데이터 분석가 (게임데이터 분석) 과제

OPGG 지원자 : 김서정

Date: 2020. 11. 26

Email: souljoin0229@gmail.com

과제1)

분석주제: 인게임 오브젝트(용, 전령)와 승률간의 관계분석

데이터: - lol 10.18 버전 KR 지역 솔로랭크 match data (약 17만건중 약 3만건 사용)

- KR 지역 솔로랭크 다이아몬드 티어 이상의 소환사가 1명이상 포함된 경기들

분석을 활용한 결과물 : - 관련 그래프 및 수치지표

- 인게임 오브젝트 및 그외 지표들을 활용한 승리예측 모델 생성

분석언어: Python, Pycaret(AutoML 라이브러리)

데이터 중, 사용한 Features :

Numeric_type_OBFeatures	Bool_type_OBFeatures	Numeric_type_Champion_Features
Dragonkills (용킬수)	Firstdragon (첫용)	Totaldamagedealttochampions (챔피언에게 가한 피해량)
Inhibitorkills (억제기킬수)	Firstinhibitor (첫억제기)	Totaldamagetaken (받은 총피해량)
Riftheraldkills (전령킬수)	Firstriftherald (첫전령)	totalminionskilled (CS)
Baronkills (바론킬수)	Firstbaron (첫바론)	Killsplusassists (챔피언처치및어시스트)
Towerkills (타워킬수)	Firstblood (퍼블)	Deaths (데스)
Teamid (진영-블루(100)/레드(200))	Firsttower (포블)	visionscore (시야점수)
Gameduration (경기시간)		

<분석과정>

1) 전체 데이터 통계 확인을 통한 문제정의

문제1: 2) 게임에서 승리하기위해서 최소 몇 개이상의 용을 먹어야할까?

- 킬한 용의 개수에 따라 승률이 어떻게 바뀌는지?

문제2: 3) 꿩 대신 닭 = 용 대신 전령? 전령의 영향력은 얼마나 될까?

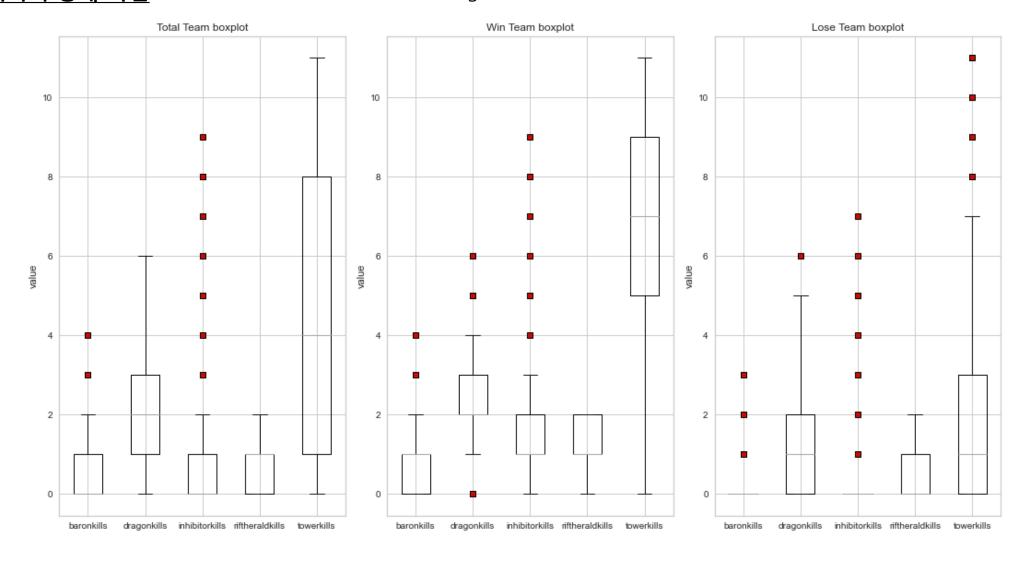
- 용과 전령의 승리와의 관계분석



각 오브젝트 및 인게임 환경지표을 활용한 승리 예측 모델

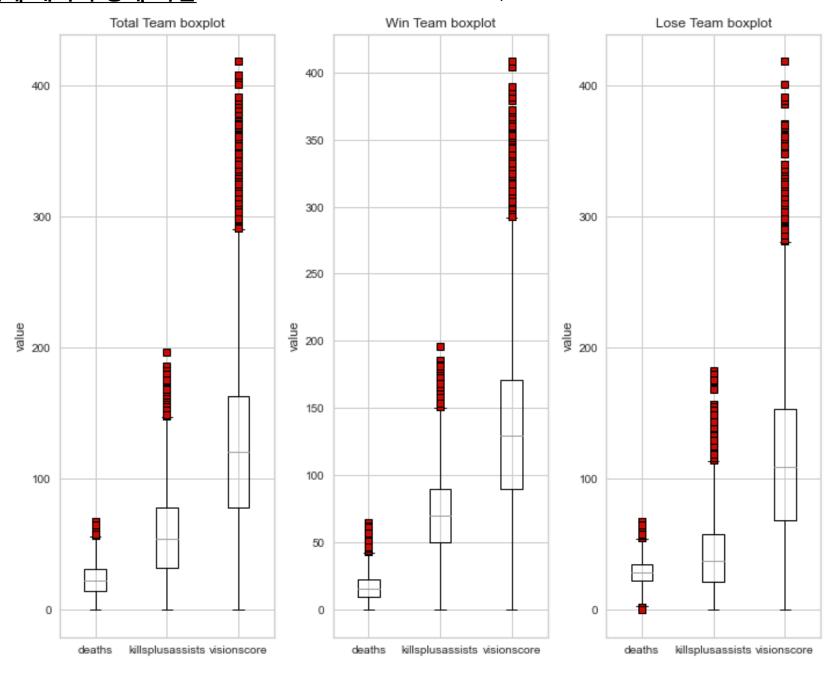
- 첫 용, 첫 전령, 첫 바론, 첫 타워, 첫 억제기 등등

1) 전체 데이터 통계 확인 - 전체팀, 승팀, 패팀의 baronkills, dragonkills, inhibitorkills, riftheraldkills, towerkills 확인



패한 팀의 baronkills, dragonkills, inhibitorkills, riftheraldkills, towerkills 모두 승팀보다 현저하게 낮음

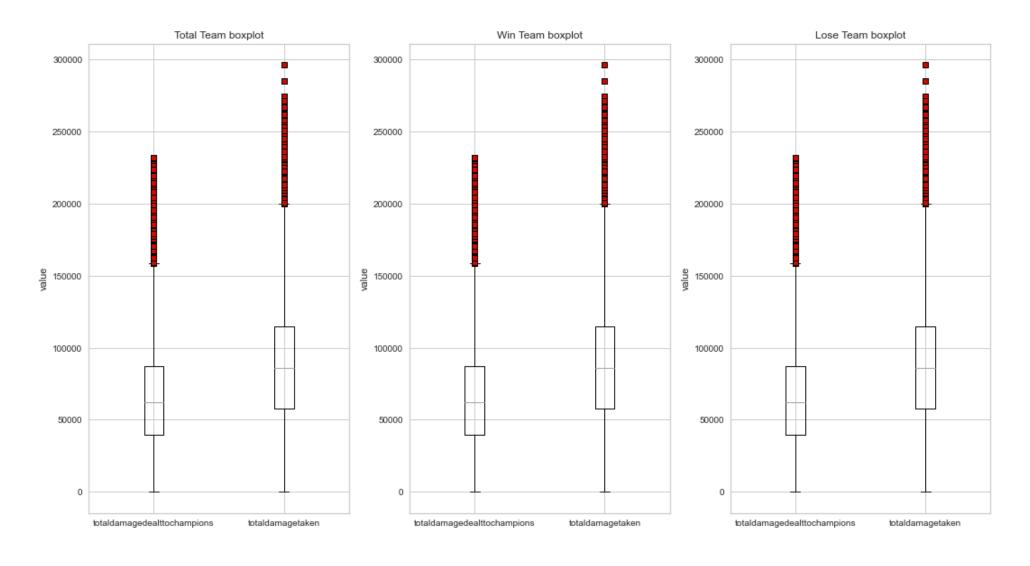
1) 전체 데이터 통계 확인 - 전체팀, 승팀, 패팀의 deaths, killsplusassists, visionscore 확인



이긴팀의 통계량이 전체팀 의 통계량과 비슷함.

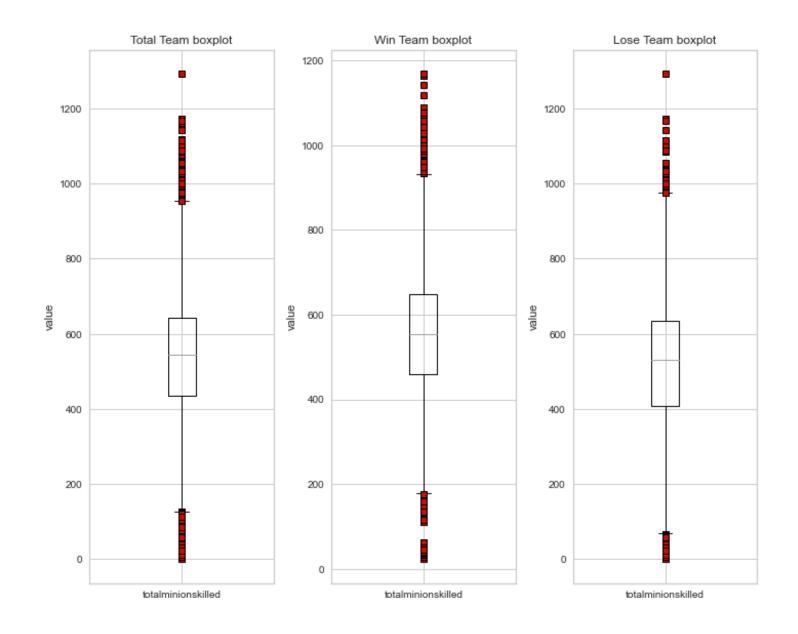
반면, 진팀은 전체팀보다 분포가 넓음

1) 전체 데이터 통계 확인 - 전체팀, 승팀, 패팀의 챔피언에게 가한 총피해량, 받은 총 피해량 확인



이긴팀, 진팀이 비슷한 평균과 분포를 나타냄

1) 전체 데이터 통계 확인 - 전체팀, 승팀, 패팀의 CS량 확인



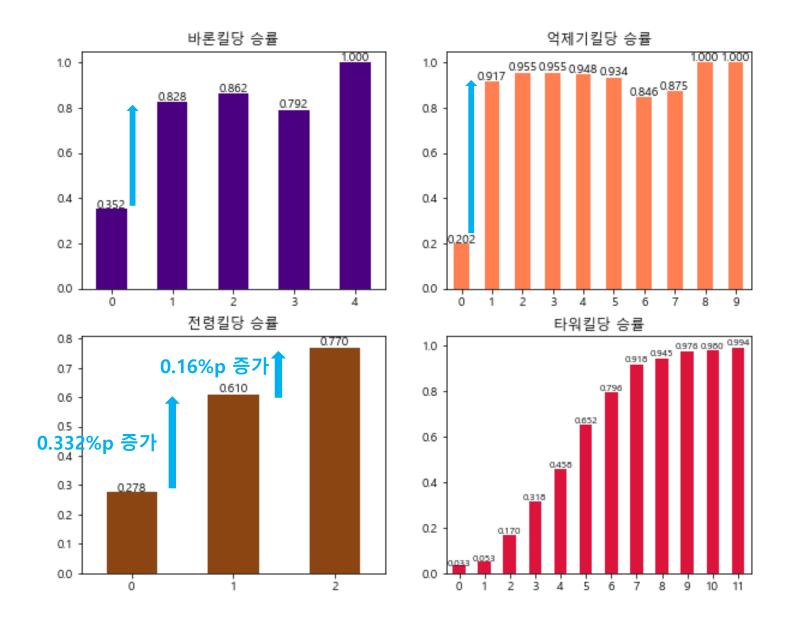
-CS의 경우, 승팀이 근소하 게 평균값이 높음

하지만, 거의 차이 없음

승패팀의 여러 속성중, baronkills, dragonkills, inhibitorkills, riftheraldkills, towerkills 를 제외하고 나머지 속성에서는 승패팀의 유의미한 차이가 보이지 않았음

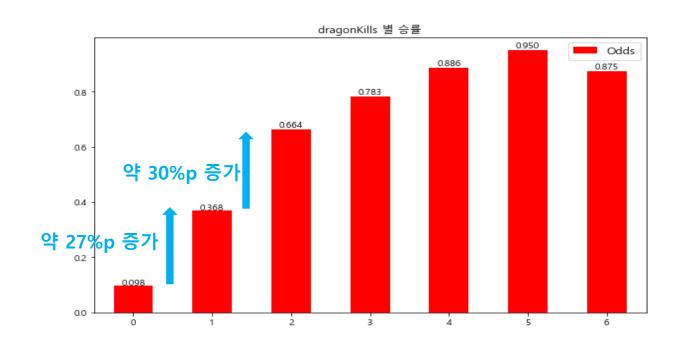
그렇다면, baronkill과 dragonkill, inhibitorkills, riftheraldkills, towerkills가 승리에 미치는 영향력은 어느정도 일까?

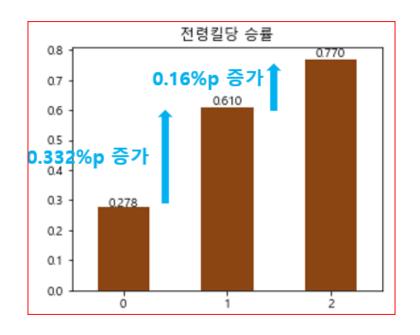
1) baronkill과 dragonkill, inhibitorkills, riftheraldkills, towerkills가 승리에 미치는 영향력



- 바론1킬, 억제기1킬부터 승률이 눈에 띄게 높아짐 (각각, 47%p, 71%p 상승)
- 타워킬, 전령킬과 승률은 정비례

1) baronkill과 dragonkill, inhibitorkills, riftheraldkills, towerkills가 승리에 미치는 영향력





전령: 0 -> 1 (0.610) / 1-> 2(0.77) 용: 0 -> 1 (0.368) / 1 -> 2 (0.664) 용보다 전령을 먹었을 때, 승률이 높아보이지만, 변화율은 용을 먹었을 때가 훨씬 크다 ⇒ 용이 전령보다 큰 변수를 창출

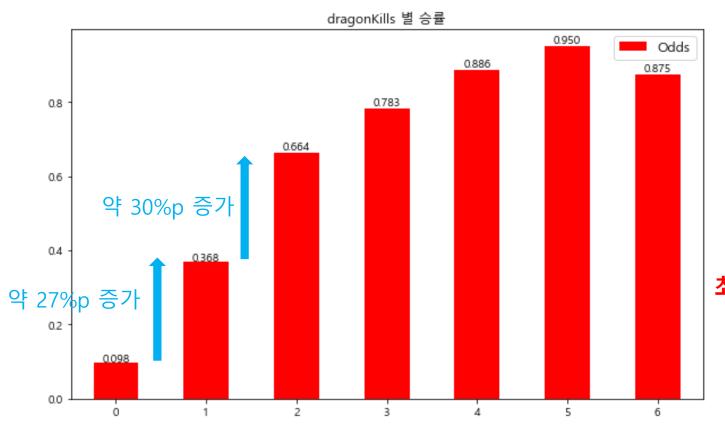


초반 게임의 변수창출 : 용 > 전령

그렇다면, 최소 몇 개의 용을 먹어야 승리를 확신할 수 있을까? 또한, 전령은 용의 대체재일까? 아니면, 어쩔수 없는 선택인걸까?

2) 게임에서 승리하기위해서 최소 몇 개이상의 용을 먹어야할까?

용의 개수에 따른 승률은? (용 n개를 먹은 전체(승+패)팀 중, 승리한 팀의 수)

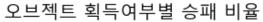


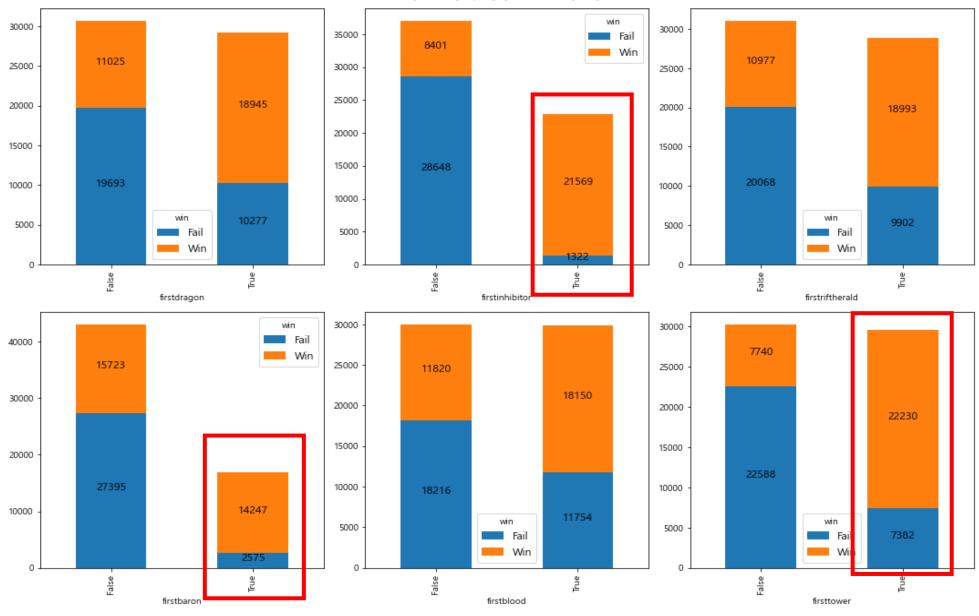
0->1, 1->2용의 승률에서 가장 크게 승률그래프가 높아지는 것을 확인

3용부터 승률이 80%부근으로 상승

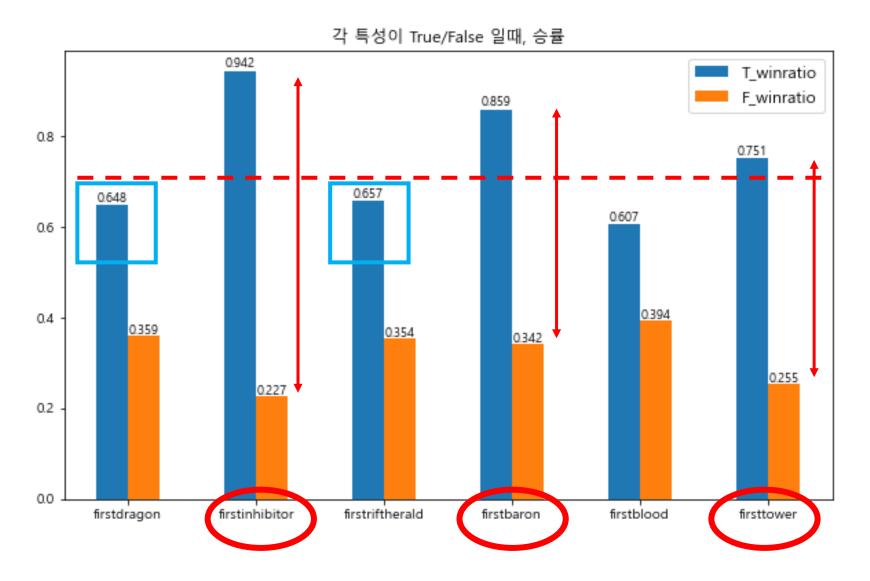
최소, 3용부터 승리를 확신하는 것이 가능할것

3) 꿩 대신 닭 = 용 대신 전령? 전령의 영향력은 얼마나 될까?





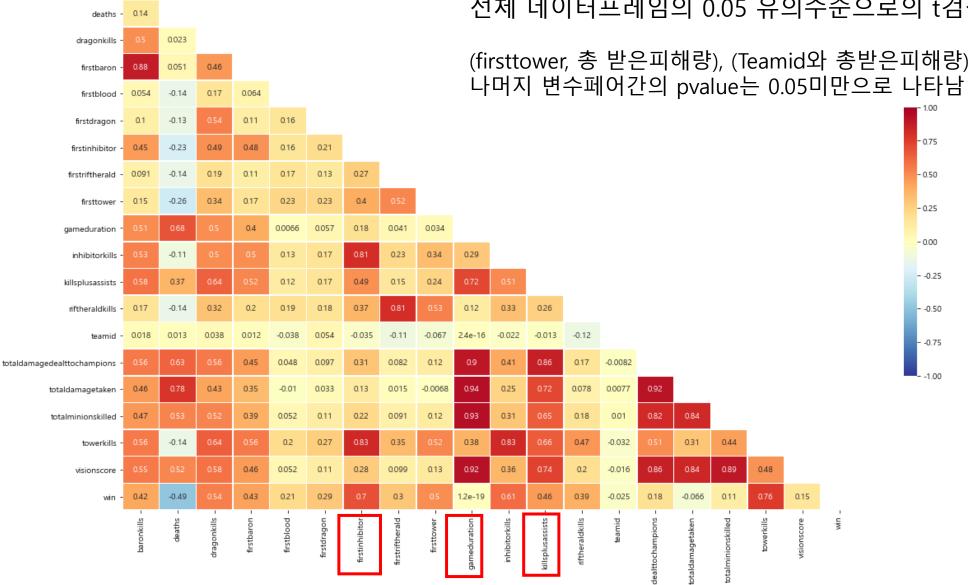
3) 꿩 대신 닭 = 용 대신 전령? 전령의 영향력은 얼마나 될까?



'첫 억제기','첫 바론', '첫 타워'가 True일 때가 False일 때보다 승률이 월등히 높은 것을 확인

다만,
'첫 용'과 '첫 전령'의 승률이
각각 0.648, 0.657로
근소한차이를 보이고 있음.
-> 첫 용 과 첫 전령이 경기에
미치는 영향이 비슷하다는 의미

3) 전체 속성간의 상관분석 baronkills



전체 데이터프레임의 0.05 유의수준으로의 t검정 결과:

(firsttower, 총 받은피해량), (Teamid와 총받은피해량)의 pvalue가 0.05이상,

3) 상관분석 결과 (용과 전령에 대하여)

Firstriftherald – Win (0.3)

FirstDragon – Win (0.29)



(첫용, 승리), (첫 전령, 승리)와의 상관관계 -> 0.3, 0.29로 비슷한 수준

Firstriftherald FirstTower (0.52)

FirstTower – Win (0.5)



전령의 전략적 사용은 직접적인 타워 또는 억제기 킬로 이어지기 때문에, 용을 선택하는 것과는 다르게 운영적측면에서의 성과와 변수를 창출하는 요인 이 될 수 있을 것.

정글러가 용 대신 전령 이라는 관념에서 전령을 유의미하게 사용할 수 있도록 라인상황, 적 정글 포지션 등을 파악하고 라이너들과 합을 맞춘다면, 용과 버금가는 혹은 용 이상의 운영적 효과를 얻을 수 있을 것이라 생각



Win 에 영향을 거의 미치지 못하는 속성: Teamid(블루진영/레드진영)

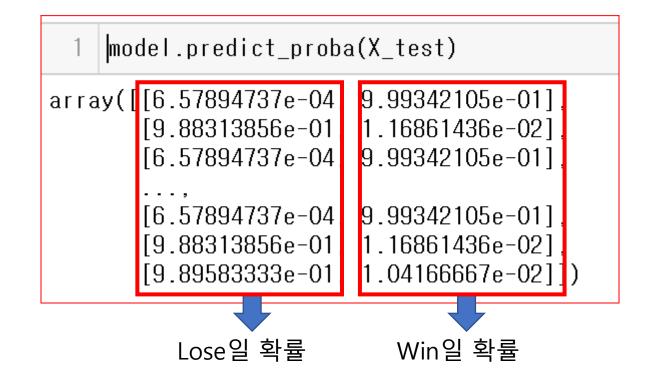
받은총피해량

게임시간

해당, 속성들을 제외하고 나머지 속성을 가지고 승리예측모델을 생성

Ex) Decision Tree

```
1 X,y=data.iloc[:,:-1], data['win']
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)
3 4 tree = DecisionTreeClassifier(max_depth=7, random_state=1)
5 model=tree.fit(X_train, y_train)
print('훈련셋 정확도 : ', model.score(X_train,y_train))
7 print('테스트셋 정확도 : ', model.score(X_test,y_test))
후려셋 정확도 : 0.9744744744744
테스트셋 정확도 : 0.9689133578022467
```



Pycaret을 이용해, Win or Lose를 예측하는 머신러닝 모델 중, Test Accuracy가 높았던 Top10 모델 생성

1	best3_2 = compare_models(n_select = 3)								
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	мсс	TT (Sec)
0	CatBoost Classifier	0.9771	0.9977	0.9803	0.9740	0.9771	0.9541	0.9541	16.5359
1	Extreme Gradient Boosting	0.9765	0.9976	0.9798	0.9735	0.9766	0.9531	0.9531	1.9516
2	Light Gradient Boosting Machine	0.9757	0.9976	0.9801	0.9715	0.9758	0.9513	0.9514	0.4389
3	Gradient Boosting Classifier	0.9741	0.9969	0.9758	0.9724	0.9741	0.9481	0.9481	5.4143
4	Ada Boost Classifier	0.9734	0.9969	0.9749	0.9721	0.9735	0.9469	0.9469	1.5918
5	Ridge Classifier	0.9712	0.0000	0.9709	0.9715	0.9712	0.9423	0.9423	0.0518
6	Linear Discriminant Analysis	0.9711	0.9957	0.9706	0.9715	0.9711	0.9422	0.9422	0.2219
7	Random Forest Classifier	0.9681	0.9934	0.9703	0.9661	0.9682	0.9363	0.9363	0.1185
8	Extra Trees Classifier	0.9653	0.9945	0.9736	0.9578	0.9656	0.9307	0.9308	0.7821
9	Decision Tree Classifier	0.9617	0.9617	0.9613	0.9620	0.9617	0.9234	0.9234	0.2127
10	Logistic Regression	0.9492	0.9867	0.9502	0.9484	0.9492	0.8983	0.8984	0.3340

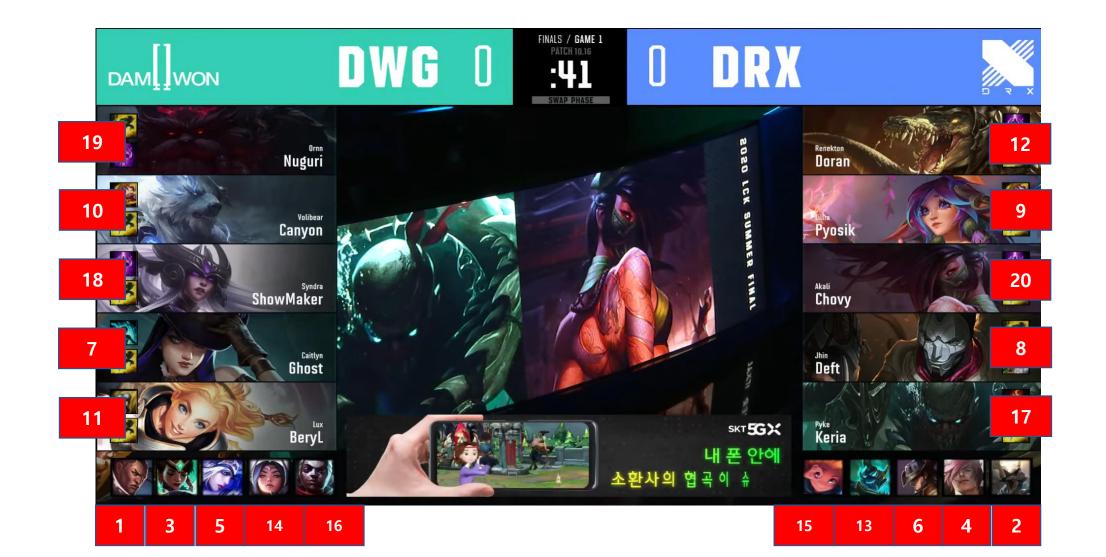
Top10 모델중, Top3의 모델을 blending하여 최적의 모델 생성

1 b	lended_2	= blend	l_model:	s(estim	ator_li	st = b	est3_2,
	Accuracy	AUC	Recall	Prec.	F1	Kappa	мсс
0	0.9784	0.0000	0.9830	0.9740	0.9785	0.9568	0.9568
1	0.9736	0.0000	0.9813	0.9665	0.9738	0.9472	0.9473
2	0.9767	0.0000	0.9779	0.9755	0.9767	0.9534	0.9534
3	0.9796	0.0000	0.9803	0.9789	0.9796	0.9591	0.9591
4	0.9768	0.0000	0.9813	0.9727	0.9769	0.9537	0.9537
Mean	0.9770	0.0000	0.9807	0.9735	0.9771	0.9540	0.9541
SD	0.0020	0.0000	0.0017	0.0041	0.0020	0.0040	0.0040

사전과제 2-1)

- 2020 LOL Champions Korea Summer Final (DWG vs. DRX) 경기 리뷰
- 2020 LCK Summer 결승전 경기를 보고, 경기의 승패에 영향을 주었던 장면 (벤픽상황, 라인전, 오브젝트 싸움 등)을 뽑고 그 장면을 뽑은 이유에 대해서 설명해주세요.

경기승패에 영향을 주었던 장면 - 제1경기 (밴& 픽)



2020 Lol Championship Korea Summer Final (DWG vs DRX)

경기승패에 영향을 주었던 장면 - 제1경기

DWG

- 케이틀린을 필두로, 바텀 카운터조합인 애쉬와 카르마를 밴
- DWG의 바텀 라인전을 세게하여 DRX의 원딜인 진의 과감한 플레이를 저지할 요량으로, 원딜 케이 틀린, 서폿 럭스를 가져옴
- 나머지 탱커조합으로 DWG 팀을 구성함
 - => DWG이 포킹 및 CC 조합으로 탑오른, 미드 신드라, 서폿 럭스

바텀, 후반게임에 승부수

DRX

- 베릴(SUP, DWG)이 서폿세트, 서폿판테 온으로 바텀싸움에서 좋은 모습을 보여 줬기 때문에, 밴
- 탑 레넥톤, 정글 릴리아픽, 그리고 헤카림(릴리아 카운터)밴으로 탑정글 주도권을 가져오고자 함
- 서폿 럭스를 예상하고 파이크를 꺼내서, 바텀압박을 하고 싶어한 듯함
- => 탑정글, 초반게임 승부수

과제2 2020 Lol Championship Korea Summer Final (DWG vs DRX)

경기승패에 영향을 주었던 장면 - 제1경기

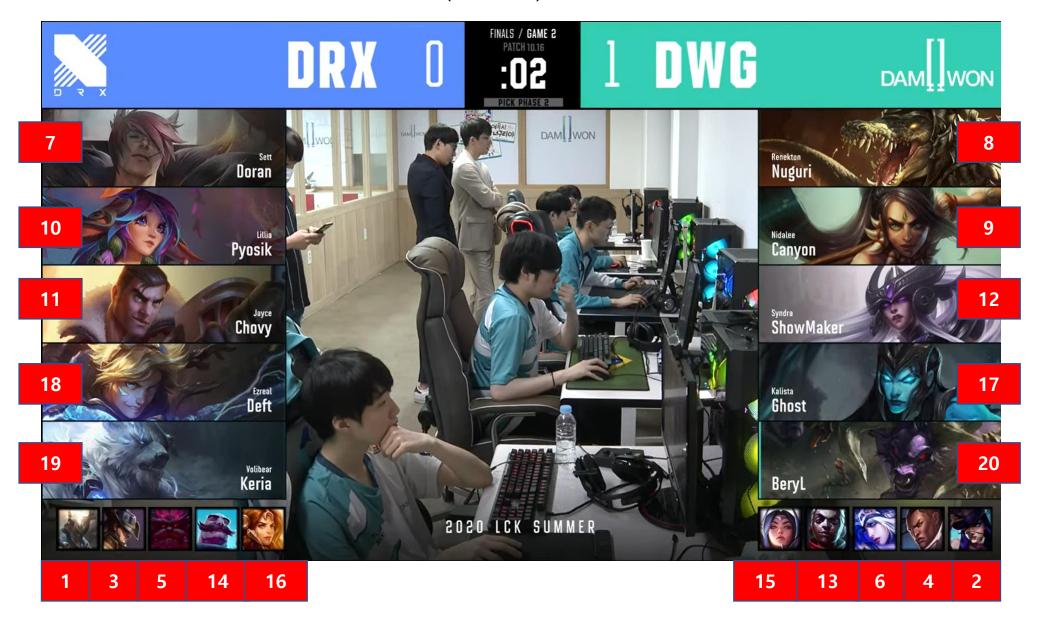
주요장면 시간	상황	선정 이유
1:13:50	미드에 담원 정글이 전령을 풀고, 3(DWG) vs 4(DRX) 대치	-럭스(DWG)속박 및 궁을 맞은 <mark>아칼리</mark> (DRX)가 플래쉬 빠짐
		- DRX 미드 1차타워 Break
1:17:01	-3번째용(바다용)을 두고 대치상황 -진(AD, DRX)궁으로 이니쉬 -오른(TOP, DWG)궁을 사용하여 진 커 튼콜 제지 및 릴리아(JG, DRX), 레넥톤 (TOP, DRX) 에어본 -> 레넥톤이 럭스(SUP, DWG)속박, 궁 맞고 딜링이 집중되면서, 볼리베어가 (JG, DWG)가 처치	-레넥톤현상금을 볼리베어가, 아칼리현상금을 케이틀린이 가져감 -> 케이틀린 성장 -바다용을 DWG이 가져감 -킬스코어: 3(DWG) vs 1(DRX)
	- 오른이 릴리아, 진을 묶어두고 상대 ->DRX의 딜링 억제, 아칼리의 어그로 및 암살 역할을 못하게끔 함	

과제2 2020 Lol Championship Korea Summer Final (DWG vs DRX)

경기승패에 영향을 주었던 장면 - 제1경기

주요장면 시간	상황	선정 이유
1:21:55	-DRX쪽 미드 2차타워두고 대치상황 -릴리아가 케이틀린 덫밟고 hp반정도 깎임 -릴리아가 이니쉬후, 존야 -> 아칼리(DRX), 레넥 톤(DRX) 한타 앞쪽 진입 - 오른(DWG)이 기둥돌진으로 진,파이크(DRX) 에어본, 레넥톤이 그 상황에서 오른을 보고 후 위로 가며 오른에 딜 이미 진입한 아칼리는 DWG의 딜링이 집중되 며 사망	-DWG이 TotalGold 역전 -케이틀린(AD, DWG)가 아칼리 처치하며 성장 -케이틀린(AD,DWG) 3코어, 진(AD,DRX) 2코어 -DWG이 4번째용(바다용) 가져감

경기승패에 영향을 주었던 장면 - 제2경기 (밴 & 픽)



경기승패에 영향을 주었던 장면 - 제2경기

DWG

담원의 레넥톤,니달리-> 탑주도권 가져옴 Drx가 세트, 릴리아 제이스 픽 -> DWG은 세트, 제 이스를 보고 서폿 세트, 탑제이스라 생각하여 미드 라인을 파이지 않기위해서 나머지밴픽을 이렐, 에 코같은 미드챔을 밴한 듯.

=> 다이브조합 구성

DRX

탑주도권은 깔끔하게 DWG을 주고, 제이스를 미드에 세워서 릴리아와 함께 미드 주도권가져 오자라는 마인드.

DWG 나머지가 바텀픽턴이었기 때문에, 무난하게 나머지 바텀 서폿밴

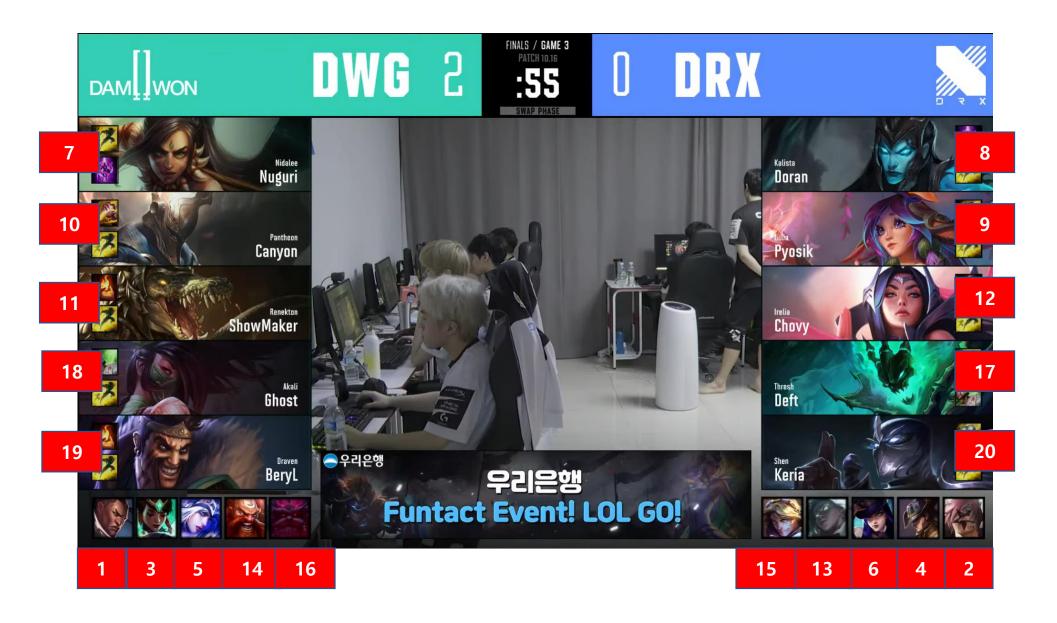
=> 포킹조합구성

주요장면 시간	상황	선정이유
2:03:01	-극 초반, 니달리(DWG)가 와드를 빼면서, 릴리아동선 체크를 못함 -> 릴리아(DRX)의 2랩 미드갱 -> 신드라(DWG)플빠짐 -니달리(DWG)의 레드 카정 but, 세트(SUP, DRX)의 시야확보로 들킴 -> 제이스(MID, DRX)와 신드라(MID, DWG)이 지원갔지만, 지원가면서 대치 -> 신드라 피가 30%이하로 깎임 -동시에, 릴리아는 탑 갱, 제이스(MID, DRX)의 플Q로 신드라 퍼블, 그리고니달리 더블킬	니달리(DWG) 플빠짐, 알리스타(SUP, DWG) 플 빠짐 킬스코어 – 2(DRX) VS 0(DWG) 골드는 DRX가 우위

주요장면 시간	상황	선정이유
2:29:17	- 니달리(JG, DWG)가 미드에 전령을 풀 - 미드 타워쪽에 정신이 팔린 DRX, 따라서 DRX 가 바텀 1차타워를 밀게됨 ??	DWG이 1차타워를 밀고 2차타워까지 데미지를 입힘
2:53:53	-DRX 블루쪽 알리스타(SUP, DWG)의 제이스 (MID, DRX)를 겨냥한 이니쉬 -> 릴리아(JG,DRX)의 칼리스타(AD,DWG)처치 - 알리스타가 이즈(AD, DRX)를 W로 블루쪽으로 밀어넣음 -> 신드라(MID, DWG)이 처치 - 레넥톤(TOP, DWG)이 적진영에 진입 -> 릴리아 처치	-DRX의 정글이 없으니, DWG의 바론 킬 - DWG이 TOP쪽 타워까지 킬

주요장면 시간	상황	선정이유
2:58:40	-4번째 용 대치-> DWG이 4번째용(바다용)을 먹고 한타시작-릴리아(JG, DRX)가 용을 먹기전 진입했으나, 실패-알리스타(SUP, DWG)가 세트(SUP, DRX)를 DWG진영에 밀어넣음 -> 칼리스타(AD, DWG)가 세트 처치, 릴리아 더블킬, 이후 진입한 볼리베어(TOP, DRX)를 니달리(JG, DWG)이 처치 - 제이스(MID, DRX)가 니달리 처치하였으나, 둘러싸여 칼리스타에 의해 사망	KILL SCORE: 11(DWG): 7(DRX) 잘 성장했던 제이스를 칼리스타가 현상금을 획득하며 킬초기화, 칼리스타 성장
2:58:40	- 담원의 바론트라이 -세트(SUP, DRX)의 진입 (궁, E)로 적 묶음. 그 사이에 릴리아가 바론을 먹고, Q로 딜링 -동시에, 담원미드(신드라)와 DRX미드(제이스)끼리 1:1 싸움- 고와중에 알리스타 WQ 그리고 신드라가 릴리아 처치 -> 이즈킬, 제이스 킬로 이어짐	- DRX 미드 1차 억제기 부숴짐, - DRX 탑 3차 타워 및 부숴짐 - 용과 민 타워 이득으로 계속 스노우볼

경기승패에 영향을 주었던 장면 - 제3경기 (밴 & 픽)



경기승패에 영향을 주었던 장면 - 제3경기 (밴 & 픽)

DWG

DWG은 바텀주도권을 무조건 양보하지 않겠다는 마인드로 고스트(AD, DWG)가 조커픽으로 드레이븐을, 베릴(SUP, DWG)이 판테온을 픽함.

DRX

DRX가 이전밴픽에서 고정적으로 꺼냈던 판테온 대신 케이틀린을 밴함. 이전경기의 상대팀의 픽이었던 레넥톤,니달리를 풀어주고, 칼리스타 릴리아을 픽, 이즈리얼, 세나를 연속으로 밴, 탑쉔을 픽하면서, 탑주도

권보다는 바텀주도권을 가져오려고 한듯.

=> 담원과 DRX의 바텀주도권 경쟁싸움

경기승패에 영향을 주었던 장면 - 제3경기

주요장면 시간	상황	선정이유
3:44:32	- DRX3인(릴리아, 칼리, 쓰레쉬) 2번째용(불용)을 치고있는 상황 (3:3대치) - 드레이븐(AD, DWG)이 오자마자 DWG의 니달 리, 아칼리가 동시에 진입, 후위에 레넥톤(TOP, DWG) 텔 - 아칼리가 불용을 먹고 릴리아(JG, DRX)를 처치 - 쉔(TOP, DRX)가 궁을 칼리스타에 썻으나, 아칼 리 궁으로 쓰레쉬(SUP, DRX), 칼리스타(AD, DRX) 처치	-미드 주도권이 없어진 이렐리아(MID, DRX)가 아칼리, 레넥톤에 의해 처치당함 - 킬스코어 : 7(DWG) VS 0(DRX) - 6000골드 이상 차이

경기 리뷰

- 그동안, 상체위주의 주도권싸움에서 이겨왔던 담원이 아니라, 바텀게임으로 경기를 터뜨린 케이스였다.
- 전령을 용을 먹는 것과 버금가는 혹은 그 이상으로, 골드를 수급하고 타워까지 밀 수 있도록 활용한 것. (담원 정글 canyon 의 장점)
- DRX의 팀 조합에서 릴리아가 과연 좋은 픽이었을까?라는 생각이 듦. 또, 2,3경기에 상대 니달리가 있는데, 굳이 릴리아를 픽한 이유가 무엇일지??
- DRX에서 릴리아의 연구가 미진했다라고 생각

사전과제 2-2)

프로 경기를 보고, 선수들의 실력을 데이터 기반으로 정량적인 측정한다고 했을때, 어떤 지표들이 필요하며 그 지표를 만들기 위해서 필요한 데이터가 무엇이 있을지 정리해주세요.

- 챔프 상성대비 CS, 골드, 경험치지표

예를들어, 애쉬 vs 케이틀린 이라고 한다면, 애쉬와 케이틀린이 바텀 원딜로 붙었을때의 평균 CS, 골드, 경험치지표들의 평균을 기준으로 +,-지표 로 바꾸는 것.



서폿을 제외한 라인의 선수들에게 보다 정확한 라인전의 기준점을 제시할 수 있음.

- 챔프 특성 대비 생존율

KDA

(총 입은 데미지량 / 챔피언에게가한 총데미지량)

OR



라인별 챔프 특성을 반영한 생존율을 알 수 있음

챔피언에게 가한 데미지량

15분마다 (체력+방어력+마법저항력)의 평균

사전과제 2-2)

프로 경기를 보고, 선수들의 실력을 데이터 기반으로 정량적인 측정한다고 했을때, 어떤 지표들이 필요하며 그 지표를 만들기 위해서 필요한 데이터가 무엇이 있을지 정리해주세요.

기타)

- 킬마다, 어시스트한 포지션 지표



다이브, 한타기여도 등을 측정하는데 사용가능

- **특별 골드 지표** 타워골드, 현상금수급



팀 또는 특정라인의 성장을 위해 한타기여 또는 다이브역량 측정하는데 사용가능