

ESTsoft AI 모델 개발자 양성과정 10기

데이터 수집 및 저장 프로젝트 기획서

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 프로젝트 기획서
- 제출 일자 : 5/28(수)
- 깃허브 경로 : <https://github.com/gigas5234/dust>
- 팀 설명 : 6조 “먼지가 되어”
- 작성 팀원 : 김남준

| | |
|----------|--|
| 프로젝트 주제 | 미세먼지 농도와 호흡기 질환 발생 간의 상관관계 분석 및 시각화 |
| 배경 및 필요성 | <p>공중보건 이슈: 미세먼지(PM10) 및 초미세먼지(PM2.5)는 천식, 비염, 만성 폐질환 등 다양한 호흡기 질환을 유발하거나 악화시킨다는 연구 결과가 다수 존재합니다.</p> <p>정책 수립 기반 마련: 지역 및 연령대에 따라 건강 영향이 다르게 나타나는 만큼, 정량적 분석이 중요합니다.</p> <p>데이터 기반 의사결정: 질병관리청, 건강보험공단 데이터와 미세먼지 데이터를 통합 분석함으로써, 의료 대응 및 환경 정책 수립의 기초 자료로 활용할 수 있습니다.</p> |
| 목적 | <p>월별 지역별 미세먼지(PM10, PM2.5) 평균값과 호흡기 질환 진료 건수를 연계 분석</p> <p>연령별·지역별 민감도 분석 (고령자, 수도권 중심)</p> <p>계절성 및 추세 시각화를 통한 공공 인사이트 제공</p> |
| 분석 범위 | 데이터 수집 및 전처리, EDA (탐색적 데이터 분석), 상관분석, 회귀분석, 시각화 및 정책 제언 |
| 데이터 전처리 | <p>날짜 및 지역 기준 통합 키 생성 (ex. 2020-05 / 서울특별시)</p> <p>결측치 처리 및 단위 통일 ($\mu\text{g}/\text{m}^3$, 명)</p> <p>계절 변수 생성 (봄/여름/가을/겨울)</p> <p>연령대, 성별 기준 집계</p> |
| 모델링 방안 | <p>상관분석: Pearson/Spearman 계수 분석</p> <p>통계적 검정: ANOVA, t-test (연령대, 성별, 지역 비교)</p> <p>회귀분석: 선형 회귀, 계절 조정 회귀</p> |
| 기대효과 | <p>과학적 가설 검증을 통한 정책 수립 기반 제공</p> <p>미세먼지 고위험군 예측 및 지역 기반 의료 대응 가이드 제시</p> <p>환경 데이터와 건강 데이터를 연계한 공공데이터 활용 사례 제시</p> |
| 역할 및 책임 | <p>프로젝트 매니저 (김남준) : 일정 관리, 전체 통합 문서 관리</p> <p>데이터 엔지니어 (김원혁) : 미세먼지 데이터 전처리 및 통합 상관/회귀 분석, 시각화 코드 구현</p> <p>데이터 분석가 (이원일) : 건강보험 진료 데이터 분석, 통계 검정</p> <p>머신러닝 엔지니어 (신지웅) : 상관/회귀 분석, 시각화 코드 구현</p> |