

2024년도 공공기관 용역과제
AI개발 수행내역서

과제명	AI기반 고혈압 예측모델 개발 및 시각화
담당자	정지운

2024년 01 월 06 일

1. 사업과제 : AI기반 고혈압 예측모델 개발 및 시각화

2. 개요 및 현황

2.1 추진배경 및 목적

- 고혈압은 전 세계적으로 가장 흔한 만성 질환 중 하나로, 심혈관 질환과 뇌졸중 등 심각한 합병증을 초래할 수 있습니다. 그러나 현재의 고혈압 진단 방법은 시간이 많이 소요되며, 정확한 예측과 모니터링이 어려운 상황입니다.
- 본 개발의 목적은 AI 기반의 고혈압 예측 모델을 개발하여 개인의 건강 데이터를 기반으로 고혈압 발생 가능성을 예측하고, 실시간 모니터링 및 시각화를 통해 사용자에게 개인화된 건강 관리 솔루션을 제공하는 것입니다.

2.2 과제 범위

과제구분		내용
AI	AI기반 고혈압예측모델 구현	원시 데이터 수집 및 데이터셋 구축
		데이터 전처리, 표준화
		예측모델 선정 및 학습
		평가지표를 활용한 모델 성능 평가
		예측모델 웹기반 시스템 구축
		테스트
시각화	실시간 고혈압 여부 데이터 시각화	대시보드 기반 시각화 구현
		실시간 데이터 수집 인터페이스 구축
		이상치 감지
		사용자 맞춤형 데이터 표출

2.3 과제 추진 방법

1) 구축 대상 선정 기준

○ 데이터 접근성 및 활용성

- 고혈압 예측 모델을 위한 원시 데이터는 쉽게 접근할 수 있으며 활용 가능한 고품질 데이터를 선정합니다.
- 데이터 출처와 수집 방법이 명확해야 하며, 예측 모델에 적합한 특성을 포함해야 합니다.

○ 데이터 품질과 신뢰성

- 정확하고 일관된 데이터가 필수적입니다.
- 불완전하거나 오류가 있는 데이터를 배제하여 모델의 성능을 최적화합니다.

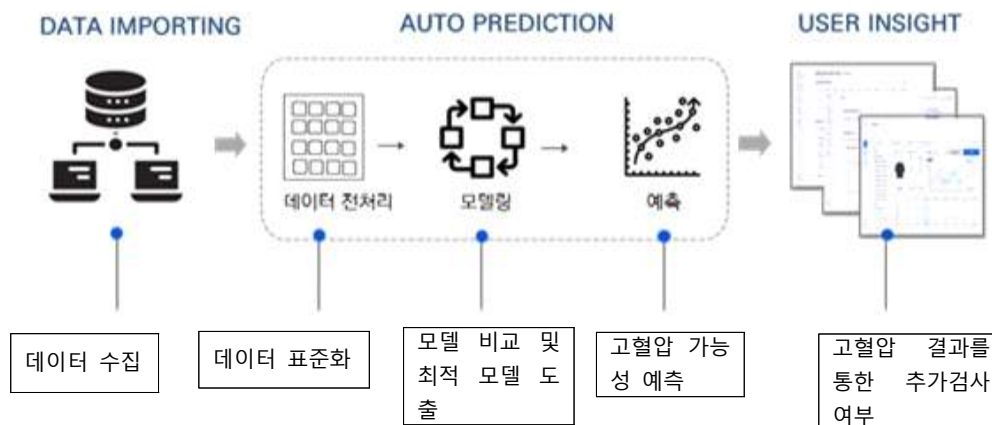
○ 다양성 및 포괄성

- 예측모델을 통해 고혈압에 상대적 기여도가 높은 지 여부
- 다양한 연령대, 성별 데이터를 포함하여 모델이 일반화된 예측 성능을 가질 수 있도록 합니다.

2) AI 예측 분석모델 적용 대상

기능	수집 데이터	예측모델인자(독립변수)	AI예측 분석 대상
고혈압 예측	- 혈압약 복용 여부, 성별, 수축기 혈압 외에 고혈압 핵심 변수가 포함된 데이터셋	- 고혈압변수 : BPMeds, diabetes, male, sysBP 등	- 고혈압 가능성

3) AI 분석모델 구축 프로세스



연구개발 주요 결과물

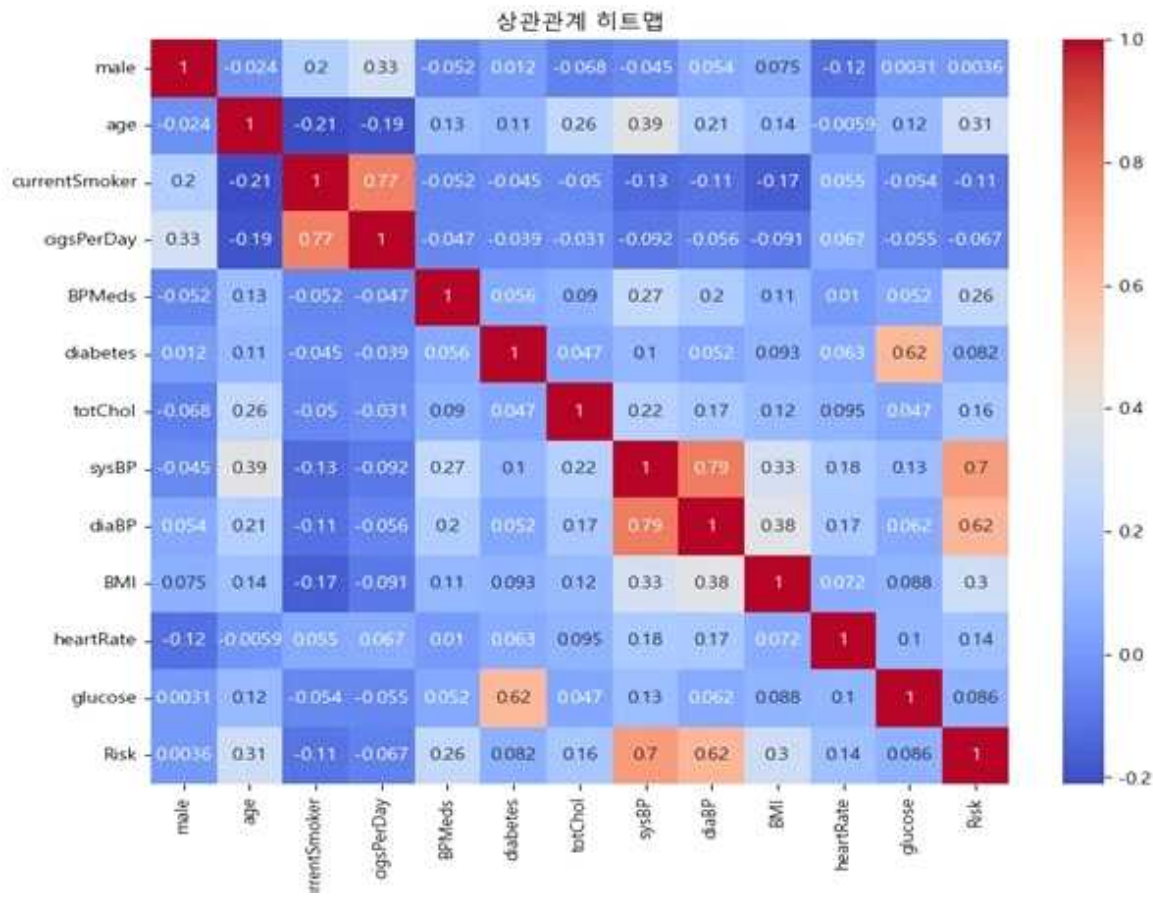
1. 데이터 수집

- 캐글 고혈압 예측 데이터(엑셀)

male	age	currentSmoker	cigsPerDay	BPMeds	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	Risk
1	39	0	0	0	0	195	106	70	26.97	80	77	0
0	46	0	0	0	0	250	121	81	28.73	95	76	0
1	48	1	20	0	0	245	127.5	80	25.34	75	70	0
0	61	1	30	0	0	225	150	95	28.58	65	103	1
0	46	1	23	0	0	285	130	84	23.1	85	85	0
0	43	0	0	0	0	228	180	110	30.3	77	99	1
0	63	0	0	0	0	205	138	71	33.11	60	85	0
0	45	1	20	0	0	313	100	71	21.68	79	78	0
1	52	0	0	0	0	260	141.5	89	26.36	76	79	1
1	43	1	30	0	0	225	162	107	23.61	93	88	1
0	50	0	0	0	0	254	133	76	22.91	75	76	0
0	43	0	0	0	0	247	131	88	27.64	72	61	0
1	46	1	15	0	0	294	142	94	26.31	98	64	1
0	41	0	0	0	0	332	124	88	31.31	65	84	1
0	39	1	9	0	0	226	114	64	22.35	85	NA	0
0	38	1	20	0	0	221	140	90	21.35	95	70	1
1	48	1	10	0	0	232	138	90	22.37	64	72	1
0	46	1	20	0	0	291	112	78	23.38	80	89	0
0	38	1	5	0	0	195	122	84.5	23.24	75	78	0
0	41	0	0	0	0	195	139	88	26.88	85	65	0
0	42	1	30	0	0	190	108	70.5	21.59	72	85	0
0	43	0	0	0	0	185	123.5	77.5	29.89	70	NA	0
0	52	0	0	0	0	234	148	78	34.17	70	113	0
0	52	1	20	0	0	215	132	82	25.11	71	75	0
1	44	1	30	0	0	270	137.5	90	21.96	75	83	1
1	47	1	20	0	0	294	102	68	24.18	62	66	0
0	60	0	0	0	0	260	110	72.5	26.59	65	NA	0
1	35	1	20	0	0	225	132	91	26.09	73	83	1
0	61	0	0	0	0	272	182	121	32.8	85	65	1
0	60	0	0	0	0	247	130	88	30.36	72	74	0
1	36	1	35	0	0	295	102	68	28.15	60	63	0
1	43	1	43	0	0	226	115	85.5	27.57	75	75	0
0	59	0	0	0	0	209	150	85	20.77	90	88	1
1	61	1	5	0	0	175	134	82.5	13.59	72	75	0
1	54	1	20	0	0	214	147	74	24.71	96	87	1

2. 데이터 분석

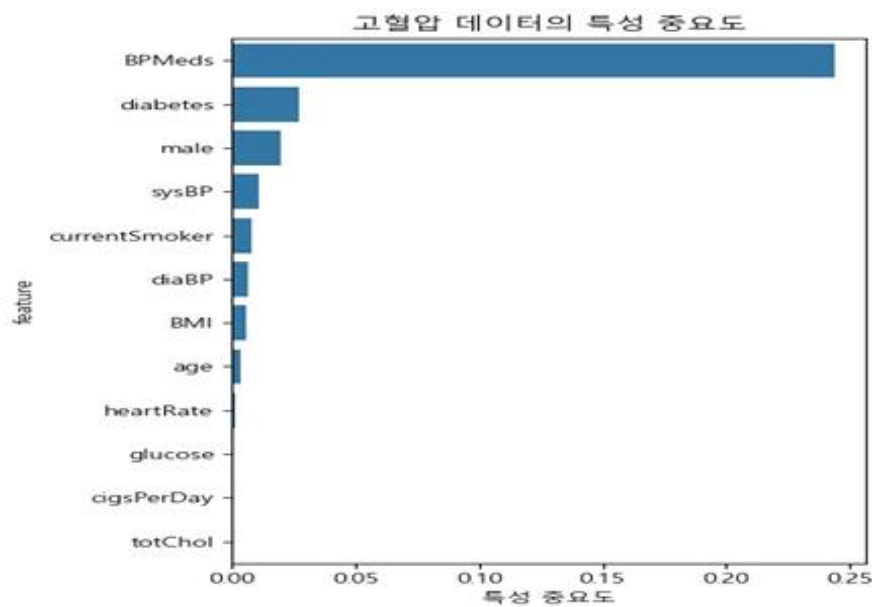
2.1 고혈압데이터 상관관계(Heatmap)



- EDA 히스토그램, 히트맵 변수별 분포를 통해 정규분포 여부, 데이터 변환(ex. 로그변환) 필요성 및 변수 간 관계를 유추
- 고혈압에 영향을 미치는 요인 분석
 - 고혈압변수 : BPMeds, diabetes, sysBP

2.2 탐색적 데이터 분석

- 고혈압 예측 데이터의 탐색 중요도



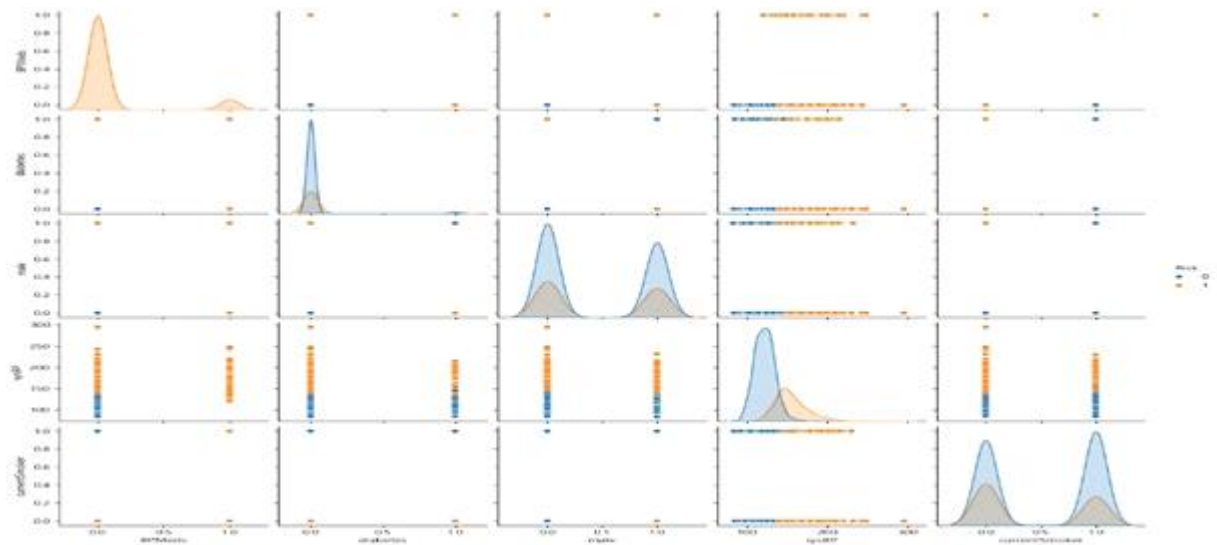
- 결측치 통계

```
# 결측치 관측
display(dataset.isnull().sum())
✓ 0.0%
```

male	0
age	0
currentSmoker	0
cigsPerDay	0
BPMeds	0
diabetes	0
totChol	0
sysBP	0
diaBP	0
BMI	0
heartRate	0
glucose	0
Risk	0

dtype: int64

○ 주요 변수별 데이터 분포(Histogram)



○ 데이터 전처리

	male	age	currentSmoker	cigsPerDay	BPMeds	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	Risk
0	1	39	0	0.00	0.00	0	195.00	106.00	70.00	26.97	80.00	77.00	0
1	0	46	0	0.00	0.00	0	250.00	121.00	81.00	28.73	95.00	76.00	0
2	1	48	1	20.00	0.00	0	245.00	127.50	80.00	25.34	75.00	70.00	0
3	0	61	1	30.00	0.00	0	225.00	150.00	95.00	28.58	65.00	103.00	1
4	0	46	1	23.00	0.00	0	285.00	130.00	84.00	23.10	85.00	85.00	0

3. 데이터 학습 및 모델정의

3.1 모델정의

○ 분류 모델 정의 : Random Forest

```
rf_model = RandomForestClassifier(n_estimators=30, random_state=42)
```

3.2 모델학습

○ 모델 학습

```
rf_model.fit(X_train, y_train)
```

✓ 0.1s

RandomForestClassifier

RandomForestClassifier(n_estimators=30, random_state=42)

3.3 모델 평가

○ 모델 평가

```
랜덤 포레스트 성능:
정확도: 0.89
분류 보고서:

```

	precision	recall	f1-score	support
0	0.91	0.93	0.92	516
1	0.84	0.80	0.82	235
accuracy			0.89	751
macro avg	0.87	0.87	0.87	751
weighted avg	0.89	0.89	0.89	751

```

혼동 행렬:
[[479 37]
 [ 46 189]]
```