**기업요구사항 기반의 문제해결 프로젝트 기획안**

**2022년 07월 28일**

| **프로젝트 조** | **4조 4ir Jordan** | |
| --- | --- | --- |
| **프로젝트 팀원** | 팀명: 4ir Jordan  팀장: (DS) 이재연  팀원: (DS) 김예은, 신혜빈 (DE) 김흥민, 박상권, 박선재, 최진호 | |
| **프로젝트 주제** | 한정판 Jordan 운동화 분석 및 타겟층별 신발 추천, 한정판 신제품 운동화 리셀 가격 예측, 셀럽의 인스타그램 Jordan 게시물 알림 서비스 | |
| **프로젝트 수행 방향**  **(주요 기능 설명)** | **사이언스** | **엔지니어링** |
| Jordan 운동화 추천 프로그램  Jordan 신제품 가격 예측  셀럽의 인스타그램 Jordan 게시물 알림 서비스 | 데이터 수집 및 전처리  파이프라인 구축  데이터베이스 구축  데이터 수집 및 처리 자동화  Jordan 운동화 분석 및 시각화  셀럽의 인스타그램 Jordan 게시물 알림 서비스 |
| **프로젝트 수행 도구** | **사이언스** | **엔지니어링** |
| Python  Jupyter notebook  Tensorflow  Keras  OpenCV  Pandas  Matplotlib | Python  Jupyter notebook  Colab  VScode  MongoDB  MySQL  Scrapy  Selenium  Spark  Pandas  scikit-learn  Airflow  Slack bot |

| **프로젝트 목적** | Jordan 운동화 분석을 통해 제품 가격 형성에 영향을 미치는 요인 도출  Jordan 입문자를 위한 신발 추천 프로그램 제공  신제품의 리셀 시장 진입 후 최대, 최소, 가장 최근 거래 내역 가격 예측  셀럽의 인스타그램 Jordan 게시물 알림 서비스 | |
| --- | --- | --- |
| **필수 기능** | **사이언스** | **엔지니어링** |
| [Jordan 운동화 추천]   * Jordan 이미지 분류 모델 학습 및 학습된 모델 층을 이용한 512차원의 벡터로 임베딩 * K-Means 군집을 이용하여 신발 추천 프로그램 작성   [Jordan 신제품 가격 예측]  엔지니어링 팀이 가져온 데이터를 XGBoost를 이용하여 신제품의 리셀 시장 최대, 최소, 가장 최근 체결 가격을 예측  [인스타그램 Jordan 게시물 알림 서비스]  게시물 사진을 Fast R-CNN 모델을 이용하여 신발 부분만 추출 후 CNN 모델을 통해 Jordan 운동화 모델을 분류하고 Jordan 모델명을 반환하는 함수 작성 후 엔지니어링 팀에 전달 | [수집]   * kream 사이트에서 제공하는 모든 한정판 리셀 제품 정보 (제품명, 모델번호, 출시일, 컬러, 발매가, 체결내역, 빠른배송, 게시물 수, 찜, 총 거래수, 사이즈, 제품 이미지, 차트 이미지 등 모든 데이터)을 Scrapy, Selenium을 활용하여 크롤링 * LuckyDraw 사이트에서 제공하는 신상품 리셀 제품 정보 (제품명, 모델번호, 출시일, 컬러, 발매가, 제품 이미지)를 Scrapy 이용를 활용하여 크롤링 * 리셀 뉴스 기사를 Selenium을 활용하여 크롤링   [저장]   * MongoDB에 모든 데이터를 저장하여 Data Lake구축   [처리]   * 데이터 정제 * 제품명에서 라인, 콜라보, 타입, 시리즈 정보 추출 * 제품 사진에서 배경을 제외하고 신발 이미지에 해당하는 픽셀의 RGB값 추출   모든 사진의 픽셀 값을 하나의 데이터프레임으로 변환  K-means를 통해 군집화를 진행하여 각 클러스터별 평균 RGB 값을 구하고, 색깔 이름 레이블링 진행  [분석]   * Jordan 운동화 분석 및 시각화를 통해 제품 가격 형성에 영향을 미치는 요인 도출   [자동화]   * Airflow를 이용한 럭키드로우 크롤링 자동화 * 데이터처리 파이프라인 구축   [인스타그램 Jordan 게시물 알림 서비스]   * Airflow를 이용하여 셀럽이 착용한 신발 알림 서비스 구현 |
| **포함 기술** | **사이언스** | **엔지니어링** |
| [Jordan 운동화 추천]   * CNN * KMeans   [Jordan 신제품 가격 예측]   * XGBoost   [인스타그램 Jordan 게시물 알림 서비스]   * CNN * Fast R-CNN | [수집]   * Scrapy * Selenium   [저장]   * MongoDB * MySQL   [처리]   * Pandas * Spark * scikit-learn   [분석]   * Matplotlib * Seaborn   [자동화]   * Airflow   [인스타그램 Jordan 게시물 알림 서비스]   * Selenium * Airflow * Slack bot |

**□ WBS 첨부 ( OR 별도로 폴더 업로드)**