|  |  |
| --- | --- |
| **프로젝트 개요서** |  |

**1. 프로젝트 정보**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **참여 프로젝트 정보** | 구분 | 세부내용 | | | | |
| 프로젝트  주제 | 빅데이터를 이용한 오피니언 마이닝 기반 창업 추천 서비스 | | | | |
| 개발 목표 | ▪ 통계 데이터를 이용한 상권 분석  ▪ 오피니언 마이닝을 이용한 창업 추천 | | | | |
| 예상 수행  기간 | 기간 : 2023.04.20 ~ 2023.05.08 | | | | |
| 수행 내용 | 1. 프로젝트 기획  1) 기획 및 시장 조사  2) 기술조사 및 사례조사  3) 프로젝트 목표 및 진행 단계 설정  4) 서비스 흐름도 및 화면 구성도 구성  5) 기획서 및 개요서 작성  2. 프로젝트 준비  1) 크롤링 공부  2) 협업 여부 확인  3. 데이터 수집 및 전처리  1) 상권 분석용 데이터 수집  2) SNS 오피니언 데이터 크롤링  3) 수집 데이터 정제  4) 데이터 정형화  4. 모델링  1) Linear Regression, LightGBM (Classifier, Regressor) 테스트 후, 가장 적합한 모델 선정  2) 학습  3) Validation 진행    5. DB Table 구성  1) 각종 Table 정의 및 구성 | | | | |
| 사용 기술 | [Front End]  - Tableau Public  [Back End]  - Python  - Colab  [Data Engineering]  - Library : Sklearn, Pandas, Numpy, Selenium, BS4  [Modeling]  - Framework : Sklearn, Pickle, Pandas, Numpy, LightGBM  [Visualization] – Seaborn, Matplotlib | | | | |
| **시스템 아키텍처&**  **서비스 흐름도** |  | | | | | |
| **예상 결과물** | Web | | App | 모델 | 시제품 | API |
|  | |  | ○ | ○ |  |

**2. 필요 기반지식 및 활용 데이터**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **필요 기반지식** | 프로젝트 진행 전 미리 알아두면 좋은 기반 지식에 관한 내용입니다. | | | |
| - 크롤링 | | - 머신러닝 | - 태블로 퍼블릭 |
| **활용 데이터** | 프로젝트 진행 시 활용할 데이터에 관한 내용입니다. | | | |
| 수집할 데이터 | 행정동코드, 상권상주인구, 상권직장인구, 상권생활인구, 상권영역, 상권점포, 상권집객시설, 상권추정매출 | | |
| 관련 URL | 서울시 열린데이터 광장 https://data.seoul.go.kr/ | | |
| 공공데이터 포털 https://www.data.go.kr/ | | |
|  | | |