



NBA 정규 리그 예측모델

-동/서부 conference 진출팀 예측-
(NBA 2010-2019 data)

김소정



Index

01 DATA 설명

02 예측 모델

03 모델 간 비교

04 결론 및 시사점

DATA 설명

리그설명

변수설명



WESTERN CONFERENCE

EASTERN CONFERENCE

진출

구단 예측



NBA champion

VS



진출

구단 예측



Teams General Traditional ▾

2010년~2019년 경기 기록

SEASON
2019-20SEASON TYPE
Regular SeasonPER MODE
Per GameSEASON SEGMENT
All Games[Advanced Filters](#)[RECENT FILTERS](#)[GLOSSARY](#)[SHARE](#)

TEAM	GP	W	L	WIN%	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB	DREB	REB	AST	TOV	STL	BLK	BLKA	PF	PFD	+/-
1 Milwaukee Bucks	65	53	12	.815	48.2	118.6	43.5	91.2	47.7	13.7	38.6	35.6	17.8	24.0	74.2	9.5	42.2	51.7	25.9	14.9	7.4	6.0	4.6	19.2	21.3	11.3
2 Los Angeles Lakers	63	49	14	.778	48.2	114.3	42.9	88.6	48.5	11.2	31.4	35.5	17.3	23.7	73.0	10.6	35.5	46.1	25.9	15.1	8.6	6.8	3.7	20.6	21.4	7.4
3 Toronto Raptors	64	46	18	.719	48.3	113.0	40.6	88.5	45.8	13.8	37.0	37.1	18.1	22.6	80.0	9.7	35.5	45.2	25.4	14.4	8.8	4.9	5.3	21.5	20.0	6.5
4 LA Clippers	64	44	20	.688	48.2	116.2	41.6	89.7	46.4	12.2	33.2	36.6	20.8	26.2	79.2	11.0	37.0	48.0	23.8	14.8	7.1	5.0	4.9	22.0	22.8	6.5
5 Boston Celtics	64	43	21	.672	48.4	113.0	41.2	89.6	45.9	12.4	34.2	36.3	18.3	22.8	80.1	10.7	35.3	46.0	22.8	13.6	8.3	5.6	5.6	21.4	20.6	6.2
6 Denver Nuggets	65	43	22	.662	48.5	110.4	41.8	88.9	47.1	10.9	30.4	35.8	15.9	20.5	77.5	10.8	33.5	44.3	26.5	13.7	8.1	4.6	4.5	20.0	20.0	3.0
7 Utah Jazz	64	41	23	.641	48.1	111.0	40.1	84.6	47.5	13.2	34.4	38.3	17.6	22.8	77.2	8.8	36.3	45.1	22.2	14.9	5.9	4.0	4.6	20.0	20.6	3.2
8 Miami Heat	65	41	24	.631	48.7	112.2	39.6	84.4	47.0	13.4	34.8	38.3	19.5	25.1	77.8	8.5	36.0	44.5	26.0	14.9	7.4	4.5	4.2	20.4	21.5	3.2
9 Houston Rockets	64	40	24	.625	48.3	110.8	40.3	85.1	47.3	10.4	29.3	35.5	19.8	24.8	79.7	8.1	34.6	42.7	21.9	13.5	7.6	5.0	4.1	18.8	22.8	2.5
9 Oklahoma City Thunder	64	40	24	.625	48.3	110.8	40.3	85.1	47.3	10.4	29.3	35.5	19.8	24.8	79.7	8.1	34.6	42.7	21.9	13.5	7.6	5.0	4.1	18.8	22.8	2.5
11 Indiana Pacers	65	39	26	.600	48.3	109.3	42.2	88.4	47.7	10.0	27.5	36.3	15.1	19.1	78.7	8.8	34.0	42.8	25.9	13.1	7.2	5.1	4.5	19.6	18.6	1.9

NBA 에서 제공하는 각 시즌 별 팀의 승률 및 다양한 DATA 확인 가능

DATA 설명

리그설명

변수설명

변수명	정의	단위
WIN	The percentage of games played that a player or team has won.	%
PTS	The number of points scored.	#
FGM	The number of field goals that a player or team has made. This includes both 2 pointers and 3 pointers.	#
FGA	The number of field goals that a player or team has attempted. This includes both 2 pointers and 3 pointers.	#
3PM	The number of 3 point field goals that a player or team has made.	#
3PA	The number of 3 point field goals that a player or team has attempted.	#
FTM	The number of free throws that a player or team has made.	#
FTA	The number of free throws that a player or team has attempted.	#
OREB	The number of rebounds a player or team has collected while they were on offense.	#

DATA 설명

리그설명

변수설명

변수명	정의	단위
AST	The number of assists -- passes that lead directly to a made basket -- by a player.	#
TOV	A turnover occurs when the player or team on offense loses the ball to the defense.	#
STL	Number of times a defensive player or team takes the ball from a player on offense, causing a turnover.	#
BLK	A block occurs when an offensive player attempts a shot, and the defense player tips the ball, blocking their chance to score.	#
BLKA	The number of shots attempted by a player or team that are blocked by a defender.	#
PF	The number of personal fouls a player or team committed.	#
PFD	The number of personal fouls that are drawn by a player or team.	#

예측모델 선정

(1)로지스틱회귀분석

(단계적 변수 선택)

(2)의사결정나무

(Gini, 엔트로피, ChiSQ)

(3)뉴럴네트워크

(neuralnet)



예측모델

로지스틱

의사결정나무

뉴럴네트워크

로지스틱 회귀분석

Step: AIC=106.86

WIN ~ PTS + FGM + FGA + X3PM + FTM + FTA + OREB + DREB + TOV +
STL + PFD

		Df	Deviance	AIC
<none>			82.862	106.86
+ BLK	1		81.577	107.58
+ AST	1		82.436	108.44
+ BLKA	1		82.596	108.60
+ PF	1		82.630	108.63
+ X3PA	1		82.741	108.74
- X3PM	1		86.768	108.77
- FTM	1		86.881	108.88
- FGM	1		86.928	108.93
- PFD	1		87.533	109.53
- PTS	1		87.703	109.70
- FTA	1		92.521	114.52
- TOV	1		106.854	128.85
- STL	1		116.682	138.68
- OREB	1		149.184	171.18
- DREB	1		155.630	177.63
- FGA	1		219.826	241.83

단계적 변수 선택 방식으로 설명력 높은
최적의 예측 변수 선택

- 단계적 선택 방식 'both' 를 이용
AIC = 106.86

- 최적의 예측 변수 생성 (11개의 변수 선택)

예측모델

로지스틱

의사결정나무

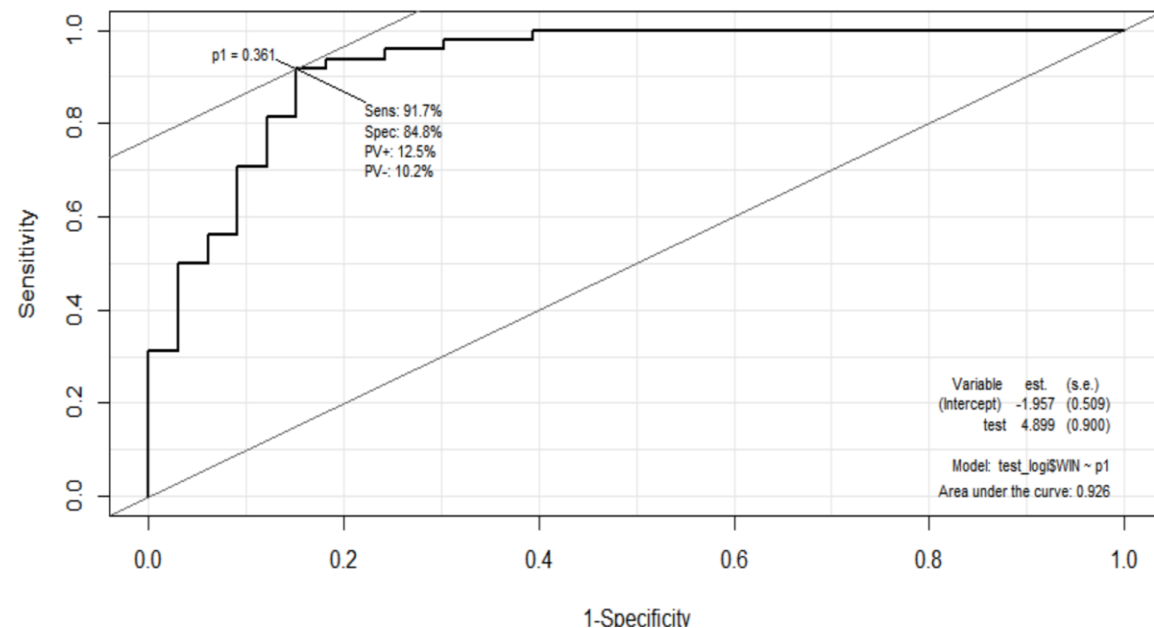
뉴럴네트워크

- Confusion Matrix and statistics

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	40 (TP)	8 (FP)
	Negative (0)	5 (FN)	28 (TN)

- Accuracy = 0.8395
- Sensitivity = 0.7778
- Specificity = 0.8889
- Precision = 0.8333

- ROC 커브 및 AUC 결과



- AUC = 92.6%

예측모델

로지스틱

의사결정나무

뉴럴네트워크

의사결정나무

대표적인 알고리즘

Gini

Entropy

ChiSQ

R 패키지

tree

Rpart

C 5.0

CHID

의사결정나무 세가지 알고리즘 중
최적의 알고리즘 선택

예측모델

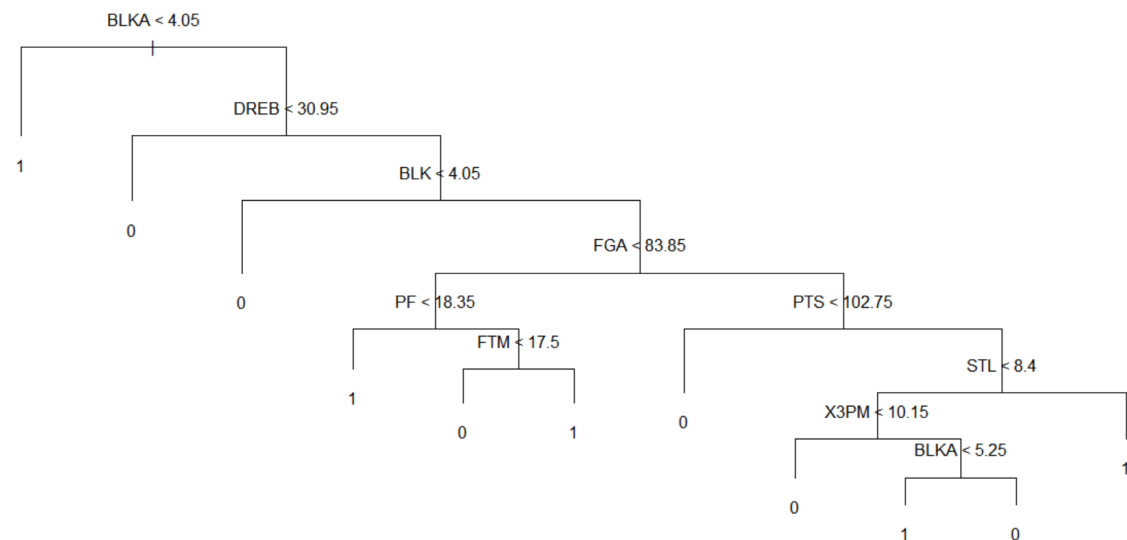
로지스틱

의사결정나무

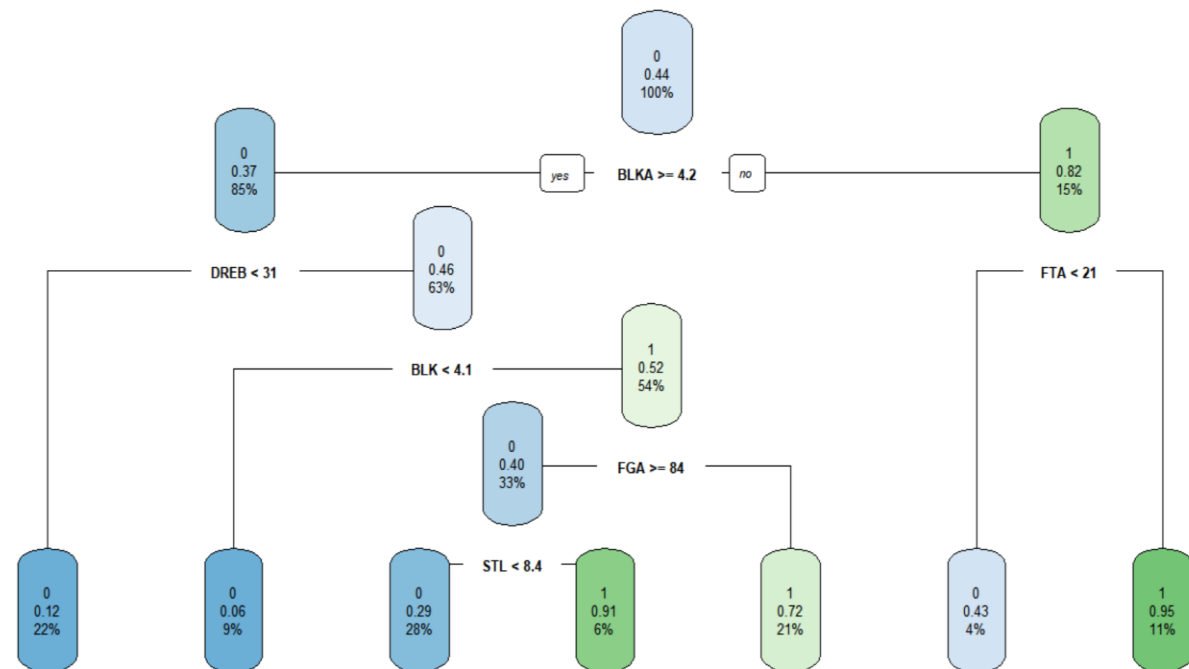
뉴럴네트워크

- 의사결정나무(Gini) – Tree, Rpart

tree



Rpart



예측모델

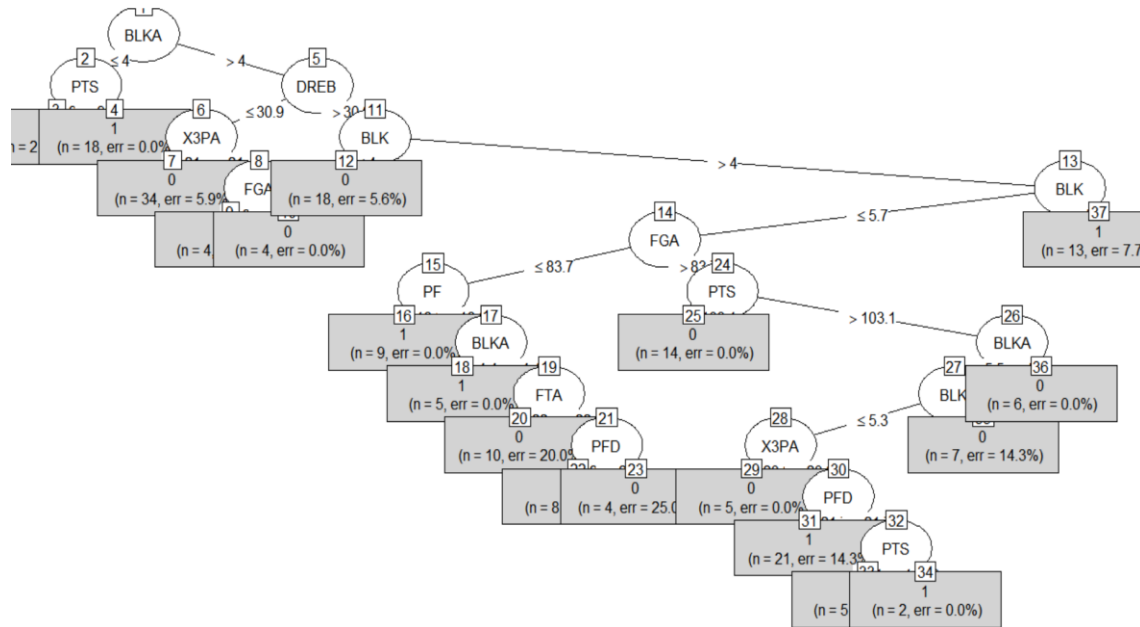
로지스틱

의사결정나무

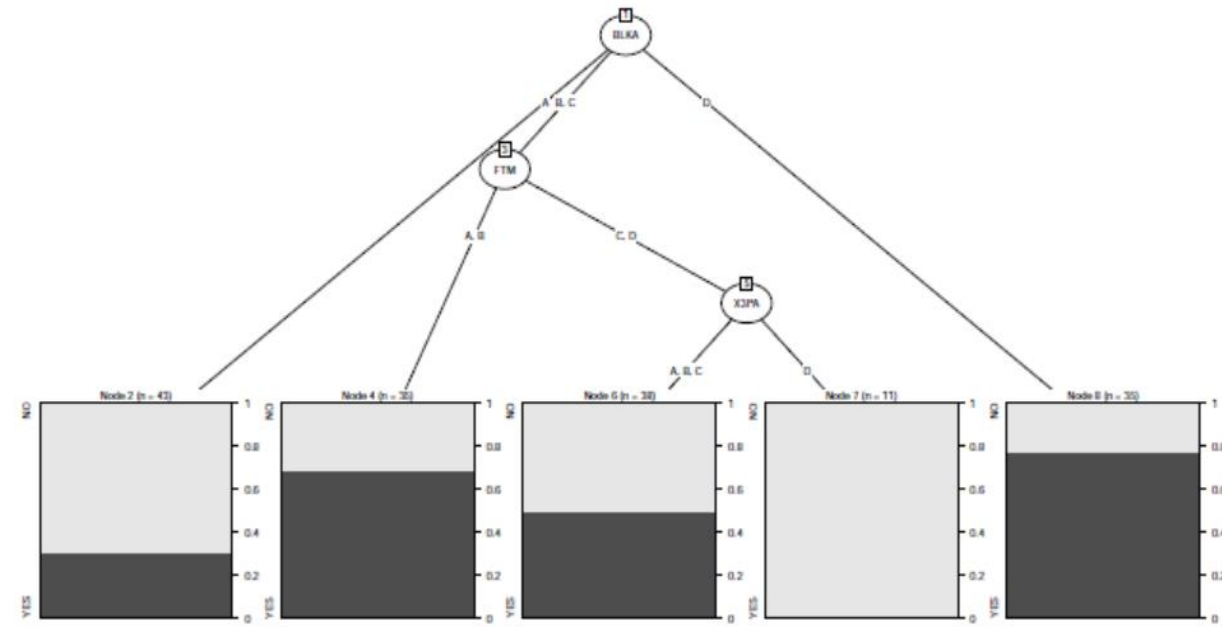
뉴럴네트워크

- 의사결정나무(엔트로피, 카이) – C4.5, CHID

C 5.0



CHID



예측모델

로지스틱

의사결정나무

뉴럴네트워크

- Confusion Matrix and statistics

tree

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	26 (TP)	5 (FP)
	Negative (0)	22 (FN)	28 (TN)

- Accuracy = 0.6667
- Sensitivity = 0.8485
- Specificity = 0.5417
- Precision = 0.8387

Rpart

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	24 (TP)	8 (FP)
	Negative (0)	24 (FN)	25 (TN)

- Accuracy = 0.6049
- Sensitivity = 0.7576
- Specificity = 0.5000
- Precision = 0.75

예측모델

로지스틱

의사결정나무

뉴럴네트워크

- Confusion Matrix and statistics

C 5.0

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	29 (TP)	9 (FP)
	Negative (0)	24 (FN)	19 (TN)

- Accuracy = 0.6543
- Sensitivity = 0.7273
- Specificity = 0.6785
- Precision = 0.7631

CHID

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	42 (TP)	18 (FP)
	Negative (0)	20 (FN)	28 (TN)

- Accuracy = 0.5555
- Sensitivity = 0.6774
- Specificity = 0.6086
- Precision = 0.7

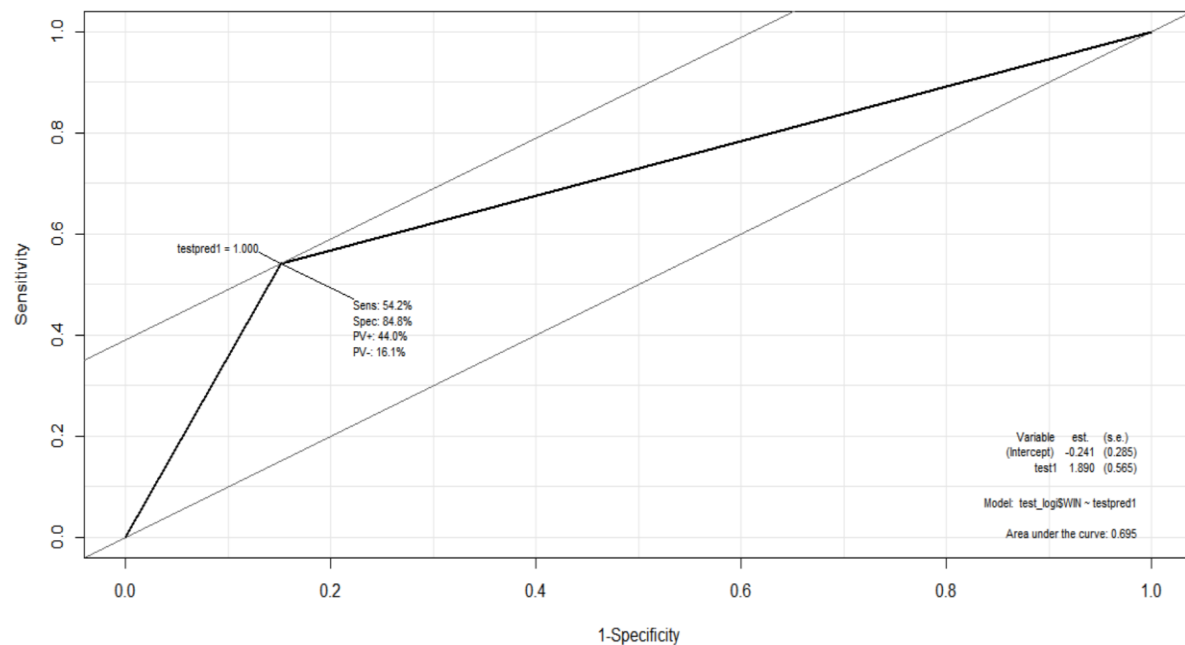
예측모델

로지스틱

의사결정나무

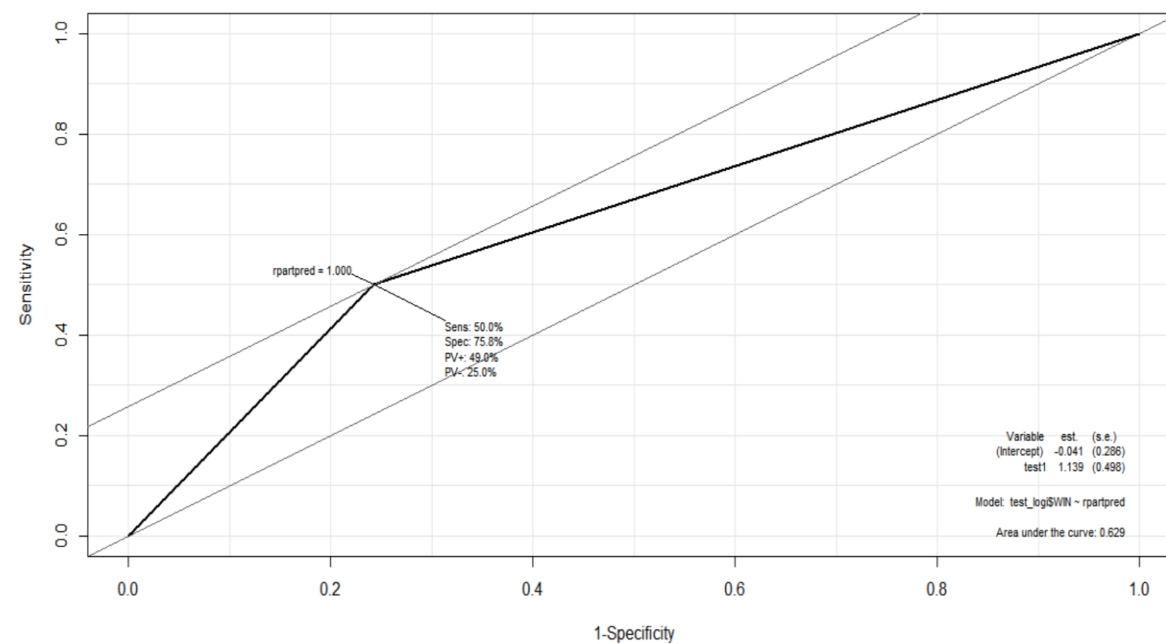
뉴럴네트워크

tree



- Accuracy = 0.6667 , AUC = 0.695

Rpart



- Accuracy = 0.6049 , AUC = 0.629

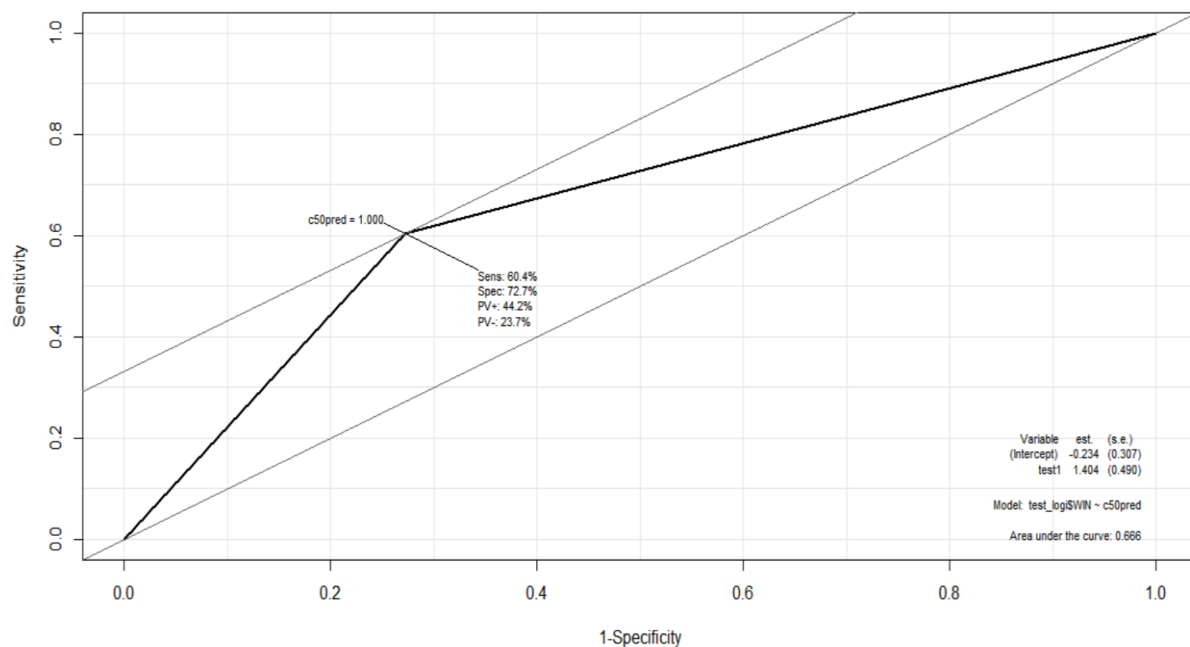
예측모델

로지스틱

의사결정나무

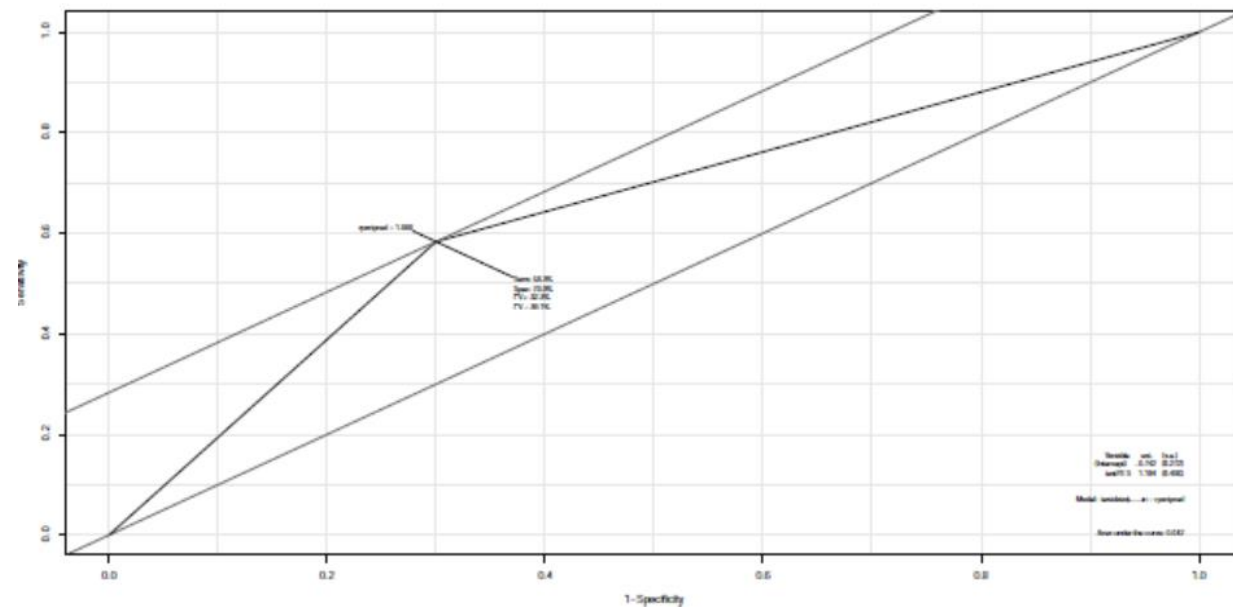
뉴럴네트워크

C5.0



- Accuracy = 0.6667 , AUC = 0.695

CHID



- Accuracy = 0.6049 , AUC = 0.629

예측모델

로지스틱

의사결정나무

뉴럴네트워크

	Gini	ChiSQ	엔트로피
Accuracy	66.67%	55.55%	65.43%
Sensitivity	84.85%	67.74%	72.73%
Specificity	54.17%	60.86%	67.85%
Precision	83.87%	70%	76.31%

의사결정나무 중 최적의 알고리즘 : Gini

- Accuracy(정확성) : 진출 구단과 진출하지 못한 구단을 정확히 예측
- Sensitivity(민감도) : 실제 진출한 구단 중 진출한 구단을 정확히 예측
- Precision(정밀도) : 진출 구단으로 예측한 것 중 실제로 진출한 구단을 예측

➡ 진출팀을 정확하게 예측하는
Gini 모델 채택

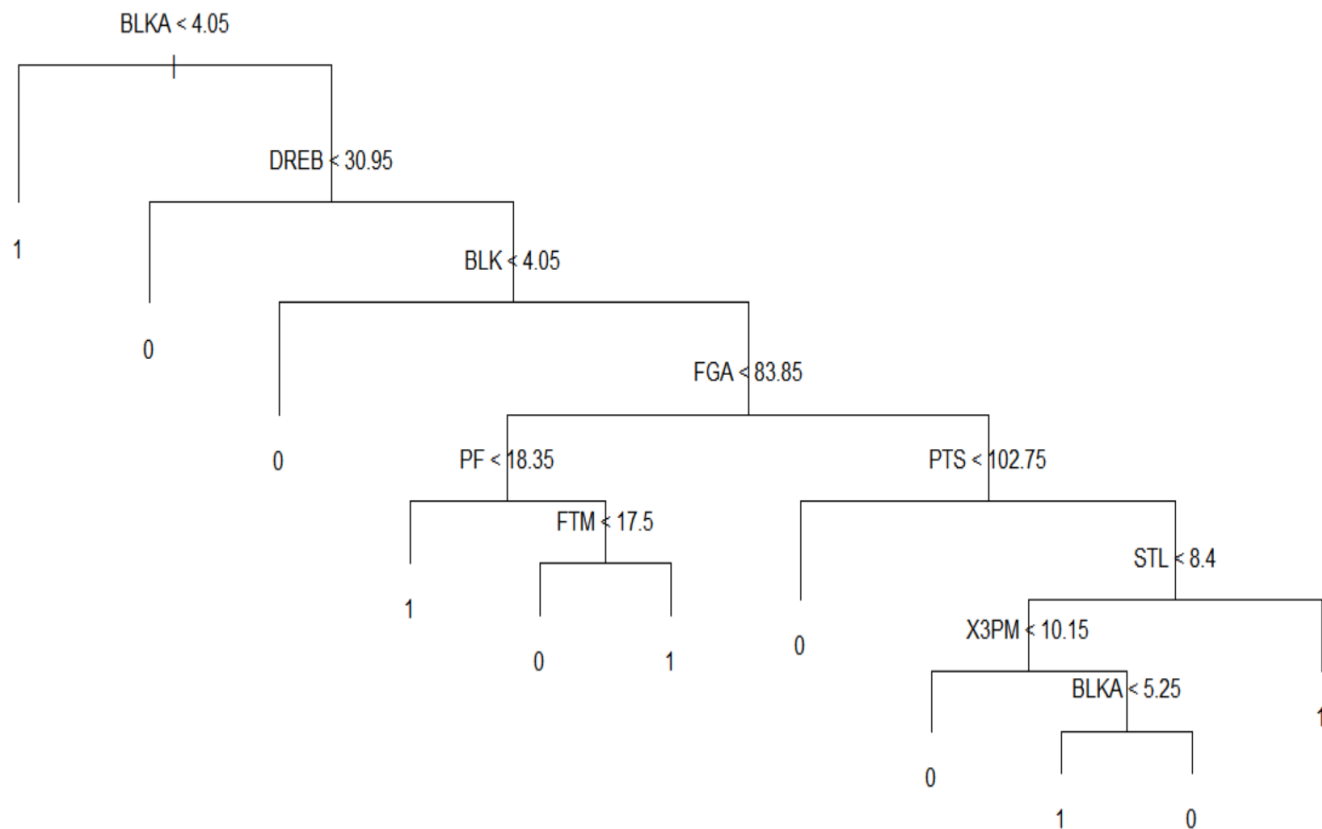
예측모델

로지스틱

의사결정나무

뉴럴네트워크

Tree 결과



진출 여부의 척도 변수

BLKA (블락 시도) 4.05 이상,
3PM (3점슛 성공) 10.15 이상,
STL(가로채기) 8.4 이상,
PTS(득점수) 102.75 이상,
FGA(야투시도수) 83.85 이상,
BLK(블락) 4.05 이상,
DREB(수비리바운드) 30.95 이상

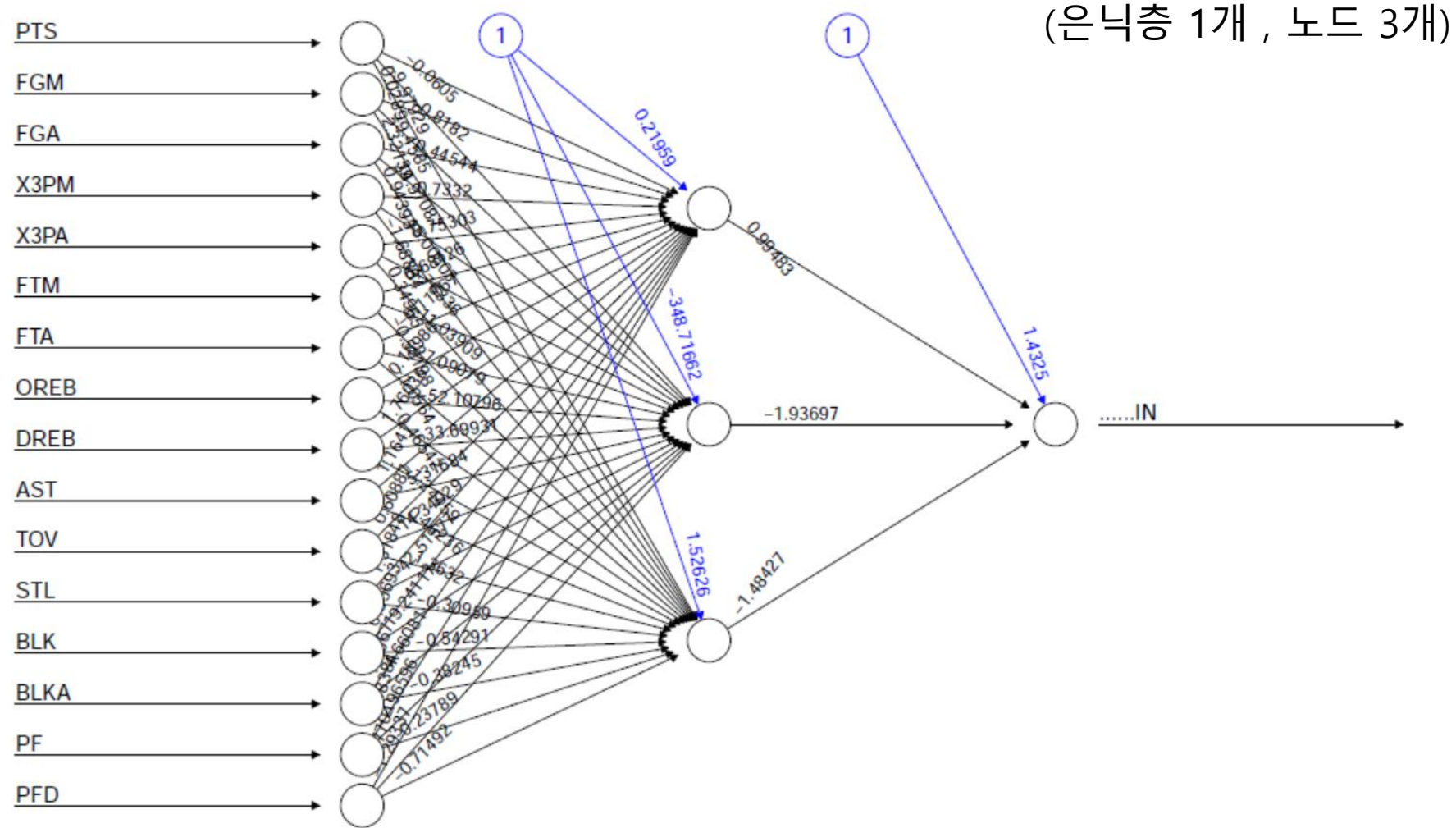
예측모델

로지스틱

의사결정나무

뉴럴네트워크

Neural Network



예측모델

로지스틱

의사결정나무

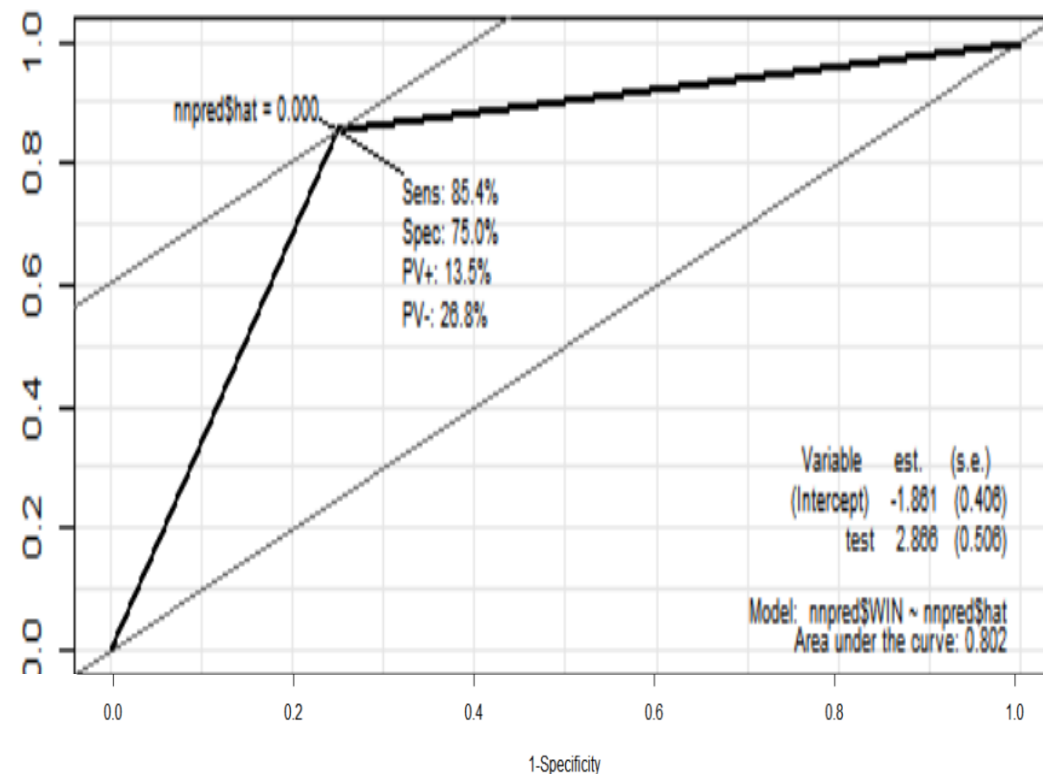
뉴럴네트워크

- Confusion Matrix and statistics

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	39 (TP)	15 (FP)
	Negative (0)	8 (FN)	47 (TN)

- Accuracy = 0.7889
- Sensitivity = 0.8297
- Specificity = 0.7580
- Precision = 0.7222

- ROC 커브 및 AUC 결과



AUC는 전체 중 약 80.2%의 면적을 차지하는 것을 확인

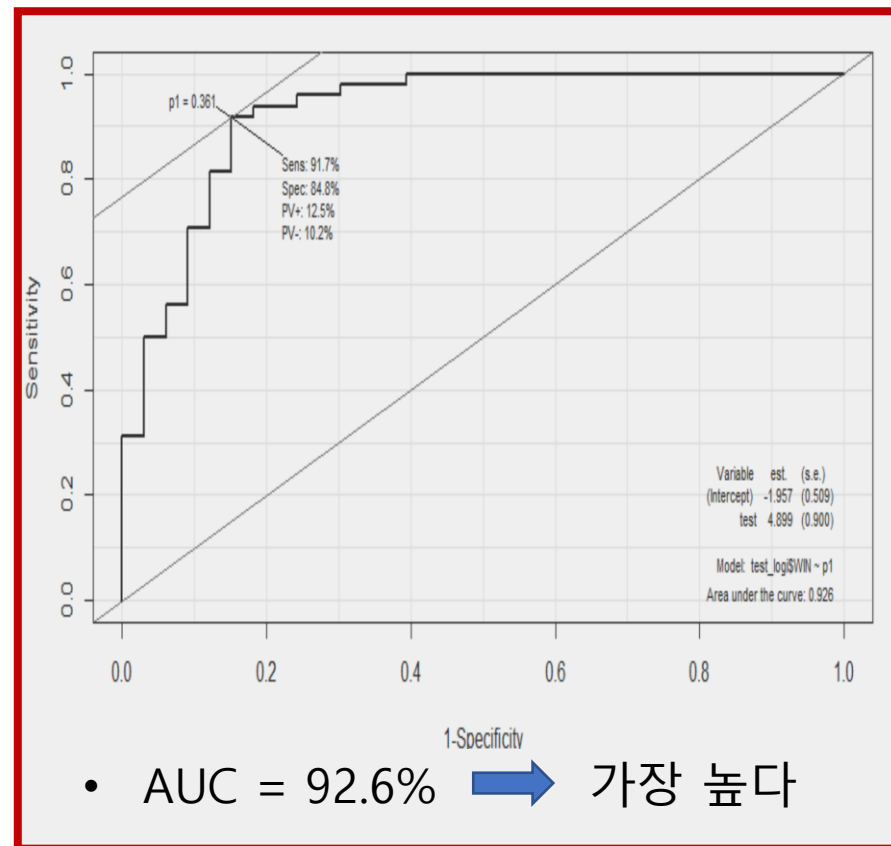
모델 간 비교

모델 간 비교

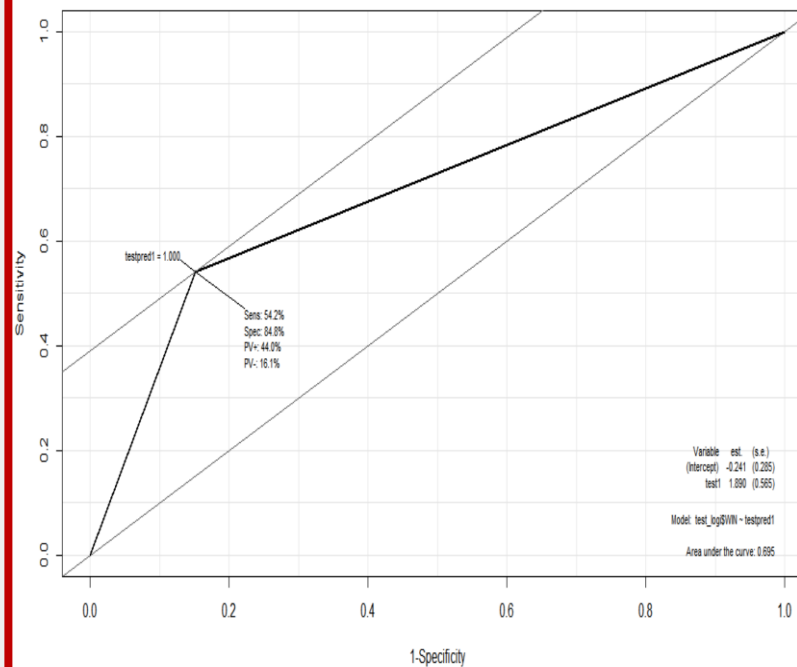
예측 분석 간 비교

모델 간 비교

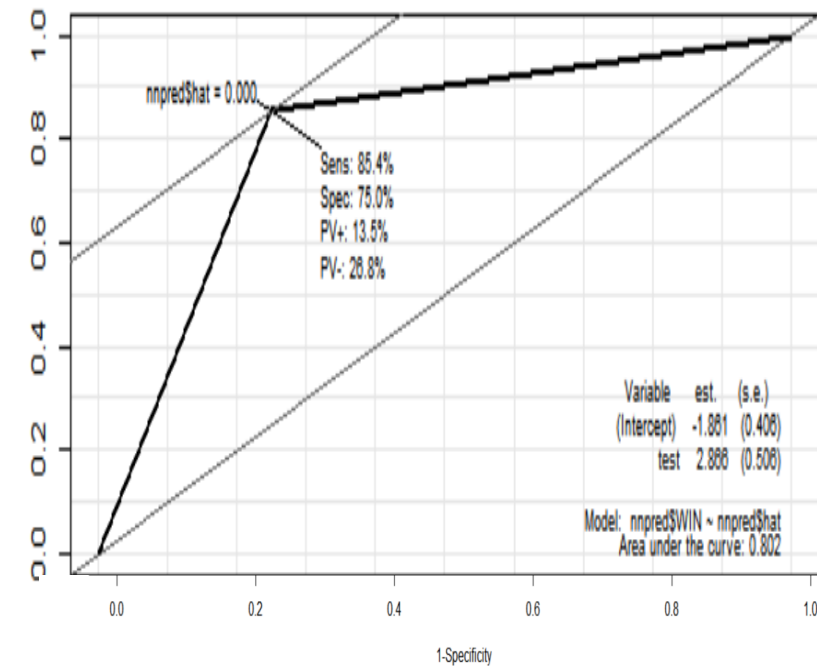
로지스틱스 회귀분석



의사결정나무



뉴럴네트워크



모델 간 비교

모델 간 비교

예측 분석 간 비교

Accuracy

진출한 구단과 진출하지 못한 구단을 올바르게 예측한 비율 ($TP+TN/TP+FP+FN+TN$)

Sensitivity

실제 진출한 구단 중 진출한 구단을 올바르게 예측한 비율 ($TP/TP+FN$)

Specificity

실제 진출하지 못한 구단 중 진출하지 못한 구단을 올바르게 예측한 비율 ($TN/TN+FP$)

Precision

진출한 구단으로 예측한 것 중 실제 진출한 구단인 비율 ($TP/TP+FP$)

<실제값과 예측값 비교 표>

		Actual Values	
		Positive(1)	Negative(0)
Predicted values	Positive (1)	TP	FP
	Negative (0)	FN	TN

모델 간 비교

모델 간 비교

예측 분석 간 비교

	로지스틱회귀 분석	의사결정나무	뉴럴네트워크
Accuracy	83.95%	66.67%	78.89%
Sensitivity	77.78%	84.85%	82.97%
Specificity	88.89%	54.17%	75.80%
Precision	83.33%	83.87%	72.22%

• 로지스틱 회귀분석

- 진출하거나 진출하지 못한 구단 예측을 가장 정확하게 예측
- 실제 진출하지 못한 구단 중 해당 구단을 가장 정확하게 예측

• 의사결정나무

- 실제 진출한 구단 중 해당 구단을 가장 정확하게 예측
- 진출한 구단으로 예측 한 구단 중 실제 진출한 구단을 가장 정확하게 예측

• 뉴럴네트워크

- 종합적으로 편중되지 않은 값을 보임
- 실제 진출한 구단 중 해당 구단을 어느정도 예측

결론 및 시사점

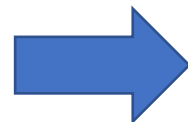
결론

예측 분석 간 비교

I. 예측 모형 선택

<각 모델 별 예측 검증 지표>

	모델
Accuracy	로지스틱 회귀분석
Sensitivity	의사결정나무
Specificity	로지스틱 회귀분석
Precision	의사결정나무



Sports TOTO 입장

구단의 진출 예측이 가장 중요

예측력을 나타내는
Accuracy가 높은 모델 선택

로지스틱회귀분석 채택



구단 팬 입장

응원하는 구단의 진출이
가장 중요

진출 예측을 나타내는
Sensitivity와 Precision이 높은
모델 선택

의사결정나무 채택

결론 및 시사점

결론

예측 분석 간 비교

II. 모델 활용 가능 사례

릴라드 "PO 기회 안줄거면 경기 뛰지 않겠다"

기사입력 2020.05.27 05:05:50 | 최종수정 2020.05.27 07:13:04



매경닷컴 MK스포츠(美 알링턴) 김재호 특파원

NBA가 2019-20시즌 재개를 앞두고 있는 가운데 포틀랜드 트레일 블레이저스 주전 가드 다미안 릴라드는 의미 없는 경기는 뛰지 않겠다고 말했다.

NBA는 현재 시즌 재개를 위한 여러가지 시나리오를 고민중이다. 모든 팀이 참가해 정규시즌을 마무리하는 방안, 플레이오프 진출권에 들어와 있는 팀들이 별도의 토너먼트를 치르는 방안, 아니면 시즌 중단 당시 양 컨퍼런스 8위 안에 들었던 16개 팀만 불러 바로 플레이오프를 치르는 방안 등이 거론되고 있다.

릴라드는 "만약 돌아왔는데 플레이오프 진출 기회가 없다면, 나는 연습도 하고 팀원들과 함께하겠지만 경기는 벤치에서 보겠다. 만약 돌아와서 7번 시드부터 12번 시드까지 토너먼트 방식으로 플레이오프 진출권을 놓고 다툰다면, 내 생각에 완벽하다고 본다"며 자신의 생각을 전했다.



중단된 리그 재개 어떻게?



NBA 입장

중단된 시점의 순위에서 재개

전반적인 정확도가 높은
로지스틱 회귀모델로
리그 시뮬레이션 활용



구단 입장

진출에 있어 공정하지 못한 방식

진출 예측이 높은
의사결정모델로
구단의 진출 여부
시뮬레이션 활용

결론 및 시사점

결론

예측 분석 간 비교

Ⅲ. 시사점

실제 구단의 사례

휴스턴 로켓츠 구단의 사례

1. 롱투(Long-two)는 버린다
2. **3 점슛 시도에 집중한다**
3. **야투(2점슛+3점슛)시도를** 포함하는 슈팅 효율성 지표를 고안

2017-2018 시즌 Western 1위 차지

의사결정나무

진출 여부의 척도 변수

BLKA (블락 시도)
3PM (3점슛 성공)
STL(가로채기)
PTS(득점수)
FGA(야투시도수)
BLK(블락)
DREB(수비리바운드)

실제 사례 분석과 데이터 분석 결과 방향성 일치,
스포츠 데이터 분석의 중요성 입증



Thank You