

데이터 분석 최종결과보고서

I. 참가자 정보

제 목	실시간 날씨 데이터를 이용한 충남,대전,세종 각 지역별 1시간 이내 교통사고확률 예측모델 개발 및 날씨 데이터와 교통사고 데이터간의 상관관계 분석	
팀 명	국민대민썬	
성 명	강민수	
연락처	휴대폰	010-9840-3002
	E-mail	daro98@naver.com

※ 제출 파일 리스트

- 1. 기상청 날씨 데이터와 교통사고 데이터 간의 상관관계 분석_국민대민썬팀.ipynb
- 2. 실시간 날씨 데이터를 기반으로 충남·대전·세종 각 지역별 1시간 이내 교통사고확률 예측모델_국민대민썬팀.ipynb
- 3.traffic_accident_prediction.pkl
- 4. 데이터 분석 최종결과 보고서_국민대민썬팀.hwp

※ 제출 파일 리스트 설명

- 1. 기상청 날씨데이터를 활용한 대전·세종·충남 지역 교통사고 분석
- 2. 기상청 날씨데이터를 활용한 대전·세종·충남 지역 교통사고 예측 모델 개발
- 3. 개발한 예측 모델
- 4. 데이터 분석 최종결과 보고서

※대부분 .ipynb 에 설명 및 개발이 되어 있습니다.
.ipynb 파일들은 코랩 기준으로 작성되었습니다.
꼭 확인 부탁드립니다.

II. 개요

○ (분석/시각화) 목적

날씨가 교통사고에 미치는 영향 : 이 중 날씨와 기상 요인은 일차적으로는 도로 환경에 큰 영향을 주며, 이차적으로는 운전자의 감정 상태나 운전 환경에도 영향을 미친다. 따라서 기상 요인과 교통 사고 발생간의 연관성을 분석하고, 그 피해를 정량화하여 정책적으로 반영할 필요성이 매우 높다. 라는 기상청 정보에 착안하여 기상 요인과 교통사고 간의 상관관계를 분석하고 연관성이 있다고 판단되면 기상청 날씨 데이터를 활용하여 미래 데이터를 예측하는 모델을 개발하기 위해 분석하였다.

○ 배경 및 필요성

교통사고 위치를 분석하여 자주 발생하는 지역등을 분석하거나 CCTV 데이터와 결합하여 추가 설치 지역 건의 같은 분석들은 이미 TAAS 교통사고분석시스템에 의해서 대부분이 분석가능하다. 따라서 실용적인 분석이나 예측모델 개발이 필요하다고 판단하였다.

기상청 데이터를 활용하여 충남,대전,세종 각 지역별 1시간 이내 교통사고 발생 여부와 발생 확률을 예측해주는 모델을 개발하기 위해 우선 기상청 날씨 데이터와 제공 받은 데이터간의 연관성이 있음을 확인할 필요성이 있었다.

○ 분석/시각화 결과 내용 요약

- 계절별 분석 : 교통사고는 여름/겨울철이 봄/가을철보다 증가하는 경향 보입니다.
- 기상청 날씨 데이터별 분석 : 현지기압(hPa), 해면기압(hPa), 시정(10m), 지면온도(°C) 네 가지 지표가 가장 뚜렷하게 교통사고와 연관이 있는걸로 확인 할 수 있습니다. 기압과 시정(ex: 날씨가 맑음 or 흐림) 같은 경우는 하늘 상태와 연관되어 있는데 운전 할 때 기압과 시정이 안 좋으면 시야확보에 어려움이 있어 교통사고 확률이 높아지는 것을 알 수 있다. 또한 지면온도 같은 경우는 겨울철에 낮아지면 도로가 얼어붙어 교통사고 확률이 높아지는 것을 알 수 있다.
- 기상청 데이터를 접목시켜서 교통사고 예측을 해보았는데 기상요인과 실제 교통사고 데이터가 밀접한 상관관계를 가지고 있음을 분석 할 수 있었음
- 위와 같은 분석결과를 토대로 실시간 일기예보를 이용한 1시간 이내 교통사고 확률 예측 모델 개발이 가능할 것이라고 판단함

III. 분석/시각화 결과 상세내용

○ 분석/시각화 결과 상세내용

	지점	지점명	일시	예측사고확률
0	129	서산	2023-01-19 00:00:00	33.97%
1	129	서산	2023-01-19 01:00:00	33.86%
2	129	서산	2023-01-19 02:00:00	30.83%
3	129	서산	2023-01-19 03:00:00	34.43%
4	129	서산	2023-01-19 04:00:00	31.13%
...
5067	239	세종	2023-02-14 05:00:00	41.27%
5068	239	세종	2023-02-14 06:00:00	42.01%
5069	239	세종	2023-02-14 07:00:00	33.28%
5070	239	세종	2023-02-14 08:00:00	62.54%
5071	239	세종	2023-02-14 09:00:00	75.65%

5072 rows × 4 columns

다음 사진은 개발한 실시간 날씨데이터로 미래데이터를 예측한 결과이다.
제공받은 데이터와 기상청 날씨 데이터를 학습시켜서 제공받은 데이터 이후의 미래 데이터를 개발한 모델로 예측한 결과이다.
모델 성능과 같은 경우는
F1 Score : 0.8577479940831054
Accuracy : 0.7910865343471248
- 1 시간 이내 교통사고발생을 5 개를 예측하면 4 개를 성공적으로 맞추는 준수한 성능을 가진 모델 개발하였고 사고확률까지 계산 가능하게 만들었다.
예측확률 뿐만 아니라 사고발생 여부를 확인할 수 있다.
개발한 모델은 **pickle** 파일로 저장을 따로 해놓았고 불러와서 사용만 하면되게 개발하였다.

자세한 분석내용은 함께 첨부한 **.ipynb** 파일에 기재하였음

○ 결과 해석 및 시사점

- 기상청 날씨 데이터별 분석 : 현지기압(hPa), 해면기압(hPa), 시정(10m), 지면온도(°C) 네 가지 지표가 가장 뚜렷하게 교통사고와 연관이 있는걸로 확인 할 수 있습니다. 기압과 시정(ex: 날씨가 맑음 or 흐림) 같은 경우는 하늘 상태와 연관되어 있는데 운전 할 때 기압과 시정이 안 좋으면 시야확보에 어려움이 있어 교통사고 확률이 높아지는 것을 알 수 있다. 또한 지면온도 같은 경우는 겨울철에 낮아지면 도로가 얼어붙어 교통사고 확률이 높아지는 것을 알 수 있다.
개발한 모델로 실시간 교통사고 방지에 도움이 될 것이라 예측된다.

○ 기대효과

개발한 모델로 각 지역별로 1시간 이내 사고발생 여부와 확률이 실시간으로 뜨기 때문에 위험지역 우선으로 순찰하여 교통사고 방지 및 빠른 대처가 가능해지고 효율적인 순찰루트 및 인력을 배분할 수 있다.

개발한 모델이 사고발생을 예측하면 10 번중 8 번이 실제 사고가 발생하기 때문에 예상 사고 발생지역에 미리 라디오 방송이나, 전광판으로 사고발생 주의를 띄우거나 해서 미리 운전자들에게도 경고가 가능하므로 사고발생을 현저히 줄일 수 있다고 예측된다.

IV. 기타

○ 건의 사항

실시간 교통 사고 예측 모델은 실시간으로 기상청 **Open Api** 를 이용하여 1시간 이내 각 지역별 사고 예측 확률 및 사고 발생 여부를 추가적으로 접목시킬 수 있다.

실시간 교통 사고 예측 모델은 시계열 데이터 예측 모델이기 때문에 시간이 지날 수록 더 많은 데이터를 학습이 가능하므로 성능이 계속 올라 갈 수 있다.

실시간 교통 사고 예측 모델은 시간 단위 데이터를 활용하였지만 기상청 일단위 데이터를 사용한다면 같은 매커니즘으로 일단위 사고 예측 확률 및 사고예측 여부 모델 개발이 가능하다.

○ 활용 데이터 및 참고 문헌 출처 (필수)

[기상청 외부데이터 출처] : <https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>

[이미지출처] : <https://www.kma.go.kr/daejeon/html/info/business01.jsp>