환자 방문 데이터 처리

1. 데이터 준비 및 전처리
   1. 차트 번호 + 날짜 기준으로 통합
      1. 동일 환자의 동일 날짜 데이터를 하나의 문서 형태로 병함
      2. 진료 및 처방 정보를 문자열로 합치고 메타데이터 생성
   2. 벡터 임베딩 생성
      1. OpenAI text-embedding-ada-002 모델로 1,536차원 임베딩 벡터 생성
      2. Qdrant 벡터 데이터베이스에 임베딩 저장
2. 차원 축소 및 클러스터링
   1. 차원 축소
      1. UMAP을 사용하여 임베딩을 저차원 공간으로 매핑
   2. 밀도(Density) 기반 클러스터링
      1. HDBSCAN을 이용하여 환자 그룹(클러스터) 자동 탐지
      2. 노이즈로 분류된 데이터는 별도 처리
   3. 메타 클러스터링
      1. 클러스터의 무게중심을(centroid) 이용하여 상위 메타 클러스터로 재분류
      2. Sihouette Score, UMAP 2차원, Andrews Curve 등 군집 품질 지표 및 시각화를 통한 클러스터링 적합성 평가
3. 환자 시계열 예측
   1. 클러스터별 방문량 시계열 구축
   2. 주간 Rolling Window XGBoost 예측
      1. 일정 기간 데이터 학습 (52주) 다음 기간 예측
      2. 시계열 특징을 External variable로 추가
4. 물류 시계열 예측
   1. 물류 시계열 구축
   2. 주간 Rolling Window XGBoost 예측
      1. 클러스터별 환자 시계열 값을 설명변수로 하여 물류 예측