

SSG 랜더스 외국인 선수 영입 전략

데이터마이닝 기반 외국인 타자 및 선발투수 후보 선정 보고서

Data: KBO/STATIZ 2023-2026 | MLB/MiLB 2023-2026 | Candidate pool
Hitter 736 / Pitcher 1,009

고려대학교 SDA

June 23, 2026

초록 (Abstract)

국문

본 보고서는 SSG 랜더스의 외국인 타자 및 외국인 선발투수 영입 문제를 하나의 데이터마이닝 의사결정 문제로 재구성한다. 목표는 성적이 좋은 선수를 단순 나열하는 것이 아니라, SSG의 반복 약점을 정량적으로 정의하고, 그 약점을 완화하면서 KBO 번역 가능성과 시장 접근성을 동시에 갖춘 후보를 찾는 것이다. 최종 검증 단계에서는 모델 지지 신호, 시장 접근성, 계약/비용 조건, 의료 확인 필요성을 함께 반영하는 통합 모델로 재구성했다. KBO/STATIZ, MLB Savant, MiLB 성적, roster/transaction, public salary/contract signal, medical status, 기존 모델 산출 테이블을 통합했고, Team Need Mining, Historical KBO Success/Failure, Similarity/Archetype, KBO Translation, 시장 접근성 모형, 의료/실행 가능성 검증 모형을 앙상블했다. 검증 결과 타자 최종 후보 3인은 Will Brennan, Dominic Fletcher, Dylan Carlson으로 수정했고, Luis Matos는 초기 발견 모델에서 강한 후보였지만 age 24와 MLB 재도전 가치 때문에 접촉 보류로 재분류했다. 투수 최종 후보 3인은 Josh Fleming, Keegan Thompson, Kolby Allard이며, 최종 접촉 1순위는 Will Brennan과 Josh Fleming이다.

English

This report reframes SSG Landers' foreign hitter and starting pitcher acquisition as a structured data-mining decision problem. Rather than ranking surface statistics, the analysis identifies SSG's repeated game-state weaknesses, converts them into player-level feature requirements, and selects candidates with both KBO translation potential and practical market access. In the final validation layer, the decision rule combines model support, market access, contract/cost constraints, and medical checks before contact priority is assigned. The revised hitter board is Will Brennan, Dominic Fletcher, and Dylan Carlson, while Luis Matos is moved to an early-screen hold because his age and remaining MLB reset value create player-side and club-rights uncertainty. The pitcher board is Josh Fleming, Keegan Thompson, and Kolby Allard. The recommended first contacts are Will Brennan and Josh Fleming.

Key Takeaways

- SSG의 보강 문제는 단순 장타 부족이 아니라 RHP game-script, run-kill avoidance, extra-out resilience가 결합된 game-state interaction 문제다.
- 후보 선별은 전체 시장에서 시작해 후보 생성 모듈, 데이터마이닝 1차 검증, 현실성 검증 단계, 최종 접촉 순위로 좁히는 funnel로 설계했다.
- 타자는 현실성 제외 조건을 강화해 Will Brennan, Dominic Fletcher, Dylan Carlson을 최종 접촉 후보군으로 제시한다.
- Matos는 초기 발견 모델에서 강한 후보였지만 age 24와 MLB 재도전 가치 때문에 최종 후보가 아니라 접촉 보류가 맞다.
- Suwinski는 BB%와 barrel upside가 있지만 실패 경고 신호와 K% 위험이 커 최종 Top 3가 아니라 watchlist가 맞다.
- 투수는 supervised classifier만으로 확정 추천하기 어렵기 때문에 command, starter floor, 의료 확인, 시장 접근성 검증을 결합했다.
- 최종 접촉 1순위는 타자 Will Brennan, 투수 Josh Fleming이다.

Contents

초록 (Abstract)	i
1 서론	1
1.1 문제 정의	1
1.2 핵심 지표	1
1.3 분석 질문	1
1.4 의사결정 구조	1
2 데이터 및 전처리	2
2.1 데이터 소스	2
2.2 분석 단위	2
2.3 후보군 생성	2
2.4 Feature Contract	3
3 방법	4
3.1 전체 모델 구조	4
3.2 Team Need Mining Model	5
3.3 Historical KBO Success/Failure Model	5
3.4 Similarity/Archetype Matching Model	5
3.5 KBO Translation Risk Model	5
3.6 시장 접근성과 의료/실행 검증	6
3.7 앙상블 weight	6
3.8 성능 평가	6
4 결과	7
4.1 SSG 보강 문제의 데이터마이닝 결과	7
4.2 타자 후보 현실성 반영 Top 3	7
4.3 타자 후보 실행 보정 Top 3	8
4.4 타자 최종 1인: Will Brennan	8
4.5 타자 후보별 차별점	9
4.6 투수 후보 현실성 반영 Top 3	9
4.7 투수 후보 실행 보정 Top 3	10
4.8 투수 최종 1인: Josh Fleming	11
4.9 투수 후보별 차별점	11
4.10 모델별 지지와 경고	11
4.11 검증 후 후보 재분류	12
5 논의	12
5.1 왜 단순 성적 순위와 다른가	12
5.2 실제 영입 협상에서 확인할 조건	13
5.3 Plan A / Plan B / Plan C	13

6	한계	13
6.1	투수 classifier 안정성이 낮은 이유	13
6.2	그 외 한계와 보완	14
7	결론	14
7.1	분석 질문별 요약 답변	14
7.2	최종 의사결정표	15
7.3	핵심 한 줄	15
A	부록	15
A.1	최종 후보 보드	15

List of Tables

1	주요 야구 지표 요약	1
2	분석 질문	1
3	데이터 구성 요약	2
4	후보 funnel	3
5	SSG 약점과 선수 feature contract	3
6	앙상블 base learner 구성	4
7	모델 누수 및 보정 검증 요약	5
8	모델 weight	6
9	모델 성능 및 사용 강도	7
10	타자 후보 현실성 반영 Top 3	7
11	타자 후보 실행 보정 Top 3	8
12	초기 발견 리드와 최종 접촉 1순위의 분리	9
13	타자 후보 차별점	9
14	투수 후보 현실성 반영 Top 3	10
15	투수 후보 실행 보정 Top 3	10
16	투수 후보 차별점	11
17	모델별 후보 지지와 경고	11
18	검증 후 재분류 결정	12
19	협상 전 확인 조건	13
20	접촉 우선순위	13
21	한계와 보완 방향	14
22	분석 질문별 결론	14
23	최종 후보 3인 및 접촉 1순위	15
24	Final candidate board summary	15

List of Figures

1	SSG weakness mining rule	3
2	타자 후보 최종 접촉 우선순위	8
3	투수 후보 초기 순위와 실행 보정 순위	10

1 서론

1.1 문제 정의

이 보고서의 대상은 “좋은 선수”가 아니라 “SSG가 실제로 데려와야 하는 선수”다. 외국인 선수 영입은 순수 선수 평가와 다르다. 특정 팀의 약점, KBO 번역 가능성, 계약 가능성, 의료 상태, 비용 조건이 동시에 충족되어야 한다. 따라서 본 분석은 후보를 점수화하는 단일 랭킹 문제가 아니라, 여러 조건을 통과하는 후보를 단계적으로 좁히는 데이터마이닝 funnel 문제로 설계했다.

1.2 핵심 지표

야구 지표에 익숙하지 않은 독자도 결과를 해석할 수 있도록 핵심 지표를 표 1에 정리한다.

Table 1: 주요 야구 지표 요약

지표	의미	실무 해석
OBP	출루율	이닝을 이어가는 능력, 상위타순 안정성
SLG/ISO	장타력 및 순장타력	출루를 득점으로 바꾸는 능력
K% / Whiff%	삼진 및 헛스윙 위험	KBO 적응 실패와 run-kill 위험
BB%	볼넷 생산성	공을 고르고 이닝을 늘리는 능력
xwOBA	타구 질 기반 예상 공격력	표면 성적보다 지속 가능성 평가에 유리
BB/9	9이닝당 볼넷	투수의 traffic 생성 위험
HR/9	9이닝당 피홈런	장타 허용 및 대량실점 위험
GS / IP	선발 등판 및 이닝	선발 지속성과 5이닝 floor 평가

1.3 분석 질문

본 보고서는 표 2의 여섯 질문을 따라 전개된다.

Table 2: 분석 질문

RQ	질문	최종 산출물
RQ1	SSG의 핵심 보강 문제는 무엇인가?	SSG weakness rule
RQ2	약점은 어떤 선수 조건으로 변환되는가?	타자/투수 feature contract
RQ3	후보군은 어떻게 생성되고 줄어드는가?	전체 후보군에서 실행 보정 board
RQ4	각 모델은 어떤 관점에서 후보를 평가하는가?	모델 구성표와 성능 평가표
RQ5	타자 및 투수 최종 후보 3명은 누구인가?	slot별 실행 보정 Top 3
RQ6	최종 접촉 1순위는 누구인가?	접촉 우선순위 Plan A

1.4 의사결정 구조

최종 결정은 data-mining ranking과 현실성 반영 접촉 순위를 분리한다. Data-mining ranking은 구조화 지표가 과거 KBO 성공/실패 패턴과 얼마나 닮았는지를 뜻하고, 최종 접촉 순위는 계약, 비용, 의료, 시장 접근성, KBO행 가능성을 함께 적용한 실행 순위다. 이 구조 때문에 초기 발견 후보가 최종 접촉 1순위가 아닐 수 있다. 본 보고서에서 Luis Matos는 타자 발견 단계에서 강한

후보지만 age 24와 MLB 재도전 가치 때문에 최종 접촉 후보군이 아니라 hold로 내려가고, Will Brennan이 현실형 접촉 1순위로 올라간다.

2 데이터 및 전처리

2.1 데이터 소스

분석에 사용한 데이터의 역할은 표 3와 같다. 모델 입력은 정량 structured data로 한정했다.

Table 3: 데이터 구성 요약

데이터 소스	기간	단위	사용 목적
KBO/STATIZ 팀/선수 데이터	2023-2026	팀, 선수, 경기, 상황	SSG 문제 정의와 KBO label 구성
SSG 상황별 경기 데이터	2026	game-state rule	Team Need Mining Model
MLB Savant	2023-2026	pitch/player- season	타자 기초 역량과 KBO translation
MiLB 성적 데이터	2025-2026 중심	player-season/ role	후보 시장 구축과 투수 starter floor
MLB roster/transaction	2025-10 이후	선수 상태	시장 접근성 모형
NPB/CPBL 공개 roster/stat seed	2026	선수/리그	시장 depth와 아시아쿼터 확장성 확인
public salary/ contract/medical	2026-06 확인	선수	시장/의료 실행 검증
기존 output tables	프로젝트 산출물	모델/후보/검증 단계	최종 보고서 재현성과 후보 board 구성

2.2 분석 단위

타자는 선수-시즌 및 최근 plate appearance 단위를 기본으로 하며, 외야 및 지명타자 role, 좌우 타, contact floor, on-base/damage profile을 함께 본다. 투수는 선수-시즌 및 최근 IP/GS 단위를 기본으로 하며, 선발 지속성, command stability, traffic damage control, KBO/ABS translation을 함께 본다.

2.3 후보군 생성

후보 시장은 MLB 40-man 경계선, DFA/outright/minor league contract 선수, AAA regular role 선수, 최근 트랜잭션 후보, KBO행 가능성이 있는 나이와 계약 상태의 선수로 구성했다. 후보 funnel은 표 4와 같다.

Table 4: 후보 funnel

단계	타자 잔존	투수 잔존	판정 의미
Step 1. 전체 구조화 시장	736	1,009	모델 입력 가능한 후보 pool 구성
Step 2. 후보 생성 모듈	16	18	Market, Team Fit, Upside Screen 통과
Step 3. 데이터마이닝 1차 검증	6	3	접근성, 표본, 부상 flag 1차 통과
Step 4. 초기 Top 3	3	3	순수 모델 점수와 반복 등장 신호
Step 5. 실행 보장 Top 3	3	3	계약, 의료, 실행 가능성 반영

2.4 Feature Contract

Team Need Mining의 핵심은 팀 약점을 선수 조건으로 번역하는 것이다. 표 5는 SSG 약점과 선수 feature의 연결이다.

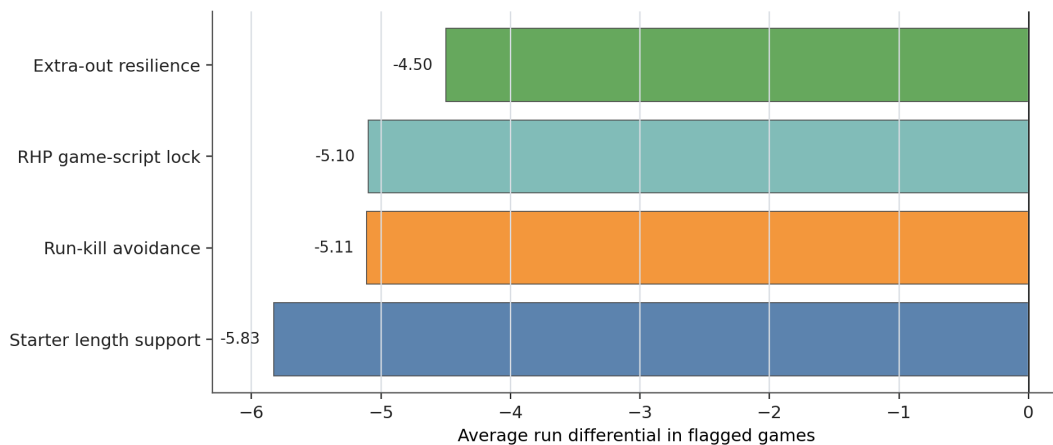


Figure 1: SSG weakness mining rule

Table 5: SSG 약점과 선수 feature contract

슬롯	SSG weakness	Player feature	해석
타자	RHP game-script unlocker	vs RHP on-base damage	우투 선발 경기에서 외야/DH가 이닝을 다시 여는 능력
타자	run-kill avoidance	low GDP and low chase risk	병살, 삼진, 약한 타구로 이닝을 끊지 않는 능력
타자	two-strike survival	two-strike contact floor	불리한 카운트에서 최소 생산성을 유지하는 능력

슬롯	SSG weakness	Player feature	해석
타자	role stability	corner OF or DH continuity	외야/DH 슬롯에서 바로 쓸 수 있는 역할 안정성
투수	traffic-command stabilizer	low free-pass volatility	추가 출루 이후 볼넷으로 이닝을 키우지 않는 능력
투수	starter length floor	five-inning floor	5이닝 이상 버틸 수 있는 최소 근거
투수	extra-out resilience	damage control after traffic	수비 실수나 추가 출루 이후 대량실점을 막는 능력
투수	ABS adaptation	zone command	ABS 환경에서 볼넷 리스크가 커지지 않는 능력

3 방법

3.1 전체 모델 구조

본 보고서는 하나의 최종 앙상블 모델을 사용한다. 다만 내부 base learner는 서로 다른 질문을 담당한다. 표 6은 각 모델의 역할이다.

Table 6: 앙상블 base learner 구성

모델명	목적	출력	사용 강도
Model A. Team Need Mining	SSG 약점을 game-state interaction으로 정의	weakness rule, feature contract	공통 기준
Model B. Historical KBO Success/Failure	입단 전 정보로 성공/실패 패턴 유사도 추정	모델 지지/실패 경고 점수	타자 후보 비교용, 투수 진단
Model C. Similarity/Archetype Matching	KBO 성공 유형 또는 SSG 필요 유형과 유사도 계산	archetype score	보조
Model D. KBO Translation Risk	MLB/MiLB 성과가 KBO에서 유지되지 않을 위험 평가	PASS/YELLOW/HOLD/RED	강함
Model E. 시장 접근성 모형	실제 영입 가능성 평가	available/conditional/blocked	제외 조건
Model F. 의료/실행 검증 모형	점수 상위 후보의 실행 불가능성 보류	PASS/YELLOW/HOLD/RED	제외 조건

3.2 Team Need Mining Model

Team Need Mining Model은 선수를 직접 고르는 모델이 아니라, 후보에게 요구되는 조건을 정의하는 모델이다. SSG의 공격 약점은 단순 외야 홈런 부족보다 RHP game-script, run-kill avoidance, extra-out resilience에서 더 구체적으로 나타난다. 투수 약점은 구위 총량보다 traffic 이후 볼넷과 장타 억제, 5이닝 이상을 버티는 선발 안정성에서 나타난다.

3.3 Historical KBO Success/Failure Model

이 모델은 과거 KBO 외국인 선수의 입단 전 데이터를 바탕으로 성공/실패 패턴 유사도를 추정한다. 타자 모델은 pre-KBO Savant feature를 사용한 Ridge Logistic Regression 계열로 구성했고, repeated stratified CV에서는 성공 AUC 0.833, 실패 AUC 0.738을 기록했다. 다만 표본이 22행으로 작고 같은 선수가 여러 시즌 반복되므로, 선수 단위 그룹 검증에서는 성공 AUC가 0.650까지 낮아졌다. 따라서 개별 선수 값은 실제 계약 성공확률이 아니라 후보 비교를 위한 모델 지지 점수와 실패 경고 점수로만 해석한다. 투수 모델은 pre-KBO MiLB feature를 기반으로 한 sparse logistic diagnostic model로 구성했으나 AUC 0.603에 그쳐 최종 추천 모델이 아니라 진단 신호로 제한했다.

Table 7: 모델 누수 및 보정 검증 요약

검증 항목	결과	해석
label/outcome 피쳐 누수	발견 없음	success, failure, first-KBO 결과, renewal, player id 계열은 학습 피쳐에서 제외
후보 입력 누수	발견 없음	후보 입력은 recent Savant/MiLB 구조화 지표로 매핑
train-candidate ID overlap	2명 발견, 최종 후보와는 무관	Jared Young, Patrick Wisdom 이 후보풀에 남았으나 최종 board에는 없음
선수 단위 그룹 검증	success AUC 0.650, failure AUC 0.500	full-fit 확률값을 절대 성공확률처럼 쓰면 과신 위험이 큼
보고서 처리	확률 표기 제거	모델 지지 점수, 실패 경고 점수, 접촉 순위로만 사용

3.4 Similarity/Archetype Matching Model

Similarity/Archetype Matching은 성적 총량이 아니라 역할 유사성을 본다. 예를 들어 Matos는 단순 파워보다 contact floor와 외야/DH role 안정성에서 SSG 필요 유형과 닮았고, Fleming은 에이스형 구위보다 traffic-command starter 유형에 가깝다.

3.5 KBO Translation Risk Model

KBO Translation Risk는 MLB/MiLB 성과가 KBO에서 유지되지 않을 위험을 평가한다. 타자는 K%, chase, whiff, 변화구 대응, run-kill risk를 본다. 투수는 BB/9, HR/9, starter continuity, pitch-quality 확인 필요성을 본다.

3.6 시장 접근성과 의료/실행 검증

시장 접근성 모형은 40-man 상태, DFA/outright, minor contract, salary signal, option status를 평가한다. 의료/실행 검증 모형은 초기 모델 점수가 높아도 즉시 전력 가능성이 낮으면 최종 접촉 후보군에서 보류한다. Carson Spiers와 Brian Van Belle이 그 사례다.

3.7 앙상블 weight

타자와 투수는 모델 신뢰도가 다르므로 weight도 다르게 둔다. 타자는 supervised success/failure model의 성능이 비교적 강해 큰 weight를 주고, 투수는 classifier 안정성이 낮아 cross-model consensus와 실행 검증을 더 강하게 반영한다.

Table 8: 모델 weight

슬롯	모델	가중치	근거
타자	historical success/failure classifier	0.40	타자 AUC가 높아 주요 base learner로 사용
타자	SSG fit translation pipeline	0.25	SSG 2026 상황별 약점과 KBO 번역 반영
타자	KBO adaptation filter	0.15	contact, chase, 변화구 대응 위험 반영
타자	market inefficiency feature model	0.15	AAA 장점과 MLB 결함의 translation gap 탐색
타자	cross-model consensus	0.05	여러 base learner 반복 등장 보너스
투수	historical success/failure classifier	0.05	AUC 0.603으로 진단 신호로만 사용
투수	SSG fit translation pipeline	0.25	선발 이닝, damage control, KBO 번역 반영
투수	KBO adaptation filter	0.25	BB/9, HR/9, starter continuity 반영
투수	market inefficiency feature model	0.20	fringe starter 시장에서 command와 이닝 가치 탐색
투수	cross-model consensus	0.25	투수 supervised model 약점을 여러 독립 신호로 보완

3.8 성능 평가

모델별 성능과 사용 강도는 표 9와 같다.

Table 9: 모델 성능 및 사용 강도

모델	대상	표본 수	주요 성능	사용 강도
hitter success classifier	타자	22	success AUC 0.833	High
hitter failure classifier	타자	22	failure AUC 0.738	High
pitcher success diagnostic	투수	49	AUC 0.603	Low/ Diagnostic
pitcher warning model	투수	49	AUC 0.603	Low/ Diagnostic

4 결과

4.1 SSG 보강 문제의 데이터마이닝 결과

SSG의 타자 보강 문제는 장타 총량이 아니라 우투 선발 경기에서 외야/DH가 이닝을 다시 열지 못하는 구조다. 따라서 후보는 단순 장타형보다 contact floor, two-strike survival, run-kill avoidance를 동시에 충족해야 한다. 투수 보강 문제는 strikeout ceiling보다 traffic 이후 볼넷과 장타를 억제하고 5이닝 이상을 버티는 선발 안정성이다.

4.2 타자 후보 현실성 반영 Top 3

타자 data-mining ranking은 과거 KBO 외국인 타자 성공/실패 패턴을 학습한 모델의 지지 신호와 실패 경고 신호를 기준으로 계산했다. 이 값은 calibration된 절대 확률이 아니라 후보 간 우선순위와 위험 방향을 비교하기 위한 모델 산출값이다. 최종 단계에서는 여기에 현실성 검증 단계를 추가했다. 즉, 40-man 또는 최근 cash trade/claim/selected contract로 구단 통제력이 강한 선수, age 25 이하로 MLB 재도전 가치가 큰 선수, OF/DH 역할과 맞지 않는 선수는 모델 신호가 좋아도 최종 접촉 후보군에서 제외했다.

Table 10: 타자 후보 현실성 반영 Top 3

접촉 순위	선수	모델 지지	실패 경고	표본 상태	주요 지지 모듈
1	Will Brennan	강함	낮음	소표본 보정	Recent DFA/outright, Contact Floor, OF
2	Dominic Fletcher	강함	낮음	소표본 보정	Minor contract, OF, independent signal
3	Dylan Carlson	중상	보통	충분	Historical, Switch-hit, OF Depth

4.3 타자 후보 실행 보정 Top 3

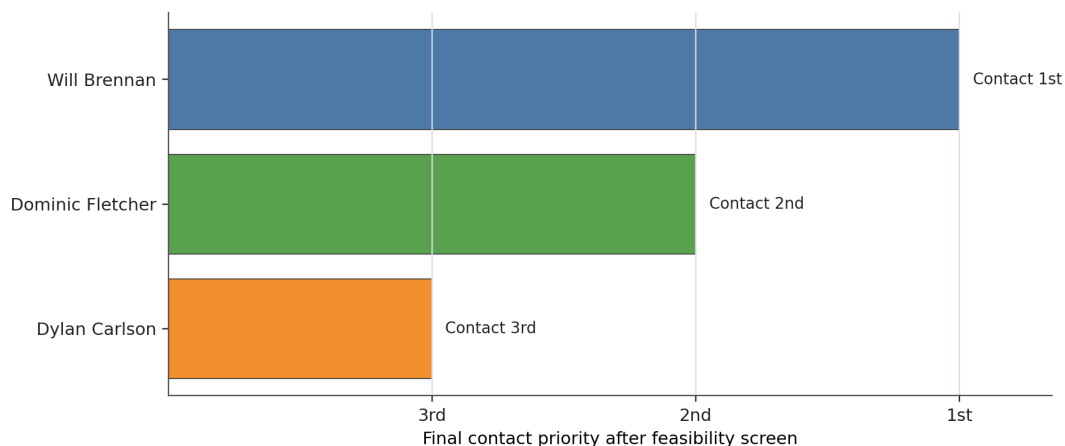


Figure 2: 타자 후보 최종 접촉 우선순위

Table 11: 타자 후보 실행 보정 Top 3

접촉 순위	선수	최종 판단	팀 필요 적합성	위험 요약
1	Will Brennan	CONTACT 1ST	DFA/outright 이후 40-man 밖으로 내려온 contact-floor OF	medical history watch는 확인 필요하나, Matos 보다 KBO 행 현실성이 높음
2	Dominic Fletcher	CONTACT 2ND	non-40man minor contract OF, 독립 후보 신호와 모델이 교차	장타 ceiling은 낮아 중심타선 해결사보다 run-kill avoidance 역할
3	Dylan Carlson	CONTACT 3RD	스위치히터 외야수로 platoon 및 OF depth 안정화에 적합	최근 소속권과 opt-out/release 조건 확인 필요

4.4 타자 최종 1인: Will Brennan

타자 부문에서 초기 발견 신호만 보면 Matos와 Nolan Jones도 강하게 보인다. 그러나 현실성 검증 단계를 앞에 두면 최종 접촉 1순위는 Will Brennan으로 바뀐다. Brennan은 28세 외야수이고, 2026년 6월 10일 DFA, 6월 17일 outright 이후 40-man 밖으로 내려온 고접근성 후보다. 통합 후보 보드에서도 모델 지지 신호가 강하고 실패 경고 신호가 낮게 나타나며, SSG가 원하는 run-kill avoidance와 contact-floor OF 조건에 잘 맞는다.

Matos를 내린 이유는 명확하다. Matos는 데이터마이닝 발견 단계에서 강하지만 24세이고, 이미 현금 트레이드로 한 차례 다른 MLB 구단이 권리를 산 이력이 있다. 이는 단순히 성적이

낮아 방출된 베테랑과 다르다. 선수와 구단 모두 MLB 재도전 옵션을 더 선호할 수 있으므로, KBO 대체 외국인 영입의 현실성 기준에서는 최종 Top 3가 아니라 접촉 보류가 맞다.

Table 12: 초기 발견 리드와 최종 접촉 1순위의 분리

항목	초기 발견 신호	현실성 검증 적용 후
Matos	초기 발견 리드, low-K contact floor	age 24, MLB reset value 때문에 HOLD
Nolan Jones	power/OBP 발견 신호	recent cash trade와 salary signal 때문에 HOLD
Will Brennan	모델 점수와 contact floor 모두 강함	recent DFA/outright로 CONTACT 1ST
Dominic Fletcher	모델 점수와 독립 후보 신호가 교차	non-40man minor contract로 CONTACT 2ND

4.5 타자 후보별 차별점

Table 13: 타자 후보 차별점

선수	강점	반대 논리	최종 판단
Will Brennan	recent DFA/outright, contact floor, 낮은 실패 경고	medical history watch 확인 필요	CONTACT 1ST
Dominic Fletcher	non-40man minor contract, OF 안정성, 독립 후보 신호와 교차	장타 ceiling은 높지 않음	CONTACT 2ND
Dylan Carlson	모델 Top 3를 통과한 스위치히터 안정형 OF	최근 소속권/release 조건 확인 필요	CONTACT 3RD
Luis Matos	모델상 contact floor와 강한 발견 신호	24세 MLB 재도전 가치가 큼	접촉 보류
Nolan Jones	출루와 장타의 결합, 강한 hard-hit signal	cash trade 및 salary/cost-share 확인 필요	COST HOLD
Jack Suwinski	BB%와 barrel 기반 upside	실패 경고와 K% 위험으로 model warning	WATCHLIST

4.6 투수 후보 현실성 반영 Top 3

투수는 supervised classifier 안정성이 낮기 때문에, 초기 모델 점수보다 현재 계약 상태와 선발 지속성을 먼저 본다. Bryse Wilson은 6월 18일 Phillies가 contract를 selected한 이후 현실성 반영 단계에서 hold로 내려간다.

Table 14: 투수 후보 현실성 반영 Top 3

현실 순위	선수	Score	주요 지지 모듈
1	Josh Fleming	80.9	non-40man, AAA starter load, BB/9 1.36, HR/9 0.51
2	Keegan Thompson	76.0	recent DFA/outright, starter load, HR/9 0.56
3	Kolby Allard	75.5	LHP, repeated DFA/outright/minor deal, HR/9 0.36

4.7 투수 후보 실행 보정 Top 3

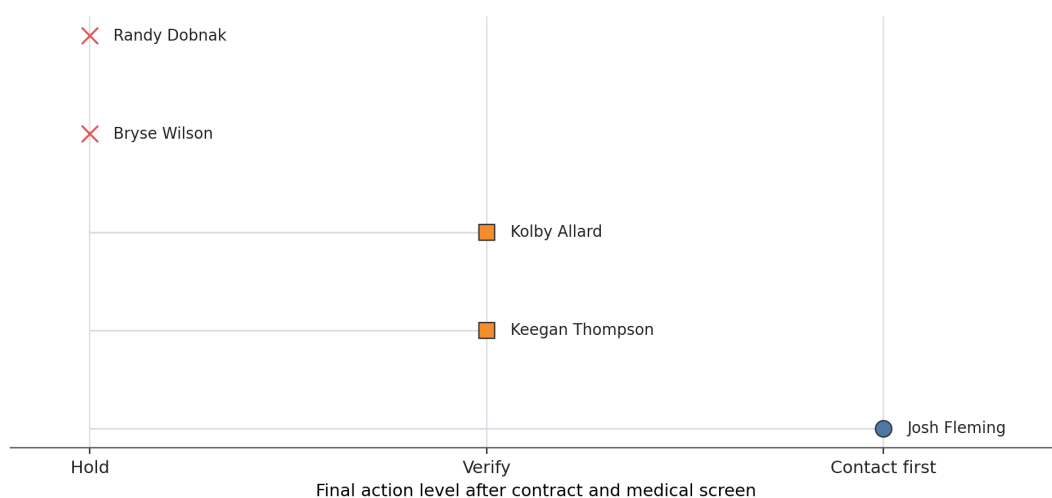


Figure 3: 투수 후보 초기 순위와 실행 보정 순위

Table 15: 투수 후보 실행 보정 Top 3

접촉 순위	선수	최종 판단	팀 필요 적합성	위험 요약
1	Josh Fleming	CONTACT 1ST	좌완 선발/스윙맨으로 traffic-command starter 조건에 가장 근접	에이스형 구위가 아니라 안정화형 선발로 정의해야 함
2	Keegan Thompson	CONTACT 2ND	recent DFA/outright high-access, starter load, HR 억제 신호	K/9 5.29라 swing-and-miss ceiling은 낮음
3	Kolby Allard	CONTACT 3RD	좌완, repeated DFA/outright, 낮은 HR/9	BB/9 3.20과 최근 소속권 확인 필요

4.8 투수 최종 1인: Josh Fleming

투수 부문은 타자보다 supervised classifier의 안정성이 낮기 때문에 초기 모델 점수만으로 결론을 내리지 않았다. command, starter floor, 의료/실행 검증, 계약 접근성을 함께 반영했고, 이 과정을 통과한 최종 접촉 1순위는 Josh Fleming이다. Fleming은 좌완 선발/스윙맨 후보이며, 낮은 HR/9, 관리 가능한 BB/9, minor contract 접근성이 결합된다. SSG 내 역할은 traffic-command starter, swingman/spot starter, HR/BB damage controller다.

Fleming을 에이스형 구위 후보로 포장하면 위험하다. 이 후보의 핵심 가치는 압도적 탈삼진이 아니라 traffic 이후 경기 붕괴를 줄이는 안정화 기능이다. 따라서 협상 전에는 최근 선발 빌드업, 구속, 구종 품질, availability를 확인해야 한다.

4.9 투수 후보별 차별점

Table 16: 투수 후보 차별점

선수	강점	반대 논리	최종 판단
Josh Fleming	좌완성, damage control, 시장 접근성	K/9 upside 제한	CONTACT 1ST
Keegan Thompson	recent DFA/outright, HR 억제, starter load	K/9 5.29로 swing-miss 부족	CONTACT 2ND
Kolby Allard	좌완, 낮은 HR/9, repeated minor deal	BB/9 3.20과 최근 소속권 확인 필요	CONTACT 3RD
Bruce Zimmermann	K/9 9.98, 67.2 IP/13 GS	HR/9 1.86이 KBO 장타 억제 조건과 충돌	WATCH
Bryse Wilson	선발 경력과 모델 진단 신호	6/18 selected contract 이후 contract blocker	HOLD

4.10 모델별 지지와 경고

Table 17: 모델별 후보 지지와 경고

모델	지지 후보	경고 후보	이유
Team Need Mining	Brennan, Fletcher, Fleming	Suwinski	run-kill avoidance와 traffic-command 조건 충족 여부
Historical Success/Failure	Brennan, Fletcher, Matos	Suwinski	타자 success/failure ranking score 차이
KBO Translation Risk	Brennan, Carlson, Fleming	Zimmermann, Suwinski	HR/9 및 K% translation risk
시장 접근성 모형	Brennan, Thompson, Fleming	Matos, Jones, Wilson	age, cash trade, selected contract 등 현실성 차이
의료/실행 검증 모형	Brennan, Fletcher, Fleming	Spiers, Van Belle	medical hold 여부

4.11 검증 후 후보 재분류

최종 보드 검증에서 가장 중요한 변화는 현실성 검증 단계의 강화다. Matos는 모델상 강하지만 age 24와 MLB 재도전 가치 때문에 최종 접촉 후보군에서 제외했다. Nolan Jones는 모델과 독립 후보 신호가 강하지만 6월 11일 cash trade와 salary signal 때문에 비용/권리 hold로 내렸다. 반대로 Brennan은 6월 17일 outright 이후 40-man 밖으로 내려온 시장 접근성 신호와 모델 지지 신호를 동시에 갖고 있어 최종 접촉 1순위로 올라간다.

Table 18: 검증 후 재분류 결정

선수	기존 위치	검증 후 위치	결정 근거
Will Brennan	model/market candidate	최종 타자 접촉 1순위	6/17 outright, OF role fit, 낮은 실패 경고
Luis Matos	발견 단계 lead	접촉 보류	age 24와 MLB 재도전 가치 때문에 즉시 KBO행 현실성이 낮음
Nolan Jones	power/OBP 발견 신호	cost/rights hold	6/11 cash trade와 salary signal 때문에 대체 외국인 비용 구조 위험
Bryse Wilson	투수 backup	contract blocker hold	6/18 selected contract 이후 current 40-man/MLB access blocker

5 논의

5.1 왜 단순 성적 순위와 다른가

단순 성적 순위는 KBO 번역 위험, 계약 접근성, 의료 상태를 반영하지 않는다. 본 보고서는 data-mining score와 최종 접촉 순위를 분리했기 때문에, 모델 신호는 좋지만 실제 영입 가능성이 낮은 선수를 설명할 수 있다. Matos는 초기 발견 단계에서 강하지만 age 24와 MLB 재도전 가치 때문에 hold로 내려간다. Nolan Jones는 강한 파워/출루 신호가 있지만 cash trade와 salary signal 때문에 비용/권리 hold다. Bryse Wilson은 투수 진단 후보였지만 selected contract 이후 contract blocker로 내려간다.

5.2 실제 영입 협상에서 확인할 조건

Table 19: 협상 전 확인 조건

조건	대상	의사결정 영향
최근 outright/release 권리	Brennan, Fletcher, Carlson	즉시 접촉 가능성 및 KBO행 수용 여부
잔여 보장액과 buyout	Jones, Suwinski	정규 외국인 계약 또는 대체 외국인 비용 구조 충족 여부
young MLB reset value	Matos, Lugo	KBO행보다 MLB 재도전을 택할 위험
최근 구속/구종 품질	Fleming, Thompson, Allard	KBO starter floor 확인

5.3 Plan A / Plan B / Plan C

Table 20: 접촉 우선순위

슬롯	Plan A	Plan B	Plan C	조건
타자	Will Brennan	Dominic Fletcher	Dylan Carlson	Brennan medical/history 확인 후 Fletcher와 병렬 접촉
투수	Josh Fleming	Keegan Thompson	Kolby Allard	Fleming 선발 빌드업 부족 시 Thompson/Allard 비교

6 한계

6.1 투수 classifier 안정성이 낮은 이유

투수 historical classifier의 안정성이 낮은 가장 큰 이유는 표본 수와 label noise다. 현재 투수 학습 표본은 49명 수준이며, KBO 입단 후 투수 성과는 타자보다 주변 환경의 영향을 더 크게 받는다. ERA와 이닝은 투수 본인의 구위뿐 아니라 수비력, 포수 리드, 불펜 승계주자 처리, 구장, 등판 간격, 시즌 중 역할 변경의 영향을 받는다. 따라서 같은 입단 전 BB/9와 HR/9를 가진 투수라도 KBO에서 받은 역할과 팀 환경에 따라 label이 달라질 수 있다.

두 번째 이유는 역할 이질성이다. 외국인 투수 표본 안에는 확정 선발, 스윙맨, 대체 선발, 불펜 전환 후보가 함께 존재한다. 이들은 모두 투수로 묶이지만 성공 기준은 다르다. 선발투수에게 중요한 starter floor와 third-time-through-order risk는 불펜 후보에게는 같은 의미가 아니다. 이 때문에 투수 classifier는 최종 추천 모델이 아니라 후보의 위험 신호를 읽는 진단 모델로 제한한다.

세 번째 이유는 pitch-quality 결측이다. KBO 번역에서 중요한 구속 추세, pitch shape, release consistency, zone command, ABS 환경에서의 called-strike dependency는 공개 집계 성적만으로 완전히 설명되지 않는다. 그래서 투수 최종 판단은 모델 산출값보다 BB/9, HR/9, 최근 선발 이닝, 의료 확인, 계약 접근성, 최근 구속/구종 확인을 함께 보는 방식으로 처리한다.

6.2 그 외 한계와 보완

Table 21: 한계와 보완 방향

한계	영향	보완
투수 KBO 외국인 표본 수	연도, 리그, 역할 변화에 민감	AUC만 보지 않고 진단 신호로 제한
투수 성과 label noise	ERA/IP가 주변 환경 영향을 받음	BB/9, HR/9, starter floor를 함께 사용
투수 역할 이질성	성공 기준이 흔들림	역할 적합성 검증을 별도 적용
pitch-quality 결측	공개 집계 성적만으로 설명 부족	구속/구종 확인을 협상 전 체크리스트로 둠
모델 보정	표본 22명 기반 점수를 절대 확률로 과장할 위험	점수보다 순위, 방향성, 검증 통과 여부를 함께 제시
salary/contract 공개 정보	잔여 부담액과 buyout이 불확실	접촉 전 agent/구단 확인
medical public signal	full medical report 부재	medical hold를 강한 제외 조건으로 운용
2026 시즌 중 데이터	SSG weakness rule이 갱신될 수 있음	주 단위 재빌드

7 결론

7.1 분석 질문별 요약 답변

Table 22: 분석 질문별 결론

RQ	요약 답변
RQ1	SSG의 약점은 단순 장타 부족이 아니라 game-state interaction 문제다.
RQ2	타자는 이닝 전환형 OF/DH, 투수는 traffic-command starter가 필요하다.
RQ3	후보는 전체 시장, 후보 생성 모듈, 데이터마이닝 1차 검증, 현실성 검증, 최종 Top 3로 축소했다.
RQ4	각 모델은 팀 필요, historical success, archetype similarity, KBO translation, 시장 접근성, 의료/실행 검증을 분담한다.
RQ5	타자 Top 3는 Brennan/Fletcher/Carlson, 투수 Top 3는 Fleming/Thompson/Allard다. Matos는 접촉 보류다.
RQ6	최종 접촉 1순위는 Will Brennan과 Josh Fleming이다.

7.2 최종 의사결정표

Table 23: 최종 후보 3인 및 접촉 1순위

슬롯	초기 발견 후보	최종 접촉 1위	최종 후보 3명	접촉 1순위
외국인 타자	Luis Matos	Will Brennan	Will Brennan, Dominic Fletcher, Dylan Carlson	Will Brennan
외국인 선발투수	Josh Fleming	Josh Fleming	Josh Fleming, Keegan Thompson, Kolby Allard	Josh Fleming

7.3 핵심 한 줄

본 보고서의 결론은 Will Brennan과 Josh Fleming이 단순히 가장 좋은 성적의 선수가 아니라, SSG의 약점 구조, KBO 번역 가능성, 시장 접근성, 비용 및 의료 검증을 함께 통과한 현실형 접촉 우선순위라는 점이다. Matos처럼 모델 점수가 높은 후보도 현실성 검증을 통과하지 못하면 최종 접촉 1순위에서 제외한다.

A 부록

A.1 최종 후보 보드

표 24는 검증 후 최종 후보 보드의 축약본이다. 전체 후보 보드 원자료는 별도 CSV 산출물로 함께 보관했다.

Table 24: Final candidate board summary

슬롯	선수	초기	접촉	최종 판단	최종 이유
foreign hitter	Will Brennan	1	1	CONTACT 1ST	6/17 outright, high market access, contact-floor OF
foreign hitter	Dominic Fletcher	2	2	CONTACT 2ND	non-40man minor contract, 독립 후보 신호와 모델 교차
foreign hitter	Dylan Carlson	3	3	CONTACT 3RD	switch-hitting OF, 모델 Top 3 통과
foreign hitter	Luis Matos	D	HOLD	발견 단계 보류	초기 발견 리드였지만 age 24와 MLB reset value 때문에 hold
foreign hitter	Nolan Jones	2	HOLD	COST HOLD	cash trade와 salary signal 때문에 비용/권리 확인 필요
foreign pitcher	Josh Fleming	1	1	CONTACT 1ST	53.0 IP/10 GS, BB/9 1.36, HR/9 0.51
foreign pitcher	Keegan Thompson	2	2	CONTACT 2ND	recent DFA/outright, HR/9 0.56
foreign pitcher	Kolby Allard	3	3	CONTACT 3RD	좌완, repeated minor deal, HR/9 0.36
foreign pitcher	Bryse Wilson	15	HOLD	CONTRACT HOLD	6/18 selected contract 이후 active MLB access blocker

References

- [1] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide.
- [2] MLB Advanced Media. Baseball Savant Statcast Search and Player Pages.
- [3] MLB Advanced Media. MLB.com player transaction logs and team transaction pages, 2025-2026.
- [4] MLB Trade Rumors and ESPN transaction pages. Minor-league contract, outright, and selected-contract reports, 2025-2026.
- [5] STATIZ. KBO team, player, and situational statistics.
- [6] Korea Baseball Organization. KBO league registration and foreign player regulations.