

다변량통계분석 기말 프로젝트

20195262 장예진, 20195263 주고은

1. 프로젝트 배경과 주제:

1) 주제정의: 약점 분석을 통한 e-sports 대회 출전을 위한 최적의 포켓몬 조합 탐색

2) 환경 분석 및 해당 주제를 선정 이유

: 매년 e-sports의 시장이 성장하면서 이에 대한 프로 리그의 전략적 분석의 중요성도 높아지고 있다. 그 중 포켓몬스터는 전 세계 미디어 믹스 총매출 1위에 달하는 세계적 문화 현상이라 불릴 만큼의 글로벌 콘텐츠이다. 매년 각국의 챔피언 트레이너들을 모아 대회를 열기 때문에 트레이너들의 개인 실력 뿐만이 아니라 전략적인 팀 분석이 승리를 이끄는 중요한 요소이다. 각 트레이너는 최대 6마리의 포켓몬으로 팀을 꾸릴 수 있으며 18개의 주타입, 18개의 보조타입, 7종류의 능력치 변수, 그 외의 포켓몬들의 성격, 특성과 같은 많은 변수를 고려해야하는 통계적 분석이 필수적이다. 그 예시로 포켓몬 월드챔피언쉽 2014 때 한국의 박세준 선수가 대회에 약해서 잘 사용되지 않는 파치리스로 팀의 약점을 보완하여 세계 1위가 되어 이슈가 된 사건이 있다. 따라서 우리는 군집분석을 통해 포켓몬의 약점을 고려한 전략적 팀 구성을 탐색해보고자 한다.

3) 연구문제 정의:

1. 대회에서는 레벨이 50으로 고정되는데 강한 포켓몬 조합이 최적의 조합인가?
 2. 특정타입에 대한 약점점수를 고려하여 군집분석을 진행하면 최적의 포켓몬 조합을 생성할 수 있는가?
- 상성과 능력치를 고려한 약점 점수
 - 평가 기준: 약점점수 = 1+각 상성 점수의 총합. 총합이 낮을수록 약점 점수가 낮음

2. 데이터 소개와 변수 정의:

1) 데이터 소개:

- The complete pokemon dataset
: 32개의 컬럼과 801개의 데이터로 이루어진 포켓몬 데이터셋
- 데이터 출처: kaggle

2) 변수 정의:

1. 상성 변수:

- against_bug: 벌레 타입 상성
- against_dark: 악 타입 상성
- against_dragon: 드래곤 타입 상성
- against_electric: 전기 타입 상성
- against_fairy: 페어리 타입 상성
- against_fight: 격투 타입 상성

```
df1.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 801 entries, 0 to 800
Data columns (total 32 columns):
#   Column              Non-Null Count  Dtype
---  -
0   abilities            801 non-null    object
1   against_bug          801 non-null    float64
2   against_dark         801 non-null    float64
3   against_dragon       801 non-null    float64
4   against_electric     801 non-null    float64
5   against_fairy        801 non-null    float64
6   against_fight        801 non-null    float64
7   against_fire         801 non-null    float64
8   against_flying       801 non-null    float64
9   against_ghost        801 non-null    float64
10  against_grass        801 non-null    float64
11  against_ground       801 non-null    float64
12  against_ice          801 non-null    float64
13  against_normal       801 non-null    float64
14  against_poison       801 non-null    float64
15  against_psychic      801 non-null    float64
16  against_rock         801 non-null    float64
17  against_steel        801 non-null    float64
18  against_water        801 non-null    float64
19  attack              801 non-null    int64
20  base_total          801 non-null    int64
21  defense             801 non-null    int64
22  hp                  801 non-null    int64
23  name                801 non-null    object
24  percentage_male     703 non-null    float64
25  sp_attack           801 non-null    int64
26  sp_defense          801 non-null    int64
27  speed               801 non-null    int64
28  type1               801 non-null    object
29  type2               417 non-null    object
30  generation          801 non-null    int64
31  is_legendary        801 non-null    int64
dtypes: float64(19), int64(9), object(4)
memory usage: 200.4+ KB
```

- against_fire: 불꽃 타입 상성
- against_flying: 비행 타입 상성
- against_ghost: 고스트 타입 상성
- against_grass: 풀 타입 상성
- against_ground:땅 타입 상성
- against_ice: 얼음 타입 상성
- against_normal: 노말 타입 상성
- against_poison: 독 타입 상성
- against_psychic: 에스퍼 타입 상성
- against_rock: 바위 타입 상성
- against_steel: 강철 타입 상성
- against_water: 물 타입 상성

2. 능력치 변수:

- Attack: 공격
- Base_total: 스탯(능력치) 총합
- Defense: 방어
- Hp: 체력
- Sp_attack: 특수공격
- Sp_defense: 특수방어
- Speed: 스피드

3. 타입(속성) 변수:

- Type1: 주타입
- Type2: 보조타입
bug, dark, dragon, electric, fairy, fight, fire, flying, ghost, grass, ground, ice, normal, poison, psychic, rock, steel, water, Nan(보조타입이 없는 경우)

4. 약점 점수: 1+ 모든 상성 점수의 총합

3) 종속 변수와 독립 변수 정의:

- 종속 변수: 약점점수
- 독립 변수: 이름,타입,상성,능력치 변수

3.데이터 탐색

1)특성변수 abilities

: 각 포켓몬의 고유한 특성을 나타냄. 본 연구에선 상성과 능력치만을 다루기에 제외시킴.

2)상성변수 against_

: 각 타입에 대한 포켓몬의 상성을 나타냄. 0,0.5,1.0,2.0 으로 이루어져있으며 숫자가 클수록 해당 타입에 약함을 나타냄.

3)능력치 변수 base_total, attack, defense, sp_attack, sp_defense, speed

:포켓몬의 능력치를 나타내며 총 능력치 합,공격,방어,특수공격,특수방어, 스피드를 나타냄.

4)이름변수 name

5)타입변수 type1, type2

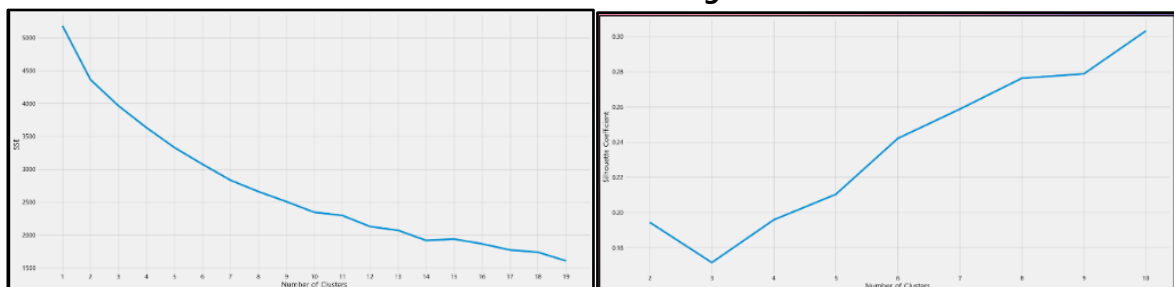
:벌레, 악, 드래곤, 전기, 요정, 격투, 불, 비행, 고스트, 풀, 땅, 얼음, 노말, 독, 에스퍼, 바위, 강철, 물 18 가지의 타입으로 이루어져있음. 각 포켓몬은 type1 과 type2 를 배정받으며 type1 만 있는 단일타입에 경우 type2 는 nan 값. 결측치 제거를 위해 '없음'으로 대체함.

<타입 상성표>

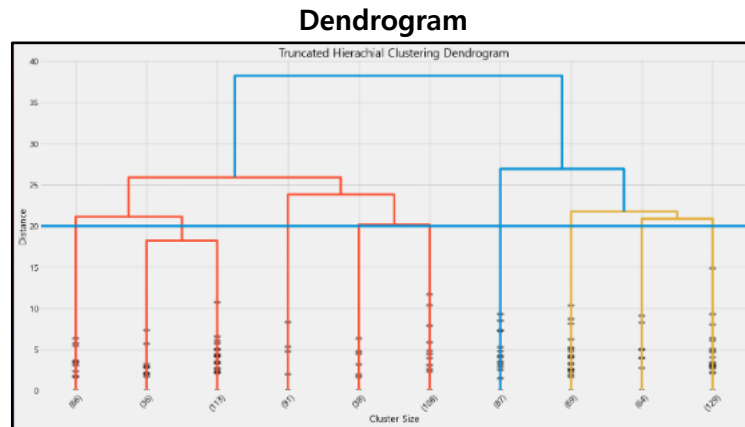
상대포켓몬 타입	효과가 굉장하다	효과가 별로인 듯하다	효과가 없다
노말	격투		고스트
불꽃	물,땅,바위	불꽃,물,얼음,벌레,강철,페어리	
물	전기,불	불꽃,물,얼음,강철	
전기	땅	전기,비행,강철	
풀	불꽃,얼음,독,비행,벌레	물,전기,물,땅	
얼음	불꽃,격투,바위,강철	얼음	
격투	비행,에스퍼,페어리	벌레,바위,악	
독	땅,에스퍼	물,격투,독,벌레,페어리	
땅	물,물,얼음	독,바위	전기
비행	전기,얼음,바위	물,격투,벌레	땅
에스퍼	벌레,고스트,악	격투,에스퍼	
벌레	불꽃,비행,바위	물,격투,땅	
바위	물,물,격투,땅,강철	노말,불꽃,독,비행	
고스트	고스트,악	독,벌레	노말,격투
드래곤	얼음,드래곤,페어리	불꽃,물,전기,물	
악	격투,벌레,페어리	고스트,악	에스퍼
강철	불꽃,격투,땅	노말,물,얼음,비행,에스퍼,벌레,바위,드래곤,강철,페어리	독
페어리	독,강철	격투,벌레,악	드래곤

타입에 따른 상성을 살펴보기 위해 공식 타입 상성표를 확인하였다. 기본 공격을 1 이라 할 때 '효과가 굉장하다'의 경우 2~4 배의 데미지, '효과가 별로인 듯하다'의 경우 0.5 의 데미지, '효과가 없다'의 경우 0 의 데미지를 상대에게 부여한다.

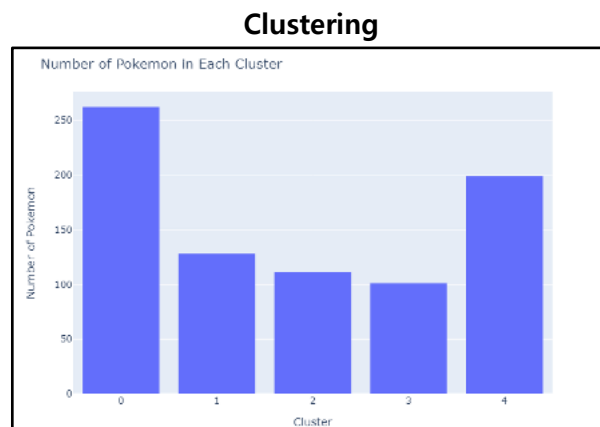
K-means clustering



군집의 수에 따른 SSE 를 나타내는 팔꿈치 그래프를 통해 k 값이 5 에서 확실하게 감소하고 실루엣 계수 또한 군집의 수가 5 이상일 때 군집화가 잘 이루어짐으로 k-means clustering 을 위한 k 값을 5 로 설정하였다.



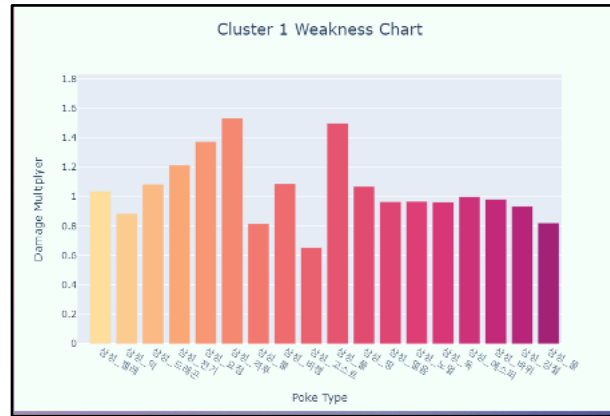
관측치의 군집화를 통해 형성된 그룹과 이들의 유사성 수준을 파악하기 위해 덴드로그램을 생성하였다. y 축은 군집간의 거리를 나타내며 수직선이 길수록 두 군집간의 거리가 멀어 그 선에 의해서 유사성이 덜함을 의미한다. 파란색 선은 baseline 을 의미하여 해당 군집이 명확하게 분리되는 중요한 거리 임계값을 나타낸다. 데이터 포인트들을 서로 유사한 그룹으로 분류하여 하단 괄호안의 숫자를 확인하여 각 분기점 아래에 있는 데이터포인트를 확인하였고 36,113 번의 cluster size 를 가진 클러스터가 가장 높은 유사도를 보임을 확인할 수 있었다.



0~4 번까지의 5 개의 클러스터를 설정한 후 각 클러스터에 속하는 포켓몬의 수를 확인하였으며 0 번인 cluster1 에 가장 많은 데이터가 존재하는 것을 확인하였다.

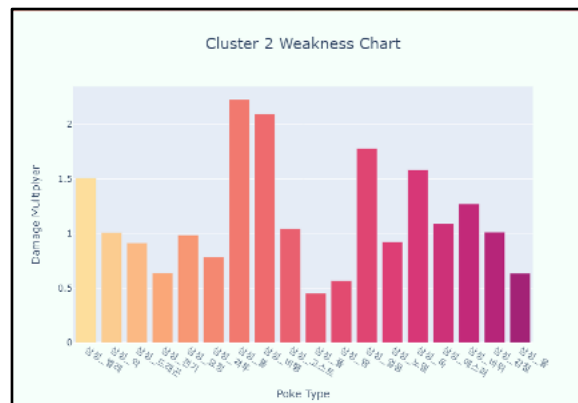
각 군집의 특징을 분석하고 군집에 속하는 포켓몬 중 legendary=0 이며 (전설의 포켓몬 아님) 총 능력치의 합 base_stat 이 600 이상인 포켓몬 5 개를 추출하였다. base_stat 을 600 으로 설정한 이유는 포켓몬에겐 총 능력치의 합인 종족값이 있는데 종족값이 600 이상일 때 일반적으로 대결에 사용 할 만한 포켓몬으로 판단되기 때문이다. 포켓몬의 특성을 고려하며 종족값이 기준치보다 낮아도 승률을 올릴 수 있으나 본 연구에선 abilities 변수를 제외시켜 분석하기 때문에 기준치를 600 으로 설정하였다.

Cluster1



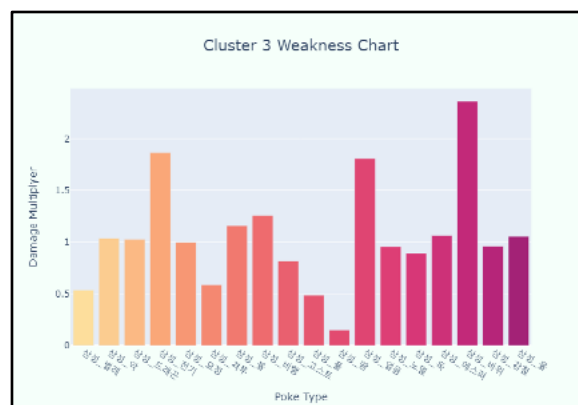
군집의 특성을 분석하니 풀, 격투, 페어리, 전기 타입에 약하며 고스트,불,물 타입에 강한 특성을 보임을 확인하였다. 추출된 포켓몬은 거북왕, 대쌍이, 게을킹, 약어리, 그레닌자였으며 공통적으로 물 타입임을 확인하였다.

Cluster2

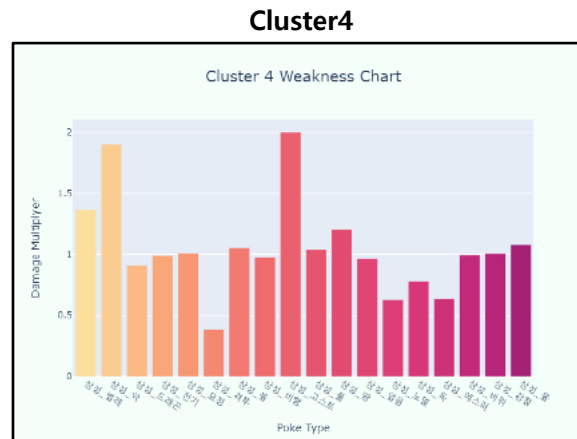


군집의 특성을 분석하니 불, 비행, 얼음타입에 약하며 풀,땅,전기 타입에 강한 특성을 보임을 확인하였다. 추출된 포켓몬은 핫삼, 이상해꽃, 뿌사이저, 나무킹, 헤라크로스였으며 공통적으로 풀,벌레 타입임을 확인하였다.

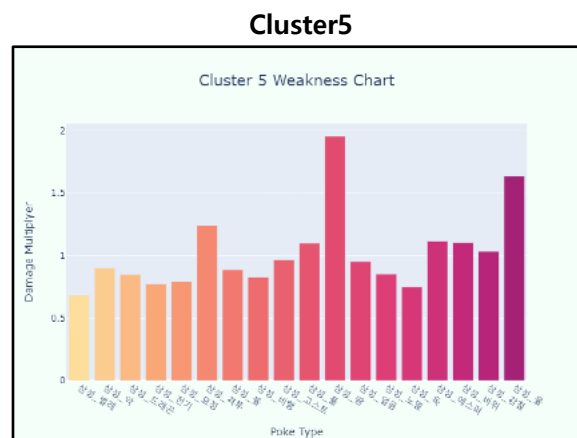
Cluster3



군집의 특성을 분석하니 바위, 얼음, 전기 타입에 약하며 땅, 격투, 벌레타입에 강한 특성을 보임을 확인하였다. 추출된 포켓몬은 리자몽, 가라도스, 프레타, 망나뇽, 보만다였으며 공통적으로 드래곤타입임을 확인하였다.



군집의 특성을 분석하니 고스트, 악, 벌레 타입에 약하며 격투, 노말, 에스퍼, 독에 강한 특성을 보임을 확인하였다. 추출된 포켓몬은 메타그로스, 팬텀, 가디안, 후딘, 엘레이드였으며 공통적으로 에스퍼 타입임을 확인하였다.



군집의 특성을 분석하니 땅, 물, 격투에 약하며 벌레, 독에 강한 특성을 보임을 확인하였다. 추출된 포켓몬은 루카리오, 강철톤, 전룡, 헬가, 번치코였으며 공통적으로 격투타입임을 확인하였다.

4. 결론 및 결과

각 군집에서 추출 된 포켓몬들 중 약점 점수가 적은 1 위들을 뽑아 타입과 상성을 보완해주는 최적의 조합을 예측하였다. 결과적으로 cluster1 의 거북왕, cluster2 의 핫삼, cluster3 의 리자몽, cluster4 의 메타그로스, cluster5 루카리오로 이루어진 조합을 생성하였다.

Name	타입 1	타입 2	체력	공격	방어	특수공격	특수방어	스피드	총스탯
거북왕	물	없음	79	103	120	135	115	78	630
햇삼	벌레	강철	70	150	140	65	100	75	600
리자몽	불	비행	78	104	78	159	115	100	634
메타그로스	강철	에스퍼	80	145	150	105	110	110	700
루카리오	격투	강철	70	145	88	140	70	112	625

서로 다른 타입으로 각자의 약점을 보완해주며 메타그로스와 거북왕은 방어와 지속력을 유지하고 햇삼, 루카리오, 리자몽은 높은 공격과 스피드로 상대를 압박할 수 있어 전체 팀의 상호보완적인 시너지를 극대화시키는 결과를 예측할 수 있었다.

이후 약점과 상성만이 아닌 능력치까지 고려한 조합을 생성해보고자 하였고 능력치를 고려한 4 가지의 역할을 배정하여 이에 맞는 포켓몬들의 조합을 생성보았다.

물리공격을 담당하는 'Physical Sweeper', 특수공격&스피드를 담당하는 'Special Sweeper', 물리공격&특수공격&스피드를 담당하여 올라운더를 수행하는 'Mixed Sweeper', 체력&방어&특수방어를 담당하는 'Tanker' 4 가지로 분류하였다. 역할에 필요한 스탯의 합산을 통해 각 군집에서 순위를 확인하였고 전체 군집에서 각 역할이 상위권인 포켓몬들로 조합을 생성하였다. 그 결과 cluster1 의 마기라스가 'Tanker', cluster2 의 나무킹이 'Mixed Sweeper', cluster3 의 프테라가 'Physical Sweeper', cluster4 의 후딘이 'Special Sweeper', cluster5 의 루카리오가 Mixed Sweeper 를 수행하는 조합이 만들어졌으며 첫번째 예측 조합과는 확연히 다른 결과를 확인할 수 있었다.

Name	타입 1	타입 2	체력	공격	방어	특수공격	특수방어	스피드	총스탯
마기라스	바위	악	100	164	150	95	120	71	700
나무킹	풀	없음	70	110	75	145	85	145	630
프테라	바위	비행	80	135	80	70	95	150	615
후딘	에스퍼	없음	55	50	65	175	105	150	600
루카리오	격투	강철	70	145	88	140	70	112	625

균형적인 스탯으로 팀의 역할 분담이 잘 이루어진 것을 확인할 수 있으나 타입 속성에서 불타입과 물타입이 없어 리스크가 생길 가능성이 있음을 예측할 수 있었다. 하지만 게임에 참가하는 포켓몬 트레이너는 최대 6 마리의 포켓몬을 소지할 수 있으므로 본 연구에서 도출한 5 개의 군집의 포켓몬을 포함하여 상황과 판단에 따라 마지막 포켓몬을 탐색한다면 리스크를 완화시킬 수 있을 것이다. 예시로 2 번째 조합의 불타입 포켓몬의 부재를 보완하기 위해 불타입 리자몽을 추가하는 방법이 있다.

정리하자면 본 연구에선 군집분석을 활용하여 각 군집의 상위 점수에 따른 조합을 선정하였으며 각 군집에 특성에 대한 필터링을 통해 종합 스탯 상위 점수에 따른 최적의 조합을 선정하였다. 이를 통해 능력치만 높은 포켓몬을 선정하는것이 최적의 방법이 아니며 각 상성과 타입을 고려한 최적의 시너지 조합의 도출이 중요하며 분석을 통해 가능함을 확인할 수 있었다.

5. 한계 및 보완점

일반적인 포켓몬 대회 기준에 맞추고자 최대한 일반화시키려 노력하였지만, 각 대회의 기준이 조금씩 상이하여 최적의 기준을 가지는 팀 조합이 약간의 오차가 있을 가능성이 있다.

또한 포켓몬은 각기 다른 성격, 특성, 기술, 소지품에 따라 배틀 방식이 달라지기 때문에 이러한 변수까지 고려한 세심한 분석의 보강이 필요하다.

실질적인 예측을 평가하기 위하여 머신러닝으로 상성 가중치와 각 능력치를 사용한 가상의 배틀 예측 모델을 개발하여 추후 보강해볼 예정이다. 데이터 분석을 통한 배틀 승률 예측은 프로 리그 뿐만이 아니라 일반 게임 이용자들이 쉽게 게임을 즐길 수 있도록 접근성을 향상시키고 팀을 꾸릴 때 승률을 올릴 수 있는 편의성을 도모할 수 있을 것이라 기대한다.

6. 참고자료 출처

1) Win Prediction in Esports: Mixed-Rank Match Prediction in Multi-player Online Battle Arena Games (e 스포츠에서의 승리 예측: 멀티 플레이어 온라인 배틀 아레나 게임에서의 혼합 순위 경기 예측)

<https://arxiv.org/abs/1711.06498>

2) "710 개 포켓몬 성격·능력 파악해야 상대를 분석하는 힘 기를 수 있어"_포켓몬 세계대회 우승한 박세준 인터뷰

<https://weekly.chosun.com/news/articleView.html?idxno=7557>

3) 2014 포켓몬 월드 챔피언십 결승_유튜브

<https://www.youtube.com/watch?v=bH0GSyyixcQ>