

# 프로젝트 제안서

# 프로젝트 개요

| 제목      | 설명   |
|---------|--|
| 프로젝트 이름 | 공공 데이터 분석을 통한 생활 정보 추천 웹 서비스   |
| 프로젝트 기간 | 2024년 10월 30일 ~ 2024년 12월 06일  |
| 프로젝트 발표 | 2024년 12월 07일  |
| 프로젝트 설명 | 국내/해외 공공 데이터 및 다양한 데이터 서비스를 활용하여 데이터 분석 서비스를 설계하고 데이터 수집, 가공, 전처리, 시각화를 통해 인사이트 도출 |

# 프로젝트 제안 배경

## ■ 빅데이터 분야 성장 현황

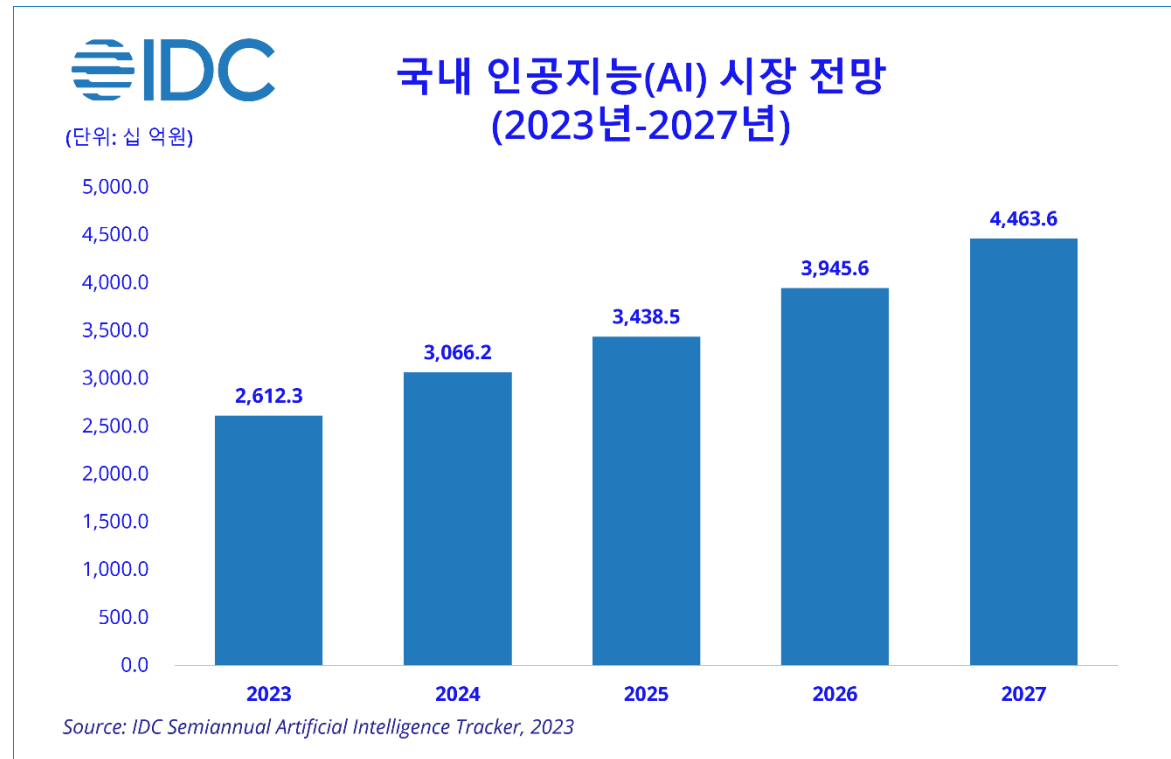
- > IT 분야 시장 분석 및 컨설팅 기관인 (한국)IDC의 조사에 따르면 2023년 국내 빅데이터 시장이 전년 대비 11.1% 성장한 2조 7054억원 규모의 매출이 발생할 것으로 예상됨
- > 같은 보고서에 따르면 특히 2027년까지 연평균 10.6% 성장하여 3조 9771억원 규모에 달할 것으로 예상됨



# 프로젝트 제안 배경

## ■ 인공지능 분야 성장 현황

- ▶ IT 분야 시장 분석 및 컨설팅 기관인 (한국)IDC의 조사에 따르면 2023년 인공지능 시장은 전년 대비 17.2% 성장하여 2조 6163억원의 매출 규모를 형성할 것으로 예상됨
- ▶ 같은 보고서에 따르면 특히 향후 5년간 연평균 성장률 14.9%를 기록하여 2027년 전체 매출은 4조 4636억원 규모에 이를 것으로 예상됨



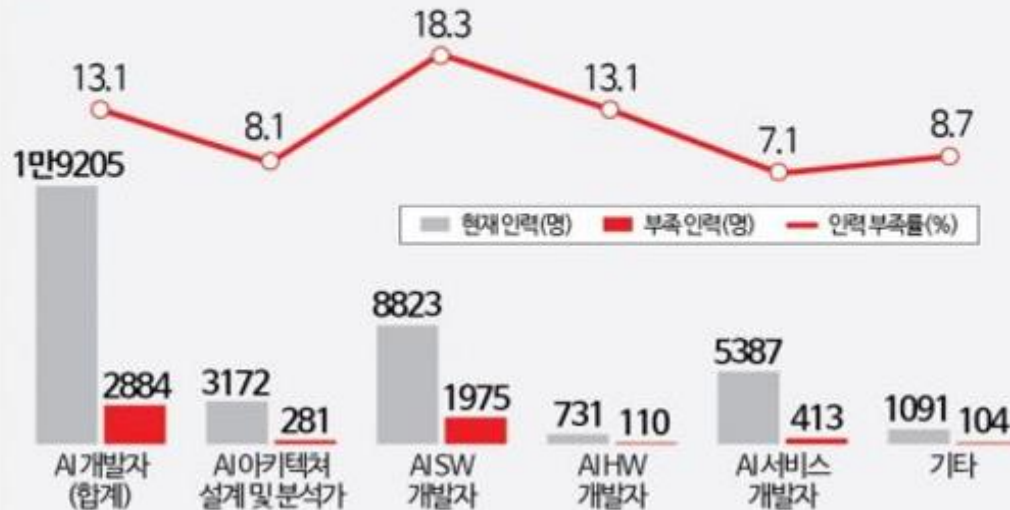
# 프로젝트 제안 배경

- 인공지능 분야 인력 수급 현황

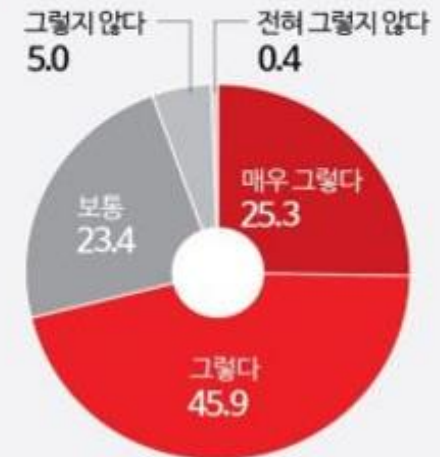
▶ 과학기술정보통신부 산하 소프트웨어정책연구소가 발표한 2021 인공지능산업실태 조사에 따르면 AI 사업자들의 71.2%가 인력 부족으로 어려움을 겪고 있는 것으로 나타나는 등 인공지능 분야 전문인력이 크게 부족한 상황

## AI개발자인력 현황

2021년 AI개발자인력 현황 (자료: 소프트웨어정책연구소)



AI사업운영상 느끼는 AI인력부족문제 (단위: %)



# 프로젝트 목표

- 프로젝트 수행을 통해 데이터 분석 및 예측 모델 개발 프로세스에 필요한 종합적인 역량 향상
  - ▷ 학습자는 실제 프로젝트 수행을 통해
  - ▷ 다양한 문제 상황을 경험하고 이를 해결하는 과정에서
  - ▷ 교육을 통해 학습한 기술 및 역량에 대한 성취도가 종합적으로 향상될 수 있으며
  - ▷ 데이터 분석 기획, 데이터 수집, 전처리, 시각화 등의 업무 프로세스에서 실질적인 역할 수행이 가능한 수준으로 역량 향상

# 프로젝트 팀 구성

| 팀  | 인원            |
|----|---------------|
| 1팀 | 이원빈, 임병남, 우병준 |
| 2팀 | 이유리, 조수연, 최현묵 |
| 3팀 | 오용석, 원석재, 이정인 |

# 적용 기술 및 도구

- 적용 기술

| 분류      | 기술                             |
|---------|--------------------------------|
| 개발 언어   | Python ( HTML, CSS )           |
| 데이터 수집  | File, Open API, Web Crawling   |
| 데이터베이스  | DB Modeling, SQL, DBMS (MySQL) |
| 데이터 전처리 | Python                         |
| 탐색적 분석  | 데이터 시각화                        |



# 적용 기술 및 도구

- 도구

| 분류        | 도구                                   |
|-----------|--------------------------------------|
| 코드 개발     | Visual Studio Code                   |
| 데이터베이스    | MySQL, MySQL Workbench, eXERD        |
| 데이터베이스 연동 | Pymysql                              |
| 형상관리      | github, git-client                   |
| 탐색적 분석    | numpy, pandas, matplotlib, seaborn 등 |

# 프로젝트 수행 요구사항

- 프로젝트는 팀 단위로 진행되며 팀 별로 3개 이상의 데이터 소스를 기반으로 데이터를 수집하고 분석하여 인사이트를 도출해야 한다
  - 전체 팀원이 참여해서 역할을 수행할 수 있도록 프로젝트의 주제와 범위 결정
  - 선택된 주제에 대해 다양한 데이터를 수집하고 유의미한 인사이트 도출
  - 각 팀 구성원은 전체 팀 주제에 포함되는 개별 분석 작업을 수행하되 각자 데이터 수집, 전처리, 시각화 과정을 수행해 볼 수 있도록 역할 분담

# 프로젝트 수행 요구사항

- 과제는 다음의 단계에 따라 작업을 수행해야 한다
  - 주제 선정 및 가설 설정
  - 데이터 수집
  - 데이터 저장
  - 탐색적 데이터 분석 및 데이터 전처리
  - 데이터 시각화

# 프로젝트 수행 요구사항

- 데이터 수집은 파일, OpenAPI, 웹크롤링 등 필요한 방법을 선택하여 수행한다
- 수집된 데이터는 파일 또는 데이터베이스에 저장한다
- 탐색적 데이터 분석 및 전처리 과정은 다음의 작업을 포함해야 한다
  - › 각 속성별 결측치 탐지 및 처리 (삭제, 대체 등)
  - › 각 속성별 기초 통계량 및 분포 특성 파악
  - › 각 속성별 이상값 탐지 및 처리
  - › 그 외 필요한 분석
  - › 비정형 데이터의 경우 가능한 범위에서 분석 및 전처리 수행

# 프로젝트 수행 산출물

- 데이터 분석 기획서 (프로젝트 발표 문서에 포함)
- 데이터 분석 수행 일정 명세 (프로젝트 발표 문서에 포함)
- 프로젝트 발표 문서
- 프로젝트 시연 영상
- 데이터 분석 수행 노트북 소스 파일
- 데이터 분석 수행 노트북 HTML 출력 파일