



환자군 별 병원 물류 수요 예측 분석 (2025 06 04)

1. 개요

2. 데이터 소개

3. 데이터 클리닝

- 환자 ID, 방문 날짜, 성별, 주상병(Principal Disease), 부상병(Secondary Disease), 처방 구분, 처방명이 들어있는 데이터
 - 이들 모두 모아서 환자의 차트(Chart)를 구성.
 - 방문일 삭제

예시

환자 ID: 1856

입원/퇴원 여부: 외래

나이: XX

Sex: X

주상병(Primary Diagnoses): 만성 인두염 (J312)

부상병(Secondary Diagnoses): 만성 후두염, 만성 편도염 (J370, J350)

Department EN

- 내복약: 록솔정(암브록솔염산염), 세토펜정 325mg[삼아약품], 스티렌투엑스
- 내시경검사: 후두경검사
- 주사약제: 트리돌주50밀리그램(염산트라마돌) (주)유한양행

- Fit 방법론. 시계열 데이터 취급하여 데이터 오염을 막음

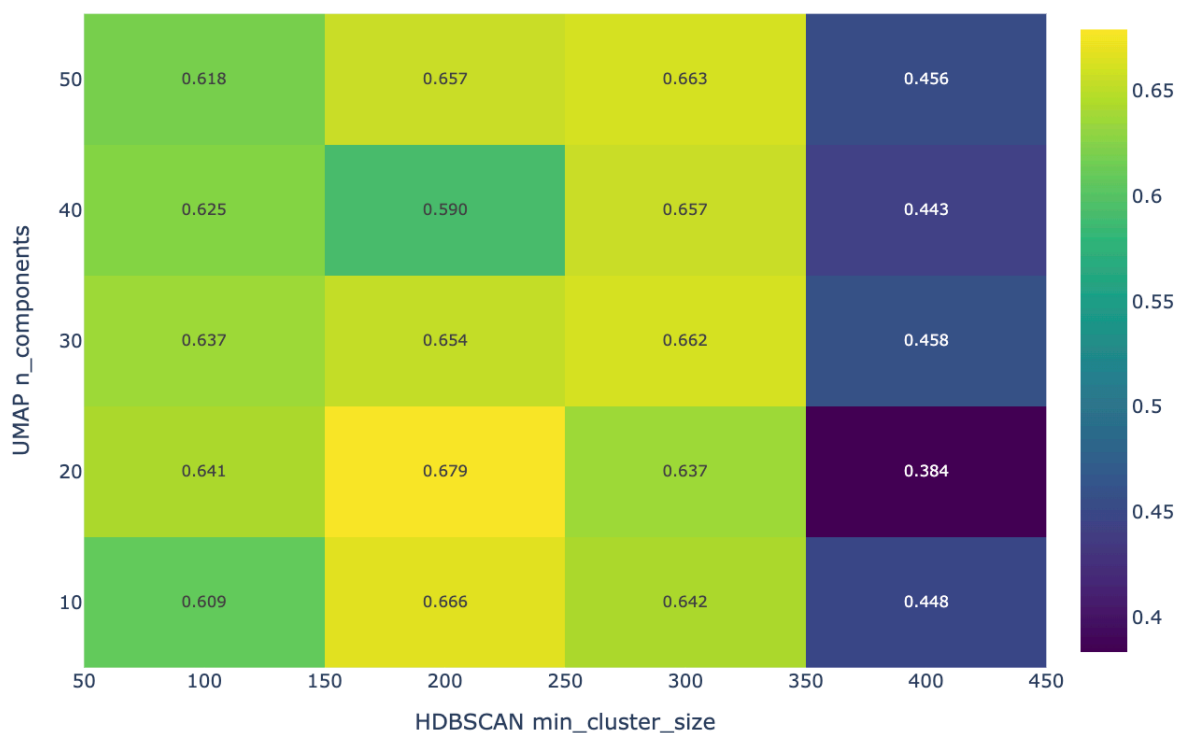
1. 2018년 Fit

2. HDBSCAN의 `approximate_predict` 를 이용하여 2019년 환자들 클러스터링
3. 2018 + 2019년 Fit
4. HDBSCAN의 `approximate_predict` 를 이용하여 2020년 환자들 클러스터링

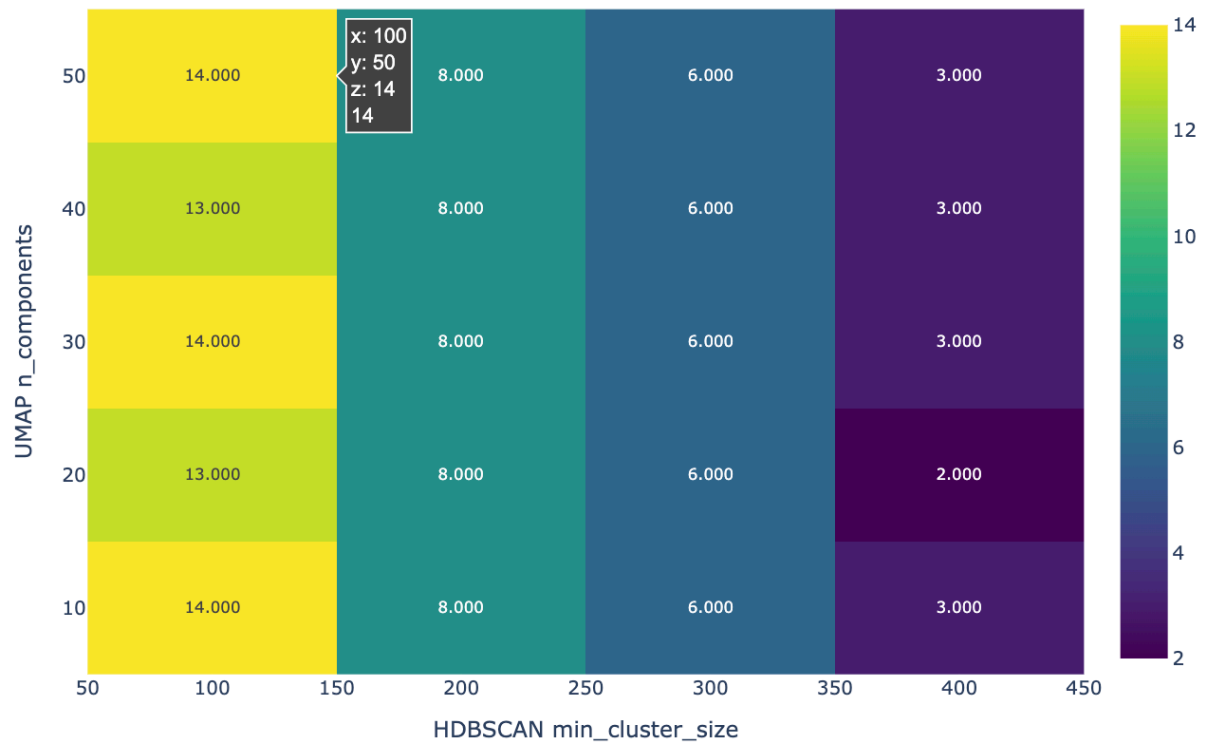
4. 머신러닝 모델링

- 환자군을 분류하기 위해 Clustering 모델 사용
 - HDBSCAN
 - Feature Selection
- 실루엣 점수로 판단한 HDBSCAN & UMAP 최적 하이퍼파라미터 찾기

Silhouette Scores for Different Parameter Combinations



- HDBSCAN & UMAP 하이퍼파라미터 별 클러스터 개수



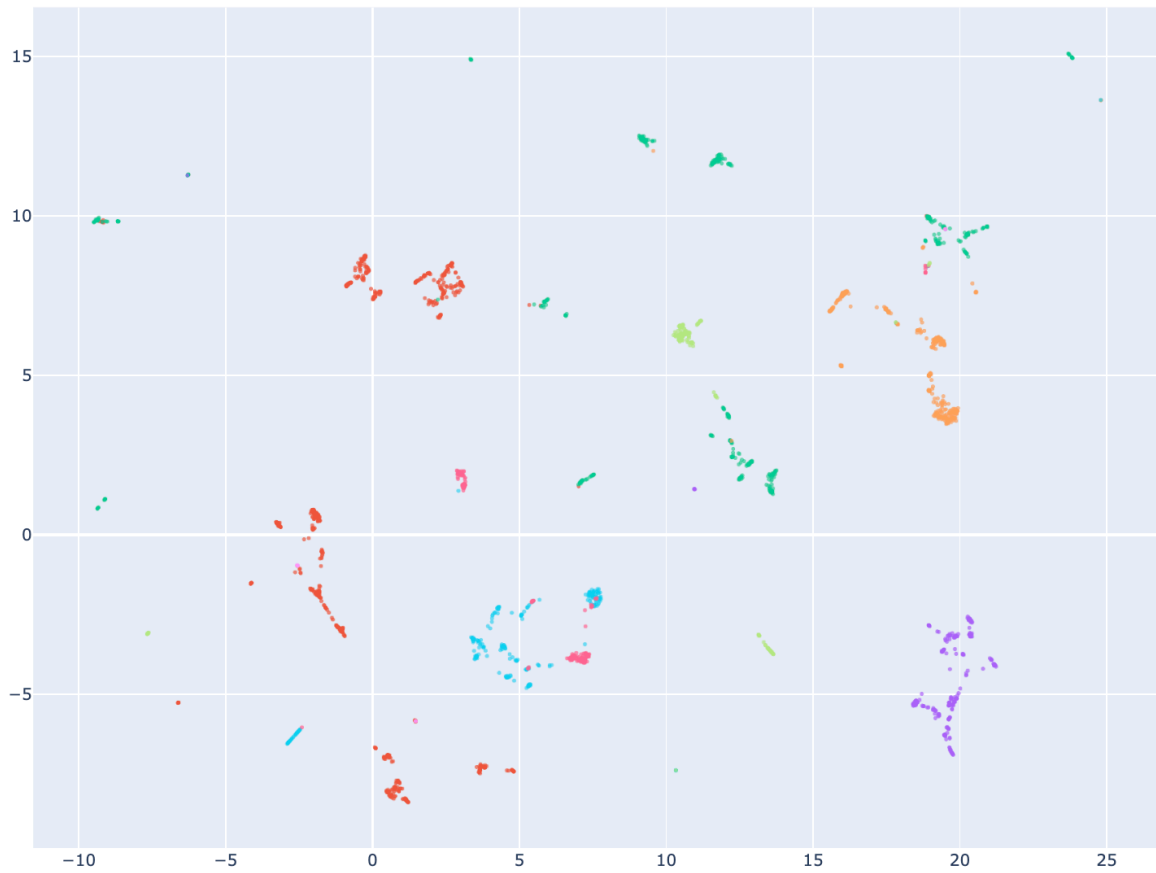
- 8개의 클러스터를 유지하면서 Silhouette Score 가 높은 UMAP (20) HDBSCAN (200) 사용
- 클러스터링 Vis (2018년 Fitting 연도)

Interactive HDBSCAN Clustering (UMAP-2D)



- 2019년도 (Approximate). 대체로 2018년도와 비슷한 집단군이 클러스터링 됨

Interactive HDBSCAN Clustering (UMAP-2D)



[patient_clusters_future.xlsx](#)

[patient_clusters.xlsx](#)

5. 결론
