

과제 수행 계획서 (조 구성)

조 이름	JB			
조 소개	집행부 인원들로 구성된 조			
운영규칙	<ul style="list-style-type: none"> ● 매주 월요일 혹은 수요일 저녁 19~21시에 정기적으로 모여 프로젝트 회의를 진행한다. ● 매주 데드라인을 설정하고 지키기 위해 노력한다. ● 데이터 분석 파트와 웹 구현 파트를 나눠 진행한다. ● 주요 사항 결정은 3명 이상 동의 하에 결정된다. 			
문제해결 전략	<ul style="list-style-type: none"> ● 데이터들 간의 상관관계를 파악하여 인사이트를 도출한다. ● 최적의 분석을 위해 여러 분석기법을 사용하여 예측 점수를 올린다. ● 복잡한 데이터를 시각화하여 직관적으로 보여주면서 데이터의 의미를 이해한다. ● 관련된 논문들을 참고하여 문제해결에 도움을 준다. <p>(참고논문: "머신러닝 기반 범죄발생 위험지역 예측", 저자 : 허선영, 김주영, 문태헌)</p>			
조원	임무	이름	연락처 (HP, E-mail)	주요 업무
	팀장	반수경	010-5344-5115 sugyeong0425@gmail.com	데이터 수집, 전처리, 분석, 예측모델과 웹 페이지 연동, 발표
	팀원	김민제	010-2725-7231 alswp7231@gmail.com	데이터 수집, 전처리, 웹 페이지 제작, ppt 발표자료 준비
	팀원	김성준	010-3865-2151 sjk5053@naver.com	데이터 수집, 전처리, 웹 페이지 제작, ppt 발표자료 준비
	팀원	최원태	010-6250-4196 jonathan0927@naver.com	데이터 수집, 전처리, 분석, ppt 발표자료 준비

※ 조 운영을 위한 규칙 예시

- 조 활동을 위해 매주 수요일 오후 7~8시 사이에 정기적으로 회의를 한다.
- 특별한 사유 없이 2회 이상 회의에 불참하는 경우, 조별 상호평가 시에 불이익을 준다.
- 주요 사항의 결정은 조 구성원 2/3 이상이 찬성해야 한다.
- 팀 구성원의 역할은 3주 단위(또는 과제 단위)로 변경한다.
- 기타 조 운영 규칙에 명시되지 않은 사항은 조 회의를 통하여 결정한다.

과제 수행 계획서 (문제 분석지)

Topic	데이터 분석을 이용한 범죄 위험 지역 알림 서비스		
조 이름	JB	날짜	10.05
조원	반수경, 김민제, 김성준, 최원태		

프로젝트 목표	데이터 분석을 통해 범죄 발생 예측 모델을 개발하고, 웹 사이트로 연동하여 서비스화 시킨다.
----------------	---

가설/해결안 (Ideas)	
<p><ideas></p> <ul style="list-style-type: none"> - 정부에서 제공하는 범죄 정보 데이터에 날씨정보(기온, 강수량, 풍속, 습도, 일조 등), 지역정보(평균 건폐율, 평균 높이, 총 건축물 수, 평균 지상층 수 등), 인구정보(총 인구수, 1인 가구 수, 여성 주거 인구 수 등) 등을 이용해 예측 모델을 개발한다. - 웹 사이트에 지도를 띄우고 예측 모델을 연동하여, 위치에 대한 범죄 발생률을 알려준다. <p><활용 기술></p> <ul style="list-style-type: none"> - 정보를 수집하고 분석하기 위한 빅데이터 기술을 활용한다. - 범죄 발생률 정보를 시각화하기 위해 HTML, CSS, JavaScript, Flask와 같은 웹 개발 기술을 활용한다. - 웹 사이트에 지도를 띄우고 위치별 범죄 발생률을 시각화하기 위해 지도 API를 활용한다. - 범죄 발생률 예측을 위해 Scikit-Learn과 Naive Bayes를 이용해 데이터 간의 상관관계를 파악하고 머신러닝 프레임워크(SVM, Decision Tree, Random Forest 등)를 활용하여 최적의 예측 모델을 한다. <p><추진 전략></p> <p>step 1. 탐색</p> <ul style="list-style-type: none"> - 정부에서 제공하는 범죄 정보 데이터, 날씨 정보, 지역 정보, 인구 정보를 수집하고 데이터의 구조의 품질을 확인한다. - 데이터 소스 및 API접근 방법을 탐구하고 데이터 수집 파이프라인을 설정한다. <p>step 2. 인사이트 발견 : 탐색과 시각화</p> <ul style="list-style-type: none"> - 데이터 시각화를 통해 지역별 범죄 발생 패턴, 날씨와 범죄 발생 간의 관계, 인구 특성과 범죄 발생률 간의 상관관계 등을 파악한다. - 인사이트 도출을 위해 통계적 분석 및 데이터 시각화 도구인 seaborn, matplotlib, folium 등을 활용한다. <p>step 3. 탐색적 분석</p> <ul style="list-style-type: none"> - 데이터 스케일링, 결측치 처리, 이상치 탐지 및 처리 등을 포함한 데이터 전처리 단계를 수행한다. <p>step 4. 실증 분석 : 문제에 대한 해답 찾기</p> <ul style="list-style-type: none"> - 선택된 모델을 웹 애플리케이션과 연동하여 사용자에게 지도 상의 범죄 발생률 예측을 제공한다. - 모델의 성능을 모니터링하고 필요한 경우 주기적으로 업데이트한다. 	

과제 수행 계획서 (문제 분석지)

<설계 추진 일정>

- ~ 10월15일 : 데이터 전처리 완료 및 데이터셋 구축
- ~ 10월29일 : 기본적인 웹 페이지 구축, 예측 모델 구축(프로젝트 50% 진행)
- ~ 11월12일 : 피드백 및 프로젝트 수정(프로젝트 80% 진행)
- ~ 11월26일 : 완성작 구축

<기대성과>

이러한 전략을 통해 정부에서 제공하는 데이터와 머신러닝 기술을 활용하여 최적의 서울시 범죄 발생률 예측 모델을 개발하고 이를 시각적으로 표시하는 웹 사이트를 팀원들과 함께 구축하여 사용자가 지도상에서 서울시 자치구별 범죄 발생률을 확인할 수 있도록 한다.

<인용문헌>

- 논문 : "머신러닝 기반 범죄발생 위험지역 예측", 저자 : 허선영, 김주영, 문태현

알고 있는 사실

(Facts)

- 서울은 전국 총 범죄 건수 중 3위를 기록한다.
- 최근에 시작된 칼부림 범죄로 인해, 폭력 및 모방범죄가 일어남으로써 시민들의 피해와 희생이 발생했다.
- 최근 전국적으로 묻지마 범죄, 또는 강력범죄에 대한 처벌 이슈가 높아짐으로써 사람들의 관심도가 높아졌다.

더 알아야 할 사항

(Learning Issues)

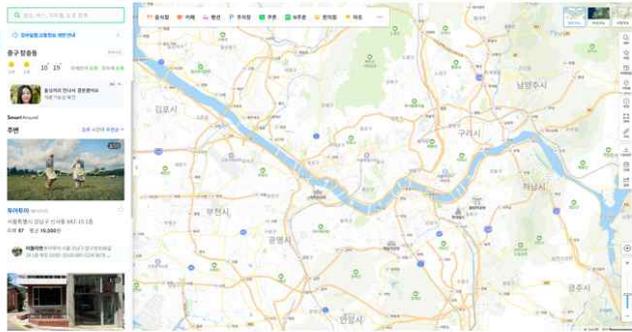
- 수집한 데이터들과 범죄률의 상관관계에 대해 알아야한다.
- 완성된 예측모델을 웹사이트로 연동시키기 위한 방법에 대해 알아보아야한다.
- 더욱 정확한 예측모델 제작을 위해 다양한 머신러닝 기술을 공부해야한다.
- 지도 api를 이용한 웹페이지 제작방법에 대해 공부해야한다.

앞으로의 계획

(Action Plan)

- 관련 데이터 수집 및 정제(데이터셋 구축)
- 데이터 분석 및 인사이트 도출
- 서울시 범죄 발생률 예측 모델을 구축한다.
- 분석한 결과를 토대로 웹 사이트 제작하여 서비스화한다.

프로젝트 실행 OUTPUT 예시



처음 화면



자치구 검색했을 때 화면