

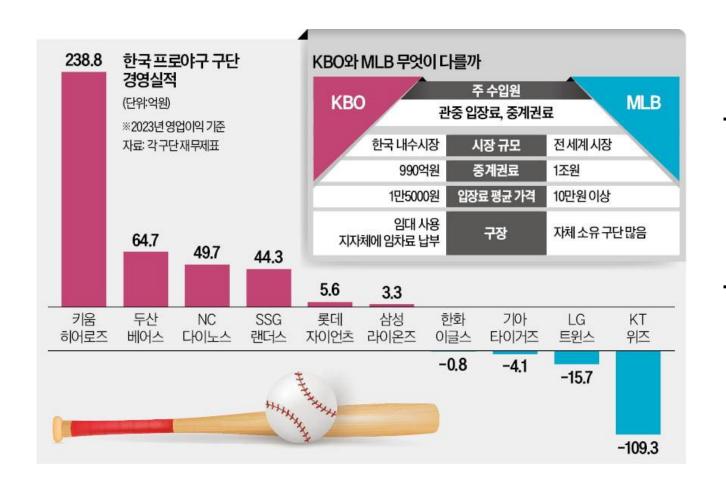
목차

주제 소개

야구 시장은 현재 한국에서도 엄청난 흥행을 끌고 있음







하지만 실제 MLB 시장의 사업성이 더 높음

시장 규모: MLB > KBO

사업성: MLB > KBO

Forbes

The Most Powerful Sports Agents

RANK A	PERSON	SPORT	MAXIMUM COMMISSIONS	CONTRACTS	AGENCY
1	Scott Boras	Baseball	\$191 M	\$3.8 B	Boras Corporation
2	Jeff Schwartz	Basketball	\$86 M	\$2.1 B	Excel Sports Management
3	Joel Wolfe	Baseball	\$58 M	\$1.2 B	Wasserman
4	Rich Paul	Basketball	\$55 M	\$1.4 B	Klutch Sports Group
5	Dan Lozano	Baseball	\$52 M	\$1 B	MVP Sports Group
6	Casey Close	Baseball	\$48 M	\$960 M	Excel Sports Management
7	Joel Segal	Football	\$42 M	\$1.4 B	WME Sports
8	Todd France	Football	\$42 M	\$1.4 B	Athletes First
9	Bill Duffy	Basketball	\$41 M	\$1 B	WME Sports
10	Sam and Seth Levinson	Baseball	\$40 M	\$810 M	ACES

스포츠 에이전트에서도 야구 Agent가 상위권에 위치

이에 따라 수많은 언론사와 사업체들이 계약 규모를 각자의 기준을 통해 예측하고자 함

현재 24-25 FA시장은 활발하게 돌아가는 중



2024-25 Top 50 MLB Free Agents With Predictions



The top 25 free agents this offseason

2) Corbin Burnes, RHP (age: 30)^

Burnes was everything the Orioles could have hoped for in 2024, giving Baltimore the workhorse ace it had been seeking. Burnes threw a gem in his lone postseason start for the Orioles, who are likely to do everything they can to keep the four-time All-Star and 2021 NL Cy Young Award winner. The right-hander should land his fifth straight top 10 Cy Young finish this year, and he will be the most sought-after starting pitcher this winter.

Potential fits: Mets, Orioles, Red Sox

The New Hork Times

Top 45 MLB free agents for 2024-25 with contract predictions, team fits: Will Soto get \$600M+?

1. Juan Soto, OF

Age: 25

B: L **T:** L **HT:** 6-2 **WT:** 225 **2024 (Yankees):** 7.9 WAR

Career: 36.4 WAR

Agent: Boras Corporation 2024 salary: \$31 million

Has Juan Soto found his permanent home with the <u>Yankees</u>? <u>Can they keep him after a strong platform year</u>? He certainly lived up to expectations in his first year in the Bronx, slashing .288/.419/.569 with 41 home runs and a league-leading 128 runs scored. He's expected to finish third in the American League MVP voting behind <u>Aaron Judge</u> and <u>Bobby Witt Jr.</u> He's met the big moments in this postseason and has logged a 1.106 OPS.

이에 따라 수많은 언론사와 사업체들이 계약 규모를 각자의 기준을 통해 예측하고자 힘

TRADERUMORS

Goal

2024-25 Top 50 MLB Free Agents
With Predictions

우리만의기준으로 24-25시즌 FA시장

12 predictions for 12 free agents

contract predictions, team fits: Will Soto get \$600M+?

데이터셋설명

1. FA시장에서의 각 선발 투수 별 Value, AGE 데이터 크롤링

MLB Free Agents

	<u> </u>								
Tracking	the status and signings o	f all MLB free agents.							
2024	✓ All Statuses	✓ All Teams	v	Pitchers (P)	✓ All Types	✓ MLB Or	nly	~	
RANK	PLAYER (94)	POS	AGE	ARM BAT/THROW	TEAM FROM	TEAM TO	YRS	VALUE	AAV
								152	\$2,302,290,000
1	Shohei Ohtani	SP/DH	30.4	R	🛕 LAA	∞ LAD	10	\$700,000,000	\$70,000,000
2	Yoshinobu Yamamoto	SP	26.2	R	JPN	∞ LAD	12	\$325,000,000	\$27,083,333
3	Aaron Nola	SP/SP2	31.5	R	🕵 PHI	🕵 PHI	7	\$172,000,000	\$24,571,429
4	Josh Hader	RP/CL	30.7	L	狗 SD	HOU	5	\$95,000,000	\$19,000,000
5	Eduardo Rodriguez	SP/SP1	31.7	L	通 DET	A ARI	4	\$80,000,000	\$20,000,000
6	Sonny Gray	SP/SP3	35.1	R	MIN	🚵 STL	3	\$75,000,000	\$25,000,000
7	Blake Snell	SP/SP3	32.0	L	狗 SD	sum SF	2	\$62,000,000	\$31,000,000
8	Shota Imanaga	SP	31.2	L	JPN	© CHC	4	\$53,000,000	\$13,250,000
9	Seth Lugo	SP	35.0	R	狗 SD	KC	3	\$45,000,000	\$15,000,000

2. 시즌별 선발 투수 Stats 크롤링



Max Scherzer Free Agent

Season	Team	Level	W	L	ERA	G	GS	QS	CG	ShO	sv	HLD	BS	IP	TBF	Н	R	ER	HR	ВВ	IBB	НВР	WP	BK	so
2008	ARI	MLB	0	4	3.05	16	7	3	0	0	0	0	0	56.0	237	48	24	19	5	21	1	5	2	0	66
2009	ARI	MLB	9	11	4.12	30	30	14	0	0	0	0	0	170.1	741	166	94	78	20	63	1	10	5	1	174
2010	DET	MLB	12	11	3.50	31	31	18	0	0	0	0	0	195.2	800	174	84	76	20	70	1	7	8	0	184
2011	DET	MLB	15	9	4.43	33	33	16	0	0	0	0	0	195.0	833	207	101	96	29	56	1	7	12	0	174
2012	DET	MLB	16	7	3.74	32	32	20	0	0	0	0	0	187.2	787	179	82	78	23	60	2	5	2	1	231
2013	DET	MLB	21	3	2.90	32	32	25	0	0	0	0	0	214.1	836	152	73	69	18	56	0	4	6	1	240
2014	DET	MLB	18	5	3.15	33	33	22	1	1	0	0	0	220.1	904	196	80	77	18	63	1	6	10	1	252
2015	WSN	MLB	14	12	2.79	33	33	23	4	3	0	0	0	228.2	899	176	74	71	27	34	2	5	10	1	276
2016	WSN	MLB	20	7	2.96	34	34	26	1	0	0	0	0	228.1	902	165	77	75	31	56	2	6	2	1	284
2017	WSN	MLB	16	6	2.51	31	31	22	2	0	0	0	0	200.2	780	126	62	56	22	55	2	11	4	0	268
2018	WSN	MLB	18	7	2.53	33	33	28	2	1	0	0	0	220.2	866	150	66	62	23	51	4	12	4	1	300
2019	WSN	MLB	11	7	2.92	27	27	17	0	0	0	0	0	172.1	693	144	59	56	18	33	2	7	0	0	243
2020	WSN	MLB	5	4	3.74	12	12	6	1	0	0	0	0	67.1	295	70	30	28	10	23	1	1	6	0	92
2021	2 Tms	MLB	15	4	2.46	30	30	18	1	0	0	0	0	179.1	693	119	53	49	23	36	0	10	2	0	236
2022	NYM	MLB	11	5	2.29	23	23	16	0	0	0	0	0	145.1	565	108	39	37	13	24	0	11	1	0	173
2023	2 Tms	MLB	13	6	3.77	27	27	14	0	0	0	0	0	152.2	621	126	65	64	28	45	0	4	2	0	174
2024	TEX	MLB	2	4	3.95	9	9	3	0	0	0	0	0	43.1	177	40	20	19	7	10	0	1	2	0	40

3. Tommy John Surgery (팔꿈치 인대 수술) 일자 Data

Tommy John Surgery List (@MLBPlayerAnalys)

										Return Date	Recovery Time			
Player	TJ Surgery Date	Team	Level	Position	Throws	Country	High School	College(s)	Age	(same level)	(months)	mlbamid	fgid	Surgeon(s)
Wen-Hui Pan	11/1/2024	PHI	A+	Р	R	Taiwan			21			808207	sa3020878	
Julian Aguiar	10/11/2024	CIN	MLB	Р	R	United States	California	Cypress	23			687924	30031	Dr. Keith Meister
Joe Musgrove	10/11/2024	SD	MLB	Р	R	United States	California		31			605397	12970	Dr. Keith Meister
Cade Feeney	10/1/2024	BOS	A+	Р	R	United States	North Dakota	North Dakota State	22			814366	sa3023228	
Christian Scott	9/25/2024	NYM	MLB	Р	R	United States	Florida	Florida	25			681035	30107	Dr. Keith Meister
Brandon Williamson	9/1/2024	CIN	MLB	Р	L	United States	Minnesota	Texas Christian	26			682227	25463	
Chad Dallas	9/1/2024	TOR	AAA	Р	R	United States	Texas	Tennessee	24			689149	sa3018180	
Nolan Perry	9/1/2024	TOR	Α	Р	R	United States	New Mexico		20			804526	sa3021476	
Carson Pierce	9/1/2024	TOR	A+	Р	R	United States	Oklahoma	Oklahoma	22			804544	sa3023179	
Jorge Mateo	8/28/2024	BAL	MLB	2B	R*	Dominican			29			622761	17273	Dr. Keith Meister
Adbert Alzolay	8/7/2024	CHC	AAA	Р	R	Venezuela			29			640470	17859	Dr. Keith Meister
Luis Medina	8/7/2024	OAK	MLB	Р	R	Dominican			25			665622	21649	Dr. Keith Meister
Cole Mathis	8/1/2024	CHC	Coll	3B	R	United States	Georgia	Charleston	20			695573		Dr. Stephen Gryzlo
River Ryan	8/1/2024	LAD	MLB	Р	R	United States	North Carolina	North Carolina	25			689981	29514	Dr. Neal ElAttrache
Luke Keaschall	8/1/2024	MIN	AA	2B	R	United States	California	San Francisco, Arizona State	21			807712	sa3022614	
Kyle Hurt	7/30/2024	LAD	AAA	Р	R	United States	California	USC	26			669165	27514	Dr. Neal ElAttrache
Ricky Tiedemann	7/30/2024	TOR	AAA	Р	L	United States	California	Golden West	21			694357	sa3018096	Dr. Neal ElAttrache
Josiah Gray	7/24/2024	WAS	AAA	Р	R	United States	New York	Le Moyne	26			680686	24580	Dr. Keith Meister
Drew Smith	7/12/2024	NYM	MLB	Р	R	United States	Texas	Dallas Baptist	30			622098	17755	Dr. Keith Meister
Luis Oritz	7/11/2024	PHI	AAA	Р	R	United States	California		28			656814	16942	
Kiefer Lord	7/3/2024	BAL	Rk	Р	R	United States	California	Washington, Carleton	21			811296	sa3023135	Dr. Keith Meister
Joev Broughton	7/2/2024	MIL	Coll	Р	1	United States	Michigan							

4. 24-25 시즌 FA시장 선발투수 리스트 (최종 예측 데이터셋)

All of the 2024-2025 free agents, position by position

STARTING PITCHERS

Corbin Burnes (30 years old, 7.2 WAR) -- declined qualifying offer

Blake Snell (32, 7.2) -- signed 5-year deal with LAD (Nov. 30)

Yusei Kikuchi (34, 6.0) -- signed 3-year deal with LAA (Nov. 27)

Max Fried (31, 5.2) -- declined qualifying offer

Nathan Eovaldi (35, 5.1)

Jack Flaherty (29, 5.0)

Nick Martinez (34, 4.9) -- accepted qualifying offer from CIN (Nov. 18)

Kyle Gibson (37, 4.0)

Justin Verlander (42, 4.0)

Nick Pivetta (32, 3.9) -- declined qualifying offer

Sean Manaea (33, 3.9) -- declined qualifying offer

Charlie Morton (41, 3.8)

Patrick Sandoval (28, 3.5)

Andrew Heaney (34, 3.5)

Kyle Hendricks (35, 3.0) -- signed 1-year deal with LAA (Nov. 7)

Clayton Kershaw (37, 3.0)

Shane Bieber (30, 2.8) -- reported 1-year deal with CLE

Max Scherzer (40, 2.8)

Patrick Corbin (35, 2.6)

José Quintana (36, 2.6)

모델구축및성능확인

모델구축과정

학습 데이터

18-19시즌 ~ 23-24시즌 FA 선수 데이터 1. FA 계약 성공 여부 예측 모델 구축

2. 계약 성공 선수의 계약 년도 예측 모델 구축

3. 계약 성공 선수의 계약 금액 예측 회귀 모델 구축

사용모델

RandomForest

실제 예측 데이터

24-25시즌 FA 선수 목록 1. FA 계약 성공 여부 예측 모델 적용

2. 계약 성공 선수의 계약 년도 예측 모델 적용

3. 계약 성공 선수의 계약 금액 예측 회귀 모델 적용

모델구축과정

사용모델 성능지표 비교

	Random Forest	Light GBM	XGBoost
MAE	0.73	0.75	0.79
МАРЕ	4.56	4.71	4.88
RMSE	0.85	0.86	0.92
R Squared	0.57	0.55	0.51

사용모델 성능지표 비교 결과

모델 별 성능의 차이는 미미한 수준이긴 하나, 네 지표 모두 Random Forest의 성능이 가장 뛰어남

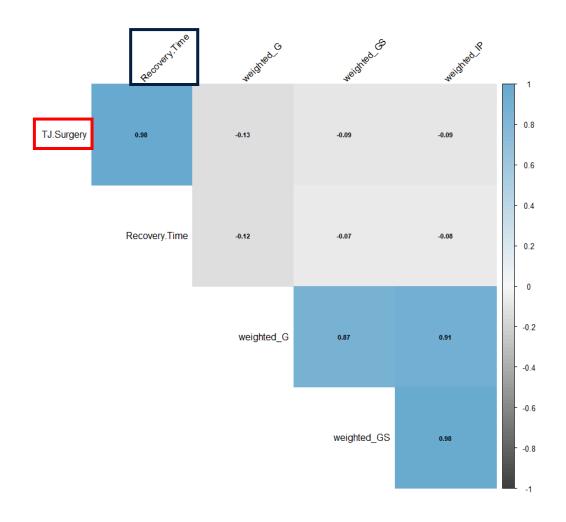
Part 3 모델사용 변수 설정

기존의 사용 가능 변수들

Season	Team	Level	W	L	ERA	G	GS	QS	CG	ShO	sv	HLD	BS	IP	TBF	Н	R	ER	HR	ВВ	IBB	НВР	WP	вк	so
2008	ARI	MLB	0	4	3.05	16	7	3	0	0	0	0	0	56.0	237	48	24	19	5	21	1	5	2	0	66
2009	ARI	MLB	9	11	4.12	30	30	14	0	0	0	0	0	170.1	741	166	94	78	20	63	1	10	5	1	174
2010	DET	MLB	12	11	3.50	31	31	18	0	0	0	0	0	195.2	800	174	84	76	20	70	1	7	8	0	184
2011	DET	MLB	15	9	4.43	33	33	16	0	0	0	0	0	195.0	833	207	101	96	29	56	1	7	12	0	174
2012	DET	MLB	16	7	3.74	32	32	20	0	0	0	0	0	187.2	787	179	82	78	23	60	2	5	2	1	231
2013	DET	MLB	21	3	2.90	32	32	25	0	0	0	0	0	214.1	836	152	73	69	18	56	0	4	6	1	240
2014	DET	MLB	18	5	3.15	33	33	22	1	1	0	0	0	220.1	904	196	80	77	18	63	1	6	10	1	252
2015	WSN	MLB	14	12	2.79	33	33	23	4	3	0	0	0	228.2	899	176	74	71	27	34	2	5	10	1	276
2016	WSN	MLB	20	7	2.96	34	34	26	1	0	0	0	0	228.1	902	165	77	75	31	56	2	6	2	1	284
2017	WSN	MLB	16	6	2.51	31	31	22	2	0	0	0	0	200.2	780	126	62	56	22	55	2	11	4	0	268
2018	WSN	MLB	18	7	2.53	33	33	28	2	1	0	0	0	220.2	866	150	66	62	23	51	4	12	4	1	300
2019	WSN	MLB	11	7	2.92	27	27	17	0	0	0	0	0	172.1	693	144	59	56	18	33	2	7	0	0	243
2020	WSN	MLB	5	4	3.74	12	12	6	1	0	0	0	0	67.1	295	70	30	28	10	23	1	1	6	0	92
2021	2 Tms	MLB	15	4	2.46	30	30	18	1	0	0	0	0	179.1	693	119	53	49	23	36	0	10	2	0	236
2022	NYM	MLB	11	5	2.29	23	23	16	0	0	0	0	0	145.1	565	108	39	37	13	24	0	11	1	0	173
2023	2 Tms	MLB	13	6	3.77	27	27	14	0	0	0	0	0	152.2	621	126	65	64	28	45	0	4	2	0	174
2024	TEX	MLB	2	4	3.95	9	9	3	0	0	0	0	0	43.1	177	40	20	19	7	10	0	1	2	0	40

......

변수 조정



최종 선정 변수

Drop 변수

Recovery.Time

TJ.Surgery

(복귀일 – Tommy John 수술일자)

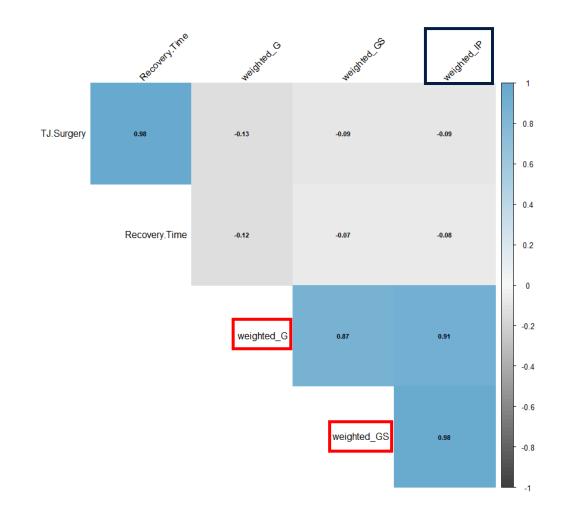
Tommy John 수술 여부 (팔꿈치 인대 수술)

* 수술 시기가 3년 이상 지난 경우 수술 안 받은걸로 취급

□ Drop 변수

□ 최종 선정 변수

변수 조정





Drop 변수

Weighted_GS

Weighted_IP

투구 이닝 수

Ex. 1회 - 3회 출전 시 투구 이닝 수 3 선발경기 수

투수가 선발투수로 등판한 경기의 수만을 의미 (1회부터 경기를 시작한 경우만 해당)

Weighted_G

출장경기 수

투수가 등판한 모든 경기의 수

■ Drop 변수

□ 최종 선정 변수

Part 3 모델사용 변수 설명

최종사용변수명추가설명

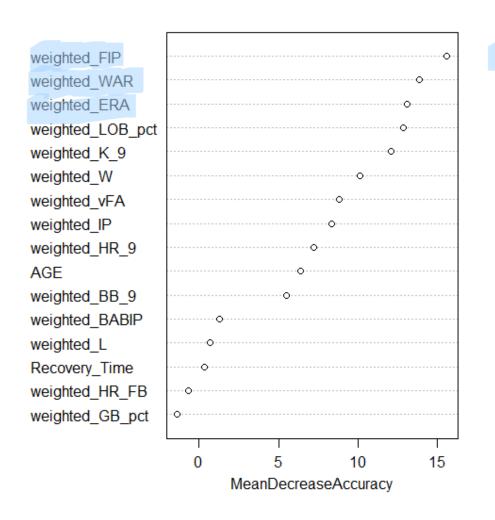
열 이름	설명
WAR	같은 포지션 다른 선수 대비 승리 기여도
FIP	수비 무관 평균 자책점
ERA	평균 자책점
Recovery_time	Tommy John 수술 이후 복귀까지 걸린 시간
AAV	총 계약금 / 계약년수
IP	투구 이닝 수
vFA	패스트볼 구속
HR_FB	플라이볼 대비 홈런 비율
HR_9	9이닝 당 홈런 허용 개수
BABIP	홈런 제외 안타 성공률
GB_pct	타구 중 땅볼 비율
LOB_pct	잔루율(주자를 내보낸 비율)
BB_9	9이닝 당 볼넷 개수
κ_9	9이닝 당 삼진 개수

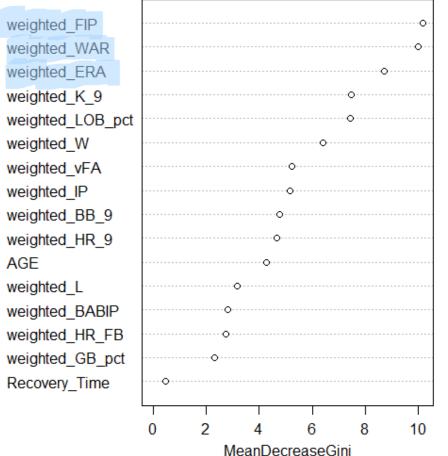
실제 사용 변수는 더 많음(나이, 승리 수, 패배 수, 계약 년 수 등..)

1. 계약 성공 여부 예측

FIP 수비 무관 평균 자책점 WAR 같은 포지션 다른 선수 대비 승리 기여도 ERA 평균 자책점

Feature Importance





성능지표

Train accuracy: 0.965

Test accuracy: 0.812

Precision: 0.811

Recall: 0.698

* 년도는 Categorize + SMOTE 진행

Category 1: 1년 계약 Category 2: 2~4년 계약 Category 3: 5년 이상 계약

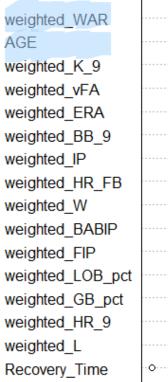
 FIP
 수비 무관 평균 자책점

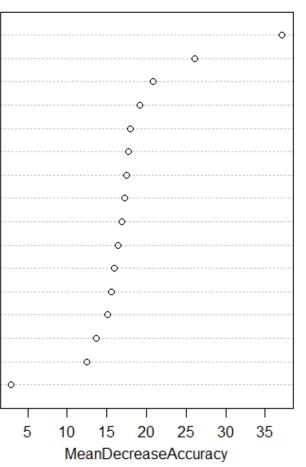
 WAR
 같은 포지션 다른 선수 대비 승리 기여도

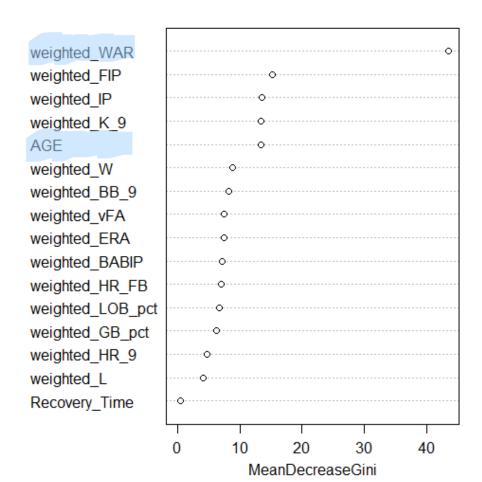
 ERA
 평균 자책점

2. 계약 년도 예측

Feature Importance







성능지표

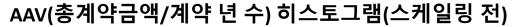
Train accuracy: 0.989

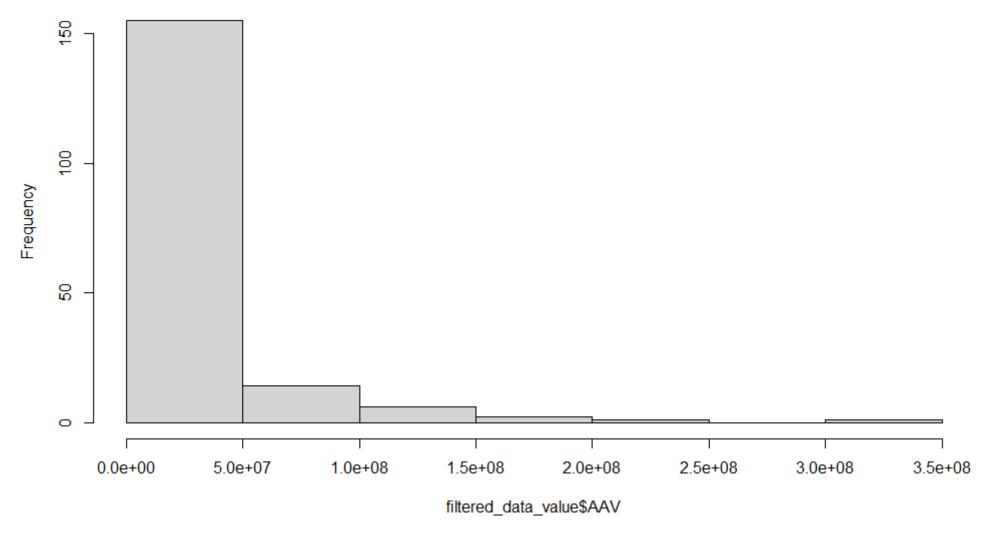
Test accuracy: 0.8

Precision: 0.838

Recall: 0.721

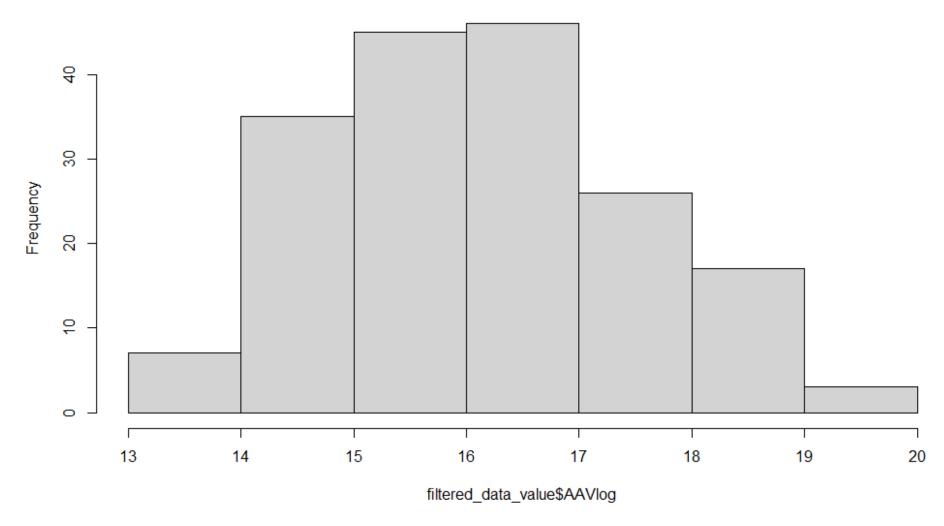
3. 몸값 예측 모델





3. 몸값 예측 모델

AAV(총계약금액/계약 년 수) 히스토그램(로그변환 후)

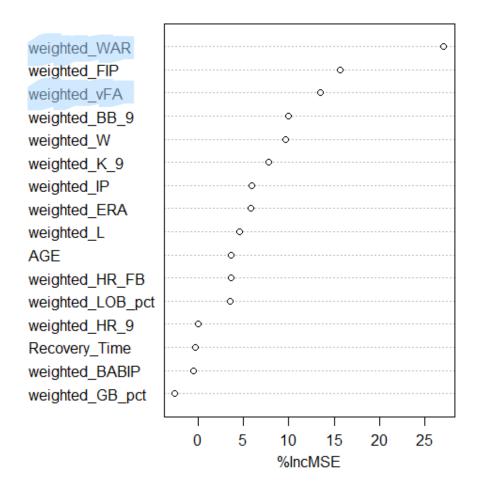


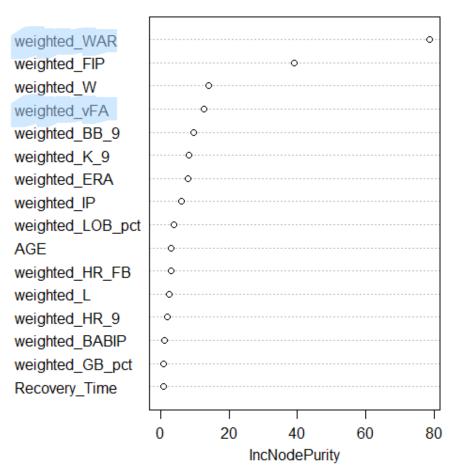
3. 몸값 예측 모델

모델별 성능지표 순위

Random Forest > LightGBM > XGBoost

Feature Importance





성능지표

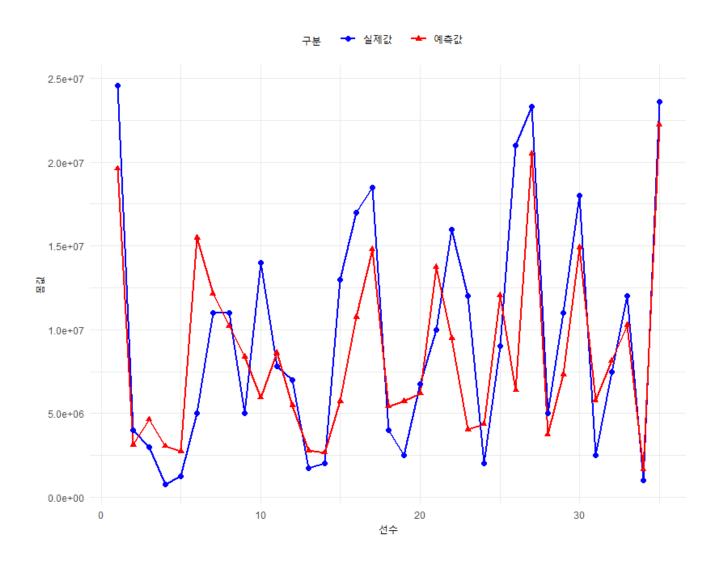
MAPE: 4.56%

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \left| \frac{Y_i - \widehat{Y}_i}{Y_i} \right|$$

예측값과 실제값이 평균적으로 4.56% 차이

3. 몸값 예측 모델

실제값 예측값 비교



실제값 예측값 비교 결과

두 그래프가 비슷한 양상을 띄는 것을 확인

예측결과확인

Part 4 모델적용결과

Category 1: 1년 계약 Category 2: 2~4년 계약 Category 3: 5년 이상 계약

Player ‡	success_pred ‡	predicted_years ‡	predicted_AAV ‡
Adam Kloffenstein	Failure	NA	NA
Adrian Houser	Failure	NA	NA
Alex Cobb	Success	category1	15896710
Alex Wood	Failure	NA	NA
Andrew Heaney	Success	category2	14107423
Anthony DeSclafani	Failure	NA	NA
Blake Snell	Success	category2	39751158
Brad Keller	Failure	NA	NA
Cal Quantrill	Success	category1	5544951
Carlos Carrasco	Failure	NA	NA
Charlie Morton	Success	category2	13765837
Chase Anderson	Failure	NA	NA
Chris Flexen	Failure	NA	NA
Clayton Kershaw	Success	category2	23906020
Colin Rea	Success	category1	4510087
Corbin Burnes	Success	category3	58159955
Dallas Keuchel	Failure	NA	NA
Domingo Germán	Failure	NA	NA

Result

총 57명의 24-25시즌 FA시장 선수 중 30명 계약 성공 예측

Part 4 모델적용결과

현재까지 나온 실제 결과와 비교

Category 1: 1년 계약 Category 2: 2~4년 계약 Category 3: 5년 이상 계약

실제 계약 결과

				18	\$404,175,000
Blake Snell	\$F	🧀 LAD	5	\$182,000,000	\$36,400,000
Yusei Kikuchi	OU	A LAA	3	\$63,675,000	\$21,225,000
Shane Bieber	:LE	C CLE	2	\$26,000,000	\$13,000,000
Frankie Montas	4IL	👛 NYM	2	\$34,000,000	\$17,000,000

모델 예측 결과

Player ‡	success_pred	‡	predicted_years	‡	predicted_AAV ‡	
Blake Snell	Success ()		category2	X	39,751,158	0
Yusei Kikuchi	Success O		category2	0	24,558,024	0
Shane Bieber	Success O		category1	X	23,679,918	X
Frankie Montas	Success O		category2	0	7,435,291	X

- 예측 성공
- 예측 실패

한계점 및 느낀점

Part 5 Limitation

- 예측 향상을 위한 세부 stats 설정에 한계가 있었음. Ex. 좌투수/우투수 등의 추가 변수 포함 가능성
- 데이터 한계: 3개년 데이터로 한정됨
- 3년치 데이터에 대해 차등적인 가중치 (직전 시즌: 0.5, 2년 전: 0.3, 3년 전: 0.2)가 자의적으로 설정됨. 실제로는 전후반기에 따라서도 weight이 다르게 설정됨
- 의도적으로 FA시기에 계약을 하지 않는 선수도 있기 때문에 해당 경우 고려 못 함.
- 3개년 이전의 부상이력은 포함하지 않음.
- 시장 상황 반영 미진: 20-21 FA 당시 선수노조 파업 등 FA 계약 지연 이슈 반영 못 함.

Part 5 후기

- 실제 관심분야 관련 주제를 진행하여 매우 흥미로웠음.
- 추후 Prediction model 생성 시 일부 변수 추가하여 예측 성능지표를 높이고자 함.
- 데이터 셋이 비교적 작은 것에 대한 아쉬움과 더불어 실제 에이전시의 선수 value 설정 기준에 관심도가 더 높아짐.

