

스마트 퍼스널 심뇌혈관질환 관리

참 가 번 호	220167	팀 명	새싹
---------	--------	-----	----

□ 공모 배경

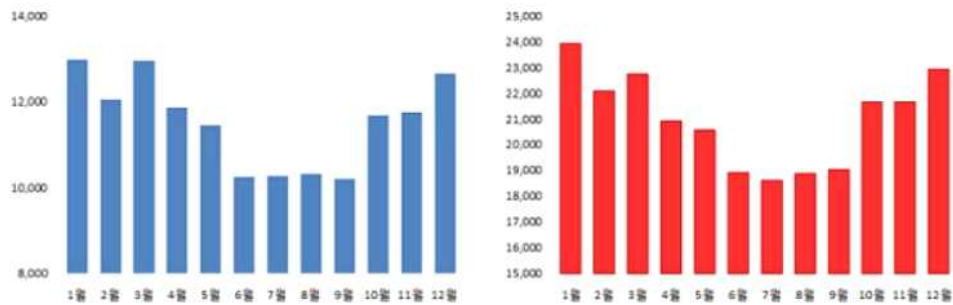


그림1. <2008년~2017년 허혈성 심장질환 월별 사망자수> 그림2. <2008년~2017년 뇌혈관질환 월별 사망자수>
[자료원] 통계청, 사망원인 통계

심뇌혈관질환은 국내 사망률 2위에 해당하는 질병으로 추운 겨울철뿐만 아니라 일교차가 큰 날과 여름에도 주의가 필요하다. 서울대 보건대학원 연구팀에 의하면 일교차가 1°C 증가할 때마다 심뇌혈관 질환자 수는 5.2% 증가하는 것으로 나타났다. [1] 하지만, "날씨"라는 단어를 검색했을 때 심뇌혈관질환과 관련된 키워드를 찾기는 쉽지 않다. 이처럼 우리는 날씨와 심뇌혈관질환의 상관관계의 심각성을 인지하지 못하고 있다. 또한, 심뇌혈관질환은 골든 타임이 있는 시간 민감성 질환으로, 병원 도착 시각에 따라 사망률에 큰 영향을 받는다. [2] 따라서 자신이 심뇌혈관질환 고위험군인지, 얼마큼의 위험인자를 보유하고 있는지 인식을 해야 한다. 그러나 병원에 자주 내원하기는 쉽지 않다. 이로 인해 최근 관심이 지속적으로 증가한 헬스케어 모바일 앱을 통해 심혈관질환의 위험성을 알리며 사용자 접근 방식의 유용성을 극대화할 필요성이 있다.

□ 분석 데이터 정의

1	tma : 날짜
2	area : 지역
3	sex : 성별(남:1, 여:2)
4	frequency : 빈도
5	ssrate : 일조율
6	sum_gsr : 합계전천일사
7	avg_pa : 평균기압
8	max_pa : 최고기압
9	min_pa : 최저기압
10	sum_rn : 합계강수량
11	avg_ta : 평균기온
12	max_ta : 최고기온
13	min_ta : 최저기온
14	avg_ws : 평균풍속
15	max_ws : 최대풍속
16	pm10 : 미세먼지
17	avg_rhm : 평균 상대습도

18	elder_rate : 고령 인구 비율
19	elder : 65세 이상 인구
20	total_pop : 지역별 전체 인구
21	fct_rhm : 예보 습도
22	fct_max_ta : 예보 최고기온
23	fct_min_ta : 예보 최저기온
24	year : 년
25	month : 월
26	day : 일
27	season : 계절
28	weekday : 요일
29	holidays : 1이면 주말, 0이면 아님
30	cold_wave : 한파, 1 : -12이하, 0 : 이상
31	heat_wave : 폭염, 1 : 33이상, 0 : 이하
32	dtr : 일교차
33	fct_dtr : 예보 데이터 일교차
34	dtr_class : 일교차 등급
35	fct_dtr_class : 예보 데이터 일교차 등급
36	pm10_class : 미세먼지 등급
37	last_class : 1 : 각 월마다 마지막 날, 0: 다른 날
38	wind_chill : 체감온도
39	stroke : 뇌졸중 가능 지수(1, 2, 3, 4등급)
40	elder_rate_class : 고령 등급(성숙인구사회, 고령화 사회, 고령사회, 초고령사회)
41	city : 1 : 수도권 & 광역시, 0 : 아닌 지역들
42	EQ : 건강 관련 삶의 질의 5가지 차원 (운동능력, 자기관리, 일상활동, 통증/불편, 불안/우울)을 활용하여 만든 지표, 1에 가까울수록 삶의 질이 높음
43	obesity:비만율
44	smoking : 흡연율
45	stress : 스트레스 인지율
46	drink : 고위험 음주율

24번부터 41번까지 파생변수이다. 기온뿐 아니라 일교차와 풍속의 영향을 더한 체감온도 변수를 만들었다. 또한, 심뇌혈관질환에 위험을 높이는 요소인 흡연율, 스트레스, 음주와 같은 생활습관에 해당하는 데이터를 추가하였다.

□ 활용 분석기법 및 모델링 결과

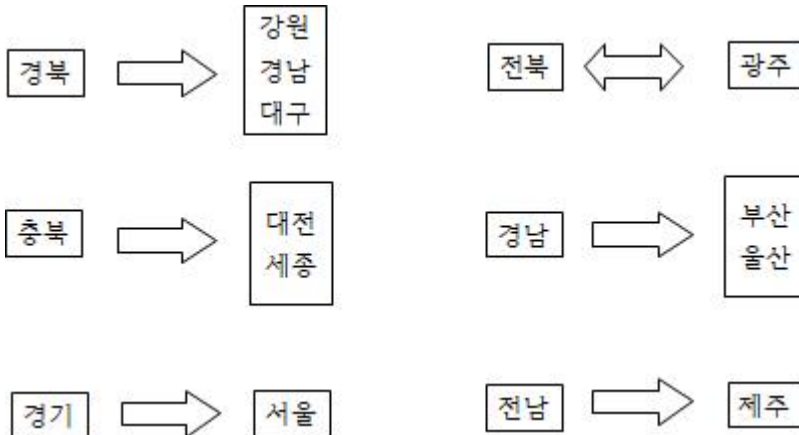
- EDA

*NA 처리

1. 풍속, 기온, 일조율, 기압, 강수량은 사전에 데이터를 합치면서 다음과 같이 NA 처리 과정을 거쳤다.
강수량: 0으로 처리함/ 일조율: 울산을 제외한 나머지 지역들은 평균으로 대체함. 울산은 지역 전체가 NA이기 때문에 0으로 대체함/ 기온, 기압, 풍속: 평균으로 대체함
2. 인구데이터는 세종시가 2012년 01월 01일부터 2012년 06월 30일까지 NA가 있기 때문에 2013년에서 2015년까지의 1월~6월의 월평균으로 대체함
3. 습도: 전북은 2013년 8월의 평균을 충북은 2015년 3월의 평균을 대입함
- 4-1. 예보 데이터 (습도, 일 최고기온) : 2012년 11월, 12월, 2015년 5월 21일~31일 때의 NA를 2012년은 2013년 데이터, 2015년은 2014년 데이터 사용함
- 4-2. 예보 데이터(일 최저기온) : 2012년 11월, 12월, 2014년 12월 강원, 2015년 5월 21~31일 때의 NA를 2012년은 2013년 데이터, 2015년은 2014년 데이터, 강원은 2015년 데이터 사용함
5. pm10 : 화살표 기준으로 왼쪽에 있는 박스가 NA를 대체하는 지역이며, 오른쪽에 있는 박스가 NA가

있는 지역을 의미한다.

우선 충북은 충남으로, 경남은 경북으로 대체한 후 아래의 그림을 기반으로 하여 pm10의 NA를 처리함



*기타처리

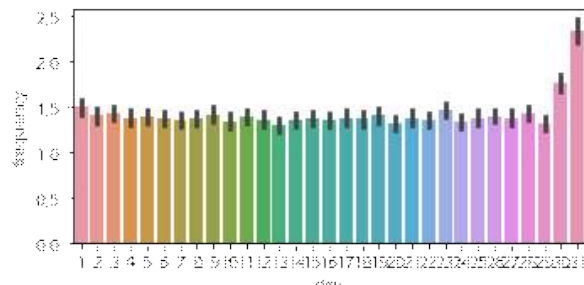
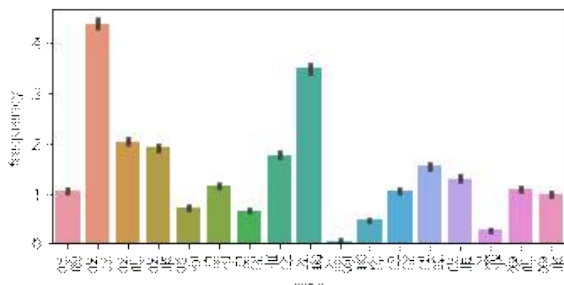
1. 예보 데이터를 사용하여 다음 날의 심뇌혈관질환 발생 빈도를 예측하는 모델을 만들기 위해서 frequency는 2012년 1월 1일에 대한 값을 제거하고, 설명변수는 2015년 12월 31일에 대한 값을 제거하였다.

*시각화

0. 먼저 frequency에 대한 분석을 진행하였을 때, frequency가 10 이상인 경우의 빈도가 상당히 낮기 때문에 일반적이지 않은 경우로 판단, 즉 이상치로 판단하여 frequency에 대한 추가적인 처리를 시도하였다. 이상치를 처리하는 방법으로 해당 이상치가 있는 행을 제거하거나 MAD(Median Absolute Deviation)를 이용하는 방안 등을 시도하였다. 예측 결과까지 반영하여서 frequency를 9 이하인 경우로 제한하는 방법을 채택하여 시각화를 진행하였다.

1. 범주형

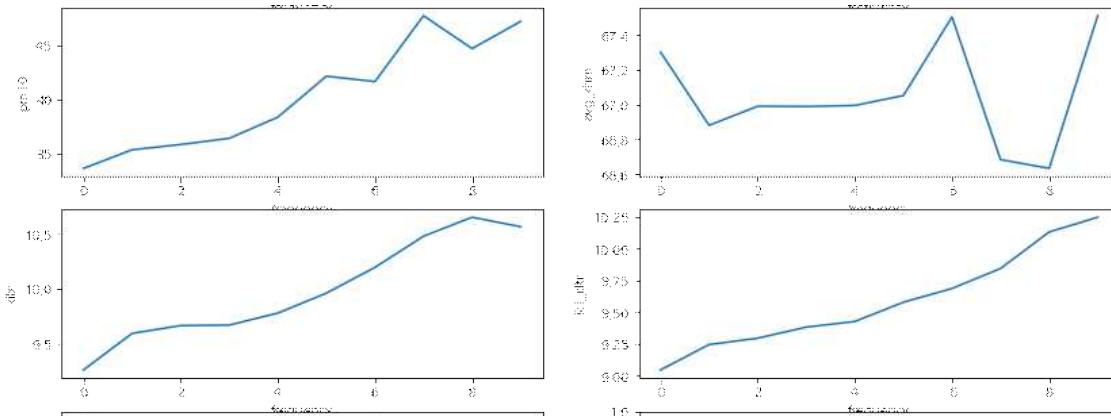
* 범주형 변수들에 대해 범주별 frequency에 대한 막대그래프를 그렸을 때 지역의 경우 경기, 서울에 대해서, day에 대해서는 30~31일과 같은 월말에 특히 높게 나타났으며, 그에 따른 결과로 last_class 역시 월말일 때 frequency가 높게 나왔다. 그 외에도 한파인 날, 일교차가 큰 경우, 미세먼지 농도가 높을 수록, 뇌졸중 가능 지수가 1, 3등급인 경우, 고령화 정도가 고령화 사회 및 초고령화 사회일 때 frequency가 높게 나온 것을 확인하였다.



2. 연속형

* 기상 변수에 대해 전체적으로 일조율, 평균기압, pm10, 일교차, 예보 일교차는 frequency에 대해 증가하는 경향을 보였으며, 평균기온, 평균풍속, 평균습도, 체감온도는 frequency에 대해 감소하는 경향을 나타냈다.

* 전체적으로 흡연/음주/EQ가 낮을수록 높은 frequency를 보이며, 스트레스는 반대의 경향을 보였다.



3. 기타

- * city 변수에 대해서 뇌졸중 가능 지수 및 EQ, stress와 같은 개인 특성 변수들은 전체적으로 수도권이고 광역시일 때가 높게 나왔다.
- * 뇌졸중 가능 지수가 높을 때는 12, 1, 2월(겨울), 뇌졸중 가능 지수가 낮을 때가 6, 7, 8월(여름)로 나타났다.

- 모델링

1. 최종 선택 모델 : LightGBM

frequency를 잘 예측하는 것이 중요하다고 생각하여, 균형 트리 분할 방식보다 예측 오류 손실을 최소화하는 구형 사상을 가지고 수행 시간도 비교적 빠른 LightGBM이 적합하다고 판단함

2. 데이터

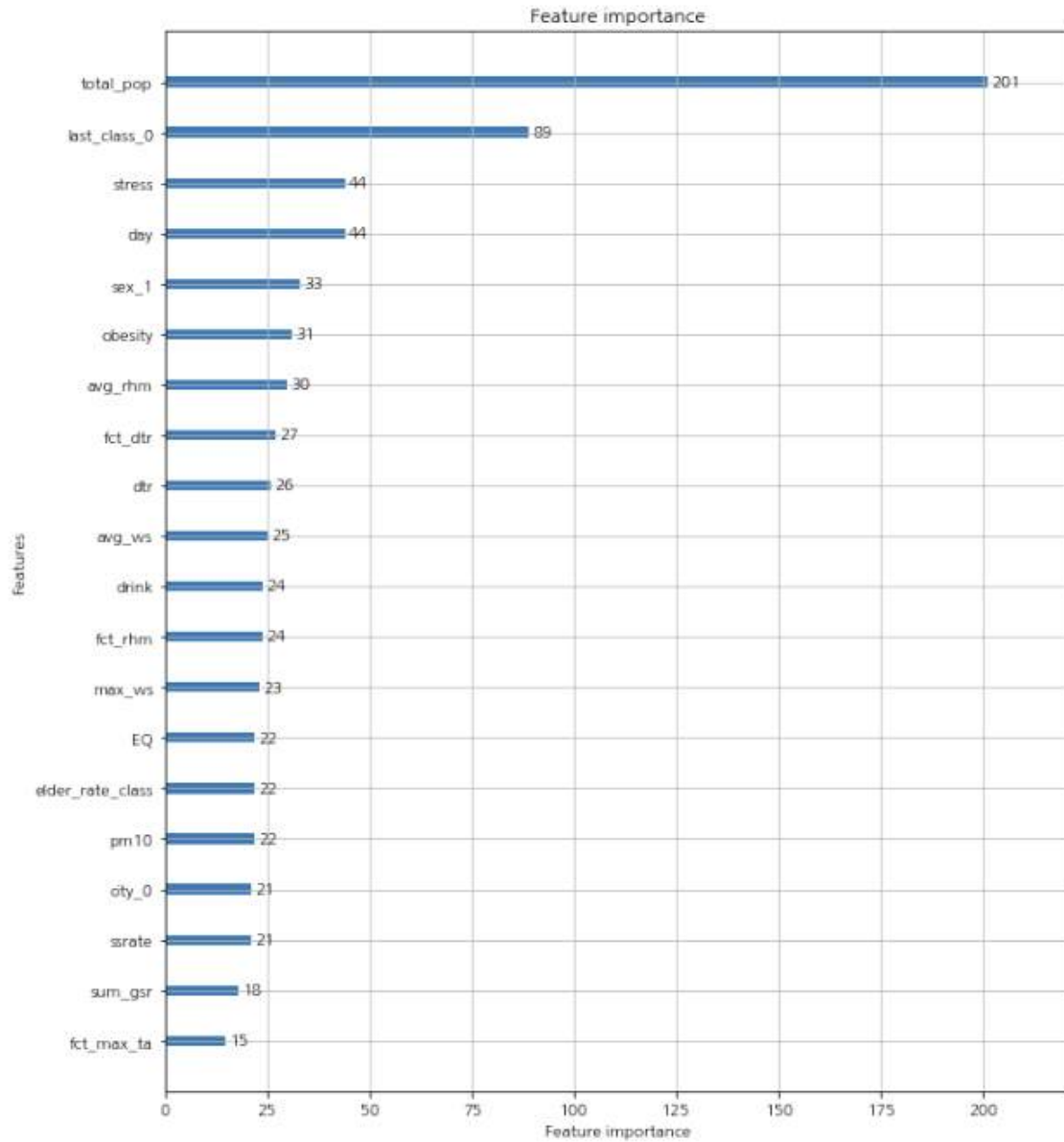
- * EDA 결과 최종 선정된 변수들에 대해 pipe line을 구축하여 한 번에 정돈된 데이터를 사용할 수 있도록 하였다.
- * 범주형 변수들에 대하여 각 변수 타입에 맞게 더미 인코딩/라벨 인코딩을 시행하였다.
- * EDA에서와 동일하게 frequency의 범위를 9 이하로 맞추어 진행하였으며, elder_rate/elder/weekday와 같은 변수들은 제거 시 모델 성능이 다소 향상되었기에 모델링에서 제외했다.

3. 하이퍼파라미터 튜닝

- * 랜덤서치를 사용하여 과적합을 방지하기 위해 트리의 깊이, terminal node에서의 샘플 개수 등에 대한 튜닝을 실시하였다.

4. 변수 중요도 :

- * 심뇌혈관질환 환자 수를 예측하면서 변수 중요도의 상위권에서는 인구적 요인 및 요일, 성별 관련 변수가 높게 측정되었으며, 개인 특성에 관한 변수는 스트레스, 비만을, 음주 관련 변수가 높게 측정됨
- * 기상 변수에서는 습도, 일교차, 풍속 관련 변수들의 중요도가 높게 측정됨
- * pm10, 풍속, 합계전천일사, 일조율, 기온 관련 변수도 중요한 역할을 함
- * EQ, 노인 인구 비율을 등급으로 나눈 변수, 수도권과 광역시가 아닌 지역 또한 중요한 역할을 함



□ 서비스 활용 방안



갤럭시 워치에서 혈압 측정 사용 모습 (사진 = 삼성전자 뉴스룸)

· 헬스케어 앱을 이용해 심뇌혈관질환 주의 경보 알림

- 코로나 19 팬데믹 이후 건강 및 웰빙이 최고의 관심사로 떠오르면서 2021년 한 해 동안 건강 및 피트니스 앱이 10%가량 성장했다. [3] 모바일 헬스는 의료산업의 단점인 시간적, 공간적 제약 등을 보완할 수 있으므로 이를 통해 다음과 같은 방식으로 알림을 보낸다.

- 1) 기존의 헬스케어 앱에 개인의 특성(나이, 흡연율, 음주 횟수 등)을 입력한다.
- 2) 사용자의 위치(거주지역)를 파악해 그에 맞는 기상예보를 통해 심뇌혈관질환 발병자 수를 예측한다.
- 3) 심뇌혈관질환 주의 경보 알림을 보낸다. (예시 : 00 님이 입력한 상태 및 기상 상황에 따르면 현재 심뇌혈관질환 위험군에 속합니다. 아침 운동과 전조증상에 주의하세요. 등)

위의 서비스를 스마트 워치에도 적용하면 스트레스 측정기능도 사용할 수 있게 되어 맞춤형 심뇌혈관질환 감지가 효과적으로 된다. 이를 통해 자신이 얼마나 심뇌혈관질환에 노출되어 있는지 인식하여 위험정도를 인지한다.

· 날씨 키워드에 심뇌혈관 지수 강조

- “날씨”는 많이 검색되는 단어 중 하나이다. [4] 심뇌혈관질환의 인식조사 결과로 우려는 크나 예방 수칙에 대해선 잘 모르는 것으로 나타났다. [5] 헬스케어 비 이용자 혹은 스마트 워치가 없는 사람들을 위해서 검색을 통해서도 심뇌혈관질환 주의 경보를 알린다. 심뇌혈관 지수를 강조하여 주의해야 할 수칙과 전조증상 등을 함께 기재한다. 이를 통해 사람들은 심뇌혈관 지수의 중요성과 깨달음을 얻게 된다.

· 서비스 홍보

- 미디어 및 홍보, 교육이 심뇌혈관질환 관련 정보습득 및 행동 변화를 일으키는 주요 원인으로 밝혀졌다. [5] 그러므로 서비스 홍보를 병행해야 한다. 심뇌혈관질환이 국내 사망률에 큰 영향을 미치는 내용을 포함하여 서비스 배포 및 적극적 활용을 독려한다.

□ 기대효과

1. 스마트한 심뇌혈관질환 관리

- 평소 자신이 심뇌혈관질환의 위험에 노출되어 있는지 위험군인지를 인식
- 의료기관 방문 없이 질병의 예방·관리·치료 디지털 헬스케어

2. 기존 심뇌혈관질환 발병자들을 위한 재발 방지

- 면밀한 후속 관리를 통해 사망의 직접적인 원인이 되는 재발을 막아 사망률을 줄임[7]
- 알림을 통해 전조증상이 나타나면 조기 진료를 받아 골든타임을 확보

3. 젊은 심뇌혈관환자 감소

- 최근 증가하고 있는 젊은 심혈관질환 환자들에게도 경각심을 심어줌

4. 사회경제적 비용 줄임

- 심뇌혈관질환 진료비를 줄여 사회경제적 손실을 막음

5. 운동장려

- 맞춤형 신체활동을 하게 되어 심뇌혈관질환을 예방하는데 중요한 운동 실천율을 높임[8]

□ 참고문헌

- [1] 신방실. (2012년03월16일). 심혈관 질환, 일교차가 클수록 증가.
KBS NEWS <https://news.kbs.co.kr/news/view.do?ncd=2450984>
- [2] 조정호. (2019년03월03일). 뇌심장혈관 질환 골든타임 사수 차재관 교수.
연합뉴스 <https://www.yna.co.kr/view/AKR20190228176700051>
- [3] 정성훈. (2022년01월27일). 앱애니"위드 코로나 시대, 건강·웰빙이 최고 떠오르는 관심사".
파이낸셜신문 <http://www.efnews.co.kr/news/articleView.html?idxno=94074>
- [4] 네이버 블로그. 2021년 네이버 월별 검색어 순위.
<https://blog.naver.com/wonjakga21/222638303388>
- [5] 최선. (2022년07월25일). 심뇌혈관질환 인식-실천 괴리..."예방 수칙 몰라".
MedicalTimes <https://m.medicaltimes.com/News/NewsView.html?ID=1148590>
- [6] 대한심혈관중재학회 채인호 이사장. (2021년06월15일). 전 세계 사망원인 1위 심혈관질환, 재발 예방에 집중해야 하는 이유.
헬스조선 뉴스 https://m.health.chosun.com/svc/news_view.html?contid=2021061100991
- [7] 김맑아. (2010년03월10일). 심혈관질환, '젊은 심장'이 더 위험하다.
헬스조선 https://health.chosun.com/site/data/html_dir/2010/03/09/2010030901306.html
- [8] 이금숙. (2015년09월18일). 심뇌혈관질환 예방에 가장 중요한 것은 운동.
헬스조선 https://m.health.chosun.com/svc/news_view.html?contid=2015091702121