

# 고혈압 분석 모델

2021-10-01

컴퓨터과학과 황승현

# 목차

- 고혈압 분석 모델 소개
- 지금까지 한 것
  - 데이터셋 수정
  - 논문 자료 제작
  - 박사님의 문제제기
    - Decision Tree의 한계
    - 해결방안
  - 랜덤 포레스트
- 앞으로 할 것
  - 유의미한 결과 도출



## 고혈압 분석 모델 소개

# 개요

- 식품영양학과 김혜림 박사님
- 사람의 나이, 영양, 식이 등 생활 패턴과 특정 질환의 상관관계 조사
  - 식이 패턴과 고혈압의 상관 관계를 집중 연구
- 고혈압 모델 제작
  - 새로운 변수(사람)의 고혈압 유병 여부 예측.
  - 현재는 정상혈압이지만, 이후 고혈압에 걸릴지 예측

# 모델 설명

- 고혈압 변수 가공
- 데이터 분리 및 전처리
- 결측값 대치
  - 기반 데이터 변수 수정
  - KNNImputer, SimpleImputer
- 변수 스케일링
  - StandardScaler, MinMaxScaler, QuantileTransformer
- 모델 제작 및 하이퍼파라미터 튜닝
  - Train, test 분리
  - Keras Tuner
  - Hyperband

# 모델 설명

- Decision tree 제작
  - DecisionTreeClassifier
  - Decision tree로 변수 중요도 추출
- 다양한 모델 설계
  - 변수 속아내고 모델 설계
  - 식이패턴을 이용한 모델 설계
  - 결측값을 모두 대치한 모델 설계



지금까지 한 것

# 데이터셋 수정

- **F1 ~ F17**
  - 식품군
  - 모델 제작에는 사용하지 않음
- **FA1 ~ FA5**
  - 식이 패턴
  - 영양학적으로 중요함.
- **AS1\_WAIST3\_A**
  - 허리둘레



# 데이터셋 정보

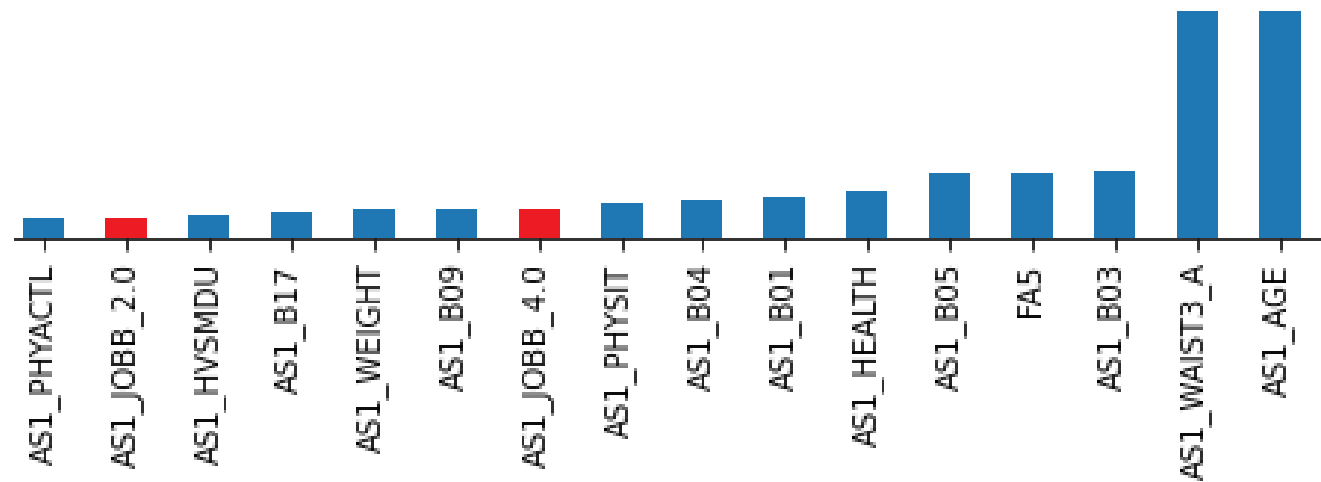
Table	Variable	Variable value	Training Dataset	Testing Dataset
AS1_03_DRSM	AS1_Drink	1	3564	909
		2	503	128
		3	3699	904
	AS1_DrDuA	0	3692	937
		1	422	112
		2	380	95
		3	273	62
		4	442	123
		5	2554	612
	AS1_TotAlc		0.043308 (0 ~ 22)	0.028645 (0 ~ 11)
	AS1_SmokeA	0	4575	1160
		1	1197	308
		2	221	58
		3	1770	415
	AS1_HvSmAm		4.507819 (0 ~ 80)	4.229985 (0 ~ 55)
	AS1_HvSmDu		6.863481 (0 ~ 61)	6.566409 (0 ~ 53)

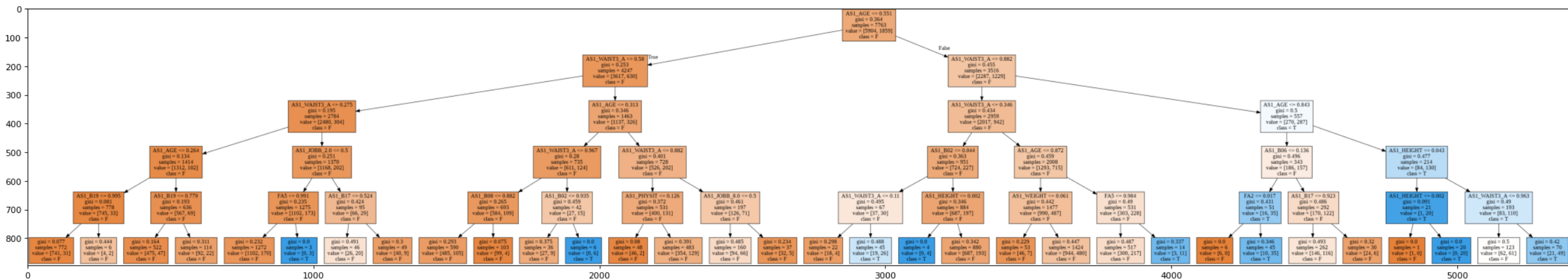
- Train 데이터, Test 데이터의 수치
- 범주형 변수
  - 각 변수, 값의 빈도
- 연속형 변수
  - 평균, 최솟값, 최댓값

## 제기된 문제

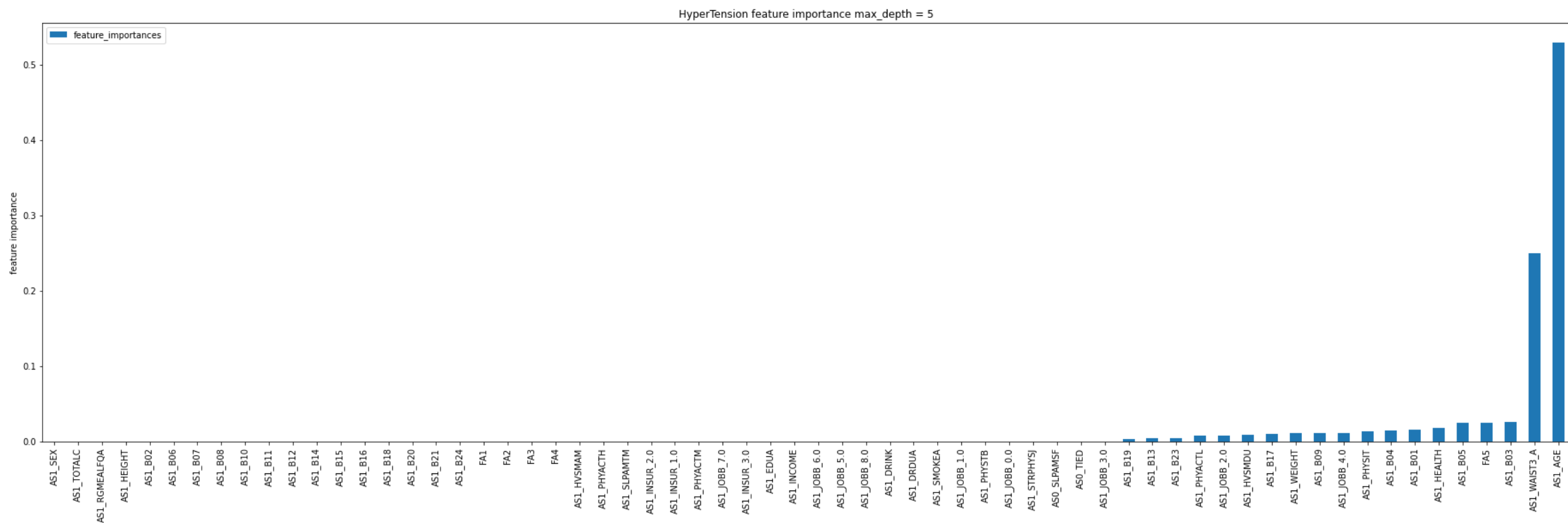
JOBB이 왜 나왔을까..?

Overfitting





Decision Tree (max\_depth = 5)



Feature Importance (max\_depth = 5)

# Decision Tree 한계

- Greedy 알고리즘
  - 최적의 트리 못 찾음
  - 일부 요소만을 사용
- 특정 요소에 과적합!!
  - 앞서 제기된 문제의 원인

# 해결방안

- Tree의 과적합을 막는 기법 사용
  - Random Forest
- ‘식이 패턴’, ‘식품 섭취 빈도’만으로 모델 제작
- 과제 설계의 오류..?

# 랜덤 포레스트

- 많은 Decision Tree로 분류 또는 평균 예측치 출력
  - Ensemble
- 배깅(bagging)
  - 여러 모델을 결합하여 과적합 줄임
- 장점
  - 월등히 높은 정확성
  - overfitting 확률 줄임, 일반화된 트리 생성
- `sklearn.ensemble.RandomForestClassifier`







앞으로 할 것

# 유의미한 결과 도출

- 해결방안 2가지 제안하기
- 2021-10-01 16:30
  - 세미나
  - 컴퓨터과학과 이수원 교수님
  - 식품영양학과 김윤아 교수님
  - 식품영양학과 김혜림 박사님

감사합니  
다~

Q&a

