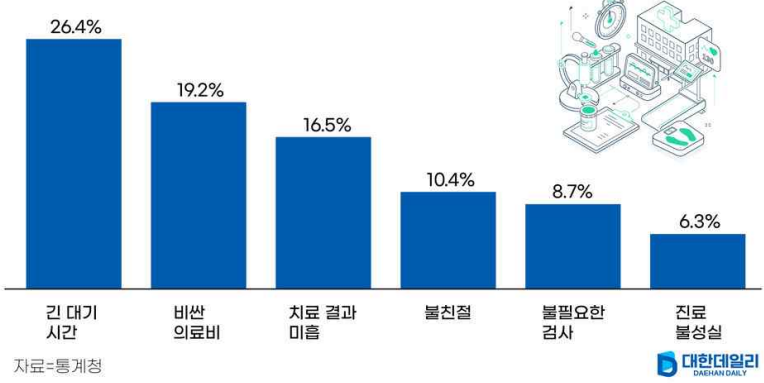



# 프로젝트 기획서

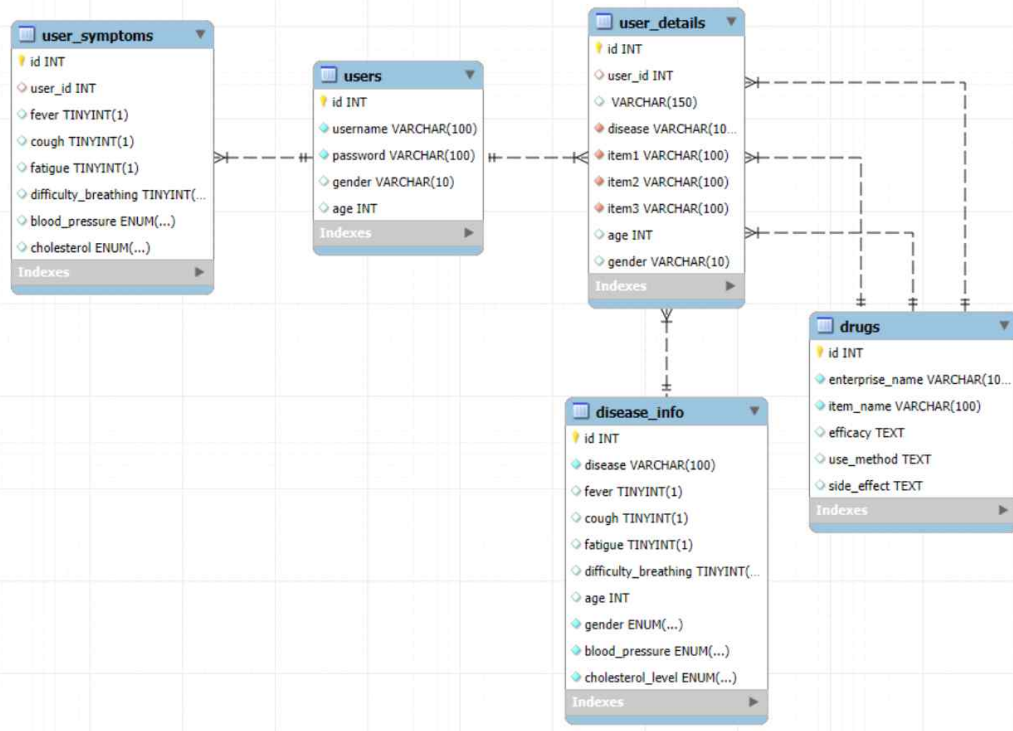
과정명	LG U+ Why Not SW캠프 6기		
작성일자	2025년 7월 1일	팀명	팀 메디멘티(MEDIMENTO)
참여인원	5명	팀장	이민진
프로젝트명	질병 예측 및 의약품 추천 서비스 (서비스명 : 메디멘토 MEDIMENTO)		
기 간	2025년 6월 26일 ~ 2025년 7월 2일		
조직구성	<ul style="list-style-type: none"> <li>- 이민진 (팀장) : 데이터 수집/전처리 및 분석, 모델링</li> <li>- 서세빈 : 화면 구성, 데이터 수집 및 서비스 개발, 발표</li> <li>- 오요셉 : 화면 구성, 데이터 수집 및 서비스 개발</li> <li>- 유지선 : 데이터 수집/전처리 및 분석, 모델링</li> <li>- 주수진 : 데이터 수집/전처리 및 분석, 모델링</li> </ul>		
프로젝트(분석) 배경	<p>&lt; 진료 대기시간 증가에 따른 사용자 불만족 &gt;</p> <p><b>의료서비스 불만족 이유</b></p>  <p>자료=통계청</p> <p>그림 1 통계청 2024년 사회조사 결과</p> <ul style="list-style-type: none"> <li>- 의료서비스에서 가장 많은 불만족을 느끼는 원인이 긴 대기시간으로 꼽힘</li> <li>- 의료서비스에서 신속성을 원하는 환자 증가</li> </ul> <p>&lt; 비대면 진료 시장 전망&gt;</p> <p>〈비대면진료 시장규모(2022~2032년)〉</p> <p>TELEHEALTH MARKET SIZE, 2022 TO 2032 (USD BILLION)</p>  <p>출처 : Precedence Research, <a href="https://www.precedenceresearch.com/telehealth-market">https://www.precedenceresearch.com/telehealth-market</a>, 2023.01.31.</p> <p>그림 2 Precedence Research 전 세계 비대면 진료 시장 규모 예측</p> <ul style="list-style-type: none"> <li>- 코로나19 팬데믹 이후, 비대면 서비스 증가 및 디지털화로 인해 전 세계적으로 비대면 진료 시장이 확대되는 현황</li> <li>- 전 세계적 비대면 진료 시장 규모가 연평균 24.13% 성장할 것으로 전망</li> </ul>		

<p>프로젝트 목적</p>	<ul style="list-style-type: none"> <li>- 경미한 증상일 경우 급히 병원을 방문하지 않아도 질병을 예측할 수 있는 서비스 제공</li> <li>- 처방 중점이 아닌 질병 예측 및 예방 서비스 제공</li> <li>- 사용자가 입력한 정보에 따라 질병을 예측하고, 질병에 맞는 약을 추천</li> <li>- 질병에 따라서 처방 없이 구매 가능한 의약품의 정보를 사용자가 쉽고 빠르게 얻을 수 있음</li> <li>- 서비스를 이용하는 사용자들의 데이터를 저장해 현재 유행하는 질병을 파악할 수 있도록 시각화 (그래프) 서비스 제공, 의료계에서 유행병 관련 정보를 수집할 수 있어 초기 예방 대처 방법 마련의 토대를 제공</li> </ul>
<p>서비스 플로우 및 서비스 화면 기획</p>	<div> <ul style="list-style-type: none"> <li>- 서비스 플로우 <ul style="list-style-type: none"> <li>▶ 시작화면</li> <li>▶ 접속화면: 1) 로그인 2) 회원가입 - 사용자 INPUT (나이, 성별)</li> <li>▶ 메인화면: 1) 증상 입력 및 질병 예측 <ul style="list-style-type: none"> <li>- 증상 선택 (열, 기침, 피로, 호흡 곤란)</li> <li>- 해당사항 선택 (혈압, 콜레스테롤)</li> <li>- 예측된 질병과 질병에 맞는 추천 약 정보 출력</li> </ul> </li> <li>▶ 통계화면: 1) 전체 사용자의 누적 데이터 시각화</li> <li>▶ 마이페이지: 1) 진단내역 - 사용자 증상, 질병, 추천 약 정보 2) 회원정보 수정</li> </ul> </li> </ul> </div> <div> </div> <p>그림 3 서비스 플로우 차트</p> <ul style="list-style-type: none"> <li>- 서비스 화면 기획</li> </ul> <div> <div> <h3>회원가입</h3> </div> <div> <h3>증상 입력 및 질병 예측</h3> <p><b>증상 입력</b> 증상을 선택해 주세요. 기침 x 발열 x</p> <p><b>해당사항</b> 혈압: 보통, <b>상상</b>, 높음 콜레스테롤: 보통, <b>상상</b>, 높음 <input checked="" type="checkbox"/> 측정된 혈압 수치는 정상입니다. <input checked="" type="checkbox"/> 측정된 콜레스테롤 수치는 정상입니다. <b>선택하기</b> 예측된 질병: 습진</p> </div> <div> <h3>전체 사용자 통계</h3> <p>~ 질병 ~ 증상 ~ 약</p> <p><b>연령대별 증상 분포 분석</b></p> </div> </div> <div> <p>서비스 화면 1</p> <p>서비스 화면 2</p> <p>서비스 화면 3</p> </div>

사용 데이터

- 공공데이터: 식품의약품안전처 의약품 개요 정보  
오픈 API를 통해 가져온 데이터
- KAGGLE 데이터: Disease Symptoms and Patient Profile Dataset  
질병 - 환자 정보 데이터 csv 데이터

ERD



분석방법

#### < 데이터 EDA 및 전처리 >

- 데이터 결측치 : X
- 탐색적 데이터 분석(EDA): 데이터 기초 통계

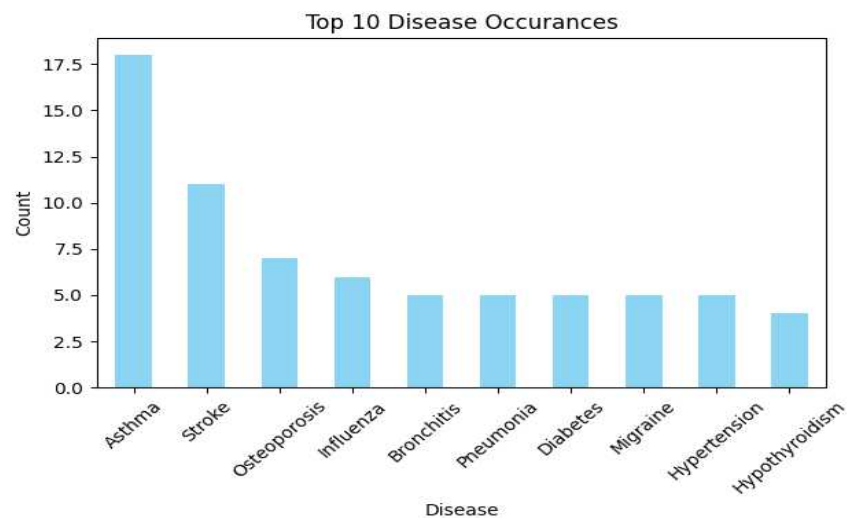


그림 8 Kaggle의 질병-환자 데이터 기반 환자 수가 가장 많은 상위 10개 질병

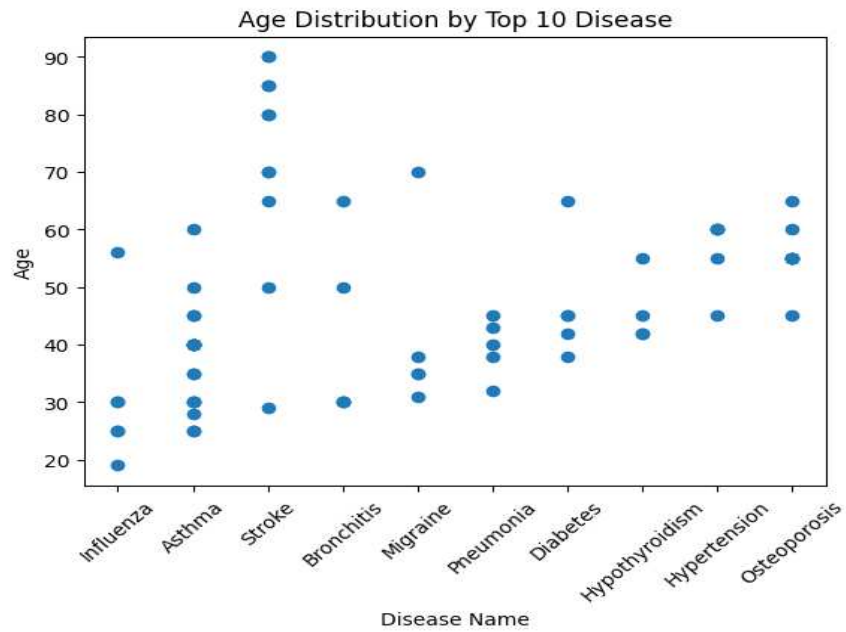


그림 9 Kaggle의 질병-환자 데이터 기반 환자 수 상위 10개 질병의 환자 나이 분포

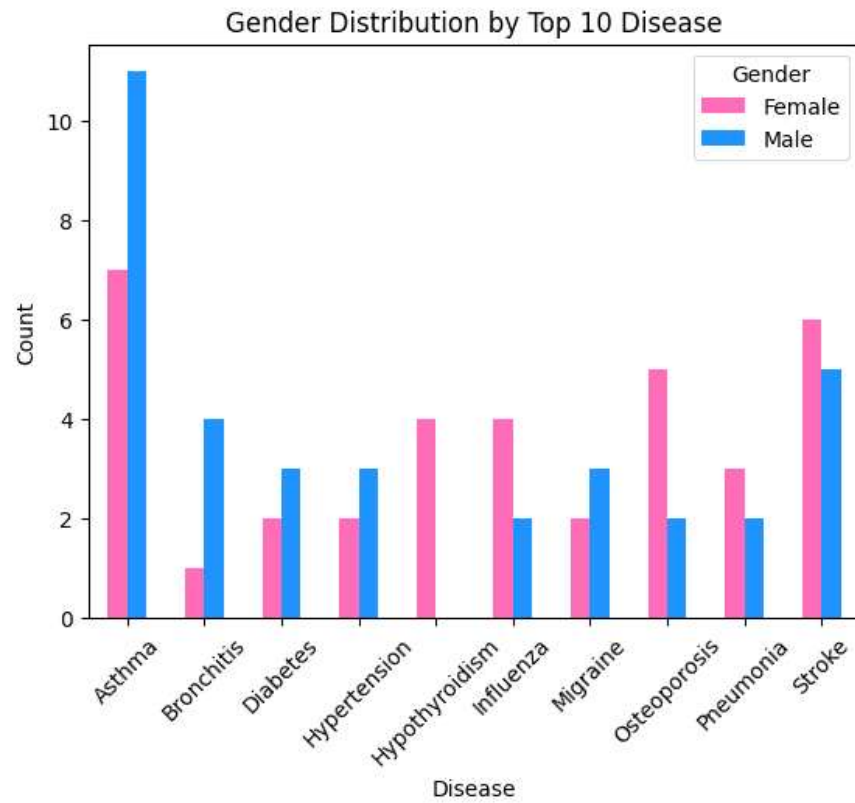


그림 10 Kaggle의 질병-환자 데이터 기반 환자 수 상위 10개 질병별 성별 분포

- 데이터 전처리 :

Kaggle의 질병-환자 데이터

▷ 수치형 변수: 나이

Scikit-learn의 preprocessing 중 StandardScaler 사용

▷ 범주형 변수

증상 여부, 성별, 혈압, 콜레스테롤:

Scikit-learn의 preprocessing 중 StandardScaler 사용

질병명:

컬럼 다양화로 인해 Scikit-learn의 preprocessing 중 LabelEncoder 사용

▷ 중요도 낮은 변수 : 발생 빈도가 1-2번이 드문 질병

등장 횟수 2회 이하인 질병을 제거

< 데이터 분석 및 모델 개발 >

- 모델 학습에 사용한 데이터: Kaggle의 질병-환자 데이터

- 예측하고자 하는 변수 (타겟 변수): 질병명(Disease)

- 독립 변수: 성별, 나이, 증상 여부(열, 기침, 피로, 호흡 곤란), 혈압, 콜레스테롤 수치

- 사용한 분류 모델: Gradient Boosting Classifier

Gradient Boosting Classifier: 의사 결정 트리를 순차적으로 연결

예측 성능 높이는 앙상블 학습 모델

Boosting 방식으로 가중치를 더해 학습

- 모델 평가지표: 정확도(accuracy)

정확하게 예측된 데이터 수 / 전체 데이터 수

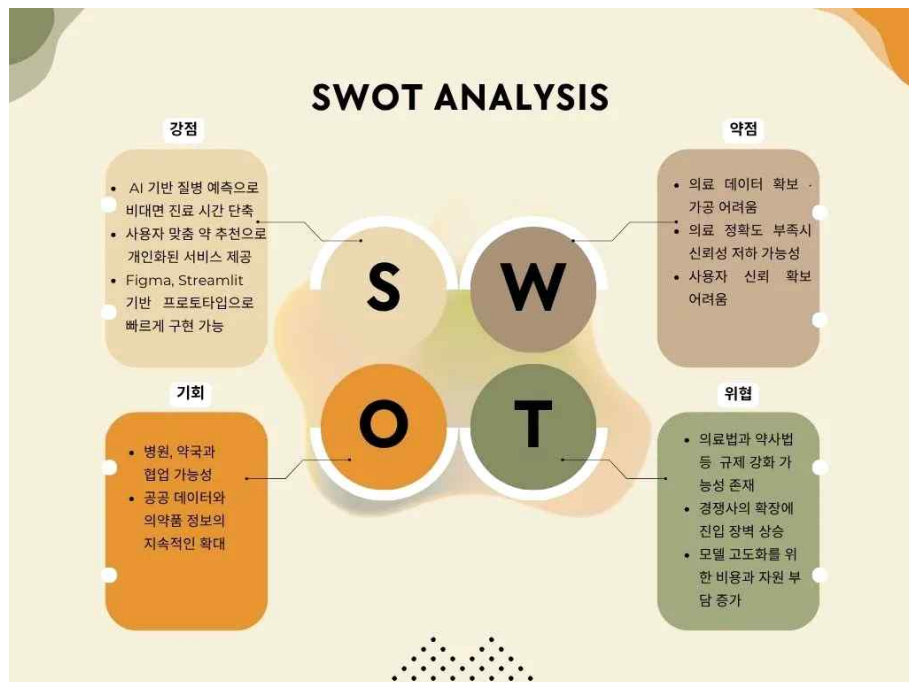
- 모델 프로세스:

▶ 사용자 입력 바탕 Gradient Boosting Classifier 모델로 질병명 예측

▶ 예측된 질병명과 사용자가 입력한 증상을 기반으로 의약품 데이터로부터 추천 약 추출

▶ 결과 : 질병명과 질병, 증상에 따른 의약품 정보

SWOT 분석



분석 결과 활용	<div>쉽고 빠른 질병 예측</div> <div><div>▶ 사용자의 기본 정보와 증상 선택으로 쉽고 빠르게 질병 예측</div><div>▶ 기존 의료서비스에서 사용자가 불만을 느꼈던 신속성 부분 보완 가능</div></div> <div>의료계 참고 자료 제공</div> <div><div>▶ 사용자가 입력한 데이터를 집계, 시각화하여 현재 어떤 증상이 많고, 어떤 질병이 유행하는 지에 관련한 인사이트 제공</div><div>▶ 의료계에서 현재 유행하는 질병 참고 참고자료로 사용 가능해 유행병 대처 방안 구상 가능</div></div>																		
추진 일정	<div>- 프로젝트 기간 : 총 5일</div> <table><tr><th>단계</th><th>일수</th><th>작업 내용</th></tr><tr><td>기획 및 요구사항 정의</td><td>1일</td><td>서비스 목표 설정, 핵심 기능 정의, 사용자 흐름 및 UX 설계</td></tr><tr><td>데이터 수집 및 전처리</td><td>2일</td><td>데이터 필드 정의, 데이터 수집, 데이터 정제 및 통합</td></tr><tr><td>예측 모델 개발</td><td>3일</td><td>질병 예측 모델 개발, 학습 및 검증, Streamlit 연동 테스트</td></tr><tr><td>시스템 구현</td><td>4일</td><td>프론트엔드 및 백엔드 개발, 머신러닝 모델 API 연동, 대시보드 개발</td></tr><tr><td>테스트 및 배포</td><td>5일</td><td>단위 테스트 및 전체 테스트, 버그 수정, 최종 배포</td></tr></table>	단계	일수	작업 내용	기획 및 요구사항 정의	1일	서비스 목표 설정, 핵심 기능 정의, 사용자 흐름 및 UX 설계	데이터 수집 및 전처리	2일	데이터 필드 정의, 데이터 수집, 데이터 정제 및 통합	예측 모델 개발	3일	질병 예측 모델 개발, 학습 및 검증, Streamlit 연동 테스트	시스템 구현	4일	프론트엔드 및 백엔드 개발, 머신러닝 모델 API 연동, 대시보드 개발	테스트 및 배포	5일	단위 테스트 및 전체 테스트, 버그 수정, 최종 배포
단계	일수	작업 내용																	
기획 및 요구사항 정의	1일	서비스 목표 설정, 핵심 기능 정의, 사용자 흐름 및 UX 설계																	
데이터 수집 및 전처리	2일	데이터 필드 정의, 데이터 수집, 데이터 정제 및 통합																	
예측 모델 개발	3일	질병 예측 모델 개발, 학습 및 검증, Streamlit 연동 테스트																	
시스템 구현	4일	프론트엔드 및 백엔드 개발, 머신러닝 모델 API 연동, 대시보드 개발																	
테스트 및 배포	5일	단위 테스트 및 전체 테스트, 버그 수정, 최종 배포																	