.6536	0.0644	0.1329	0.0310	0.0206	679
Model Averaged		0.6492	0.6865	0.0587	0.1305	0.0335	0.0177	679

1. Neural network
2. Logistic Regression
3. Bootstrap
4. Boosted
5. Discriminant
6. KNN
7. Ensemble models

Deep Diving into the Best 2

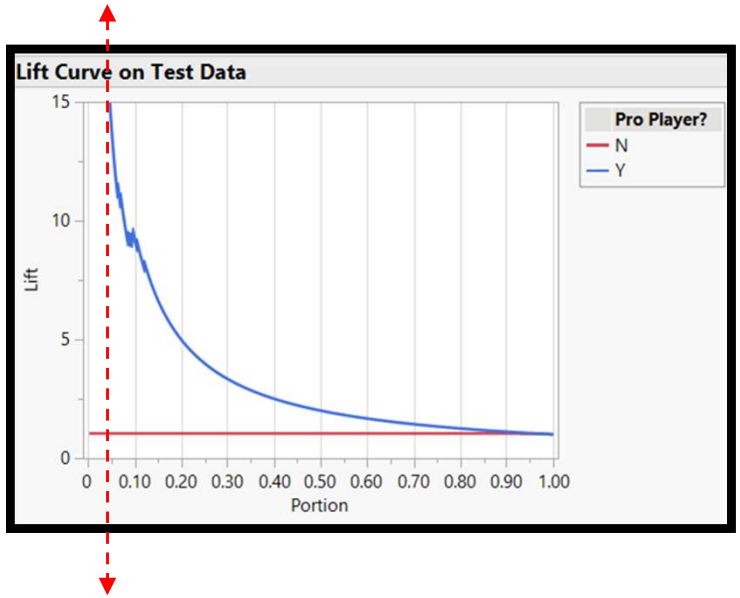
Gaussian managing complex relationships between predictors & target

Measures	Logit + Gaussian + Bootstrap Ensemble	2 layered Gaussian NN
Accuracy of 1	100%	89%
Sensitivity	48%	63%
Mis-classification rate	1.8%	1.8%
F1 score	65%	74%

Selecting NN meeting business requirement

2 layered Gaussian NN Model!

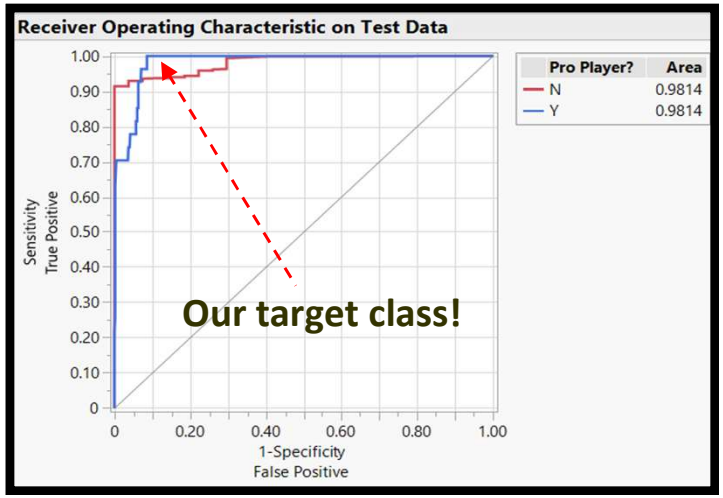
Lift curve



At 2.7% baseline, LIFT ~ 15

Response rate ~ 41%

AUC for ROC curve:

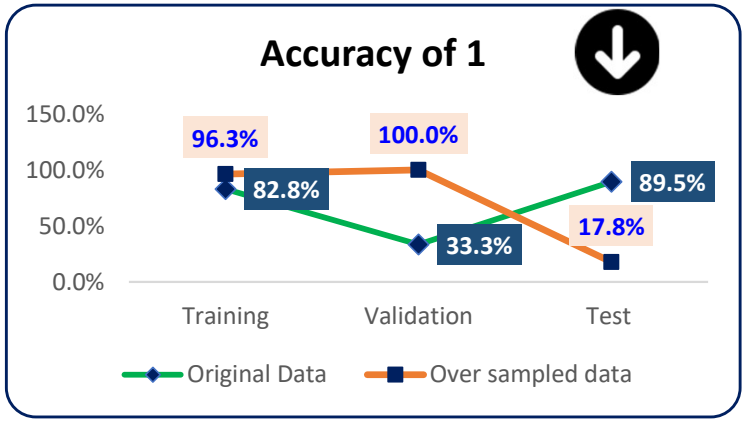


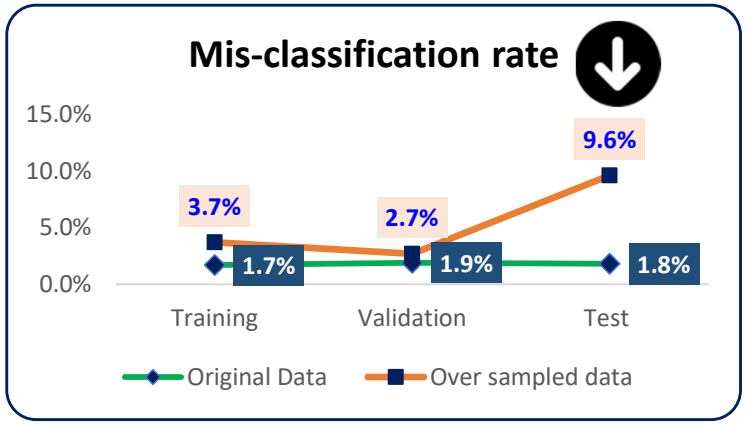
AUC 98%

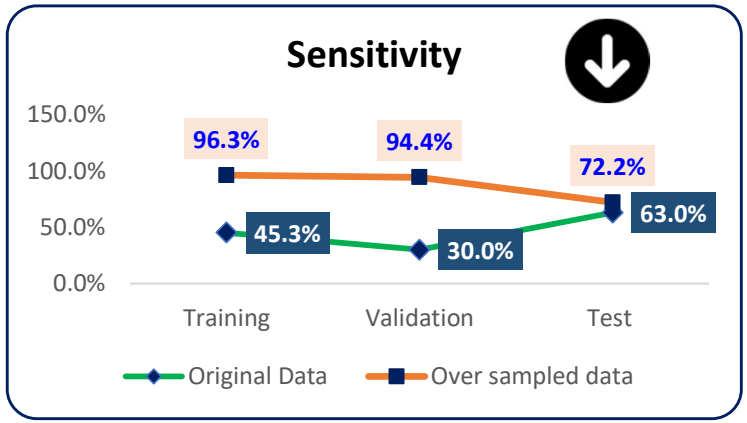
Tuning the model to better determine the rare event

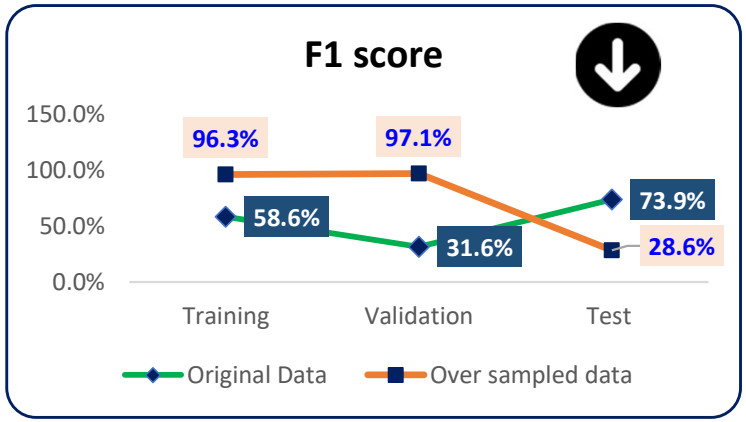
Over sampling:

Model overfits to training & validation



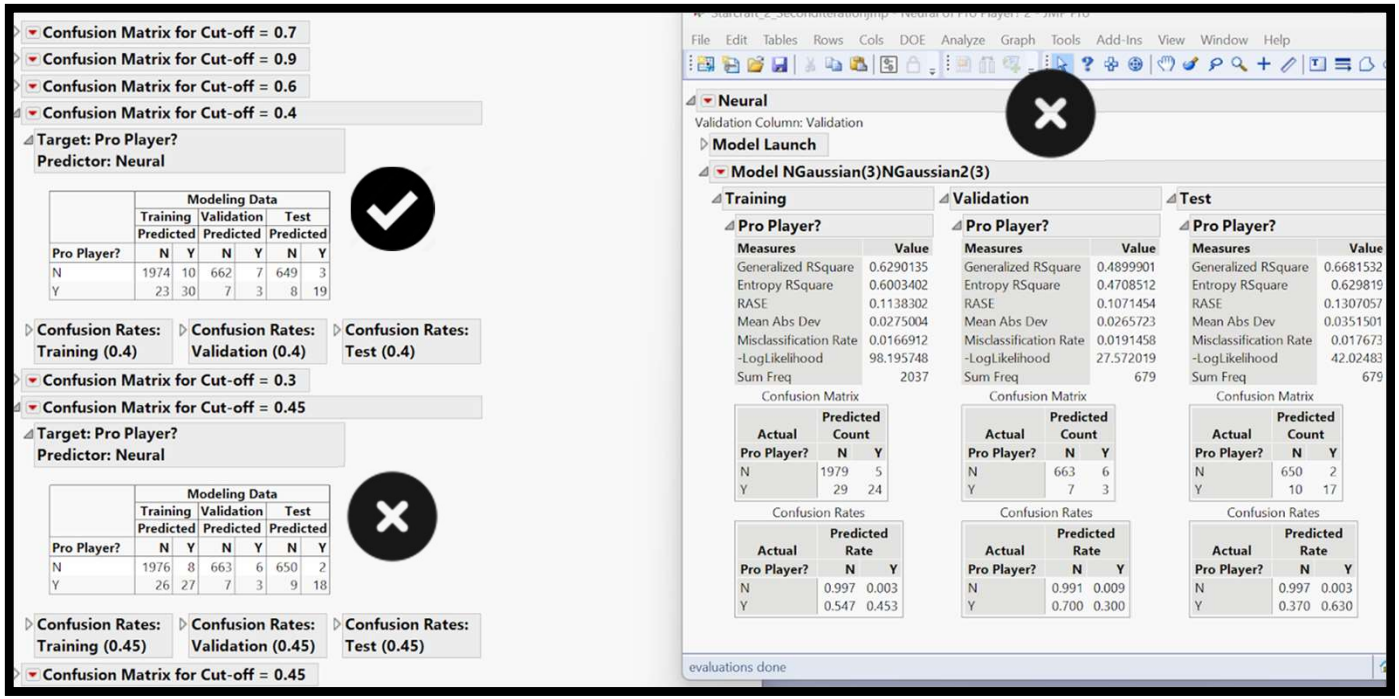






Modulating for the best prediction

Determining the optimal cut – off!



Closing with the Best Model

OPTIMAL -> 2 layered 3 node Gaussian + 0.4 cut off

Accuracy of 1: 86%

Sensitivity: 70%

Complexity: HIGH

Overall Accuracy: 98.4%

Business Recommendations

01

Identifying Potential Elite Players and Recruiting

02

Invest in Training and Development of Players

03

Optimizing Talent Management Efforts

Largest Prize Pools

	Tournament Name	Prize Money
1.	» WCS 2017 Global Finals	$700,000.00
2.	» WCS 2018 Global Finals	$700,000.00
3.	» WCS 2019 Global Finals	$700,000.00
4.	» IEM Season XVI - Katowice (SC2)	$500,000.00
5.	» IEM Season XVII - Katowice (SC2)	$500,000.00
6.	» WCS 2016 Global Finals	$500,000.00
7.	» WESG 2016 (SC2)	$402,000.00
8.	» IEM XIV - Katowice 2020 (SC2)	$400,200.00
9.	» IEM XII - World Championship (SC2)	$400,000.00
10.	» IEM XIII - Katowice 2019 (SC2)	$400,000.00
11.	» WESG 2017 (SC2)	$400,000.00
12.	» IEM Season XV - World Championship (SC2)	$250,000.00
13.	» IEM XI - World Championship (SC2)	$250,000.00
14.	» WCS 2012: Global Finals	$250,000.00
15.	» WCS 2013 Global Finals	$250,000.00
16.	» WCS 2014 Global Finals	$250,000.00
17.	» WCS 2015 Global Finals	$250,000.00
18.	» WESG 2018 (SC2)	$240,000.00
19.	» WCS 2015 Season 1 Premier League	$217,000.00
20.	» WCS 2015 Season 2 Premier	$217,000.00
21.	» WCS 2015 Season 3 Premier	$217,000.00
22.	» SK Planet Proleague Regular Season	$186,749.65
23.	» GSL Super Tournament 2011	$185,832.56
24.	» GSL Open Season 2	$176,942.60
25.	» GSL Open Season 1	$176,621.52

[more]

Top Players

	Player ID	Player Name	Total (Game)
1.	Serral	Joona Sotala	$1,215,486.05
2.	Maru	Cho, Sung Choo	$1,160,811.65
3.	Rogue	Lee, Byung Ryul	$1,074,615.88
4.	Dark	Park, Ryung Woo	$986,994.93
5.	INnoVation	Lee, Shin Hyung	$794,621.00
6.	TY	Jun, Tae Yang	$698,044.11
7.	Zest	Joo, Sung Wook	$688,532.59
8.	sOs	Kim, Yoo Jin	$664,979.39
9.	soO	Eu, Yoon Su	$632,226.81
10.	Stats	Kim, Dae Yeob	$622,727.09
11.	Neeb	Alex Sunderhaft	$589,644.76
12.	Reynor	Riccardo Romiti	$569,040.76
13.	ByuN	Byun, Hyun Woo	$532,514.33
14.	Solar	Kang, Min Soo	$511,019.38
15.	Classic	Kim, Doh Woo	$510,507.61
16.	MC	Jang, Min Chul	$509,771.79
17.	herO	Kim, Joon Ho	$489,724.98
18.	Life	Lee, Seung Hyun	$475,900.67
19.	PartinG	Won, Lee Sak	$474,313.54
20.	SpeCial	Juan Carlos Lopez	$454,774.51
21.	Polt	Choi, Sung Hoon	$451,943.24
22.	Scarlett	Sasha Hostyn	$442,078.37
23.	MMA	Moon, Sung Won	$407,169.17
24.	Trap	Cho, Sung Ho	$406,462.53
25.	Mvp	Jung, Jong Hyun	$404,994.83

[more]

References

References for data, information and images used in the presentation are as below:

- Data source: https://data.world/scuttlemonkey/starcraft-2-skillcraft-1
- https://starcraft2.com/en-us/media#screenshots-1
- https://www.youtube.com/watch?v=3frQjtyogMs
- https://www.youtube.com/watch?v=7YFENdjm5Lw
- https://www.forbes.com/sites/brettknight/2022/05/06/the-most-valuable-esports-companies-2022/?sh=431072d1599f
- https://www.esportsearnings.com/games/151-starcraft-ii
- https://news.blizzard.com/en-gb/starcraft2/13235353/winner-takes-all-iem-world-championship
- https://www.beyondgames.biz/32134/the-future-of-esports-growth/#:~:text=The%20esports%20market%20is%20estimated,events%20and%20new%20platform%20launches.

“The work contained and presented here is our work and our work alone.”

Appendix

Eyeballing the Data Set

Variable	Descriptions	Type of information
Pro Player?	Rank in the game (Yes/No): where Yes predicts a person who ranks as GrandMaster (Rank 7) and Professional leagues (Rank 8)	categorical
LeagueIndex	Rank in the game:Bronze, Silver, Gold, Platinum, Diamond, Master, GrandMaster, and Professional leagues ,coded 1-8 (further these are ranks such as in Chess Masters, Grandmasters, etc)	categorical
GameID	Unique ID number for each individual in each game	continuous
Age	Age of each player	continuous
HoursPerWeek	Reported hours spent playing per week	continuous
TotalHours	Reported total hours spent playing	continuous
APM	Action per minute	continuous
SelectByHotkeys	Number of unit or building selections made using hotkeys per timestamp	continuous
AssignToHotkeys	Number of units or buildings assigned to hotkeys per timestamp	continuous
UniqueHotkeys	Number of unique hotkeys used per timestamp	continuous
MinimapAttacks	Number of attack actions on minimap per timestamp	continuous
MinimapRightClicks	Number of right-clicks on minimap per timestamp	continuous
NumberOfPACs	Number of PACs per timestamp (Please refer to the paper below that explains visual and screen fixations)	continuous
GapBetweenPACs	Mean duration in milliseconds between PACs	continuous
ActionLatency	Mean latency from the onset of a PACs to their first action in milliseconds	continuous

ActionsInPAC	Mean number of actions within each PAC	continuous
TotalMapExplored	The number of 24x24 game coordinate grids viewed by the player per timestamp	continuous
WorkersMade	Number of SCVs, drones, and probes trained per timestamp	continuous
UniqueUnitsMade	Unique unites made per timestamp	continuous
ComplexUnitsMade:	Number of ghosts, infestors, and high templars trained per timestamp	continuous
ComplexAbilitiesUsed:	Abilities requiring specific targeting instructions used per timestamp	continuous

➔ “Pro Player? “ is the target variable for this data set

- Eyeballing the data set, defining the target variable, and checking the data type for each variable

Data Pre-processing: Reducing Dimensionality

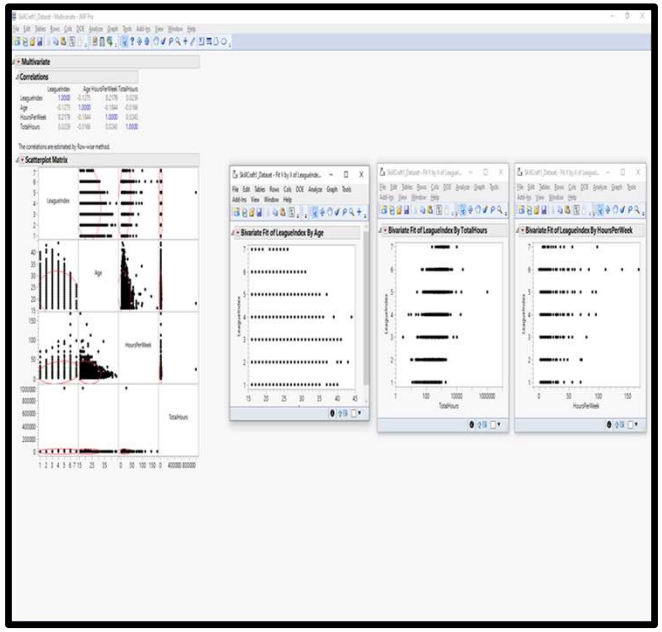
A. **Reducing row dimensionality:** As all our data rows contained individual player details, we wanted to use the complete data set and avoid any sampling to be done on the same. Hence, apart from excluding the missing data rows, we have used the complete data set

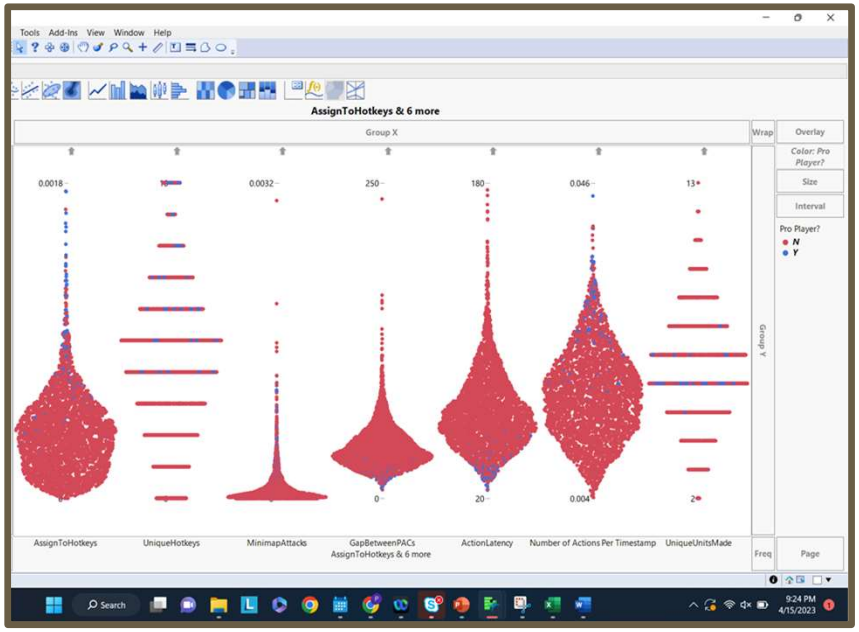
A. **Reducing column dimensionality:** We reduced the complexity of our columns by determining the important predictors using data exploration and running a few models. The log worth was examined to find predictors with high degree of importance to train the model using the same and make predictions.

- By addressing column dimensionality, we have improved the overall prediction of target class drastically. The target variable's categories have been reduced to 2 to minimize the complexity.

Data Exploration: Correlations

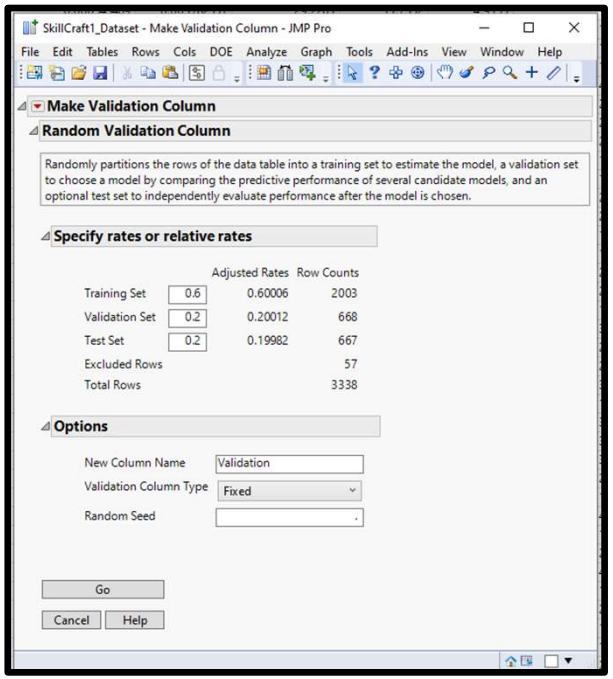
Visual representations have formed major part of our exploration (checking correlation of predictors to target)

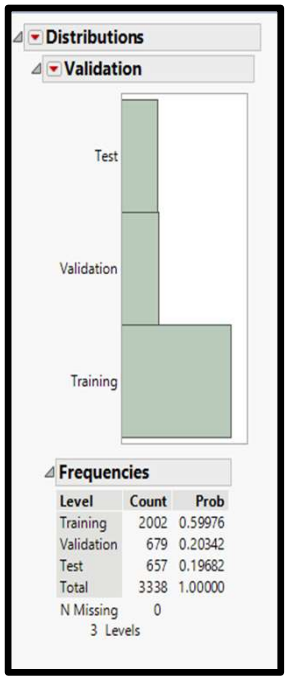




- Checking correlations between our predictor variables and target variable

Data Partition



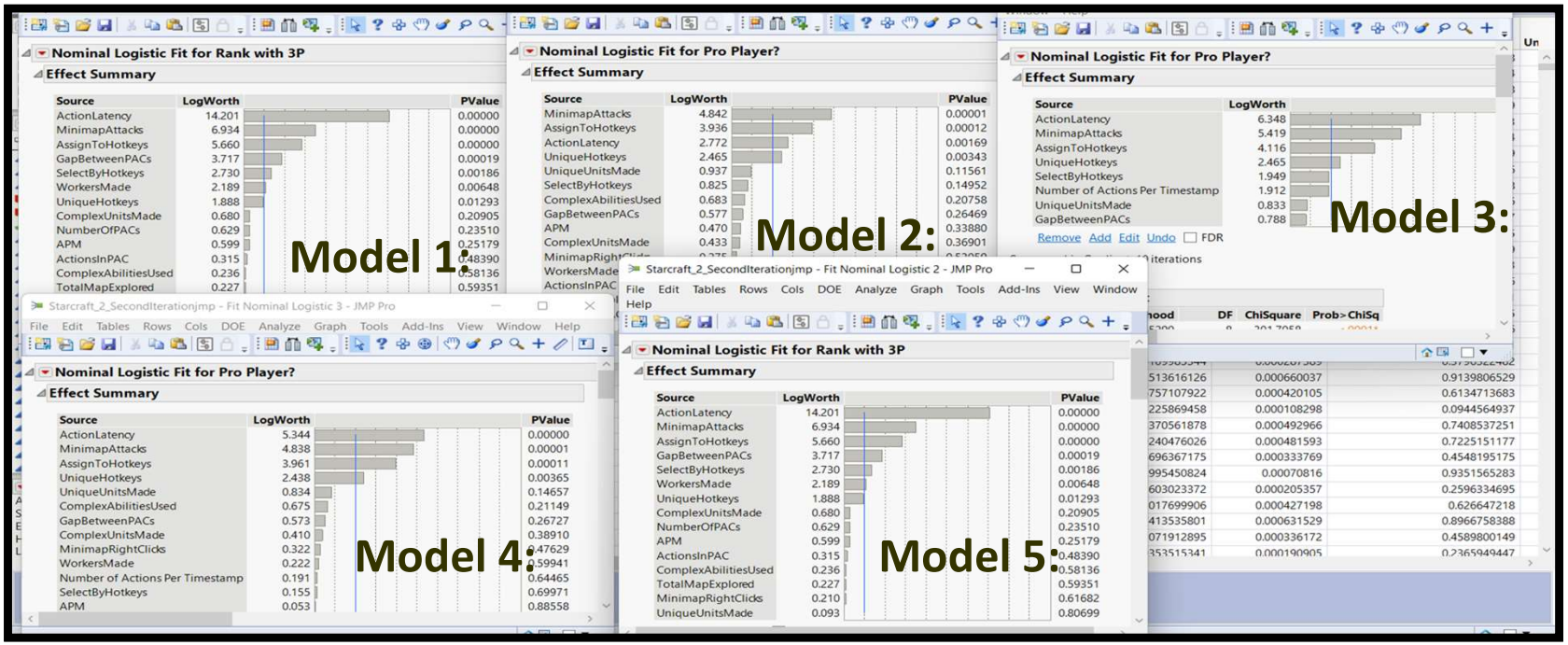


- Dataset was partitioned into 3 sections: Training, Validation and Test
- The weightage for the 3 partitions are as follows:
  - Training - 60%
  - Validation - 20%
  - Test - 20%

- Partitioning the data to get it ready to model

Modeling - Different types and comparison

Selecting the predictor variables



- We ran logistic regression to determine our key variables. Through above 5 different models, we could conclude on selecting 8 variables out of 20 that had the highest log worth.

Modeling - Different types and comparison

Variable	Selection
GameID	
LeagueIndex	
Age	
HoursPerWeek	
TotalHours	
APM	
SelectByHotkeys	Yes
AssignToHotkeys	Yes
UniqueHotkeys	Yes
MinimapAttacks	Yes
MinimapRightClicks	
NumberOfPACs	
GapBetweenPACs	Yes
ActionLatency	Yes
ActionsInPAC	
No. of actions per time stamp	Yes
AggregatedlActionsin PAC	
TotalMapExplored	
WorkersMade	
UniqueUnitsMade	Yes

Nominal Logistic Fit for Pro Player?

Effect Summary

Source	LogWorth	PValue
ActionLatency	6.348	0.00000
MinimapAttacks	5.419	0.00000
AssignToHotkeys	4.116	0.00008
UniqueHotkeys	2.465	0.00343
SelectByHotkeys	1.949	0.01125
Number of Actions Per Timestamp	1.912	0.01225
UniqueUnitsMade	0.833	0.14699
GapBetweenPACs	0.788	0.16311

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.6140	0.4508	0.6508	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6423	0.4699	0.6881	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0466	0.0421	0.0584	$\sum$ -Log(ρ[j])/n
RASE	0.1100	0.1105	0.1330	$\sqrt{\sum(y[j]-\rho[j])^2/n}$
Mean Abs Dev	0.0248	0.0250	0.0328	$\sum$ |y[j]-ρ[j]|/n
Misclassification Rate	0.0137	0.0162	0.0221	$\sum$ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Results of Logit model

- The above image shows the 8 key variables finalized along with the logistic regression model pruned with those 8 key variables

Model 0 - Logit Model (No Transform)

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.3085	0.3274	0.3333	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.5407	0.5653	0.5764	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.7173	0.7022	0.7085	∑ -Log(ρ[j])/n
RASE	0.4948	0.4913	0.4927	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4438	0.4382	0.4390	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.3279	0.3343	0.3505	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Lack Of Fit

Parameter Estimates

Effect Likelihood Ratio Tests

Confusion Matrix

Training

Actual Rank with 3P	Predicted Count 1	2	3
1	397	225	4
2	146	772	76
3	4	213	200

Actual Rank with 3P	Predicted Rate 1	2	3
1	0.634	0.359	0.006
2	0.147	0.777	0.076
3	0.010	0.511	0.480

Validation

Actual Rank with 3P	Predicted Count 1	2	3
1	146	77	1
2	49	236	33
3	0	67	70

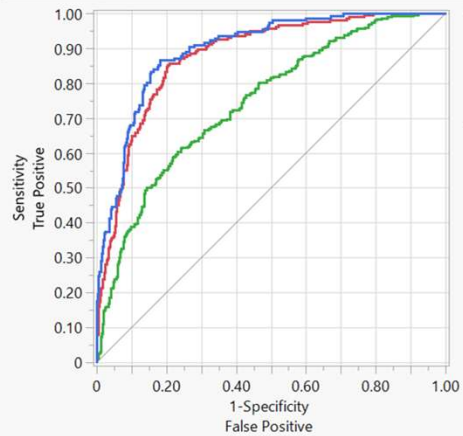
Actual Rank with 3P	Predicted Rate 1	2	3
1	0.652	0.344	0.004
2	0.154	0.742	0.104
3	0.000	0.489	0.511

Test

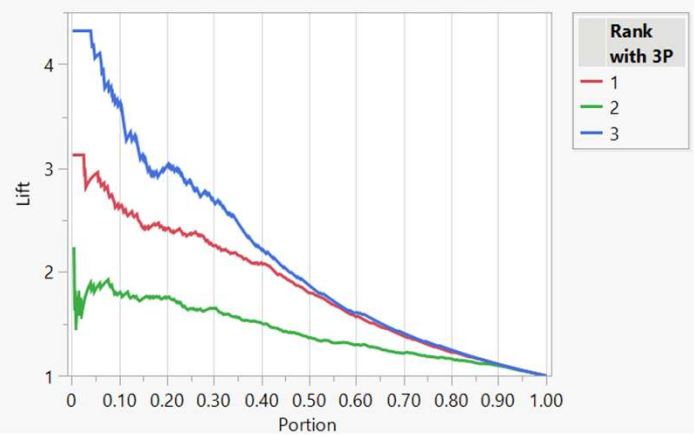
Actual Rank with 3P	Predicted Count 1	2	3
1	142	74	1
2	50	224	31
3	1	81	75

Actual Rank with 3P	Predicted Rate 1	2	3
1	0.654	0.341	0.005
2	0.164	0.734	0.102
3	0.006	0.516	0.478

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This logit model was run with the target variable having 3 categories. It led to a misclassification rate of 0.3505

Model 1 - Logit (Transform, Interaction Term)

Nominal Logistic Fit for Pro Player?

Iterations

Whole Model Test

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.5959	0.4911	0.6209	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6251	0.5106	0.6603	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0495	0.0397	0.0645	∑ -Log(ρ[j])/n
RASE	0.1152	0.1060	0.1402	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0277	0.0258	0.0370	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0165	0.0150	0.0271	∑ (ρ[j]≠ρMax)/n
N	1995	665	663	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	1980	98.839678	197.6794
Saturated	1994	0.000000	Prob>ChiSq
Fitted	14	98.839678	1.0000

Parameter Estimates

Effect Likelihood Ratio Tests

Confusion Matrix

Training

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	27	26
N	7	1935

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.509	0.491
N	0.004	0.996

Validation

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	3	7
N	3	652

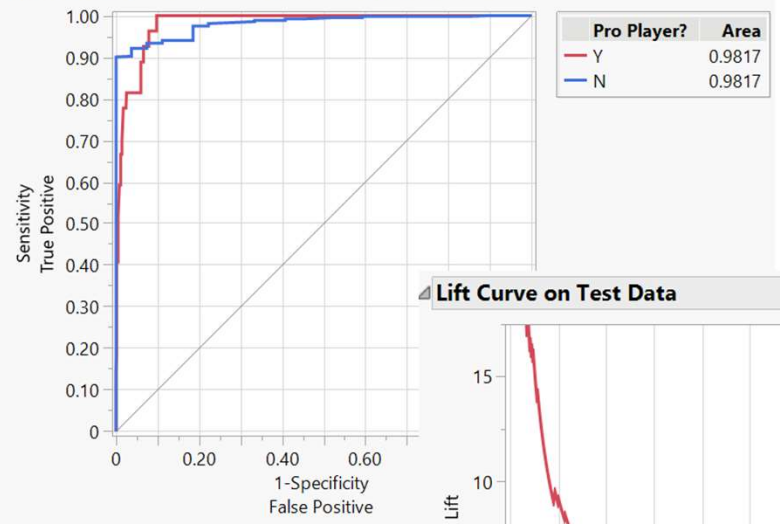
Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.300	0.700
N	0.005	0.995

Test

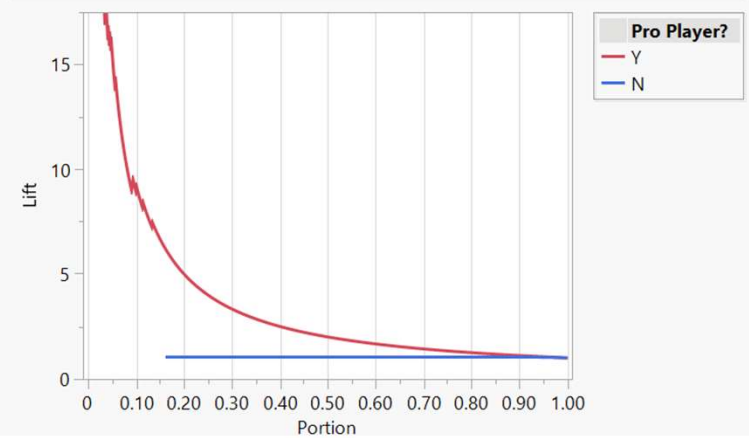
Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	12	15
N	3	633

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.444	0.556
N	0.005	0.995

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is an iteration of a logit model with transformed predictor variables to account for outliers and an interaction term
- From Model 1 onwards, all models were run with the target variable having 2 categories (Ref Slide 13)

Model 2 - Logit (No Transform, Interaction Term)

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.6198	0.4330	0.6513	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6479	0.4518	0.6886	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0459	0.0435	0.0583	∑ -Log(ρ[j])/n
RASE	0.1084	0.1127	0.1331	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0243	0.0249	0.0324	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0123	0.0162	0.0250	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Lack Of Fit

Parameter Estimates

Effect Likelihood Ratio Tests

Confusion Matrix

Training

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	32	21
N	4	1980

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.604	0.396
N	0.002	0.998

Validation

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	3	7
N	4	665

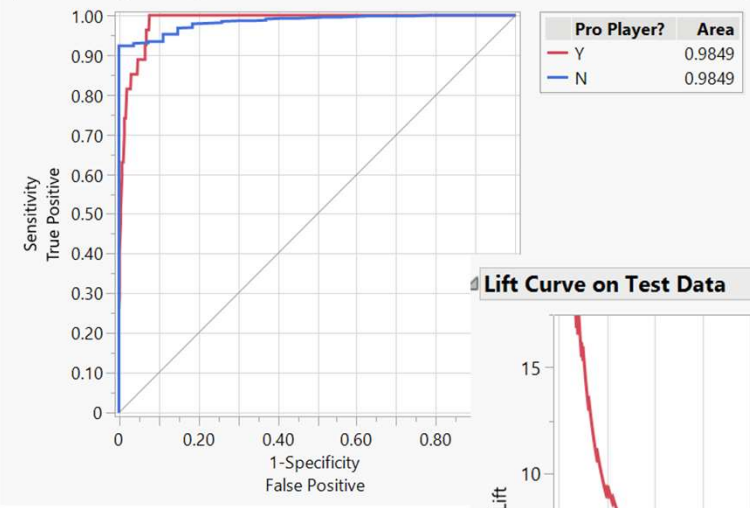
Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.300	0.700
N	0.006	0.994

Test

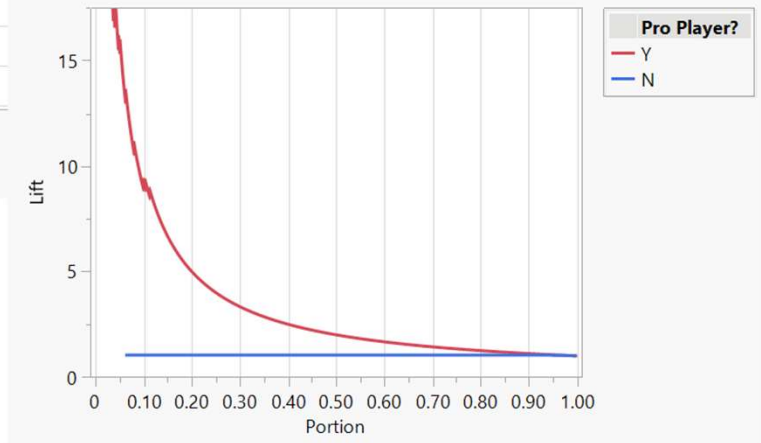
Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	15	12
N	5	647

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.556	0.444
N	0.008	0.992

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is an iteration of a logit model with no transformations and an interaction term. The misclassification rate is at 0.0250

Model 3 - Logit (No Transform,No Interaction Term)

Nominal Logistic Fit for Pro Player?

Whole Model Test

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.6213	0.4378	0.6535	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6494	0.4568	0.6906	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0457	0.0431	0.0579	∑ -Log(ρ[j])/n
RASE	0.1073	0.1123	0.1332	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0240	0.0247	0.0323	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0123	0.0162	0.0250	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	2021	93.048196	186.0964
Saturated	2036	0.000000	Prob>ChiSq
Fitted	15	93.048196	1.0000

Confusion Matrix

Training

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	31	22
N	3	1981

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.585	0.415
N	0.002	0.998

Validation

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	3	7
N	4	665

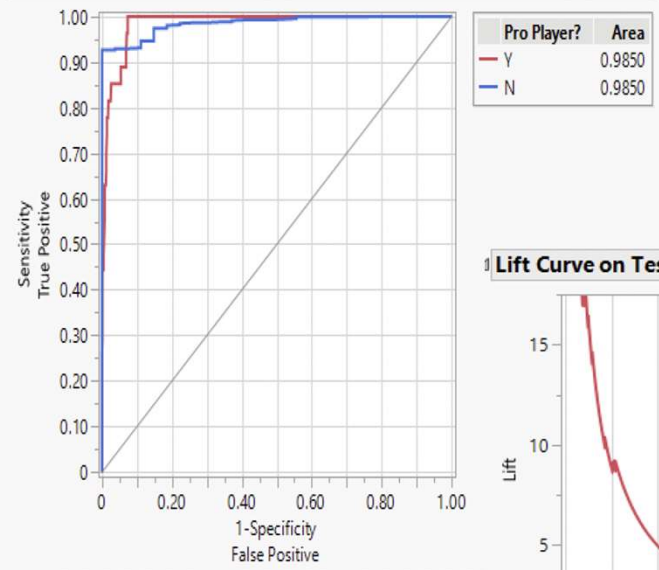
Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.300	0.700
N	0.006	0.994

Test

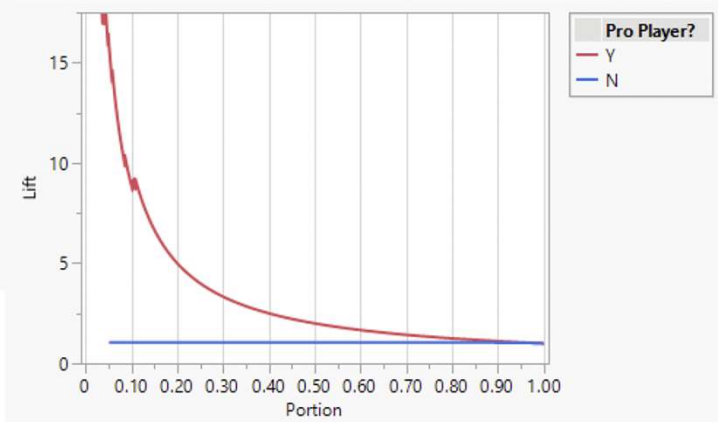
Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	15	12
N	5	647

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.556	0.444
N	0.008	0.992

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is an iteration of a logit model with no transformation and no interaction term. The misclassification rate is at 0.0250

Model 4 - Logit (No Transform,Interaction term)

Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.6140	0.4508	0.6508	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.6423	0.4699	0.6881	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0466	0.0421	0.0584	∑ -Log(ρ[j])/n
RASE	0.1100	0.1105	0.1330	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0248	0.0250	0.0328	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0137	0.0162	0.0221	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Lack Of Fit

Parameter Estimates

Effect Likelihood Ratio Tests

Confusion Matrix

Training

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	29	24
N	4	1980

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.547	0.453
N	0.002	0.998

Validation

Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	3	7
N	4	665

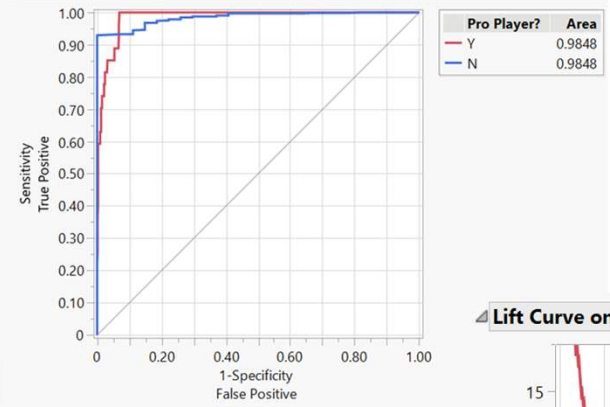
Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.300	0.700
N	0.006	0.994

Test

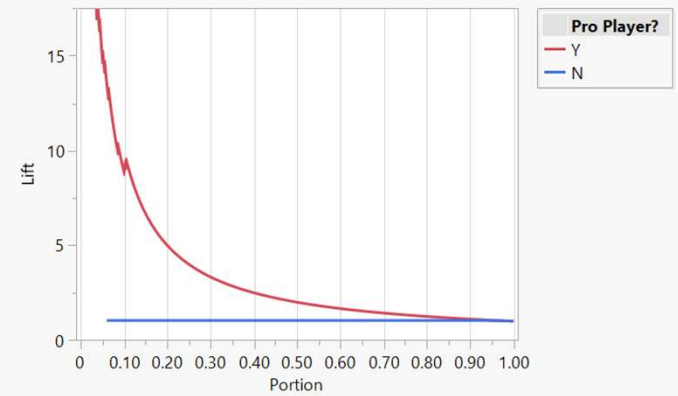
Actual Pro Player?	Predicted Count Y	Predicted Count N
Y	16	11
N	4	648

Actual Pro Player?	Predicted Rate Y	Predicted Rate N
Y	0.593	0.407
N	0.006	0.994

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is an iteration of Model 2 with only 8 predictor variables. The misclassification rate is at 0.0221

Model 5 - Neural Network TanH(3)

Neural

Validation Column: Validation

Model Launch

Model NTanH(3)

Training

Pro Player?

Measures	Value
Generalized RSquare	0.6694024
Entropy RSquare	0.642028
RASE	0.1064215
Mean Abs Dev	0.0233556
Misclassification Rate	0.012273
-LogLikelihood	87.953135
Sum Freq	2037

Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	1979	5
Y	20	33

Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.997	0.003
Y	0.377	0.623

Validation

Pro Player?

Measures	Value
Generalized RSquare	0.5086264
Entropy RSquare	0.489449
RASE	0.1059047
Mean Abs Dev	0.0226399
Misclassification Rate	0.0132548
-LogLikelihood	26.602951
Sum Freq	679

Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	666	3
Y	6	4

Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.996	0.004
Y	0.600	0.400

Test

Pro Player?

Measures	Value
Generalized RSquare	0.6335272
Entropy RSquare	0.593707
RASE	0.1421692
Mean Abs Dev	0.0343602
Misclassification Rate	0.0250368
-LogLikelihood	46.124454
Sum Freq	679

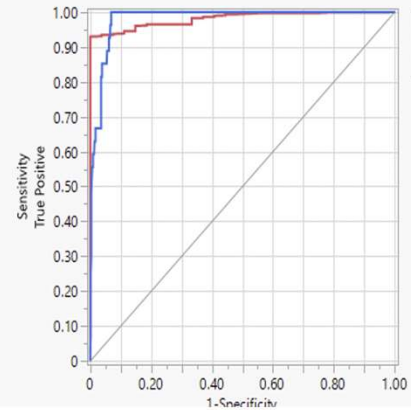
Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	648	4
Y	13	14

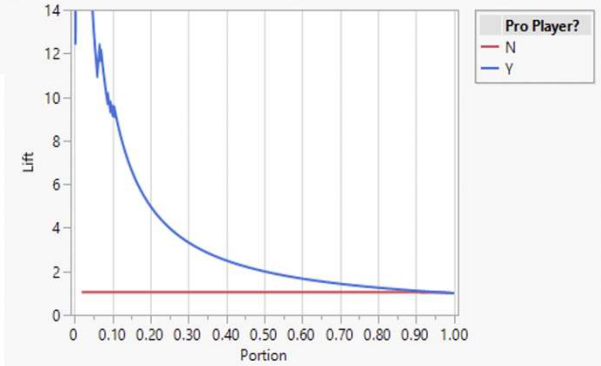
Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.994	0.006
Y	0.481	0.519

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is a neural network model with 3 TanH nodes. The misclassification rate is at 0.0250.

Model 6 - Neural Network TanH(1)L(1)G(1)

Model NTanH(1)NLinear(1)NGaussian(1)

Training

Pro Player?

Measures	Value
Generalized RSquare	0.6837078
Entropy RSquare	0.6568946
RASE	0.1061083
Mean Abs Dev	0.0228311
Misclassification Rate	0.0127639
-LogLikelihood	84.300438
Sum Freq	2037

Confusion Matrix

Actual Pro Player?	Predicted Count N	Predicted Count Y
N	1977	7
Y	19	34

Confusion Rates

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.996	0.004
Y	0.358	0.642

Validation

Pro Player?

Measures	Value
Generalized RSquare	0.4829103
Entropy RSquare	0.4637999
RASE	0.107579
Mean Abs Dev	0.0236459
Misclassification Rate	0.0162003
-LogLikelihood	27.939433
Sum Freq	679

Confusion Matrix

Actual Pro Player?	Predicted Count N	Predicted Count Y
N	665	4
Y	7	3

Confusion Rates

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.994	0.006
Y	0.700	0.300

Test

Pro Player?

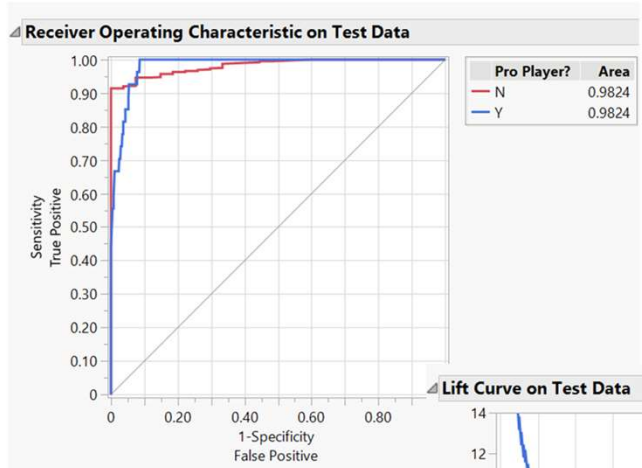
Measures	Value
Generalized RSquare	0.677575
Entropy RSquare	0.6397211
RASE	0.134102
Mean Abs Dev	0.0319869
Misclassification Rate	0.0235641
-LogLikelihood	40.900695
Sum Freq	679

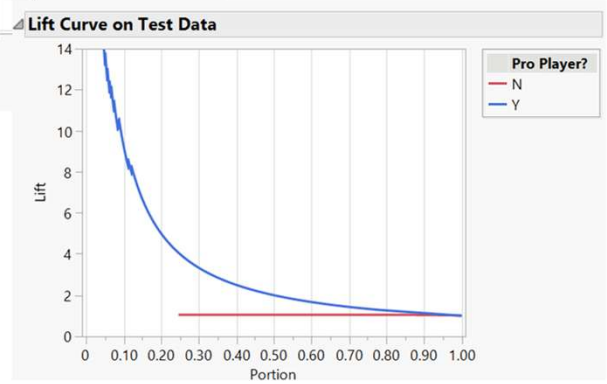
Confusion Matrix

Actual Pro Player?	Predicted Count N	Predicted Count Y
N	648	4
Y	12	15

Confusion Rates

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.994	0.006
Y	0.444	0.556





- This is a neural network with 1 TanH node, 1 Linear node, and 1 Gaussian node. The misclassification rate is 0.0236.

Model 7 - Neural NTanh(1)Linear(2)

Neural

Validation Column: Validation

Model Launch

Model NTanH(1)NLinear(2)

Training

Pro Player?

Measures	Value
Generalized RSquare	0.6444222
Entropy RSquare	0.616195
RASE	0.1098927
Mean Abs Dev	0.0252158
Misclassification Rate	0.0137457
-LogLikelihood	94.300255
Sum Freq	2037

Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	1979	5
Y	23	30

Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.997	0.003
Y	0.434	0.566

Validation

Pro Player?

Measures	Value
Generalized RSquare	0.4895982
Entropy RSquare	0.4704607
RASE	0.1071474
Mean Abs Dev	0.0247337
Misclassification Rate	0.0147275
-LogLikelihood	27.592365
Sum Freq	679

Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	666	3
Y	7	3

Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.996	0.004
Y	0.700	0.300

Test

Pro Player?

Measures	Value
Generalized RSquare	0.6980475
Entropy RSquare	0.6613509
RASE	0.1302366
Mean Abs Dev	0.0328992
Misclassification Rate	0.0235641
-LogLikelihood	38.445173
Sum Freq	679

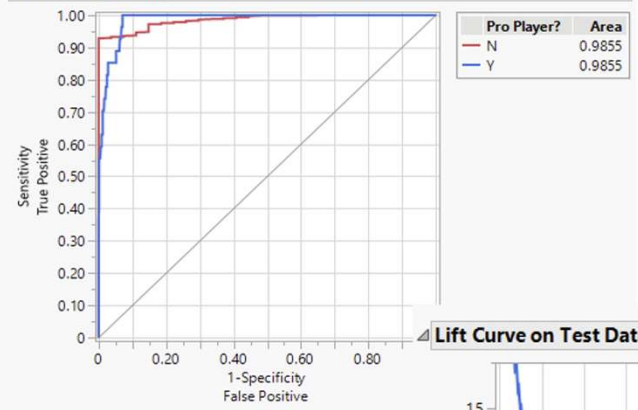
Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	648	4
Y	12	15

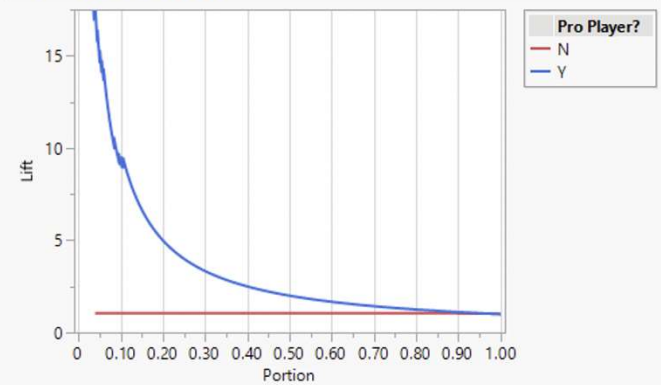
Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.994	0.006
Y	0.444	0.556

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is a neural network with 1 TanH node, and 2 Linear nodes. The misclassification rate is 0.0236.

Model 8 - Neural Network G(3)

Model NGaussian(3)

Training — **Pro Player?**

Measures	Value
Generalized RSquare	0.640916
Entropy RSquare	0.6125819
RASE	0.1102947
Mean Abs Dev	0.0243163
Misclassification Rate	0.0142366
-LogLikelihood	95.187976
Sum Freq	2037

Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	1977	7
Y	22	31

Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.996	0.004
Y	0.415	0.585

Validation — **Pro Player?**

Measures	Value
Generalized RSquare	0.5351508
Entropy RSquare	0.5160106
RASE	0.1046406
Mean Abs Dev	0.0220384
Misclassification Rate	0.0162003
-LogLikelihood	25.218927
Sum Freq	679

Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	666	3
Y	8	2

Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.996	0.004
Y	0.800	0.200

Test — **Pro Player?**

Measures	Value
Generalized RSquare	0.6535854
Entropy RSquare	0.6145728
RASE	0.1329321
Mean Abs Dev	0.0309917
Misclassification Rate	0.0206186
-LogLikelihood	43.755659
Sum Freq	679

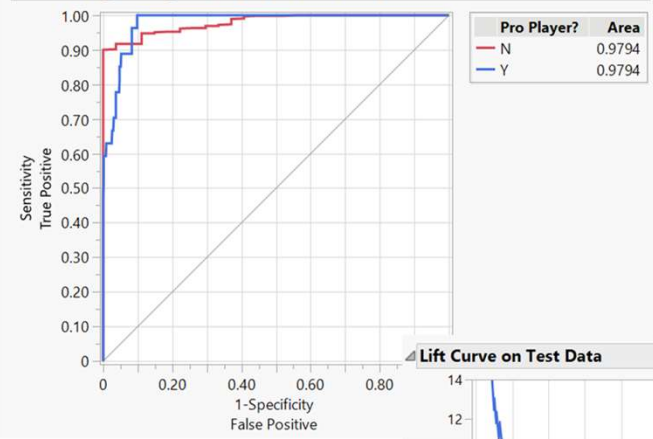
Confusion Matrix

Actual Pro Player?	Predicted Count N	Y
N	651	1
Y	13	14

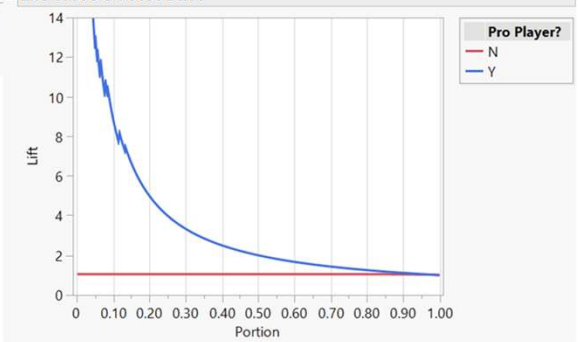
Confusion Rates

Actual Pro Player?	Predicted Rate N	Y
N	0.998	0.002
Y	0.481	0.519

Receiver Operating Characteristic on Test Data



Lift Curve on Test Data



- This is a neural network with 3 Gaussian nodes and 1 hidden layer. The misclassification rate is at 0.0206

Model 9 - Neural Network G(3)G(3)

Model NGaussian(3)NGaussian2(3)

Training

Pro Player?

Measures	Value
Generalized RSquare	0.6290135
Entropy RSquare	0.6003402
RASE	0.1138302
Mean Abs Dev	0.0275004
Misclassification Rate	0.0166912
-LogLikelihood	98.195748
Sum Freq	2037

Confusion Matrix

Actual	Predicted Count	
Pro Player?	N	Y
N	1979	5
Y	29	24

Confusion Rates

Actual	Predicted Rate	
Pro Player?	N	Y
N	0.997	0.003
Y	0.547	0.453

Validation

Pro Player?

Measures	Value
Generalized RSquare	0.4899901
Entropy RSquare	0.4708512
RASE	0.1071454
Mean Abs Dev	0.0265723
Misclassification Rate	0.0191458
-LogLikelihood	27.572019
Sum Freq	679

Confusion Matrix

Actual	Predicted Count	
Pro Player?	N	Y
N	663	6
Y	7	3

Confusion Rates

Actual	Predicted Rate	
Pro Player?	N	Y
N	0.991	0.009
Y	0.700	0.300

Test

Pro Player?

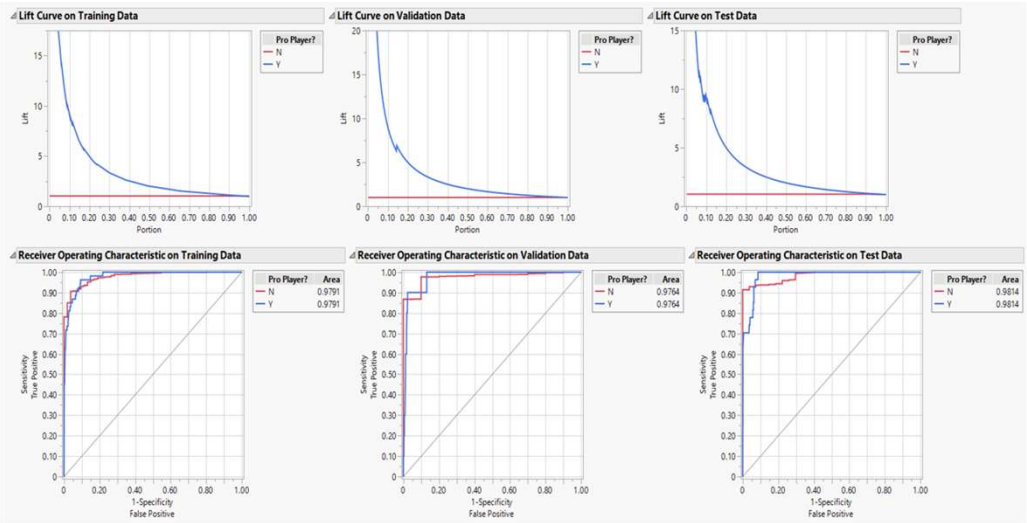
Measures	Value
Generalized RSquare	0.6681532
Entropy RSquare	0.629819
RASE	0.1307057
Mean Abs Dev	0.0351501
Misclassification Rate	0.017673
-LogLikelihood	42.02483
Sum Freq	679

Confusion Matrix

Actual	Predicted Count	
Pro Player?	N	Y
N	650	2
Y	10	17

Confusion Rates

Actual	Predicted Rate	
Pro Player?	N	Y
N	0.997	0.003
Y	0.370	0.630



- This a neural network with 2 hidden layers of 3 Gaussian nodes. The misclassification rate is 0.018.

Model 10 - Naive Bayes

Naive Bayes

Fit Details

Pro Player?

Training			Validation			Test		
Count	Misclassification Rate	Misclassifications	Count	Misclassification Rate	Misclassifications	Count	Misclassification Rate	Misclassifications
2037	0.06186	126	679	0.07216	49	679	0.06480	44

Confusion Matrix

Training

Actual Pro Player?	Predicted Count N	Predicted Count Y
N	1868	116
Y	10	43

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.942	0.058
Y	0.189	0.811

Validation

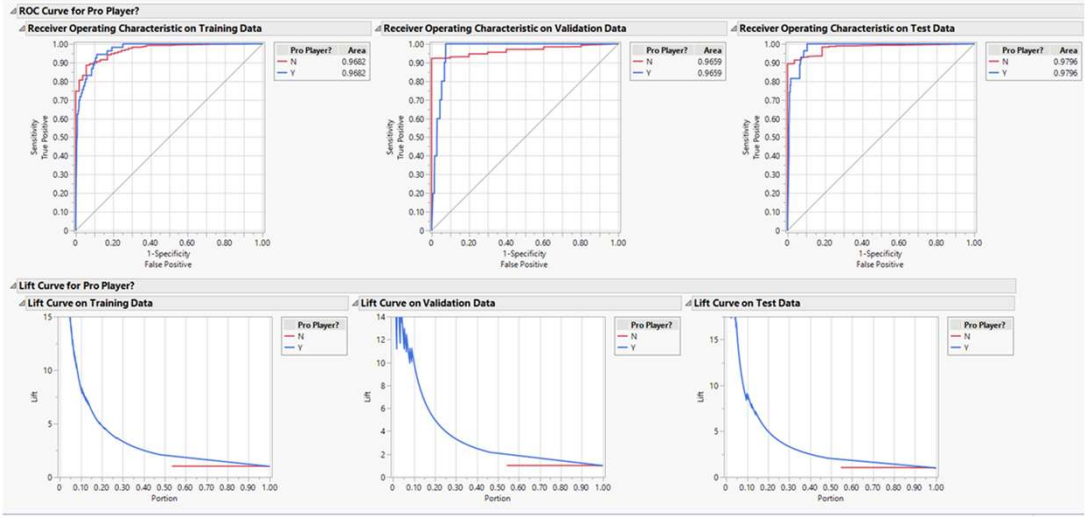
Actual Pro Player?	Predicted Count N	Predicted Count Y
N	621	48
Y	1	9

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.928	0.072
Y	0.100	0.900

Test

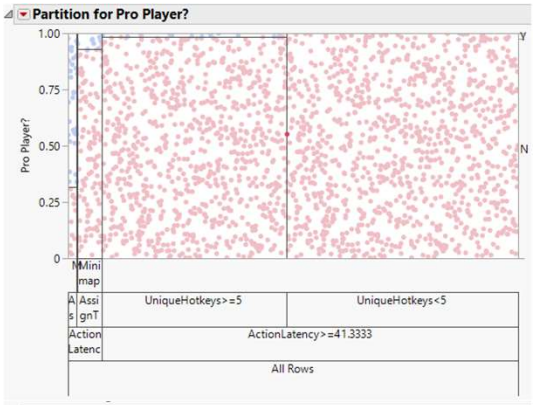
Actual Pro Player?	Predicted Count N	Predicted Count Y
N	613	39
Y	5	22

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.940	0.060
Y	0.185	0.815



- This is a Naive Bayes model. The misclassification rate is 0.065.

Model 11 - Decision Tree (UnPruned)



Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.4779	0.2259	0.3871	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.5080	0.2395	0.4272	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0630	0.0594	0.1025	∑ -Log(ρ[j])/n
RASE	0.1205	0.1189	0.1670	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0294	0.0286	0.0450	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0167	0.0177	0.0339	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Confusion Matrix

Training

Actual Pro Player?	Predicted Count N	Predicted Count Y
N	1972	12
Y	22	31

Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.994	0.006
Y	0.415	0.585

Validation

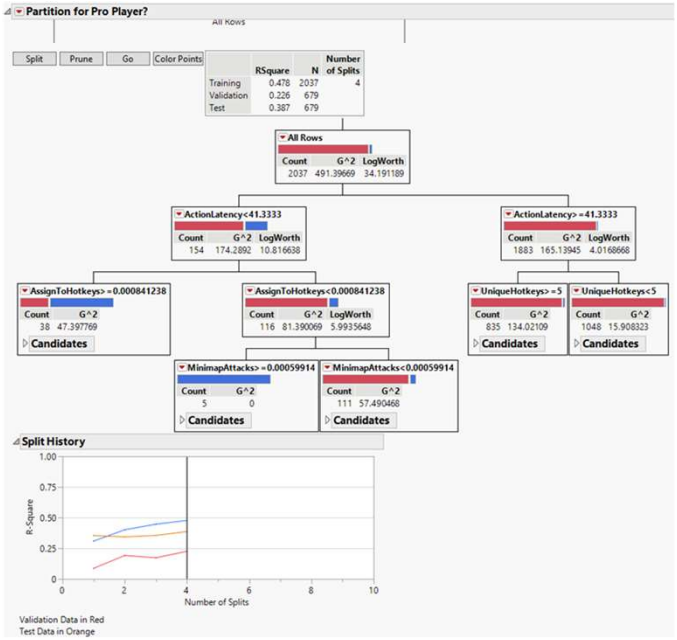
Actual Pro Player?	Predicted Count N	Predicted Count Y
N	664	5
Y	7	3

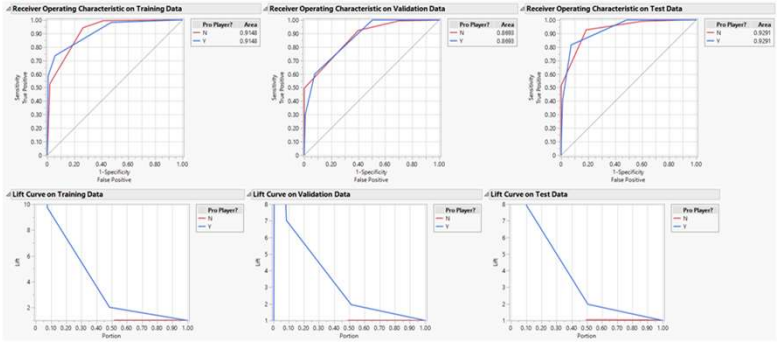
Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.993	0.007
Y	0.700	0.300

Test

Actual Pro Player?	Predicted Count N	Predicted Count Y
N	645	7
Y	16	11

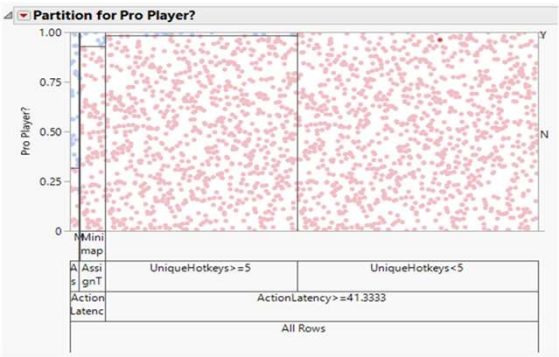
Actual Pro Player?	Predicted Rate N	Predicted Rate Y
N	0.989	0.011
Y	0.593	0.407

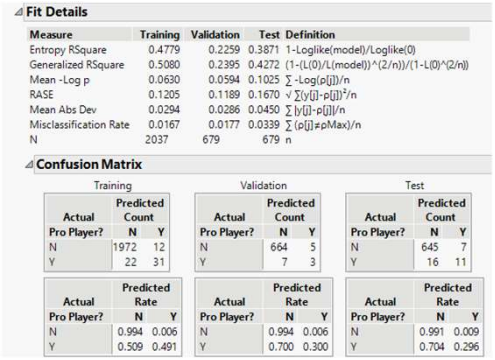




- This is an unpruned decision tree model. The misclassification rate 0.039.

Model 12 - Decision Tree (Pruned)





Fit Details

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.4779	0.2259	0.3871	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.5080	0.2395	0.4272	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0630	0.0594	0.1025	∑ -Log(ρ[j])/n
RASE	0.1205	0.1189	0.1670	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0294	0.0286	0.0450	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0167	0.0177	0.0339	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Confusion Matrix

Training

Actual Pro Player?	Predicted Count N	Y
N	1972	12
Y	22	31

Actual Pro Player?	Predicted Rate N	Y
N	0.994	0.006
Y	0.509	0.491

Validation

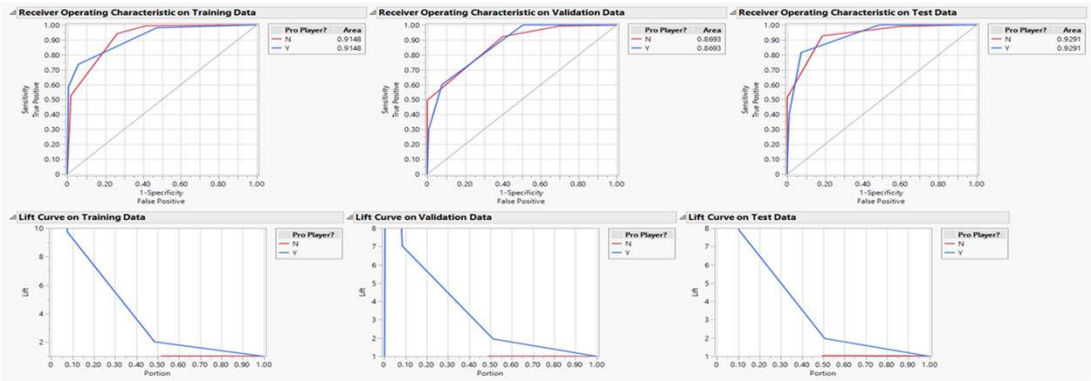
Actual Pro Player?	Predicted Count N	Y
N	664	5
Y	7	3

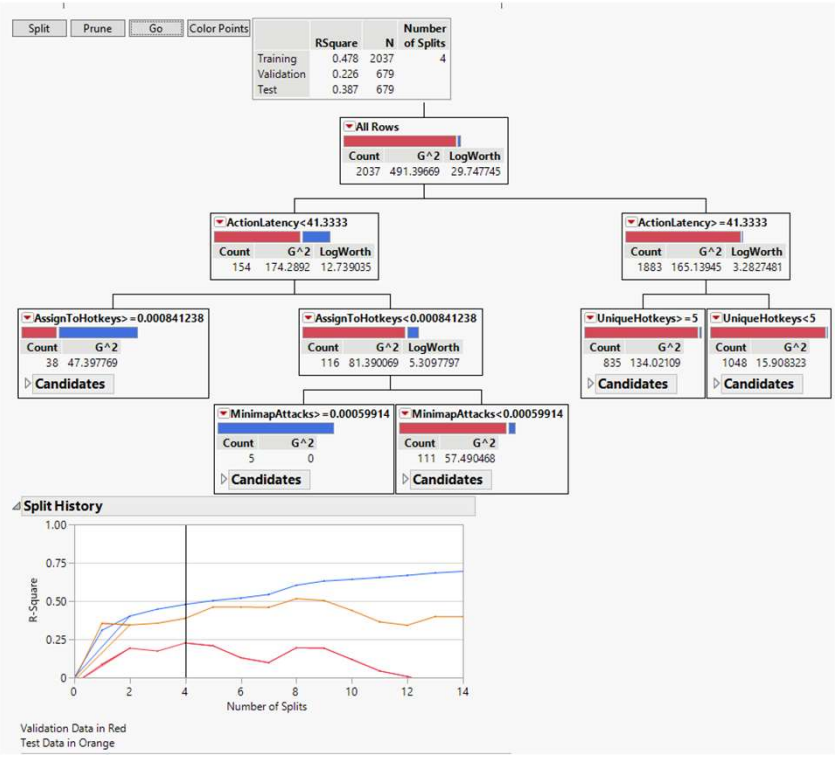
Actual Pro Player?	Predicted Rate N	Y
N	0.994	0.006
Y	0.700	0.300

Test

Actual Pro Player?	Predicted Count N	Y
N	645	7
Y	16	11

Actual Pro Player?	Predicted Rate N	Y
N	0.991	0.009
Y	0.704	0.296





- This is a decision tree model with 4 splits. The misclassification rate is at 0.0339.

Model 13 - Boosted Tree

Boosted Tree for Pro Player?

Specifications

Target	Pro Player?	Number of training rows:	2037
Validation Column:	Validation	Number of validation rows:	679
Number of Layers:	45	Number of test rows:	679
Splits per Tree:	13		
Learning Rate:	0.122		
Overfit Penalty:	0.0001		

Overall Statistics

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.8934	0.4054	0.4002	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.9045	0.4240	0.4407	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0129	0.0456	0.1003	∑ -Log(ρ[j])/n
RASE	0.0483	0.1080	0.1583	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0108	0.0210	0.0356	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0015	0.0147	0.0295	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Confusion Matrix

Training

Actual Pro Player?	Predicted Count N	Y
N	1984	0
Y	3	50

Actual Pro Player?	Predicted Rate N	Y
N	1.000	0.000
Y	0.057	0.943

Validation

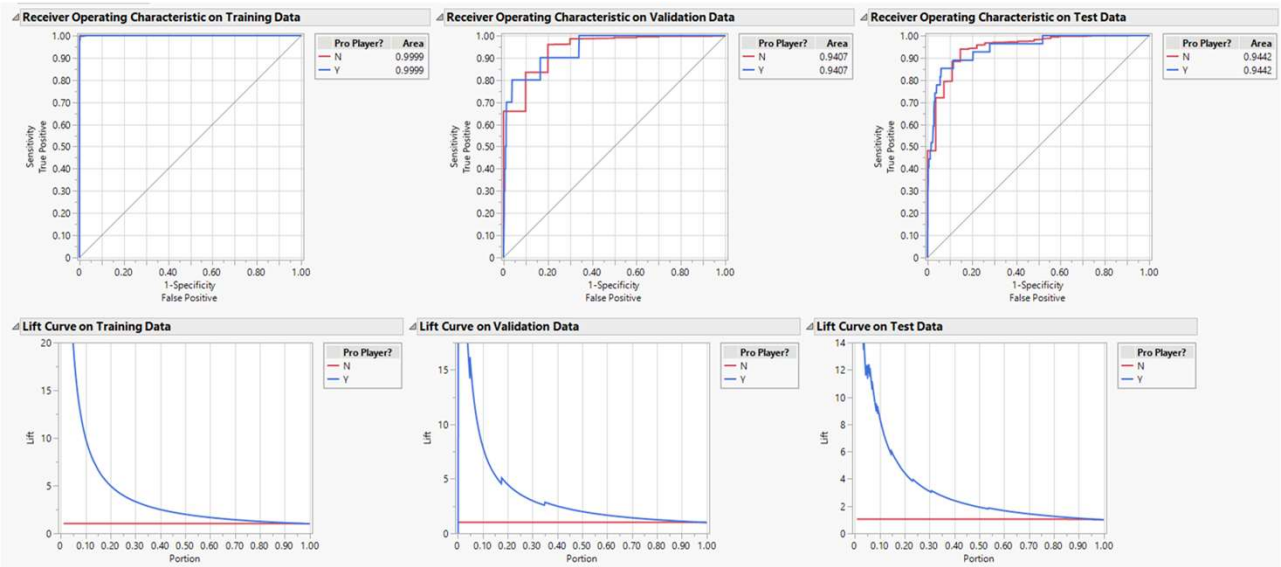
Actual Pro Player?	Predicted Count N	Y
N	666	3
Y	7	3

Actual Pro Player?	Predicted Rate N	Y
N	0.996	0.004
Y	0.700	0.300

Test

Actual Pro Player?	Predicted Count N	Y
N	648	4
Y	16	11

Actual Pro Player?	Predicted Rate N	Y
N	0.994	0.006
Y	0.593	0.407



- This is a boosted tree model. The misclassification rate is 0.0295.

Model 14 - Bootstrap Tree

Bootstrap Forest for Pro Player?

Specifications

Target	Pro Player?	Training Rows:	2037
Validation Column:	Validation	Validation Rows:	679
		Test Rows:	679
Number of Trees in the Forest:	13	Number of Terms:	8
Number of Terms Sampled per Split:	6	Bootstrap Samples:	2037
		Minimum Splits per Tree:	10
		Minimum Size Split:	2

Overall Statistics

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.8258	0.5397	0.5625	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.8427	0.5587	0.6032	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0210	0.0353	0.0732	∑ -Log(ρ[j])/n
RASE	0.0751	0.1027	0.1393	√∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0165	0.0225	0.0351	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0079	0.0162	0.0236	∑ (ρ[j]≠ρMax)/n
N	2037	679	679	n

Confusion Matrix

Training

Actual Pro Player?	Predicted Count N	Y
N	1984	0
Y	16	37

Actual Pro Player?	Predicted Rate N	Y
N	1.000	0.000
Y	0.302	0.698

Validation

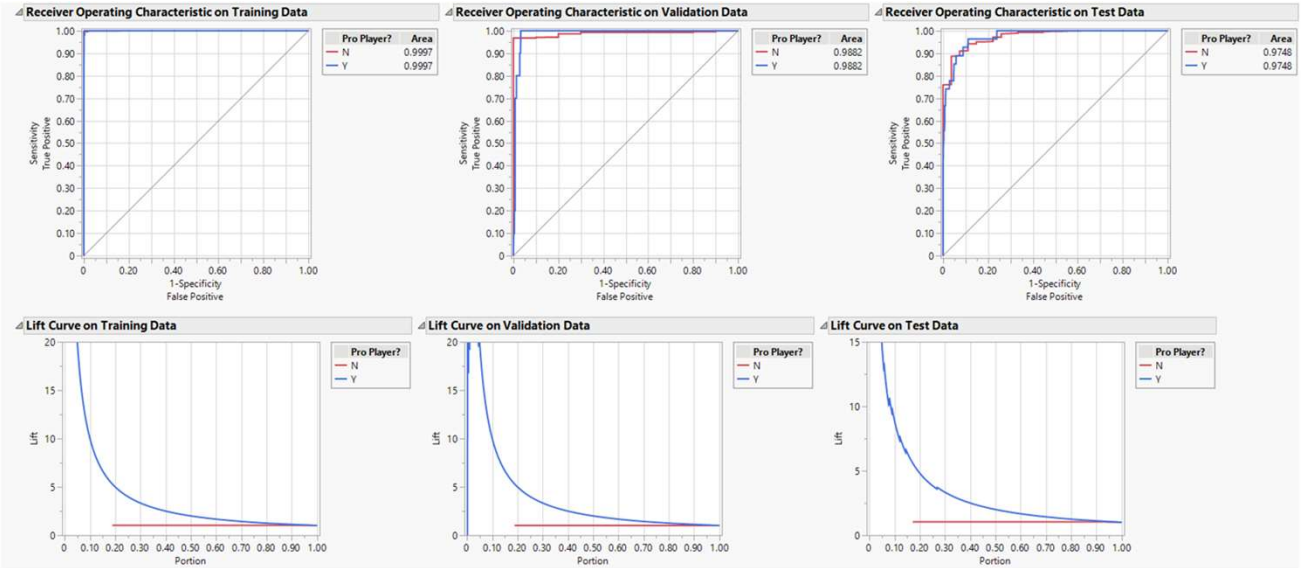
Actual Pro Player?	Predicted Count N	Y
N	667	2
Y	9	1

Actual Pro Player?	Predicted Rate N	Y
N	0.997	0.003
Y	0.900	0.100

Test

Actual Pro Player?	Predicted Count N	Y
N	652	0
Y	16	11

Actual Pro Player?	Predicted Rate N	Y
N	1.000	0.000
Y	0.593	0.407



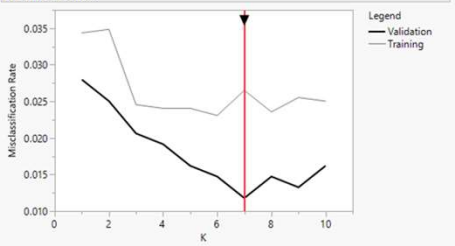
- This is a bootstrap tree model. The misclassification rate is 0.0236.

Model 15 - K-NN

K Nearest Neighbors

Pro Player?

Model Selection



Training

K	Count	RSquare	Misclassification Rate	Misclassifications
1	2037	-0.9701	0.03436	70
2	2037	-0.3599	0.03486	71
3	2037	-0.0929	0.02455	50
4	2037	0.05019	0.02405	49
5	2037	0.13492	0.02405	49
6	2037	0.18941	0.02307	47 *
7	2037	0.23155	0.02651	54
8	2037	0.27424	0.02356	48
9	2037	0.29056	0.02553	52
10	2037	0.30401	0.02504	51

Validation

K	Count	RSquare	Misclassification Rate	Misclassifications
1	679	-1.9625	0.02798	19
2	679	-0.9049	0.02504	17
3	679	-0.5153	0.02062	14
4	679	-0.3056	0.01915	13
5	679	-0.1478	0.01620	11
6	679	-0.0612	0.01473	10
7	679	0.05697	0.01178	8 *
8	679	0.08292	0.01473	10
9	679	0.13531	0.01325	9
10	679	0.16284	0.01620	11

Test

K	Count	RSquare	Misclassification Rate	Misclassifications
1	679	-0.4307	0.03535	24
2	679	0.01333	0.03387	23
3	679	0.18562	0.03093	21
4	679	0.30191	0.02798	19
5	679	0.33762	0.02946	20
6	679	0.36436	0.02946	20
7	679	0.37676	0.02946	20
8	679	0.41788	0.02946	20
9	679	0.44559	0.02651	18 *
10	679	0.45733	0.02798	19

Confusion Matrix for Best K=7

Training

Actual Pro Player?	Predicted Count N	Y
N	1969	15
Y	39	14

Actual Pro Player?	Predicted Rate N	Y
N	0.992	0.008
Y	0.736	0.264

Validation

Actual Pro Player?	Predicted Count N	Y
N	668	1
Y	7	3

Actual Pro Player?	Predicted Rate N	Y
N	0.999	0.001
Y	0.700	0.300

Test

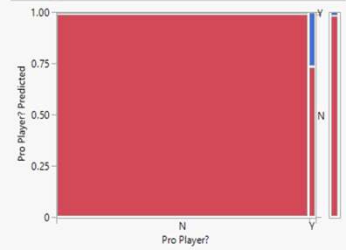
Actual Pro Player?	Predicted Count N	Y
N	651	1
Y	19	8

Actual Pro Player?	Predicted Rate N	Y
N	0.998	0.002
Y	0.704	0.296

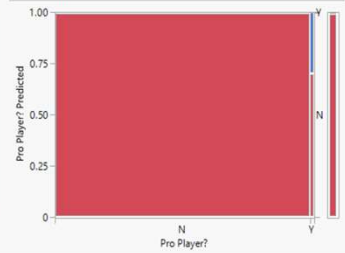
Mosaic Plot

Mosaic Plot for K=7

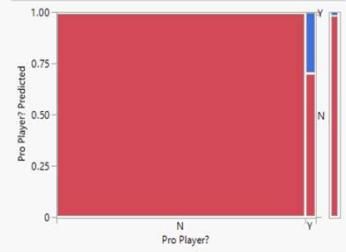
Training



Validation



Test



- This is a K-NN model with a sensitivity rate of 0.026 on test partition.

Model 16 - Discriminant Analysis

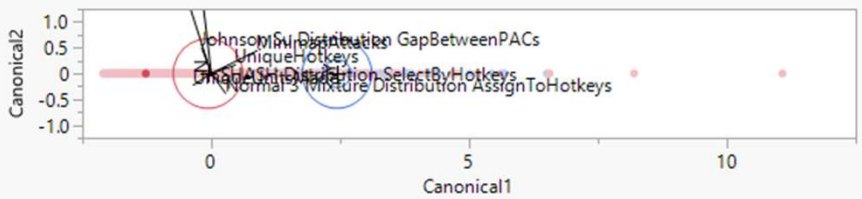
Starcraft_2_SecondIterationjmp - Discriminant by Pro Player? - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

Discriminant Analysis

Discriminant Method: Linear
Classification: Pro Player?
Validation: Validation

Canonical Plot



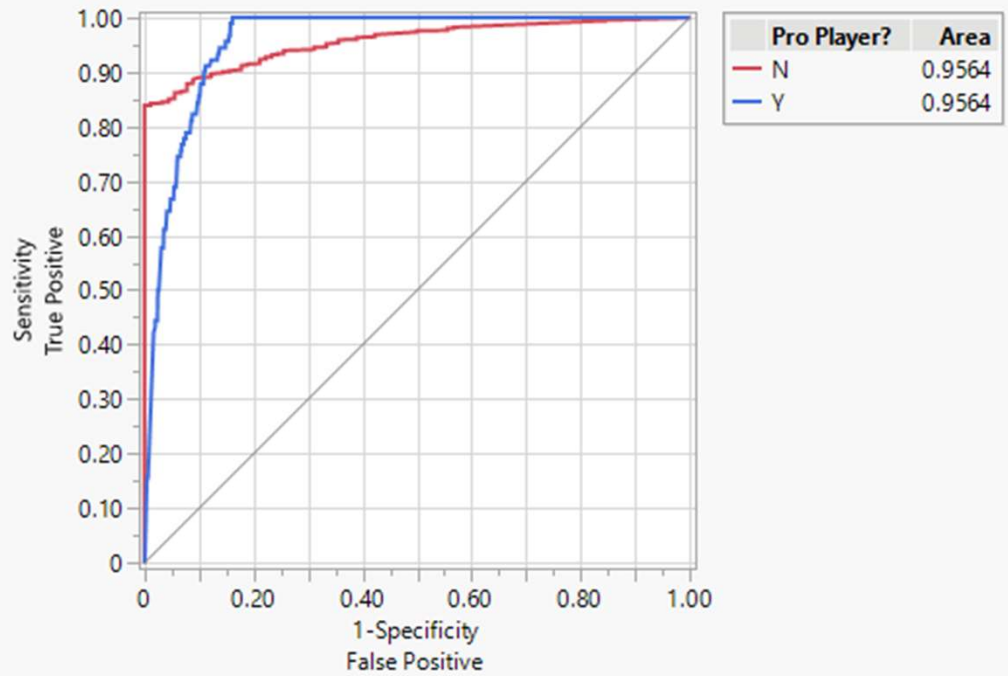
Discriminant Scores

Score Summaries

Source	Count	Number Misclassified	Percent Misclassified	Entropy RSquare	-2LogLikelihood
Training	2037	201	9.86745	-1.3971	1177.94
Validation	679	69	10.1620	-3.6297	
Test	679	60	8.83652	-0.6860	

Groups

Receiver Operating Characteristic on Training Data



- This is a Discriminant Analysis with a misclassification rate of 8.84% on test partition.

Model 17: Ensemble 1 - NN1,Bootstrap

Where(Format(:Validation) == "Test")

Model Comparison

Predictors

Measures of Fit for Pro Player?

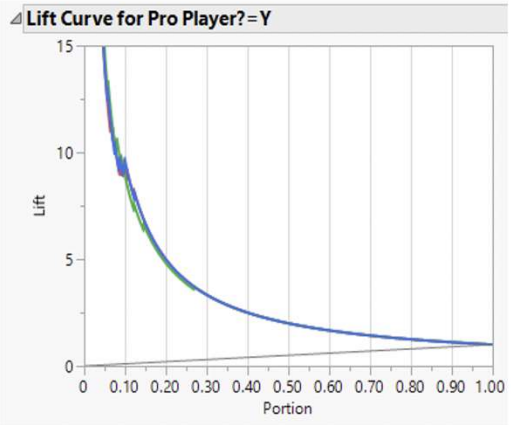
Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N
Neural Model NGaussian(3)NGaussian2(3) - Save Formulas		0.6298	0.6682	0.0619	0.1307	0.0352	0.0177	679
Bootstrap Forest		0.5625	0.6032	0.0732	0.1393	0.0351	0.0236	679
Model Averaged		0.6355	0.6736	0.0609	0.1320	0.0351	0.0206	679

Predictor Model Averaged

Actual Pro Player?	Predicted Count N	Y
N	652	0
Y	14	13

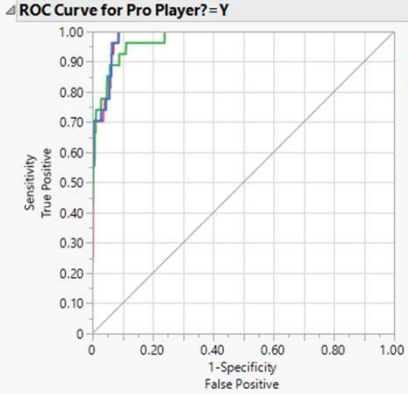
Actual Pro Player?	Predicted Rate N	Y
N	1.000	0.000
Y	0.519	0.481

Lift Curve for Pro Player?=Y



Predictor	Creator
Probability(Pro Player?=Y)	Neural Model NGaussian(3)NGaussian2(3) - Save Formulas
Prob(Pro Player?==Y)	Bootstrap Forest
Pro Player? Y Avg Predictor By Validation	Model Averaged

ROC Curve for Pro Player?=Y



Predictor	Creator	AUC
Probability(Pro Player?=Y)	Neural Model NGaussian(3)NGaussian2(3) - Save Formulas	0.9814
Prob(Pro Player?==Y)	Bootstrap Forest	0.9748
Pro Player? Y Avg Predictor By Validation	Model Averaged	0.9822

- This is an Ensemble Model combining Model 9 and Model 14. This model has a misclassification rate at 0.0206.

Model 18: Ensemble 2 - NN1, NN2, Bootstrap, Logit

Model Comparison Validation=Test

Predictors

Measures of Fit for Pro Player?

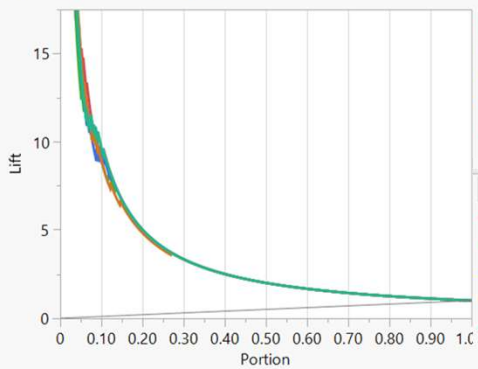
Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate
Fit Nominal Logistic		0.6508	0.6881	0.0584	0.1330	0.0328	0.0221
Neural Model NGaussian(3) - Save Formulas		0.6146	0.6536	0.0644	0.1329	0.0310	0.0206
Neural Model NGaussian(3)NGaussian2(3) - Save Formulas		0.6298	0.6682	0.0619	0.1307	0.0352	0.0177
Bootstrap Forest		0.5625	0.6032	0.0732	0.1393	0.0351	0.0236
Model Averaged		0.6492	0.6865	0.0587	0.1305	0.0335	0.0177

Predictor Model Averaged

Actual Pro Player?	Predicted Count N	Y
N	652	0
Y	12	15

Actual Pro Player?	Predicted Rate N	Y
N	1.000	0.000
Y	0.444	0.556

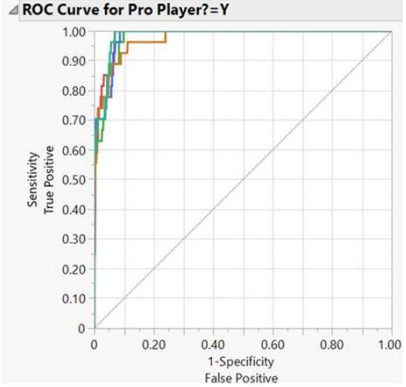
Lift Curve for Pro Player?=Y



Predictor	Creator
Prob[Y]	Fit Nominal Logistic
Probability(Pro Player?=Y)	Neural Model NGaussian(3) - Save Formulas
Probability(Pro Player?=Y) 2	Neural Model NGaussian(3)NGaussian2(3) - Save Formulas
Prob(Pro Player?==Y)	Bootstrap Forest
Pro Player? Y Avg Predictor By Validation	Model Averaged

Predictor	Creator	AUC
Prob[Y]	Fit Nominal Logistic	0.9848
Probability(Pro Player?=Y)	Neural Model NGaussian(3) - Save Formulas	0.9794
Probability(Pro Player?=Y) 2	Neural Model NGaussian(3)NGaussian2(3) - Save Formulas	0.9814
Prob(Pro Player?==Y)	Bootstrap Forest	0.9748
Pro Player? Y Avg Predictor By Validation	Model Averaged	0.9848

ROC Curve for Pro Player?=Y



- This is the second Ensemble model which combines Model 4, Model 8, Model 9 and Model 14. The misclassification rate is at 0.0177.