**[데이터 사이언스]**

**문제해결 빅데이터 활용 프로젝트 기획안**

기획안 작성일자 : 2022.08.03

| 조 명 | 4조 : 보람상조 |
| --- | --- |
| 조 원 | 조장: 우성균 조원: 김민철, 이성민, 이유경, 안소연 |
| 프로젝트 주제 및 개요 | 운전상 주위환경정보를 바탕으로 사망사고 예측, 교통사고를 유발하는 요인 분석 |
| 프로젝트 수행 방향 | 1. EDA  - 연령대별에 따른 사고율의 상관관계가 존재하는지 분석  - 기상요소에 따른 사고율의 상관관계가 존재하는지 분석  - 사고율에 대한 종단적 분석(교통사고율, 그로인한 사망률, 자차소유율등의 변화 통계 분석)  - 운전면허 난이도에 따른 사고율의 상관관계가 존재하는지 분석  - 운전면허 시험직후 2년간 사고율의 변화가 존재하는지 분석  - OECD 국가별 면허 난이도에 따른 사고율,그에 따른 사망률 분석  2.머신러닝-분류  - 각 교통사고의 여러요인들을 토대로 교통사고가 발생했을시 피해자가 사망에 이를지, 아닐지를 예측하는 분류모델 생성 및 학습  - 피처:사고유형,법규위반,기상상태,도로상태,가해자차종,피해자 차종,피해자 연령,피해자 성별 등  - 레이블: 생존:0, 사망:1 |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  - 각 기능이 아닌 주제별로 역할을 나누기로 하였고, 본인이 맡은 주제가 끝나면 다른 사람을 도우는 방식을 채택하였습니다. 그렇기 때문에 추후 변동사항이 존재할 수 있습니다  - 8월 10일부터 다 함께 머신러닝 성능 향상을 위한 업무를 하도록 하고, 그 전까지 저는 존재하는 분류모델들을 구현해보고, 정리하는 시간을 갖도록 계획하였습니다  - 지금 계획된 EDA뿐만 아니라, 교통사고와 인과관계가 강할 것 같아보이는 요인들에 대하여 추가적으로 더 조사할 수 있습니다.  조장 : 우성균  -머신러닝 분석을 위한 가이드라인 준비, 스케줄 조율  조원1: 이성민  - EDA:사고율에 대한 종단적 분석  조원2: 김민철  - EDA: 연령대별에 따른 사고율 분석  조원3: 이유경  - EDA:운전면허 난이도에 따른 사고율 분석  조원4: 안소연  - EDA:기상요소에 따른 사고율 분석 |
| 데이터셋 출처 | - EDA: 사고율에 대한 종단적 분석(이성민)  - 도로교통공단: 사망교통사고 정보(2022)  - 한국도로공사 공공데이터포털: 고속도로 중심선 위치  - TAAS: OECD 국가별 사고 통계(2018), 도로형태별 교통사고(2020~2020)  - EDA: 연령대별에 따른 사고율 분석(김민철)  - TAAS:연령층별 교통사고(2020~2021)  - KOSIS:연령층별 운전면허 소지자 현황(2020~2021), 총조사인구(2020)  - EDA: 기상요소에 따른 사고율 분석(안소연)  - KOSIS: 도로종류별\_기상상태별\_교통사고\_최근5년 전체데이터(2017~2021)  -EDA: 운전면허 난이도에 따른 사고율 분석(이유경)  - 공공데이터포털: 서울특별시 운전면허 시험 통계(2020)  - TAAS: 국가별 사고대비 사망자 통계(2019-독일,멕시코,스웨덴,일본,한국)  - 머신러닝:운전상 주위환경정보를 바탕으로 사망사고 예측(우성균)  - TAAS: 교통사고 분석 시스템 |
| 프로젝트 추진 일정 | ● 일정  7/28 ~ 8/2. 주제 수립. 교통사고의 부상심각성을 높이는 요인, 교통사고를 유발하는 요인 분석  8/3 ~ 8/9. 데이터 수집, 데이터 전처리, 데이터 EDA 완료, 머신러닝- 분류모델 들을 모두 구현해보고 정리하여 가이드라인을 만들어놓음(본인:성균)  8/10~8/12. 모델 성능 향상을 위한 머신러닝 학습  8/13~8/16. 프로젝트 발표 준비 |