**프로젝트 기획안**

**2021년 10 월 11 일**

**과정명 : 데이터 사이언스/엔지니어링 전문가**

|  |  |
| --- | --- |
| 팀 명 | Day-car(데이카) |
| 팀 원 | 팀장: 안기용  팀원: 왕기영, 강민정, 김문혁, 안성준 |
| 프로젝트 주제 | 중고차 시장 시세에 따른 중고차 가격 예측 |
| 가상 시나리오 | **시나리오**  - 중고차 가격 예측을 통해 중고차 판매자를 상대로 차량의 상태에 따른 시세를 제공하고자 함  - 사이언스 팀에는 다양한 데이터를 분석하여 사업 타당성을 검토하고 구축된 학습데이터를 이용 학습하여 모형화 작업을 진행한 후 서비스로의 연계를 위한 모델 결과 도출  - 엔지니어링 팀에는 중고차의 가격, 상태 등의 다양한 데이터를 수집 및 가공하여 학습용 데이터 구축  - 웹서비스개발팀(사이언스 및 엔지니어링 통합)은 중고차 상태를 입력하면 그에 대한 예상 가격이 나올 수 있게 사이트를 구축하고 DB와 연동하여 검색 시 시세와 시각적인 요소가 나올 수 있게 함 |
| 프로젝트 수행 방향 | ● 데이터 수집 및 구축 개요  데이터 수집 : 여러 중고차 사이트를 크롤링하여 데이터를 수집하고 전처리 후 DB에 적재  ● 주요 분석 내용  가격, 연식, 모델 등의 정보를 학습하여 중고차의 옵션에 따른 가격을 예측  ● 기대 효과  시장에서 중고차 구매 혹은 판매에 대한 예상 가격에 대한 파악이 가능하며, 전반적 중고차 시장의 시세를 파악 가능할 것으로 기대 |
| 역할 분담 및 일정 | ● 역할분담  데이터 엔지니어링팀 : 데이터 전처리 및 수집하고 DB 구축과 웹서비스 구축  데이터 사이언스팀 : DB 저장된 데이터를 이용하여 최적의 모델링 기법을 찾아 학습하여 결과 도출 및 시각화  ● 일정  10/05 ~ 10/11: 주제 선정 및 일정 수립  10/12: 주제 발표  10/13 ~ 10/15: 데이터 수집  10/18 ~ 10/20: 데이터 전처리  10/20 ~ 10/21: DB 구축  10/21 ~ 10/27: 데이터 분석 및 결과 도출  10/28 ~ 11/08: 웹서비스 구축  11/09 ~ 11/11: 발표 준비  11/12: 프로젝트 발표 |
| 프로젝트 수행 도구 | ● 빅데이터 수집, 저장 시스템  - Hadoop, Spark, MySQL  ● 분석기술  - Sklearn  ● 시각화 기술  - Matplotlib, Seaborn  ● 웹서비스 기술  - Django |