

캡스톤 디자인 #3

머신러닝, 딥러닝을 활용한
프리미어리그 Big6 순위 예측 모델



CONTENS

- 1 DATETIME 전처리
- 2 파생변수 생성 (승점)
- 3 FTR 전처리
- 4 시각화 & 분석

DATETIME 전처리

DateTime	HomeTeam	AwayTeam
2021-04-25T12:00:00Z	Wolves	Burnley
2021-04-25T14:00:00Z	Leeds	Man United
2021-04-25T19:00:00Z	Aston Villa	West Brom
2021-04-26T20:00:00Z	Leicester	Crystal Palace
2021-04-30T20:00:00Z	Southampton	Leicester

☑ ISO 8601 형식을 따르는 시간 표기



국제 협정 시간대를 사용하여 영국현지 시간 반영

DATETIME 전처리

DateTime	HomeTeam	AwayTeam
2021-04-25 13:00:00	Wolves	Burnley
2021-04-25 15:00:00	Leeds	Man United
2021-04-25 20:00:00	Aston Villa	West Brom
2021-04-26 21:00:00	Leicester	Crystal Palace
2021-04-30 21:00:00	Southampton	Leicester











☒ 국제 협정 시간대를 따르는 시간 표기



일관된 시간대 사용으로 시계열 데이터 분석에 용의

CONTENS 02

승률 VS 승점

순위	팀	경기수 ▲	승 ▲	패 ▲	무 ▲	승률 ▼
1	 LG	144	86	56	2	0.606
2	 KT	144	79	62	3	0.560
3	 SSG	144	76	65	3	0.539
4	 NC	144	75	67	2	0.528
5	 두산	144	74	68	2	0.521
6	 KIA	144	73	69	2	0.514
7	 롯데	144	68	76	0	0.472
8	 삼성	144	61	82	1	0.427
9	 한화	144	58	80	6	0.420
10	 키움	144	58	83	3	0.411

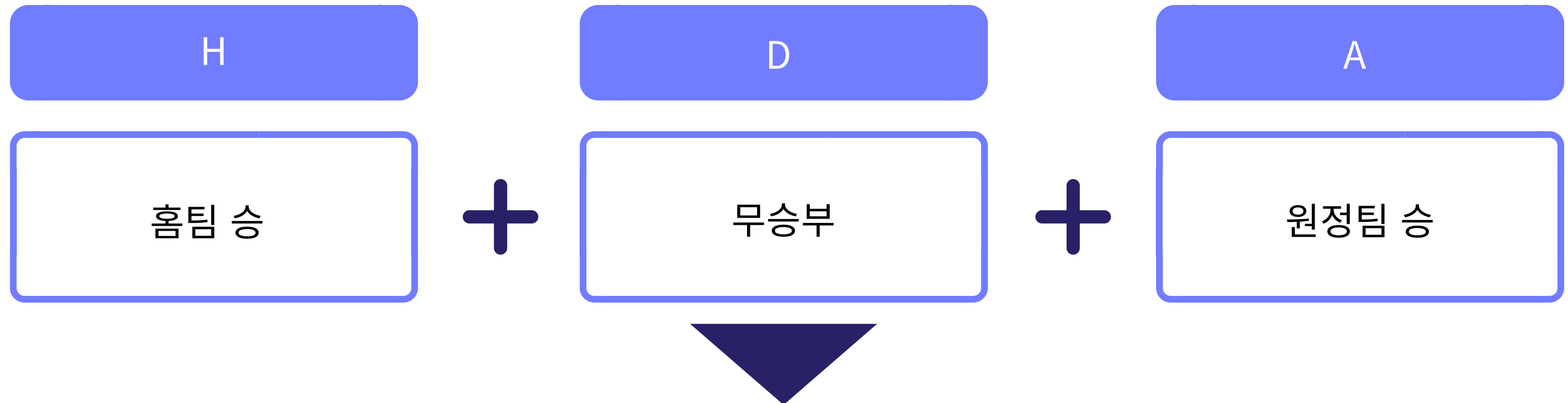
야구 순위 = 승률

순위	팀	경기수	승점 ▼	승 ▲	무 ▲	패 ▲
1	 맨체스터 시티 FC	38	89	28	5	5
2	 아스널 FC	38	84	26	6	6
3	 맨체스터 유나이티드 FC	38	75	23	6	9
4	 뉴캐슬 유나이티드 FC	38	71	19	14	5
5	 리버풀 FC	38	67	19	10	9
6	 브라이튼 앤 호브 알비온 FC	38	62	18	8	12
7	 아스톤 빌라 FC	38	61	18	7	13
8	 토트넘 홋스퍼 FC	38	60	18	6	14
9	 브렌트포드 FC	38	59	15	14	9
10	 풀럼 FC	38	52	15	7	16

축구 순위 = 승점

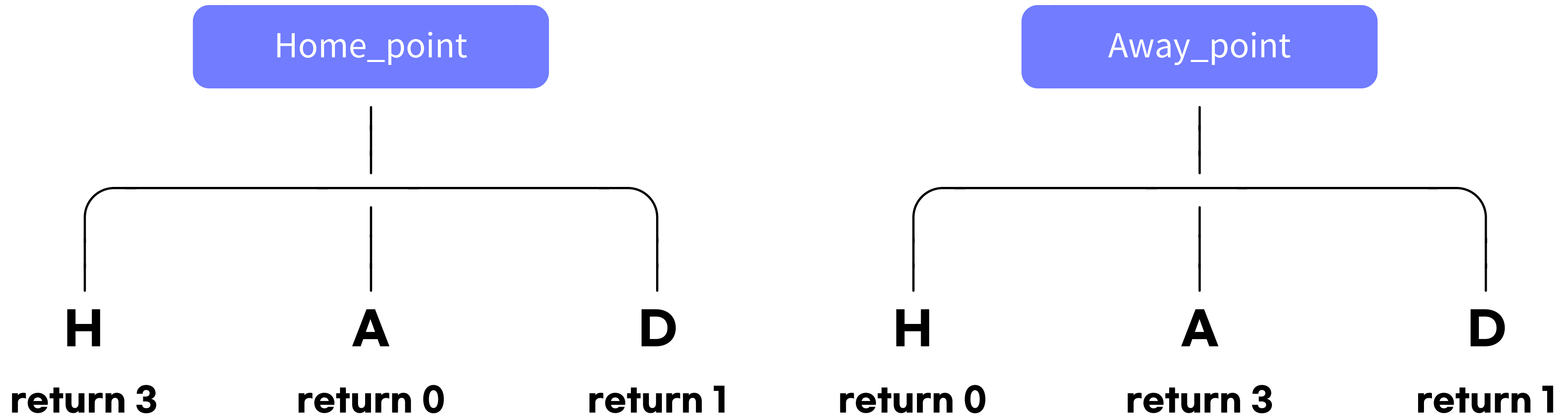
파생변수 생성 (승점)

FTR = Full Time Result



FTR를 활용하여 Big6 팀들간 경기의 승점 계산

파생변수 생성 (승점)



☑ 왜 두 함수의 연산은 서로 반대인가?

파생변수 생성 (승점)

Home Team

홈팀인 아스널
FTR은 A
홈팀인 아스널은 패배

Away Team

원정팀인 아스널
FTR은 A
원정팀인 아스널이 승리

	HomeTeam	AwayTeam	FTR
2xxx-xx-xx	Arsenal	Man United	A
2xxx-xx-xx	Man City	Arsenal	A

파생변수 생성 (승점)

Arsenal_Points

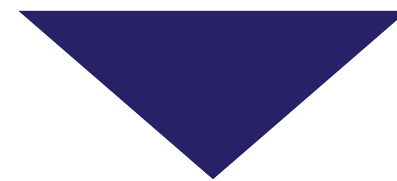
Chelsea_Points

Liverpool_Points

Man City_Points

Man United_Points

Tottenham_Points



각 팀의 승점 측정

파생변수 생성 (승점)

Arsenal_Points	Chelsea_Points	Liverpool_Points	Man City_Points	Man United_Points	Tottenham_Points
279	327	299	243	324	213
279	327	299	243	327	213
279	330	299	243	327	213
282	330	299	243	327	213
282	330	302	243	327	213

☒ 전체 경기결과를 한번에 계산하는 것이 타당한가?

파생변수 생성 (승점)

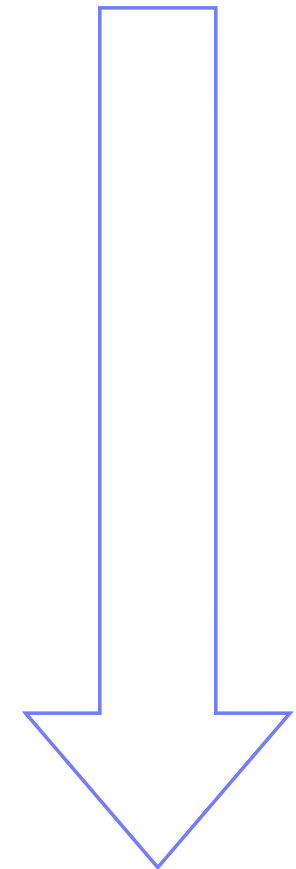
시즌별로 데이터를 나눈 후 승점을 계산한다

00-01시즌부터 20-21시즌까지 나눈다

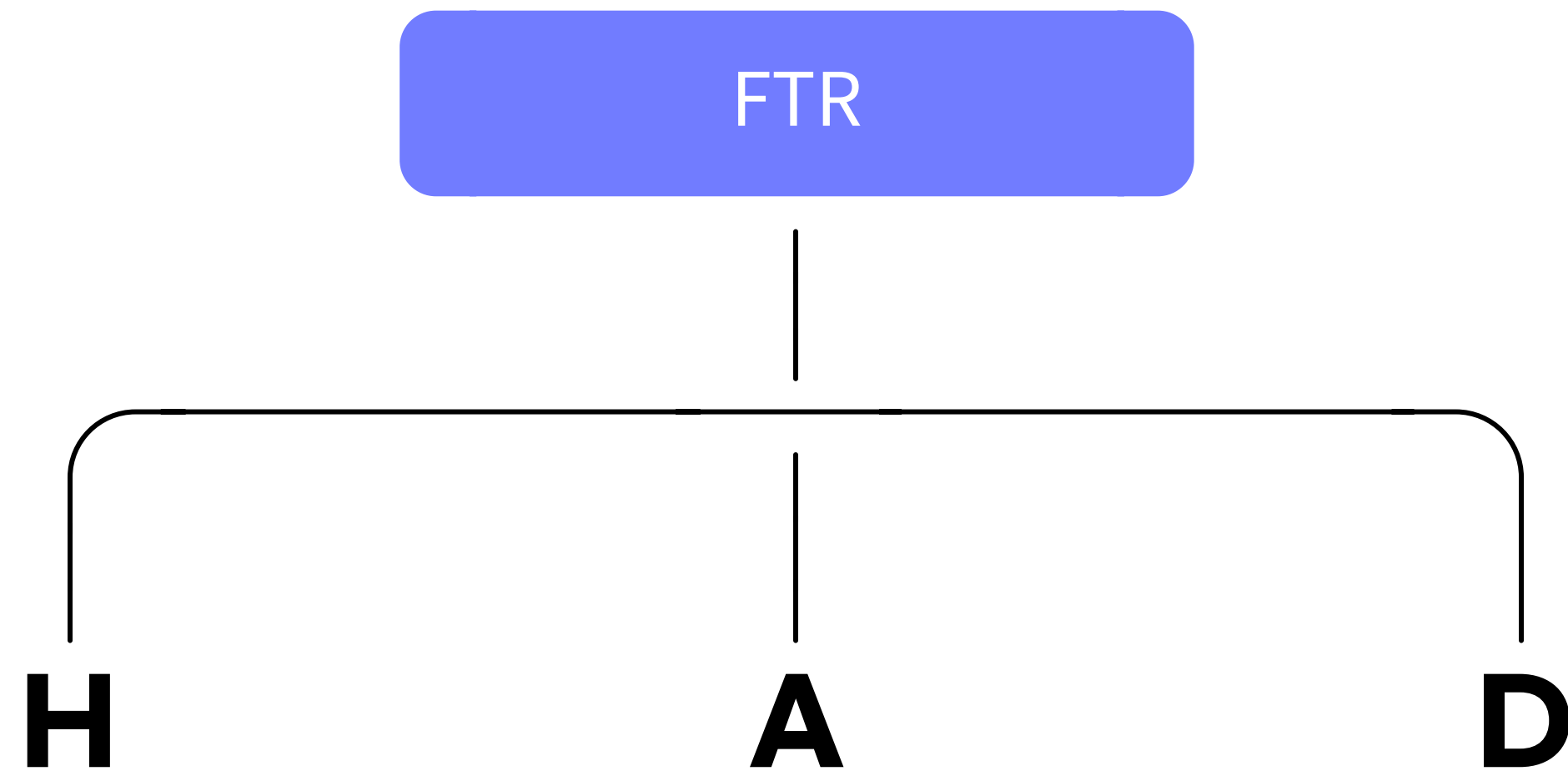
나눈 시즌들을 각각 csv파일로 변환한다

csv파일을 모두 합쳐 season.csv 생성

START

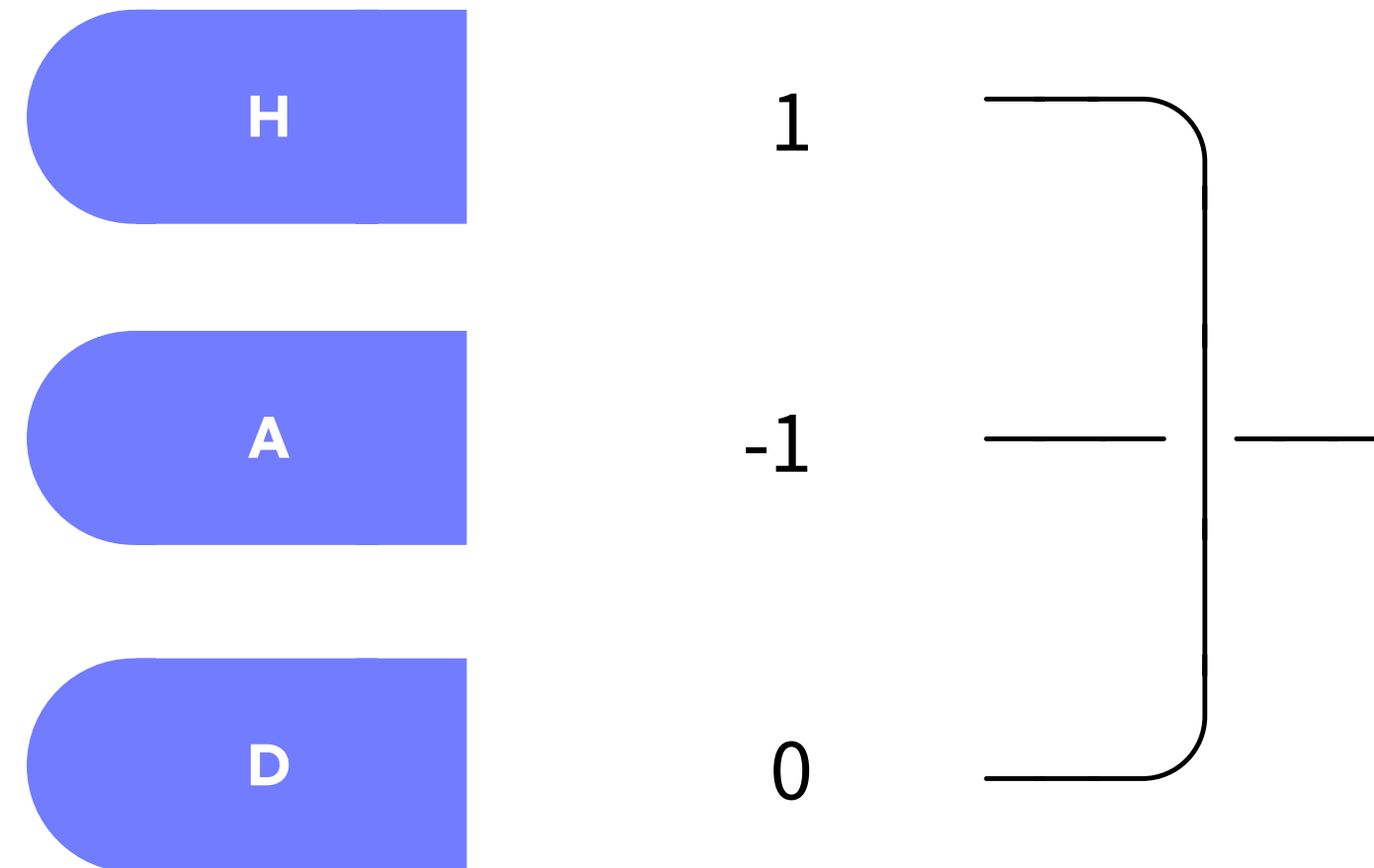


FTR 전처리



☑ dtype : object 전처리 필요

FTR 전처리



**홈팀이 이겼을때 더 가중치를
주는 방향으로 전처리 진행**

**THANK
YOU**

