**기업요구사항 기반의 문제해결 프로젝트 기획안**

**2022년 10월 14 일**

|  |  |  |
| --- | --- | --- |
| **프로젝트 조** | **1조** | |
| **프로젝트 팀원** | 팀명:  팀장: 손지수  팀원: (사이언스) 김경현, 김태훈, 이지훈 (엔지니어링) 이상엽, 이서정 | |
| **프로젝트 주제** | 입양 희망자의 성향과 니즈를 파악해 잘 맞는 유기견을 매칭 시켜주는 서비스 | |
| **프로젝트 수행 방향**  **(주요 기능 설명)** | **사이언스** | **엔지니어링** |
| 유기동물을 군집화하여 사용자가 원하는 항목을 고르면 그에 가까운 유형의 유기동물을 매칭시켜 줌  로그인 시 작은 설문 항목을 만들어 이용자의 주거환경, 경제상태, 입양 시 중점적으로 보는 부분(견종크기, 털빠짐...)등을 조사할 수 있도록 함 | 유기견 정보와 견종별 특성 데이터를 수집하고 구현될 웹 서비스의 로그 데이터를 수집하여 분석 모델에 활용할 수 있도록 데이터 파이프라인을 구축함 |
| **프로젝트 수행 도구** | **사이언스** | **엔지니어링** |
| - Jupyter Notebook  - PyCharm | - VS Code  - Jupyter Notebook  - Docker |
| **협업 도구**  - GIT : 병합 및 버전 관리, 기본 템플릿 공유  - Trello : 일정, 업무 진행 상황 파악, 기획서, 설계도, 회의록 정리  - Slack : 채널을 통한 팀 별 소통, 화면 공유 및 음성 대화, 개인 메시지  **데이터**  - 농림축산식품부 동물보호관리시스템 유기동물 정보 API, 네이버 지식 백과 견종별 특성 데이터 | |

|  |  |  |  |
| --- | --- | --- | --- |
| **프로젝트 목적** | * 유기견 파양률 감소 * 유기견 입양 문화 활성화 * 입양희망자가 유기견의 성향을 잘 파악할 수 있도록 정보 제공 | | |
| **필수 기능** | **사이언스** | **엔지니어링** |  |
| 1.군집화  - 가장 군집화를 잘할 수 있는 알고리즘 선정  2.실루엣 분석  - 군집화 평가지표  - 각 군집 간의 거리가 얼마나 효율적으로 분리되어 있는지 분석  3.추천시스템  - 추후 설문이나 로그등을 통해 데이터가 확보되면 그것을 바탕으로 이용자에게 유기견을 추천 | 1. 유기견 정보, 견종별 특성 데이터 수집 - HDFS  - 유기동물 정보 API 활용  - 네이버 지식 백과 견종별 특성 크롤링  2. Data Warehouse, Data Mart 설계 및 구축 - Oracle ATP  - ERD Cloud 활용 DB 테이블 설계  - Oracle ATP 계정에 DW/DM 구축  3. 데이터 파이프라인 구축  - 각 ETL 과정을 클래스화 하여 app으로 구현  - Airflow 배치관리로 파이프라인 자동화  4. 웹 서비스 개발  - Rest API 서버 개발- Django Rest Framework 활용  5. 실시간 로그 데이터 수집  - Kafka 활용 |  |
| **포함 기술** | **사이언스** | **엔지니어링** |  |
| - Python  - Scikit-learn | - Python, PySpark, Django  - Spark, Hadoop, Oracle, Kafka |  |

**□ WBS 첨부 ( OR 별도로 폴더 업로드)**