



【서식 3-1】 캡스톤디자인 과제 수행 결과보고서 (※ 학생 작성)_ 팀용

기업연계형 캡스톤디자인 교과목 과제 수행 결과보고서

과제 유형	■ 기업연계기반					
과제명	공공의료 데이터 기반 고위험 의료비 예측 및 시각화 시스템					
팀명	이유라/황유림					
수강 교과목명	캡스톤디자인 I		교과목 학수번호	DCSS451(00)		
교과목 담당교수	소 속	컴퓨터소프트웨어	성 명	서민석		
	E - mail	mins@korea.ac.kr	교내전화	044-860-1379		
지도교수	소 속	컴퓨터소프트웨어	성 명	한창희		
	E - mail	changheehan@korea.ac.kr	교내전화	044-860-1346		
산업체 참여 인력(PM)	소 속	(주)엘콤	성 명	이문하		
	E - mail					
산업체 역할 (자문내용)						
구분	성명	학과	학년	학번	E - mail	
참여 학생	팀장	이유라	컴퓨터융합소프트웨어 (바이오시스템의과학부)	3	2024270655	yuraira0114@korea.ac.kr
	팀원	황유림	컴퓨터융합소프트웨어	3	2024270636	riiiiiimk@korea.ac.kr



*이중전공의 경우 본 소속학과(이중전공)으로 표기

위와 같이 규정에 의해 과제를 완료하였음을 결과보고서로 제출합니다.

2026. 05. 21.

지도교수: 한창희

대표학생: 이유라


 (인 문하 서명)

 (인 문하 서명)

고려대학교 세종 SW중심대학사업단 귀하



<p>작품과제명</p>	<p>공공의료 데이터 기반 고위험 의료비 예측 및 시각화 시스템</p>
<p>과제 개요</p>	<ul style="list-style-type: none"> ○ 과제 선정 배경 <ul style="list-style-type: none"> - 고령화와 만성질환 증가로 인해 개인의 의료비 부담이 지속적으로 증가하고 있으며, 질병 발생 이후 치료하는 방식보다 사전에 건강 위험을 예측하고 관리하는 예방 중심 서비스의 필요성이 커지고 있다. - 기존 건강관리 서비스는 운동 기록, 식단 관리, 단순 건강정보 제공에 머무르는 경우가 많아 사용자가 자신의 건강 상태가 향후 의료비나 질병 위험에 어떤 영향을 주는지 구체적으로 파악하기 어렵다는 한계가 있다. - 최근 인공지능과 데이터 분석 기술이 발전하면서 건강검진 정보, 생활습관, BMI, 흡연 여부 등 다양한 데이터를 활용해 개인별 의료비와 건강 리스크를 예측할 수 있는 기반이 마련되고 있다. - 특히 공공의료 데이터를 활용한 머신러닝 기반 분석 기술은 단순 건강 정보 제공을 넘어 사용자 맞춤형 의료비 예측 및 고위험군 분석이 가능하다는 점에서 AI 헬스케어 분야의 활용 가능성이 높아지고 있다. ○ 과제의 필요성 <ul style="list-style-type: none"> - 사용자는 자신의 현재 건강 상태를 바탕으로 미래에 발생할 수 있는 의료비 부담과 질병 위험도를 사전에 확인함으로써 생활습관 개선이나 건강검진 참여 등 예방적 행동을 할 수 있다. - 기존 건강관리 시스템은 일반적인 건강 통계나 단순 정보 제공 중심으로 구성되는 경우가 많아 개인별 위험 요인을 구체적으로 분석하는 데 한계가 있다. - 본 과제는 단순 평균 통계가 아니라 머신러닝 모델을 활용하여 연령, BMI, 흡연 여부, 운동 습관 등 여러 요인을 종합적으로 분석함으로써 사용자 개인에게 더 적합한 맞춤형 예측 결과를 제공할 수 있다. - 예측 결과를 수치와 그래프로 시각화하여 제공함으로써 의료 지식이 많지 않은 일반 사용자도 자신의 건강 위험 수준과 의료비 증가 원인을 쉽게 이해할 수 있으며, 건강관리의 필요성을 직관적으로 인식할 수 있다.



<p>과제 내용</p>	<ul style="list-style-type: none"> ○ 과제 구성 <ul style="list-style-type: none"> - 건강검진 및 생활습관 관련 데이터를 수집한 뒤 결측치 처리, 이상치 확인, 범주형 데이터 변환 등 전처리 과정을 수행하여 머신러닝 모델 학습에 적합한 형태로 정리한다. - 전처리된 데이터를 기반으로 One-Hot Encoding 및 Logistic Regression 알고리즘을 적용하여 의료비 고위험군 예측 모델을 구축하고, Accuracy, Precision, Recall, ROC-AUC 기반 성능 평가를 수행한다. - 사용자가 연령, 성별, BMI, 흡연 여부 등의 정보를 입력하면 예상 의료비와 건강 리스크 결과를 실시간으로 확인할 수 있도록 Streamlit 기반 웹 인터페이스와 시각화 화면을 구현한다. ○ 과제 주요 특징 <ul style="list-style-type: none"> - 사용자의 기본 정보와 생활습관 데이터를 반영하여 개인별 예상 의료비를 산출하므로 단순 평균값이 아닌 맞춤형 예측 결과를 제공할 수 있다. - BMI, 흡연 여부, 연령 등 의료비 증가와 관련 있는 주요 요인을 함께 분석하여 사용자가 어떤 요소가 건강 위험과 비용 증가에 영향을 주는지 확인할 수 있다. - 기존 건강관리 서비스가 단순 건강 기록 및 정보 제공 중심이었다면, 본 시스템은 공공의료 데이터를 기반으로 머신러닝 모델을 활용하여 예상 의료비와 고위험군 여부를 함께 분석한다는 점에서 차별성이 있다. - 또한 예측 결과를 그래프와 표 형태로 시각화하여 사용자가 자신의 건강 위험 요인과 의료비 증가 원인을 직관적으로 이해할 수 있도록 구현하였다. - Streamlit 기반 웹 서비스 형태로 구현하여 별도의 복잡한 설치 없이 브라우저에서 사용할 수 있으며, 실시간 예측 기능과 데이터 시각화를 통해 실제 활용 가능한 AI 헬스케어 서비스 형태로 구현하였다.
--------------	--

결과물의
활용방안 및
기대효과

본 과제의 결과물은 사용자가 자신의 건강 정보와 생활습관 데이터를 입력하면 예상 의료비와 건강 리스크를 실시간으로 분석·예측해주는 웹 기반 헬스케어 시스템 형태로 구현된다. 사용자는 연령, BMI, 흡연 여부, 운동 여부 등의 정보를 입력하여 자신의 건강 상태가 향후 의료비에 어떤 영향을 미치는지 직관적으로 확인할 수 있으며, 이를 통해 질병 발생 가능성과 건강관리의 필요성을 사전에 인지할 수 있다. 또한 결과를 단순 수치가 아닌 그래프와 차트 형태로 시각화하여 제공함으로써 의료 지식이 없는 일반 사용자도 쉽게 이해하고 활용할 수 있도록 한다.

특히 기존 건강관리 서비스가 단순 건강 기록 및 정보 제공 중심이었다면, 본 시스템은 AI 기반 의료비 예측과 건강 리스크 분석을 동시에 제공함으로써 예방 중심 건강관리 의사결정을 지원할 수 있다는 점에서 차별성이 있다. 또한 공공의료 데이터를 기반으로 사용자 맞춤형 분석 결과를 제공하여 실제 의료·헬스케어 분야에 활용 가능한 AI 서비스 형태로 확장 가능성을 가진다.

산업적 측면에서는 병원, 보험사, 헬스케어 플랫폼 등 다양한 분야와 연계 가능한 AI 기반 건강관리 시스템으로 확장할 수 있다. 병원에서는 환자의 생활습관과 건강 상태를 분석하여 맞춤형 건강관리 및 예방 프로그램 제공에 활용할 수 있으며, 보험 분야에서는 건강 위험도 기반 맞춤형 보험 상품 추천이나 의료비 예측 자료로 활용 가능하다. 또한 향후 웨어러블 기기나 건강검진 데이터와 연동할 경우 실시간 건강 모니터링 및 개인 맞춤형 디지털 헬스케어 서비스로 발전시킬 수 있다.

학술적 측면에서는 머신러닝 기반 의료비 예측 모델 성능 비교와 건강 데이터 분석 결과를 기반으로 캡스톤디자인 경진대회, 교내 학술 발표, 포스터 발표 등의 형태로 활용할 수 있으며, AI와 바이오·헬스케어를 융합한 실질적인 프로젝트 사례로서 연구 가치 또한 가진다. 특히 데이터 전처리, 모델 학습, 예측 성능 평가, 시각화 및 웹 서비스 구현 과정을 모두 포함하고 있어 향후 AI 헬스케어 분야 연구 및 프로젝트의 기초 자료로 활용 가능하다.

기대효과 측면에서는 사용자가 자신의 건강 상태를 데이터 기반으로 객관적으로 이해하고 예방 중심의 건강관리 습관을 형성할 수 있도록 도움을 줄 수 있다. 또한 단순 건강 정보 제공을 넘어 AI 기반 예측 결과를 활용하여 개인 맞춤형 건강관리 방향을 제시할 수 있으며, 의료비 증가 위험 요소를 조기에 인식함으로써 건강관리 효율 향상과 의료비 절감 효과를 기대할 수 있다. 더불어 프로젝트 수행 과정에서 Python 기반 데이터 분석, 머신러닝 모델 개발, 웹 서비스 구현 등 실무 중심 기술을 직접 경험함으로써 AI·데이터·헬스케어 융합 분야의 문제 해결 능력과 실무 역량을 향상시킬 수 있다.

수행 방법	구분	성명	과제 참여 내용(역할)
	팀장	이유라	데이터 전처리 및 머신러닝 기반 의료비 예측 프로그램 구현
	팀원	황유림	데이터 정리 및 결과 화면 구성 보조
	팀원		
	팀원		
	팀원		
	팀원		

[지도교수] 한광희교수님

머신러닝 기반 공공의료 데이터 리스크 분석 시스템

2024270655 이유라 & 2024270636 황유림

BACKGROUND

공공의료 체계에서 고비용 위험군을 빠르게 식별하기 어려움
연령·성별·행위코드 기반 위험 예측 및 시각화 시스템 개발

DATA

분석 데이터	고위험 비율	기준 진료비
228,520건	20.1%	573,226원

BUILD

데이터 및 전처리
HIRA 공공의료 CSV 4종을 수집하고 분석용 데이터셋으로 통합
결측치 제거, 연도별 평균 진료비 계산, 상위 1% 이상치 제거
상위 20% 비용 구간을 high risk로 정의해 라벨링

모델 및 시스템
One-Hot Encoding과 Logistic Regression으로 분류 모델을 학습
비용 구간 결정을 입력에서 개인에 라벨 누수를 방지
Streamlit 기반 화면에서 위험 확률을 바로 확인할 수 있게 구성

RESULT

성능 요약

Accuracy	Precision	Recall	ROC-AUC
0.957	0.847	0.961	0.990

핵심 성과
고위험군 비율은 전체의 20.1% 수준으로 나타남
65세 이상 연령군에게 평균 진료비가 상대적으로 높게 관망됨
특정 행위코드는 연도 평균 진료비가 700만원 이상으로 관망됨

주요 시각화

SCREEN

EXPECTATION

기대 효과
공공의료 데이터의 복잡한 비용 패턴을 시각적으로 요약함
고위험군 의료비 발생 가능성을 감량적으로 예측해 분석 근거를 제공함
후속 연구에 대해 질병군 파악률 변수 추가 확장 가능한 기반을 만든다

요약
연구명 - 공공의료 데이터는 양대한 의료비 예측 도구가 부족함
개발명 - 비용 위험 라벨을 직접 정의하고 예측하게 연결함

결과물