

3.pdf

1.

KBO

'KBO

, KBO

2.

2024 2025 KBO

11

(Feature)

1

4,500 5,500

3.KBO

3.1.

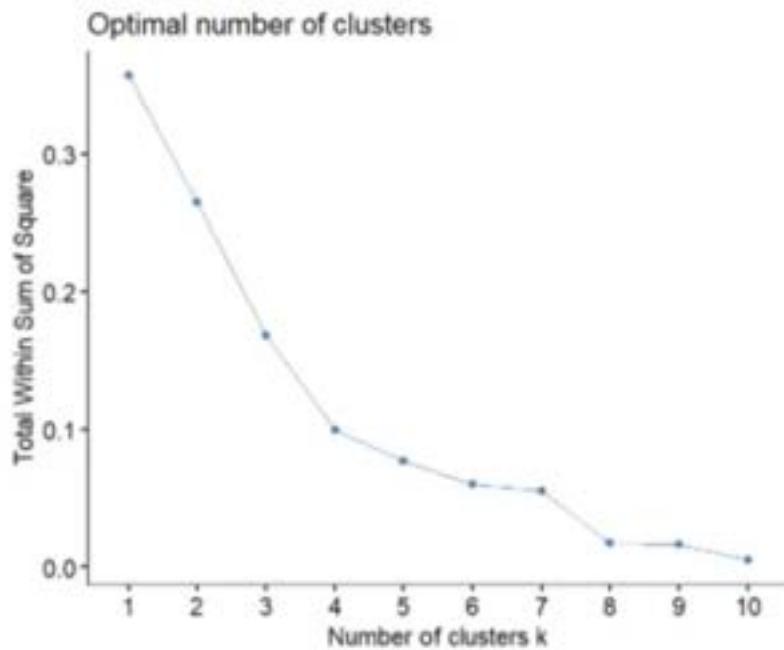
k

k=4,

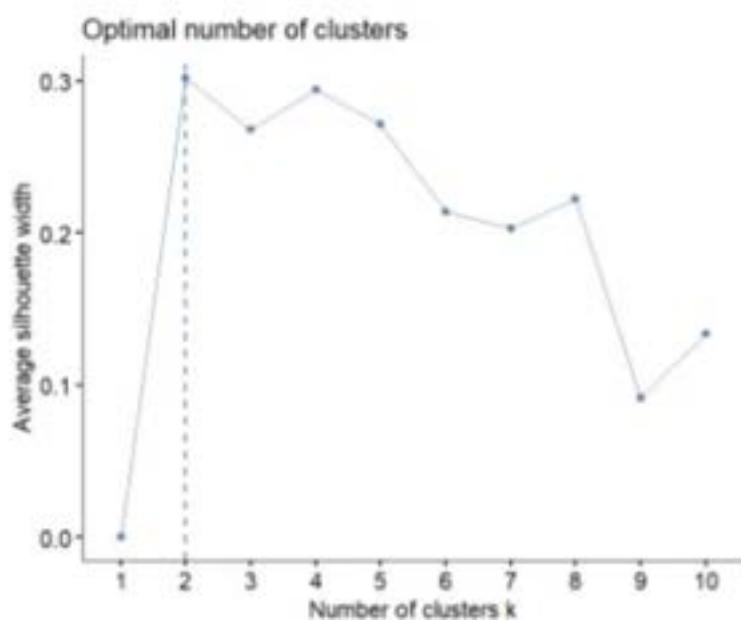
k=2

. k=2

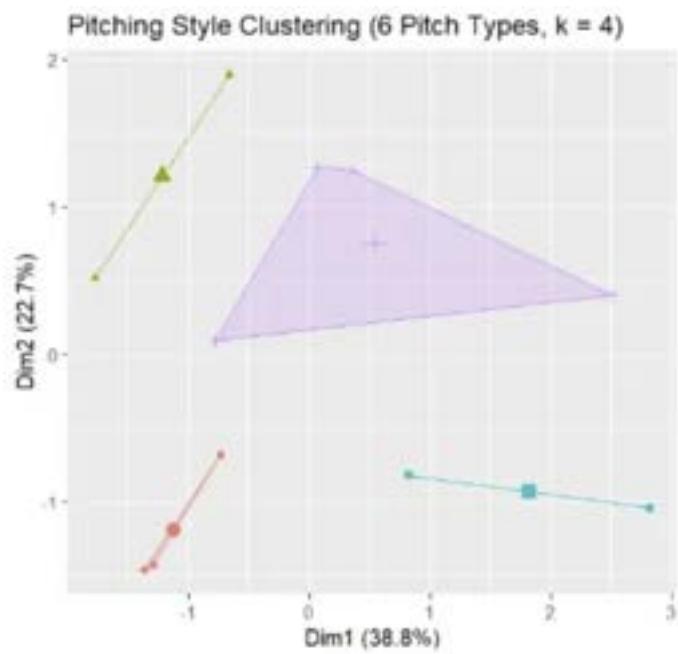
k=4



[]



[]

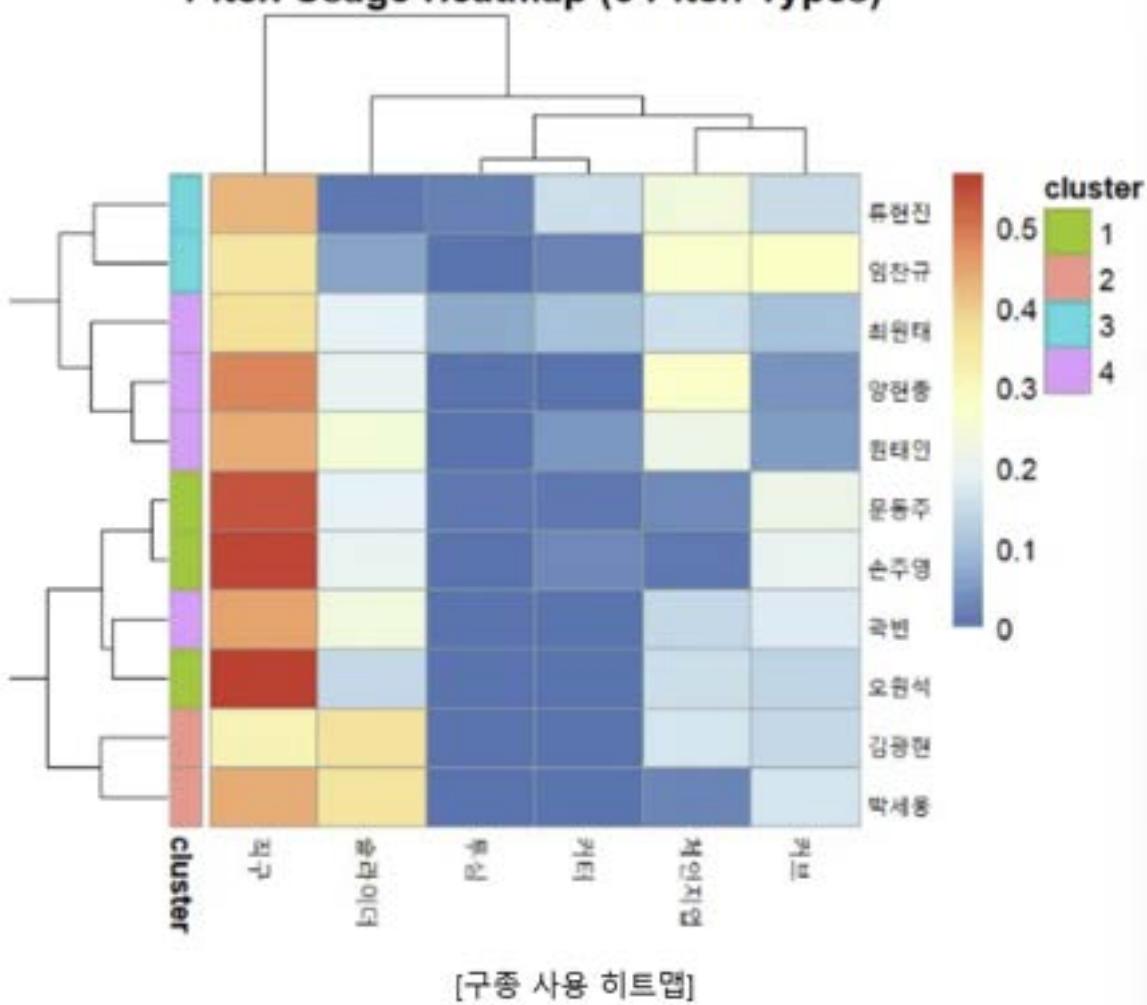


[]

	cluster	슬라이더	직구	체인지업	커브	투심	커터
	<fct>	<db7>	<db7>	<db7>	<db7>	<db7>	<db7>
1	1	0.180	0.554	0.0635	0.186	0.00360	0.0135
2	2	0.369	0.379	0.0961	0.156	0	0
3	3	0.0364	0.392	0.254	0.215	0.0111	0.0913
4	4	0.216	0.435	0.198	0.0950	0.0184	0.0379

[]

Pitch Usage Heatmap (6 Pitch Types)



[구종 사용 히트맵]

[]

Cluster 1: (, ,)

Cluster 2: + (,)

2

Cluster 3: (,)

Cluster 1

Cluster 4: / (, , ,)

3.2.

11

2024-2025

, Power, Arsenal, Tempo, Impact, Command
Stamina, Stress Relief 7

11

3

power impact

2

Arsenal Stress Relief

. Arsenal Tempo, Deception

. Power, Command, Impact

, Stamina

3.3.

XGBoost

XGBoost

2

XGBoost

(1)', (0)'

pitcher	슬라이더	직구	체인지업	커브	커터	투심	포크
곽빈	18.14	34.36	19.38	28.12	0	0	0
김광현	41.6	22.18	9.99	25.55	0	0	0.69
류현진	1.89	45.91	23.5	10.93	15.32	1.83	0.61
문동주	15.16	42.91	3.46	17.2	0.18	0.27	20.83
박세웅	28.18	32.29	1.41	19.77	0	0	18.35
손주영	14.46	47.21	0.63	21.37	2.16	0	14.18
양현종	24.9	36.42	32.96	5.72	0	0	0
오원석	12.54	51.38	12.13	22.15	0	0.08	1.71
원태인	24.97	44.26	21.63	3.27	5.67	0.2	0
임찬규	5.12	28.46	35.33	26.85	3.58	0	0.66
최원태	10.73	38.48	20.77	21.9	5.06	3.05	0

: 단순 사용 비율이 가장 높은 구종

[2]

1, 2, 3

1순위				2순위				3순위			
투수	순위	구종	성공률	투수	순위	구종	성공률	투수	순위	구종	성공률
손주영	1순위	직구	0.6398	임찬규	2순위	직구	0.6584	곽빈	3순위	체인지업	0.6276
김광현	1순위	슬라이더	0.6218	박세웅	2순위	슬라이더	0.5819	손주영	3순위	슬라이더	0.5984
원태인	1순위	직구	0.5982	류현진	2순위	체인지업	0.5805	양현종	3순위	슬라이더	0.5879
류현진	1순위	직구	0.5944	손주영	2순위	커브	0.5799	오원석	3순위	슬라이더	0.5637
최원태	1순위	직구	0.5867	곽빈	2순위	커브	0.5793	임찬규	3순위	커브	0.5578
오원석	1순위	직구	0.5785	원태인	2순위	슬라이더	0.5736	박세웅	3순위	커브	0.5558
임찬규	1순위	체인지업	0.5715	양현종	2순위	체인지업	0.5649	최원태	3순위	체인지업	0.5509
양현종	1순위	직구	0.5683	김광현	2순위	커브	0.5621	김광현	3순위	직구	0.5368
박세웅	1순위	직구	0.5636	오원석	2순위	커브	0.5052	류현진	3순위	커터	0.5308
문동주	1순위	직구	0.5405	문동주	2순위	포크	0.4947	원태인	3순위	체인지업	0.5249
곽빈	1순위	직구	0.5122	최원태	2순위	커브	0.4287	문동주	3순위	커브	0.5022

[]

3.4.

1

6

3

1.

:

2.

:

(6),

3.

:

4.

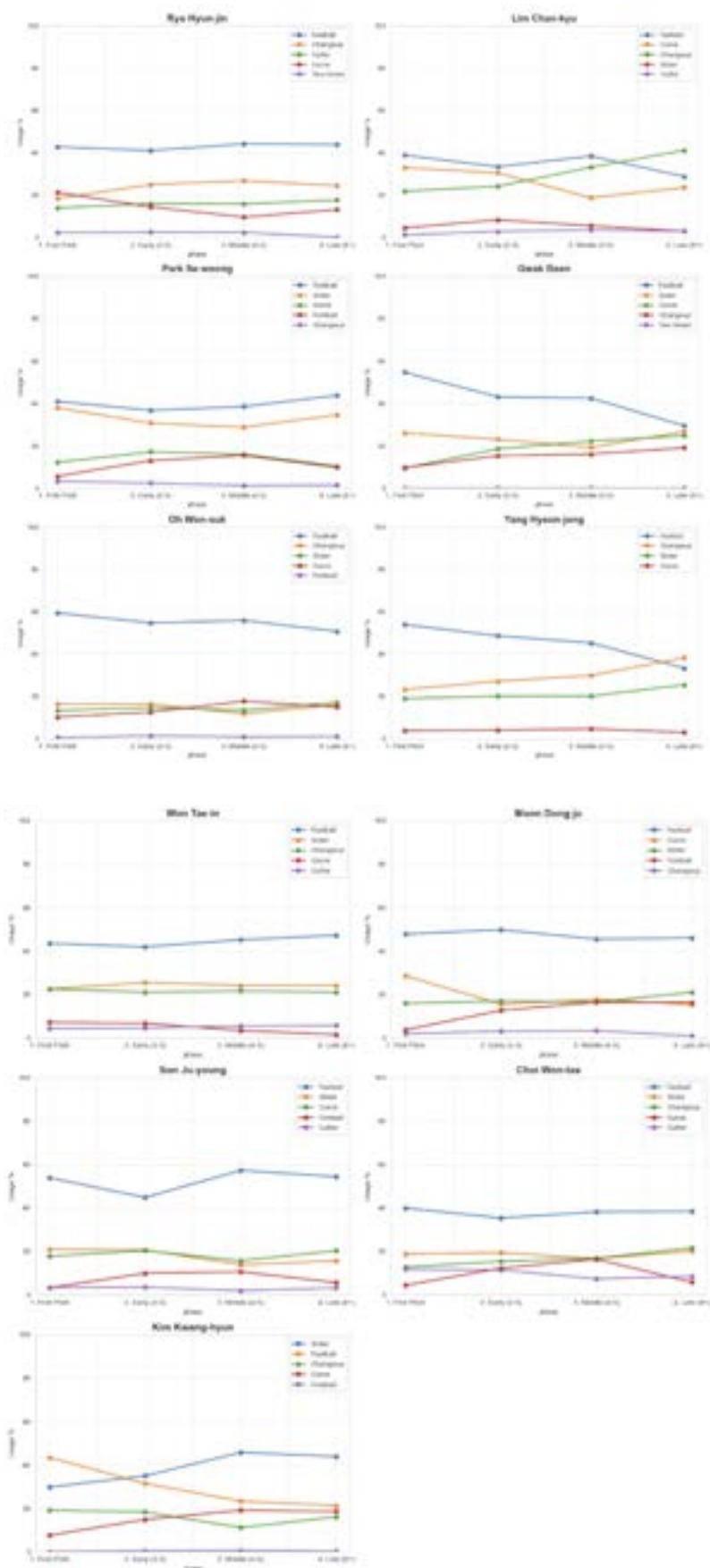
: (

) - (

, 0

Pitcher	First_Pitch_Fastball_Rate	Late_Count_Weapon	Late_Weapon_Usage_Rate	Fastball_Diff_(late - first)
Ryu Hyun-jin	42.8 Fastball		43.9	1.1
Lim Chan-kyu	38.9 Changeup		41.2	-10
Park Se-woong	41 Fastball		43.8	2.8
Gwak Beom	54.7 Fastball		29.5	-25.2
Oh Won-suk	59.4 Fastball		50.7	-8.7
Yang Hyeon-jong	51.9 Changeup		38.2	-20.6
Won Tae-in	41.5 Fastball		47.1	3.7
Moon Dong-ju	47.8 Fastball		46.1	-1.7
Son Ju-young	53.9 Fastball		54.4	0.5
Choi Won-tae	40 Fastball		38.4	-1.6
Kim Kwang-hyun	41.4 Slider		43.9	-22.1

[]



[

54.7%

29.5%

25.2%P

10%P

3.7%P

1

1.1%P

3.5.

XGBoost

shift

compute_sample_weight

F1-Score

직전투구 기반

Pre+Rec

pitcher	Acc	Pre	Rec	F1	max_depth	learning_rate	subsample	colsample_bytree	min_child_weight	reg_lambda
양현종	0.482252	0.469124	0.47391	0.470712	6	0.05	0.7	0.8	3	1
문동주	0.464497	0.479257	0.504492	0.455542	6	0.1	0.7	0.7	1	5
김광현	0.466447	0.453326	0.456786	0.435445	3	0.1	0.6	0.7	1	1
임찬규	0.440835	0.443548	0.440004	0.434607	4	0.05	0.8	0.7	3	10
곽빈	0.472956	0.432376	0.436834	0.431351	4	0.03	0.7	0.9	5	10
손주영	0.469626	0.436572	0.493068	0.429473	3	0.1	0.8	0.8	3	10
오원석	0.42464	0.436527	0.476012	0.420364	3	0.1	0.6	0.7	1	1
박세웅	0.423423	0.438553	0.463511	0.414706	3	0.1	0.6	0.7	1	10
원태인	0.428571	0.416744	0.415341	0.409618	3	0.1	0.8	0.8	3	10
류현진	0.43018	0.430641	0.454296	0.404829	6	0.1	0.7	0.7	1	5
최원태	0.415014	0.416895	0.422995	0.396058	6	0.1	0.7	0.7	1	5

[]

타석 기반

Pre+Rec

pitcher	Acc	F1	Pre	Rec	subsample	max_depth	learning_rate	colsample_bytree
양현종	0.491803	0.48301	0.481162	0.493493	0.8	4	0.1	0.7
최원태	0.496823	0.462739	0.474214	0.489958	0.8	4	0.03	0.7
오원석	0.464871	0.454236	0.466792	0.503812	0.8	4	0.03	0.8
곽빈	0.475596	0.451523	0.451118	0.46863	0.7	4	0.03	0.7
김광현	0.478818	0.447113	0.453863	0.461183	0.7	5	0.1	0.8
손주영	0.481598	0.431352	0.439125	0.479861	0.8	4	0.03	0.7
박세웅	0.443439	0.427015	0.443324	0.471387	0.7	4	0.03	0.8
임찬규	0.431557	0.426959	0.434623	0.430817	0.7	5	0.05	0.8
문동주	0.418206	0.414174	0.422819	0.477485	0.7	4	0.03	0.8
원태인	0.423764	0.400089	0.406784	0.404727	0.8	6	0.03	0.7
류현진	0.432896	0.396907	0.410975	0.424749	0.8	4	0.1	0.8

[]

당해이닝 기반

Pre+Rec

pitcher	Acc	F1	depth	lr	sub	col	lambda
양현종	0.512568	0.507859	4	0.1	0.6	0.7	1
김광현	0.495567	0.467136	4	0.05	0.7	0.6	1
곽빈	0.484677	0.46187	5	0.01	0.6	0.7	1
최원태	0.481576	0.452164	4	0.01	0.9	0.7	5
오원석	0.4637	0.451878	4	0.01	0.9	0.7	5
박세웅	0.451584	0.438614	5	0.01	0.6	0.7	10
원태인	0.443743	0.425339	5	0.05	0.8	0.9	5
임찬규	0.425287	0.423222	8	0.05	0.6	0.7	1
문동주	0.420844	0.417745	4	0.03	0.6	0.9	1
손주영	0.45531	0.415103	4	0.01	0.7	0.9	5
류현진	0.422805	0.389133	5	0.1	0.6	0.7	10

[]

당해게임 기반

Pre+Rec

pitcher	Acc	F1	Pre	Rec	subsample	reg_lambda	max_depth	learning_rate	colsample_bytree
양현종	0.535519	0.536217	0.535216	0.560361	0.7	10	4	0.03	0.7
김광현	0.495567	0.464618	0.471166	0.480978	0.7	1	4	0.03	0.7
곽빈	0.476731	0.464081	0.464393	0.500954	0.7	10	4	0.05	0.8
오원석	0.4637	0.447411	0.452413	0.486257	0.7	10	6	0.03	0.7
박세웅	0.460633	0.444991	0.449855	0.489271	0.8	10	4	0.1	0.8
문동주	0.44723	0.442503	0.460851	0.518051	0.8	10	4	0.05	0.8
최원태	0.468869	0.438428	0.454196	0.471358	0.7	1	4	0.1	0.7
손주영	0.488959	0.437032	0.442891	0.502438	0.8	10	6	0.03	0.7
임찬규	0.433647	0.431148	0.433829	0.434033	0.7	10	4	0.03	0.7
원태인	0.435331	0.405538	0.405135	0.406786	0.7	10	6	0.03	0.8
류현진	0.411705	0.401237	0.437046	0.461153	0.7	10	4	0.05	0.8

[]

API

2024-2025 KBO

406,479

96%

, XGBoost

(Log Loss)

==== Classification Report (Validation) ===				
	precision	recall	f1-score	support
Fastball	0.98	0.98	0.98	43123
Breaking	0.93	0.96	0.94	23774
Offspeed	0.96	0.92	0.94	13968
accuracy			0.96	80865
macro avg	0.96	0.95	0.95	80865
weighted avg	0.96	0.96	0.96	80865

==== Classification Report (Validation) ===				
	precision	recall	f1-score	support
Fastball	0.98	0.98	0.98	43123
Breaking	0.93	0.96	0.94	23774
Offspeed	0.96	0.92	0.94	13968
accuracy			0.96	80865
macro avg	0.96	0.95	0.95	80865
weighted avg	0.96	0.96	0.96	80865

[XGBoost

]

, F1-Score

60%, 48%, 34%
 ()

==== Classification Report (Holdout) ====				
	precision	recall	f1-score	support
Fastball	0.62	0.58	0.60	43123
Breaking	0.43	0.55	0.48	23774
Offspeed	0.40	0.30	0.34	13968
accuracy			0.52	80865
macro avg	0.49	0.48	0.48	80865
weighted avg	0.53	0.52	0.52	80865

[

XGBoost

]

(PCA)

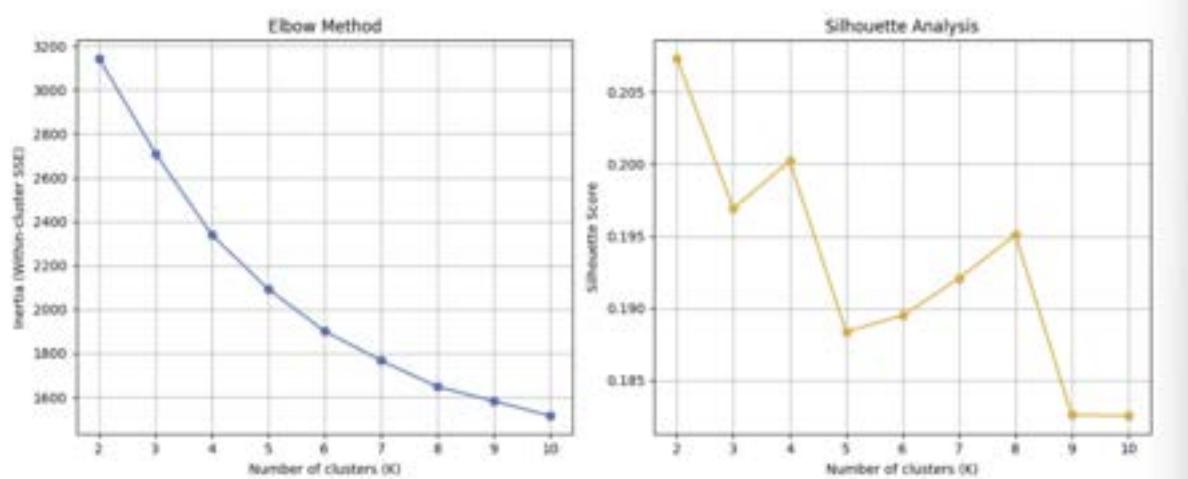
. PCA

CLR(Centered Log-Ratio)

PCA

,
silhouette

7



[

]

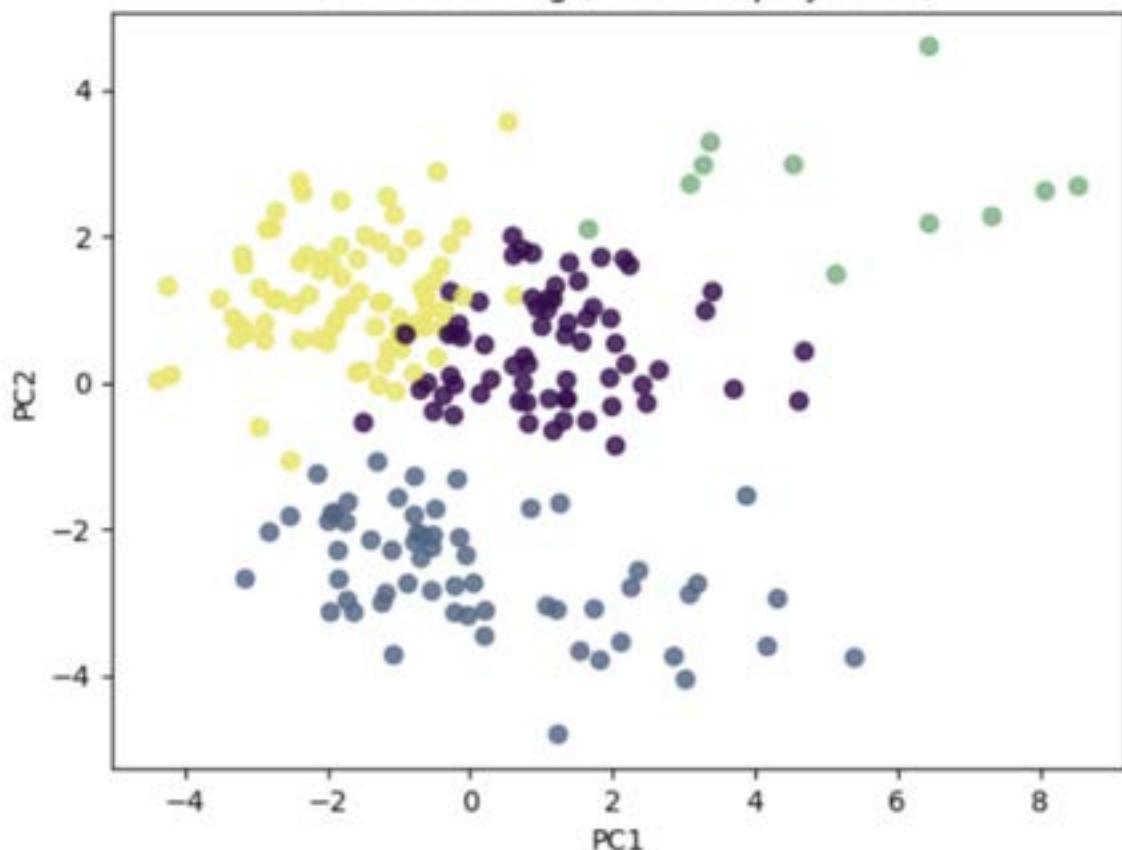
pitcherId	PC1	PC2	PC3	PC4	PC5	PC6	PC7	cluster
50030	3.704672	-0.08985	-0.83542	2.698763	-2.94006	0.815168	-1.76113	0
50106	-2.71544	2.342735	-1.96556	0.990573	0.703939	0.202142	-1.16497	3
50109	0.05381	-2.74106	-1.0045	-2.31285	0.664448	-0.06848	-0.96462	1
50354	-1.62459	-3.14009	-1.60456	-1.60243	-0.88959	0.39819	-0.1447	1
50393	1.203678	1.322219	-0.06931	-0.37522	-0.42695	-1.57533	0.397951	0
50441	0.878547	1.767839	-0.95061	0.604371	-0.69095	-0.5788	0.836715	0
50464	-0.14173	-2.11825	-0.98652	-2.68994	0.03103	0.50791	-0.28915	1
50556	0.131538	1.104241	1.617899	0.729894	-0.80046	1.84783	-0.4487	0
50662	-2.87657	0.811855	-0.98496	2.068991	-1.40262	0.221723	-0.23368	3
50812	0.858416	-1.72024	-1.19866	0.575678	-0.23638	-0.91658	1.475395	1
50859	-0.02081	-3.18666	0.034484	-1.70569	1.257464	-0.63974	-0.24698	1
50904	-0.27448	1.242853	2.246166	-1.32163	1.745691	0.445754	1.113257	0
51111	-2.51903	-1.82775	0.612049	-0.55001	-0.54029	0.273842	0.503184	1
51230	-1.06039	2.294035	-0.44217	-0.13719	0.844419	-0.70788	0.211432	3
51264	-1.93252	-1.78707	0.741528	-2.24237	0.771232	-1.81262	-0.81809	1

[]

PC1	구종 비율(CLR) + 구속/무브먼트 일부 기여도 높음
PC2	vx, vy, vz 등 속도 성분 기여가 큼
PC3	speed_std, az, z 위치 변수 영향 큼
PC4	x/z 위치 + 가속도 혼합
PC5~7	특정 구종 비율 또는 물리량이 부분적으로 작용

[]

PCA4 clustering (PC1-PC2 projection)



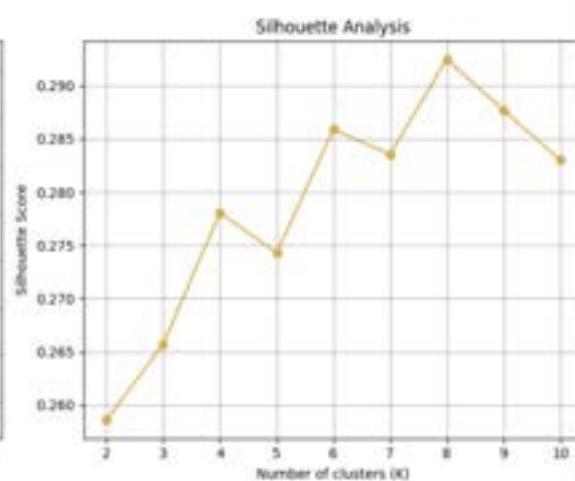
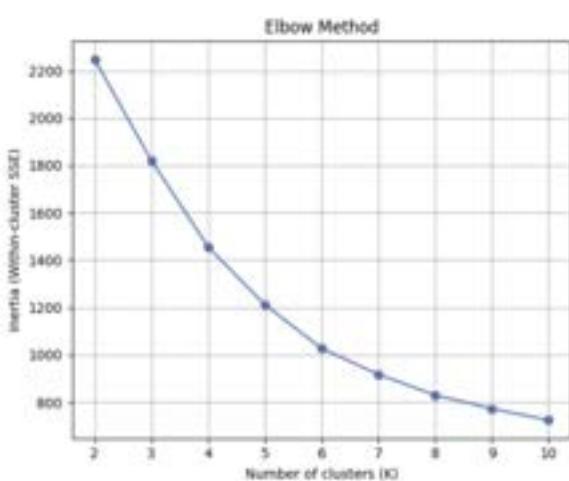
[PC1-PC2

PCA4

]

/ / / / .

4



[]

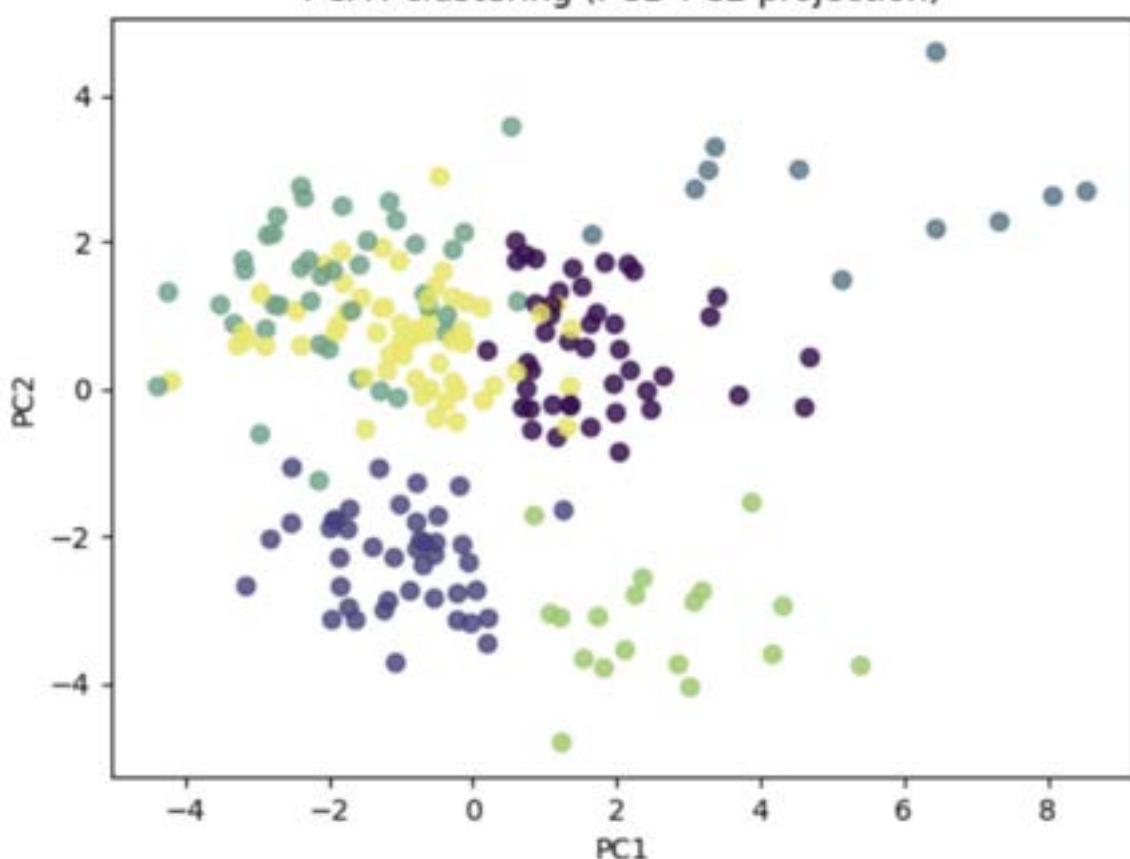
pitcherId	PC1	PC2	PC3	PC4	cluster
50030	3.704672	-0.08985	-0.83542	2.698763	0
50106	-2.71544	2.342735	-1.96556	0.990573	3
50109	0.05381	-2.74106	-1.0045	-2.31285	1
50354	-1.62459	-3.14009	-1.60456	-1.60243	1
50393	1.203678	1.322219	-0.06931	-0.37522	0
50441	0.878547	1.767839	-0.95061	0.604371	0
50464	-0.14173	-2.11825	-0.98652	-2.68994	1
50556	0.131538	1.104241	1.617899	0.729894	5
50662	-2.87657	0.811855	-0.98496	2.068991	3
50812	0.858416	-1.72024	-1.19866	0.575678	4
50859	-0.02081	-3.18666	0.034484	-1.70569	1
50904	-0.27448	1.242853	2.246166	-1.32163	5
51111	-2.51903	-1.82775	0.612049	-0.55001	1
51230	-1.06039	2.294035	-0.44217	-0.13719	3
51264	-1.93252	-1.78707	0.741528	-2.24237	1
51454	-0.20748	-3.14555	0.844605	0.302913	1
51516	-1.89631	-1.77864	0.398631	-1.67373	1
51594	-2.95286	-0.61461	-1.34581	-0.80486	3
51648	-1.95456	-3.13572	0.85443	-1.5419	1
51713	-0.67048	1.28798	-1.90037	1.448729	3
51715	-0.77127	-2.17923	0.804743	-2.08767	1

[]

PC1	Vz0_kmh_mean, z0_m_mean 등 구속/상하 무브먼트 중심
PC2	Vx0_kmh_mean, x0_m_mean 등 좌우 움직임 중심
PC3	speed_std, az_ms2_std, 커브 비율 등 변동성/결정구 성향
PC4	공의 높낮이 조절과 세부 제구 스타일 반영

[]

PCA4 clustering (PC1-PC2 projection)



[PCA4

]

5.

XGBoost

(Feature)

(Graph Modeling)