**프로젝트 기획안**

**2022년 12월 20일**

**과정명 : 프로젝트형 빅데이터 분석 서비스 개발**

| 팀 명 | 건강 삼총사 |
| --- | --- |
| 팀 원 | 팀장: 박민정  팀원:서대훈, 손지호 |
| 프로젝트 주제 | 시도별 심장 질환 위험군 분류를 통한 의료 수준 점검 |
| 프로젝트 수행 방향  (주요 기능 설명) | **● 데이터 설명**  **심장 질환**  [2021 BRFSS Data(SAS Transport Format)](https://www.cdc.gov/brfss/annual_data/annual_2021.html)  **개인 건강 정보**  [국민건강보험공단 건강검진정보](https://www.data.go.kr/data/15007122/fileData.do#tab-layer-openapi)  **지역별 의료기관 현황**  [보건의료빅데이터개방시스템 의료통계정보](https://opendata.hira.or.kr/op/opc/selectMedInfoSvcList.do)  [119 구급대 현황](https://kosis.kr/statHtml/statHtml.do?orgId=156&tblId=TX_15602_A002&vw_cd=MT_ZTITLE&list_id=156_15602_A001&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)  **● 주요 분석 및 모델링 내용**  **심장 질환 분류 모델**   1. 심장 질환이 일어나는 특성 파악 2. 1의 결과를 바탕으로 심장 질환 분류 모델 작성   **시도별 분석**   1. 건강검진정보를 심장 질환 분류에 적용    1. 개인별로 심장 질환 위험군을 분류한다    2. 지역별로 심장 질환 위험군으로 분류된 사람의 수를 도출한다 2. 의료 수준 분석    1. 1-b를 바탕으로 지역별 의료 수준과 심장 질환 위험군 수를 분석한다    2. a를 토대로 의료 수준과 관련된 의견을 제시한다 |
| 역할 분담 및 일정 | **● 역할분담**  박민정: 자료 조사 및 전처리, 시각화, 발표자료 작성  서대훈: 자료 조사 및 전처리, 특성 선정, 발표자료 작성  손지호: 자료 조사 및 전처리, 모델링, 발표자료 작성  **● 일정**   |  | 9 | 12 | 13 | 14 | 15 | 16 | 17 | 20 | 21 | 22 | 23 | 26 | 27 | 28 | 29 | 30 | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 주제 탐색 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 데이터 정리 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 주제 선정 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 시각화/변수 선정 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 데이터 모델링 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 발표 준비 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| 프로젝트 수행 도구 | **● 전처리 기술**  - R foreign  - Python BeautifulSoup, urllib  - Python Pandas  **● 시각화 기술**  - R ggplot2  - Python folium  **● 모델링 기술**  - Python Scikit-Learn |