

(4주차) 데이터 전처리 및 웹페이지 개발

✓ 2023.05.15 ~ 05.21 🔗

✓ 5.16. (이상현)

- 프로젝트 빠르게 끝내고 추가로 발굴할 아이디어 있는지 찾아보기

✓ 5.16. (한상곤)

- 지도 학습에 관련 된 건 1try 반페이지 정도...? => 여기서 얻은 걸 적어라!
- 어떤걸 만들었다 => 미비점 보완한거 => 최종 완성한거
- 모델 결정
 - 지금까지 한것 : 데이터 가져왔고, 전처리 했고, 목표 설정했고!
 - 분석단계 필요 : 비지도 학습(분석자의 가치판단)
 - 가속도 등등을 이용해서 횡수를 카운트 하고 분류를 한다
 - 이상치를 어떻게 찾지?! => 핸즈온 비지도 학습 "이상치 탐지"
 - t-SNE, PCA(커널), SVD를 써서 변동량이 제일 많은게 뭐예요?
 - PCA → 클러스터링 (k-mean), 이상치 탐색은 어떤값이 튀는지를 판단하는거
 - 시각화!!!!!!!!!!!!!!
 - 예측을 해봅시다!
 - 위험을 알려주는 서비스를 하는거 어때?
 - 보통 이시간 여기구간에서 과속 하거든?? 그니까 조심해~!
 - 백엔드 뭐하지?!
 - 일단 로그인기능 만들고
 - 파일 업로드 하면 분석해주는.. 뭐 ... 그런거.... 어때...? ...
 - 파일 업로드 양식은 내가 정해서 이대로만 넣어라!!!! ...

✓ 5.17. (홍봉희)

- 요구사항 상세히 기술 할 것
 - 서비스 요구사항 : 어떤 서비스를 하겠다?
 - 하루치 노선에 대한 위험구간 분석한 거
 - 특정 차량별로 나누어서 위험구간 분석한 거
 - 해당 노선에서? 차량에서? 어느 시간대에 위험이 많이 발생했는지?
 - 지도에서 노선 클릭하면 언제 발생했는지 같은...?
 - 데이터 요구사항 : 어떤 데이터를 분석해서 표출하겠나?

✓ 5.19. (홍봉희)

- 문서는 데이터 쪽만 보완해봐라~~(ㄱ)
- 통계적인 방법으로 급가속 등 판단하지 말고 비지도학습, 머신러닝으로 위험운전 판단?
- x축 시간, y축을 가속도, 속도 등으로 시각화 해보기

✓ 5.19. (김경민)

- 우리가 라벨링 한 결과를 지도에 몇번 발생했는지 뿌렸잖아?
- => 그 구간에서 발생 할지 안할지 예측모델 만들수 있을것 같다

- 운전기사가 그 구간을 지날때 위험운전 발생 확률(혹은 발생 할지 아닐지) 정도를 띄워주는 서비스
- 공모전도 해보자 ㅋㅋㅋ [국가교통 데이터 오픈마켓](#)
- 공모전 방향은 버스 정류장에 도착예측시간 보여주는것 처럼 해당차량이 위험운전을 많이 하는 차량인지 알려주는 서비스로
- 버스 기사들에 대한 교육효과가 있을 것으로 기대됨!

리액트에 지도 띄우기 (5/17)

- 차량 한대의 하루치 데이터 중 급가속, 급출발, 급정지 표시해서 json 데이터로 저장 후
- 리액트에서 json 파일 읽어와서 지도에 마크까지 표시하는거 해봄
- 어떤 데이터를 어떻게 띄울지는 고민을 계속 해보는 중...

급가속, 급감속, 급출발, 급정지 판별 (5/17)

- 판단기준에 따라 초당 데이터를 이용하여 판별하고 각 구분별 컬럼 추가
- 급가속 등 위험 운전시 1, 그 외 0

	급가속	급출발	급감속	급정지
count	5.948212e+06	5.948212e+06	5.948212e+06	5.948212e+06
mean	5.099684e-03	1.758848e-03	1.644864e-03	1.475401e-03
std	7.122975e-02	4.190173e-02	4.052356e-02	3.838261e-02
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
50%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
75%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
max	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00

리액트에 지도 띄우기 (5/18)

- 마크 클러스터링
- 클러스터링 숫자 찍으면 관련된 데이터 들고올수 있는것 까지 확인
- 나중에 스프링부트에서 데이터 받을때 시간정보를 분할 한 걸 받으면
- 시간별 카운트도 할 수 있겠지?!
 - 데이터 받을 양식 생각하기
 - 날짜, 초단위로 된 시간을 2시간 정도 구간 단위로 잘라서 받자~~!!
 - 그리고 필요한게 gps_x, gps_y, 급가속, 급출발, 급감속, 급정지 4종류 칼럼!
 - 인덱스로 접근할 만한거
 - 차트는 [JavaScript Stacked Column Chart – ApexCharts.js](#) 이거 쓸 예정!

차원축소, PCA (5/18)

- 의미없음.

탐색적 가시화 (5/18)

- 가속도, 차량속도, 급가속, 급감속, 급출발, 급정지 데이터로 시각화
 - 가속도가 같은 값을 가져도 차량속도에 따라서 급가속과 급출발, 급감속과 급정지 로 나눌 수 있다.

