

건강검진 결과의 이해와 활용도 향상을 위한

검진자 집단분류 및 맞춤 진료 제시

2조 김정준, 박상은, 박민규, 박정빈, 이현재

문제 정의

■ 분석 배경: 기존 건강검진 결과지의 문제점

• 모호함

- 종합판정 결과가 구체적이지 않음
- 자신의 건강상태가 어느 위치에 속하는지 알 수 없음

• 후속조치가 없음

- 건강검진 후 어떠한 진료를 받아야 하는지 알 수 없음

"건강검진 이상소견 발견해도 치료 연계 느슨...사후관리 강화 필요"

건보공단 국가검진 효율화 방안 연구, 의료진 수가검진항목 조정 제안

기사입력시간 21-01-16 09:50

직장인 10명 중 7명, 건강검진 결과지 "이해 못했다"

장진숙 기자 | 승인 2020.03.26 13:17 | 댓글 0

1)

일반건강검진 결과통보서			
수검자 성명		주민등록번호	
검진일	2022.09.30	검진장소	<input checked="" type="checkbox"/> 내원 <input type="checkbox"/> 출장
건강검진 종합소견			
판정 - <input type="checkbox"/> 정상 <input checked="" type="checkbox"/> 정상B(경계)			
<input type="checkbox"/> 일반 질환의심 <input type="checkbox"/> 고혈압·당뇨병 질환의심 <input type="checkbox"/> 유질환자			

▲ 현 건강검진
결과통보서 형태

2)

건강검진 데이터 군집화를 통해 그룹별 특성 파악

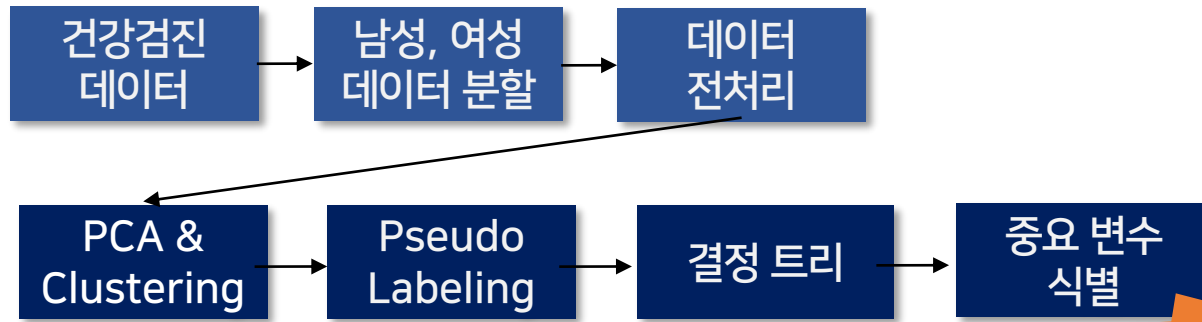
1) 장진숙, '직장인 10명 중 7명, 건강검진 결과지 "이해 못했다"', 데이터숨, 2020.03.26

2) 정승원, "건강검진 후 이상소견 발견해도 치료 연계 느슨...사후관리 강화 필요", MEDI:GATE NEWS, 2021.01.16

분석 과정

■ 이전 진행계획

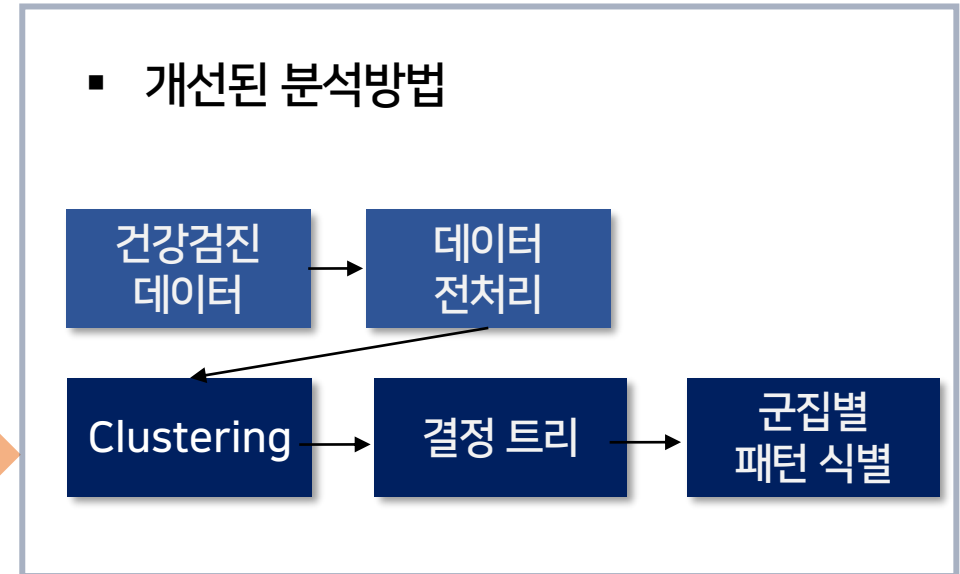
1. 중요 변수 식별



2. 중요 변수들만을 사용해서 그룹별 패턴 식별

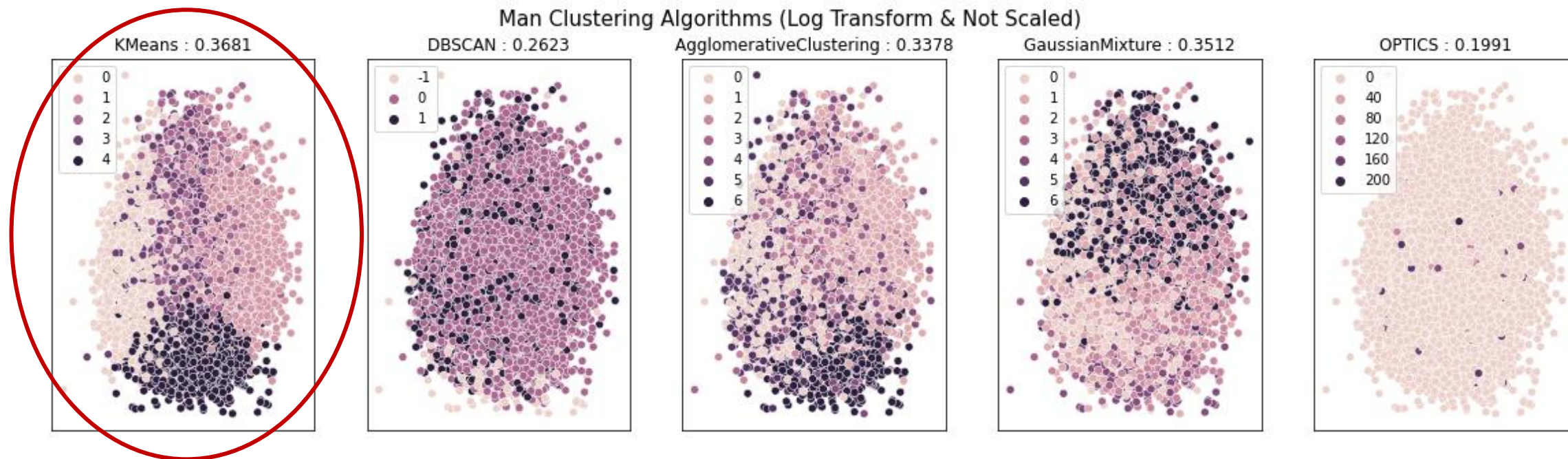


■ 개선된 분석방법



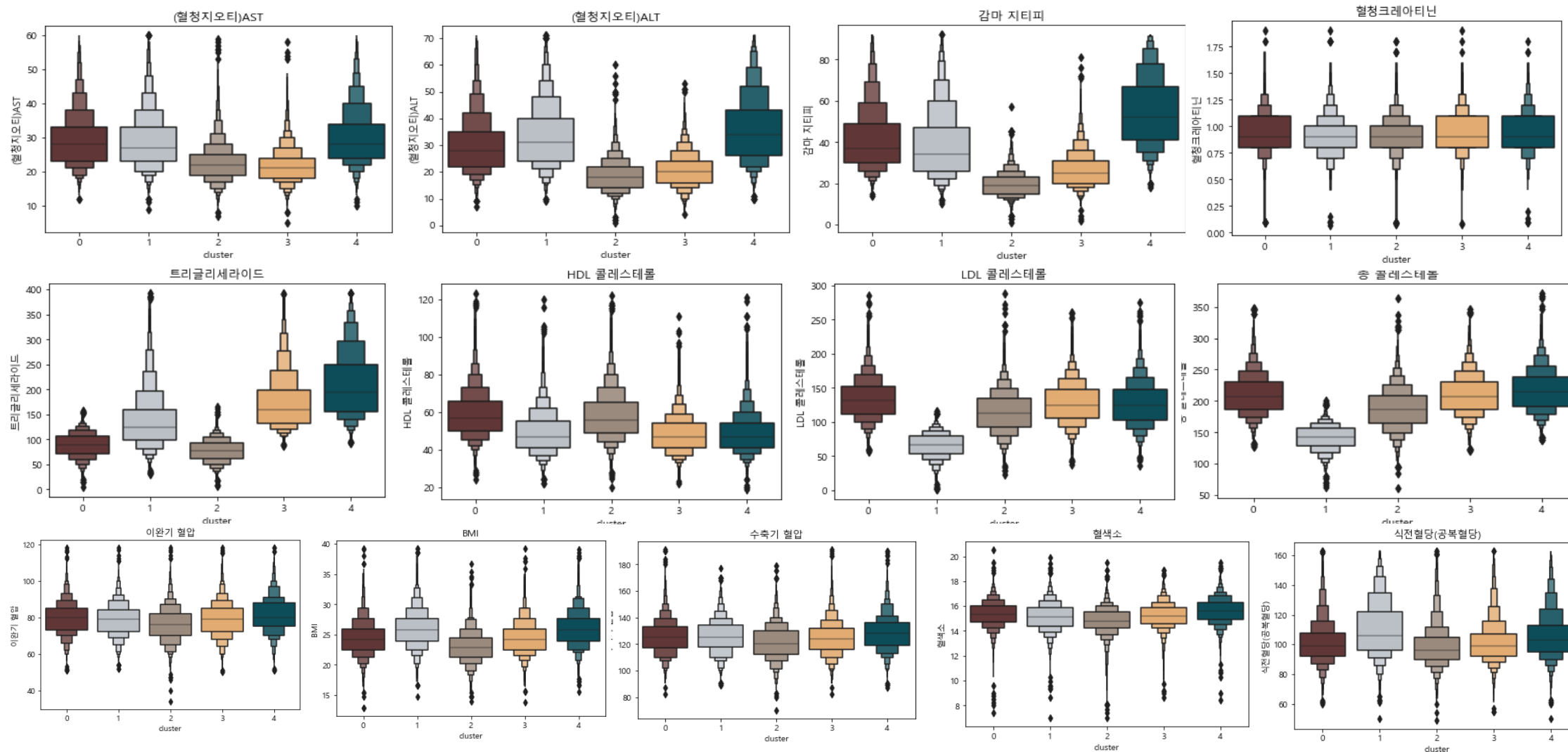
클러스터링 결과

- 휴리스틱 접근
 - Log Transform, Scaled 여부를 바꿔가며 여러 클러스터링 방법 수행, 실루엣 계수가 가장 높은 클러스터 선정
 - > Log Transform을 수행하고, Scaled은 수행하지 않은 Kmeans 클러스터 선정



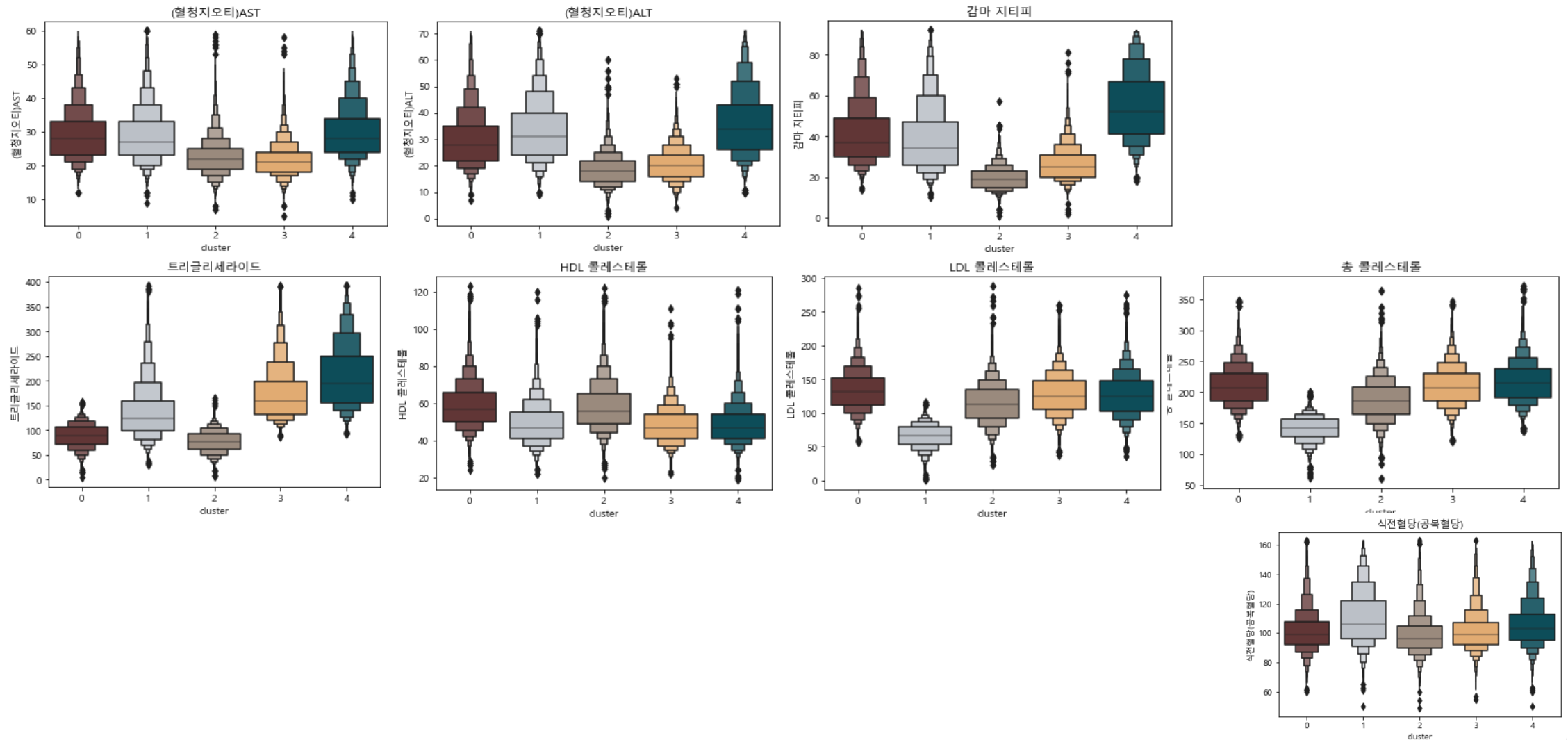
2. 분석 결과

클러스터링 결과



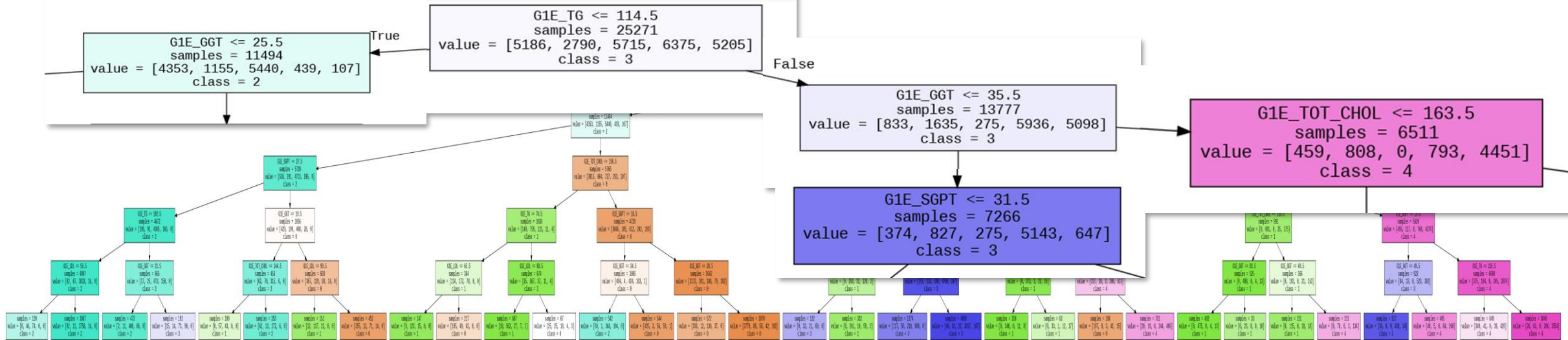
2. 분석 결과

클러스터링 결과



클러스터링 결과

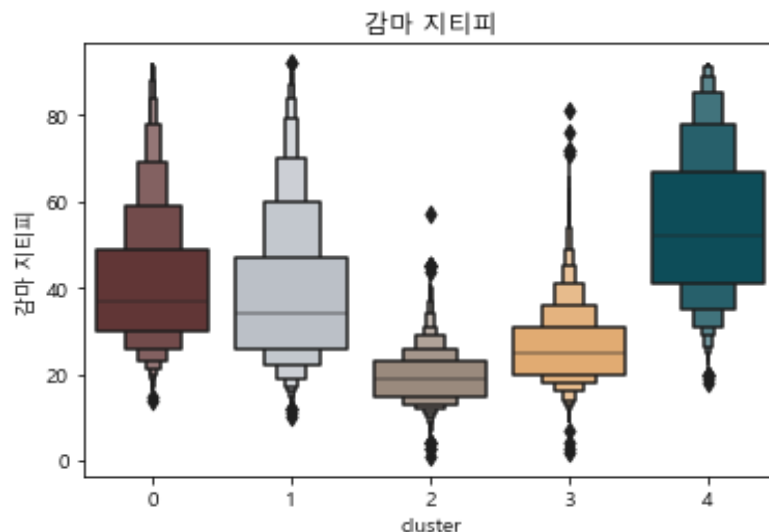
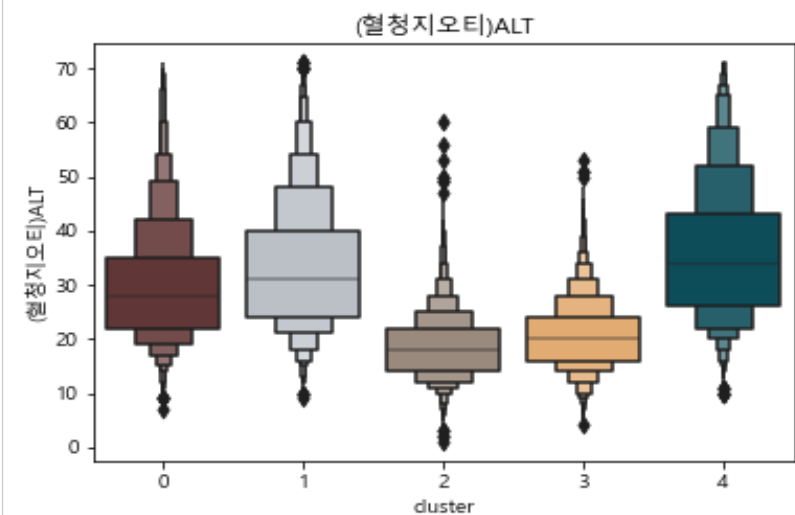
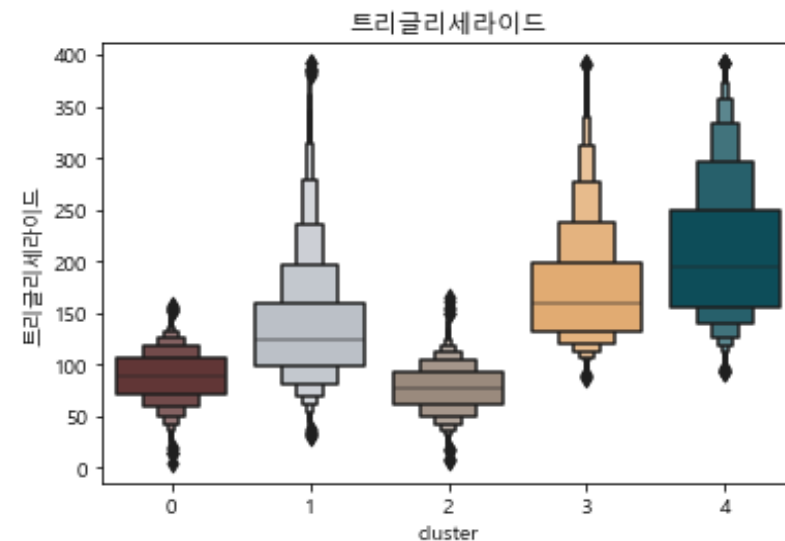
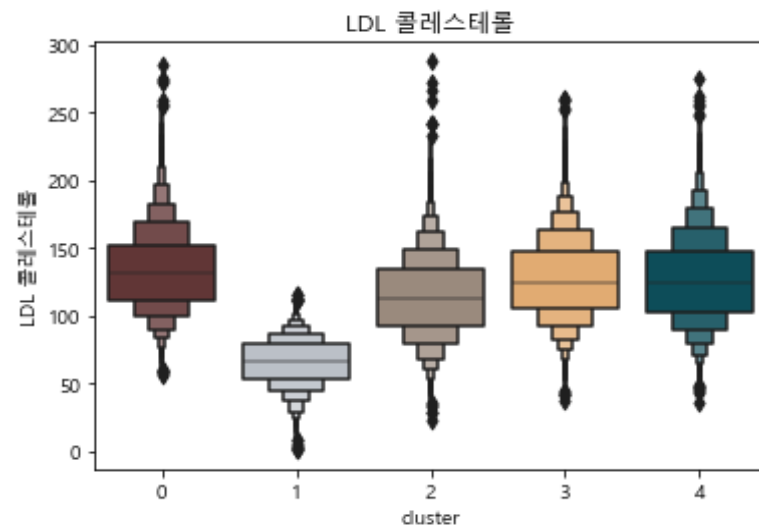
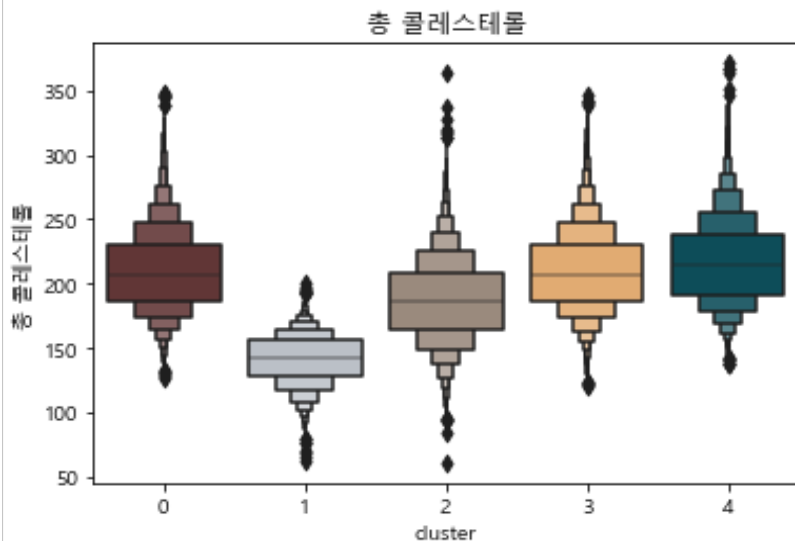
결정 트리



- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4

- G1E_TG(트리글리세라이드): 중성지방, 이상지질혈증에 관여
- G1E_GGT(감마지티피): 간장질환에 관여
- G1E_TOT_CHOL(총콜레스테롤): 이상지질혈증에 관여
- G1E_SGPT(혈청지피티, ALT): 간장질환에 관여
- G1E_LDL(LDL콜레스테롤): 이상지질혈증에 관여

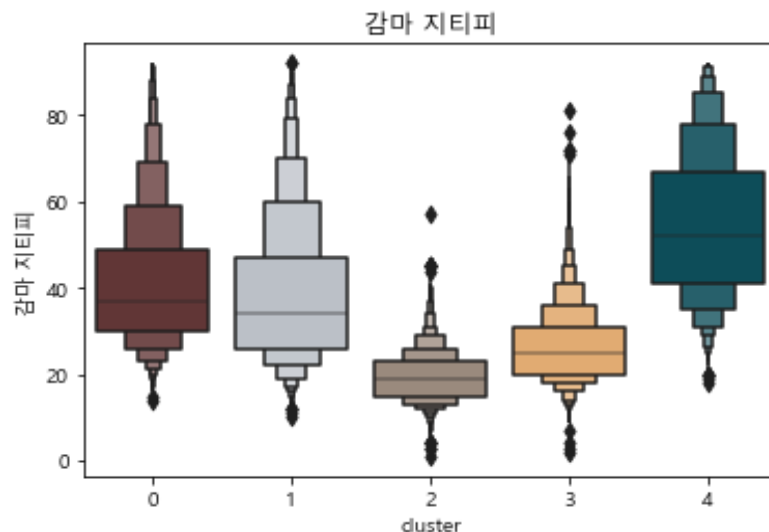
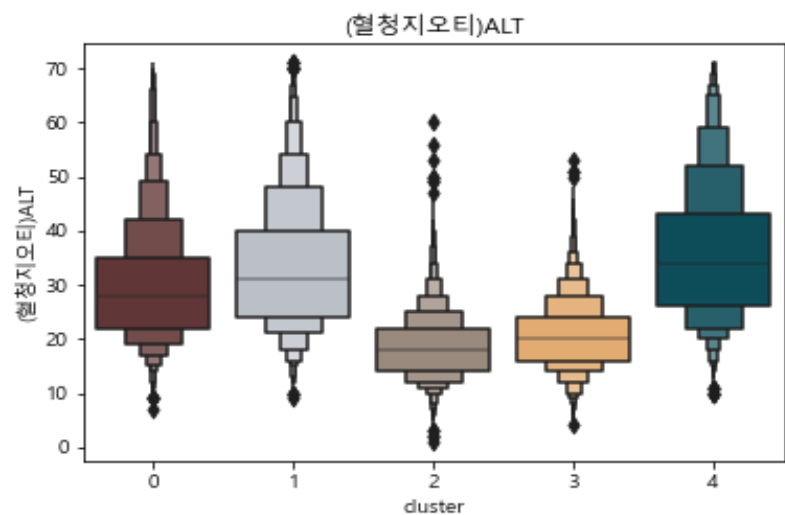
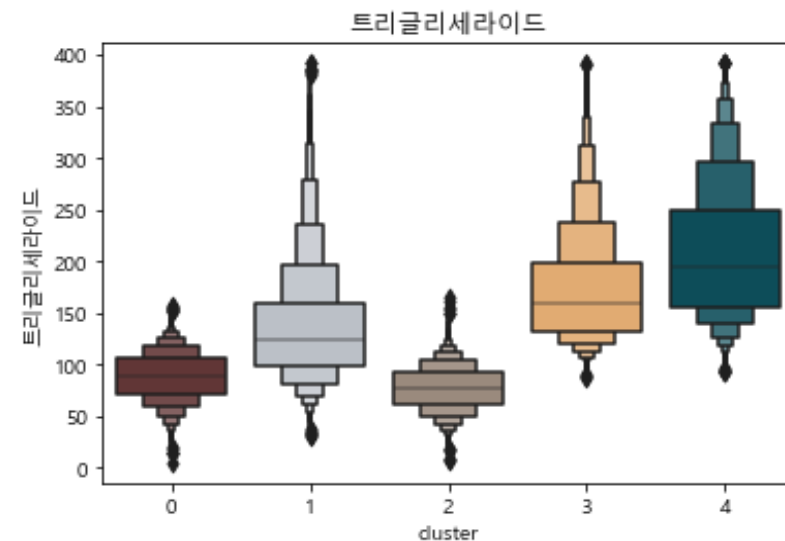
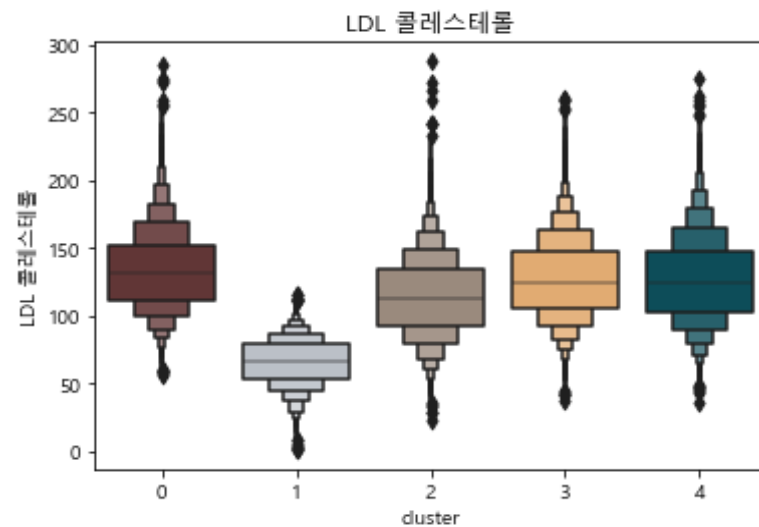
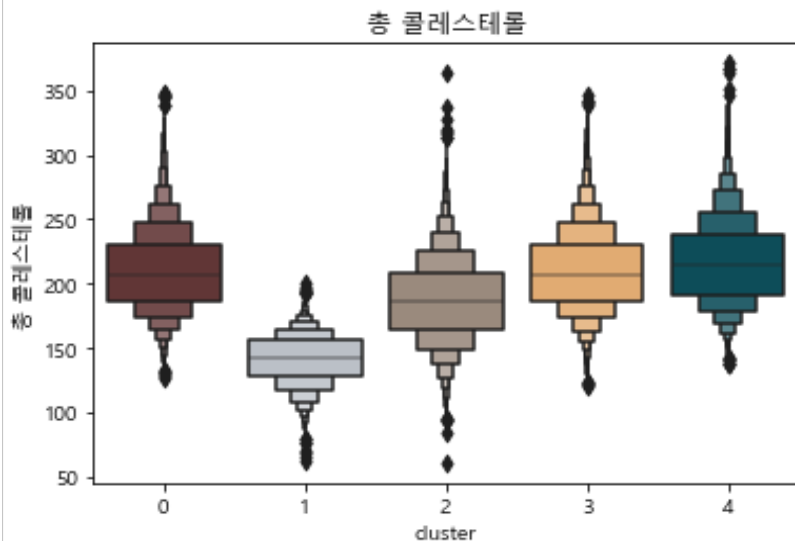
클러스터링 결과



군집 패턴 식별 결과 :

군집 2: 정상(건강) 집단
군집 0: 높은 간수치 수준
군집 1: 좋은 콜레스테롤 수준,
높은 간수치 수준

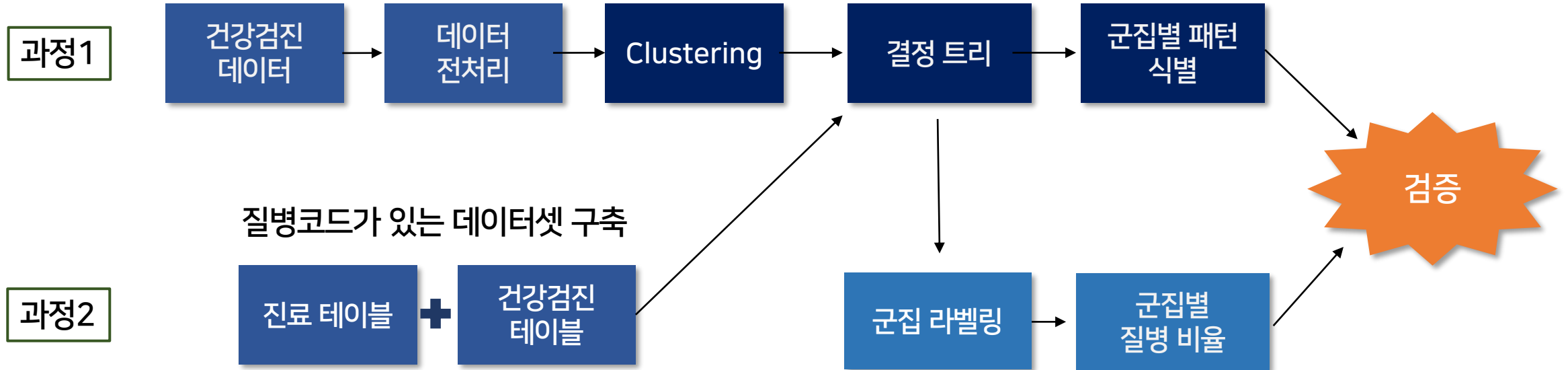
클러스터링 결과



군집 패턴 식별 결과 :

군집 3: 높은 중성지방 수준
군집 4: 높은 중성지방 수준,
높은 간수치 수준

과정추가



클러스터별 특징 파악 후 유질환자 데이터로 군집 패턴 검증

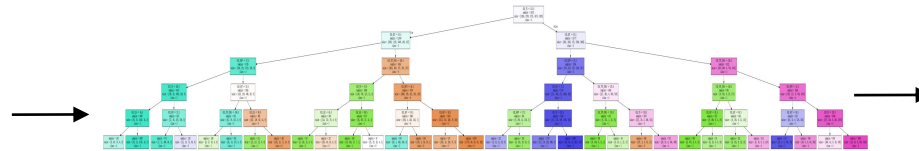
분석결과

- 과정 1의 결정 트리에 데이터셋 적용 -> 클래스 라벨링 -> 클래스별 간 질환 코드(K7_) 비율 확인
-> 과정 1의 군집 패턴 식별 결과대로 군집4의 간 질환 비율이 높다

질병코드가 있는 데이터셋 구축



1217 rows x 19 columns



```
pred = dt_clf.predict(X)
df_merge['cluster'] = pred
df_merge['cluster'].value_counts()
```

2	438
3	297
4	222
0	203
1	57



군집	0	1	2	3	4
총 환자 수	203	57	438	297	222
간질환(K7_)	20	11	42	28	38
비율	9.8%	19.3%	9.6%	9.4%	17.1%

문제해결

■ 건강검진 결과의 문제

• 모호함

- 종합판정 결과가 구체적이지 않음
- 자신의 건강상태가 어느 위치에 속하는지 알 수 없음 → 환자 분류를 위한 군집 정보 추가 제공

• 후속조치가 없음

- 건강검진 후 어떠한 진료를 받아야 하는지 알 수 없음 → 클러스터별 질병 매칭을 통한 건강검진 후속조치

검진자의 건강상태 파악을 위한 군집 추가 제공 및
관련 질병 정보 제공 → 건강검진 이해도 및 활용성 향상

의의 및 한계점

- 의의
 - 결과변수가 없는 빅데이터를 비지도 학습을 통한 의미 창출 시도
 - 별개의 데이터셋을 이용해 모델의 결과 검증
- 한계점
 - 모든 클러스터의 패턴을 질병을 통해 검증하지 못했음
 - 높지 않은 클러스터링 성능

감사합니다 :D