

SON GIRAK PORTFOLIO.

2023.03

반갑습니다!

한 걸음 한 걸음씩
성장하는,
손가락입니다.

WORK PORTFOLIO.



한 걸음씩 성장하는
성장형 개발자, 손기락입니다.



CALL. +82 10. 6669. 2361
MAIL. rlfkr1234@naver.com
GIT. <https://github.com/sgr1118>
NOTION. <https://www.notion.so/AI-NLP-59314840165748878d1550f481c7e1b1>

학력

2012. 03 ~ 2015. 02 세명고등학교 졸업
2015. 03 ~ 2022. 02 안동대학교 경제학 졸업

경력

2019. 08 ~ 2019. 11 (사)한국경제개발연구원
2020. 01 ~ 2020. 02 안동시청 도시디자인과
2020. 10 ~ 2020. 12 경북도청 빅데이터과
2021. 03 ~ 2022. 06 소상공인 시장진흥공단
2021. 11 ~ 2022. 06 에이스리서치 연구원

자격증

2019. 02 한국사능력검정시험 고급
2019. 10 빅데이터 준분석가
2020. 06 컴퓨터활용능력 1급
2020. 11 사회조사분석사 2급

기타 경험

2020. 01 ~ 2020. 06 교내 취업 서포터즈
2020. 08 ~ 2020. 09 빅데이터 인턴십 교육 수료
2022. 06 ~ 2022. 12 AIFEL 부트캠프 수료

사용
프로그램



활용 능력



프로젝트명

문자 텍스트 데이터를 사용한 요약문 생성 AI 개발 – 4 page

프로젝트 개요

AI허브 문자 데이터를 사용한 대화 상황요약 AI 서비스입니다.
문자 메시지에서 각 화자의 대화의 상황을 요약하여 빠르게 결론을 낼 수 있습니다.
각종 문자 텍스트를 요약할 수 있어 정보의 습득을 효율적으로 할 수 있습니다.

DATE : 2022.11 ~ 2022.12

프로젝트명

영화관 입장권 데이터 시각화 – 6 page

프로젝트 개요

영화 배급사의 입장에서 영화관 관객 수 데이터 셋을 활용하여
데이터 속에서 유의미한 관계를 도출하고자 하였습니다..
13~22년 영화관 입장권 통합 전산망 데이터를 사용했습니다.
DATE : 2022.07 ~ 2022.07

프로젝트명

예천군 불법주정차 완화 데이터 분석 – 8 page

프로젝트 개요

예천군 불법주정차 완화를 위한 주차장 최적입지를 선정하는 프로젝트입니다.
불법 주정차가 많이 발생하는 지역을 도출하고 건축물 용도에 따라 가점을 부여했습니다.
이후 이 둘의 상관관계를 보고 주차장이 신설되어야 할 최적입지를 도출했습니다.

DATE : 2020.10 ~ 2020.12

문자 텍스트 데이터를 사용한 요약문 생성 AI 개발

프로젝트 개요

AI허브 문자 데이터를 사용한 대화 상황요약 AI 서비스입니다.
문자 메시지에서 각 화자의 대화의 상황을 요약하여 빠르게 결론을 낼 수 있습니다.
각종 문자 텍스트를 요약할 수 있어 정보의 습득을 효율적으로 할 수 있습니다.

동기 및 문제

기존 음성 기록을 해주는 어플의 성능과 한계점을 개선 하고자 함.
회의록, 문자 데이터 같은 텍스트 데이터를 사용하여 요약문을 생성하여 빠른 상황 이해를 돕고자 함.

솔루션

- 분석 과정에서 발생한 문제점 및 해소 방법입니다.
- 1. 개인 정보 데이터 스페셜 토큰화
 - 문자 데이터 속에서 개인 정보가 라벨링이되어있기 때문에 모델에 스페셜 토큰으로 지정하여 학습하게 만듦.
 - 2. 메타 데이터 활용
 - 각 문자의 대화 유형을 모델이 인식할 수 있게 #Category#를 각 문자데이터 앞에 삽입
 - 3. 생성 전략 적용
 - Beam Search를 적용하여 성능 개선

과정 및 프로세스

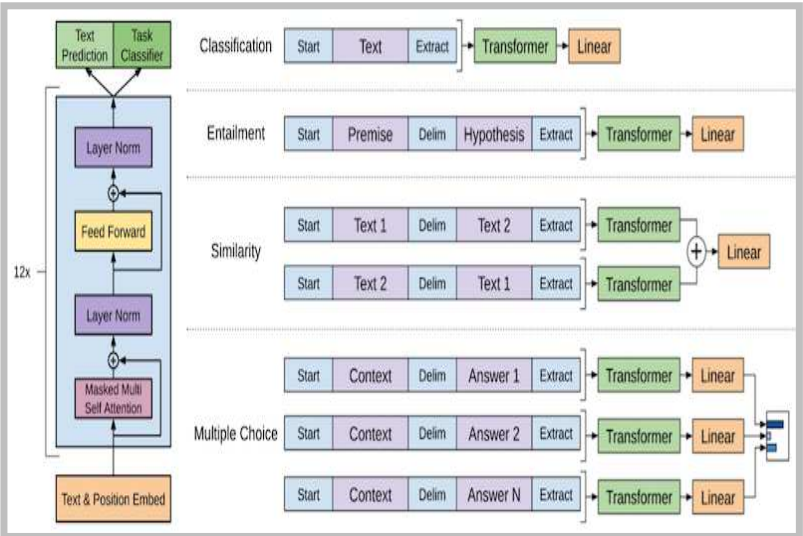
- 조사 과정은 다음과 같은 순서로 진행 되었고 전 과정을 참여하였다.
- 1. AI 허브 문자 메시지 데이터 수집
 - 2. 학습/검증용 데이터 셋 구성
 - 3. 데이터 EDA 수행 (결측, 중복, 전처리 기준 파악)
 - 4. 데이터 전처리 수행 및 모델 성능 지표 결정 (결측, 중복, 개인정보 데이터, 반복문자, 정규표현식 사용)
 - 5. 모델링 코드 구축 (KoGPT-2)
 - 6. 원본 데이터를 사용하여 모델 성능 평가 후 전처리를 수행한 성능 분석
 - 7. 모델 성능 피드백을 받은 후 개선 작업 수행
 - 8. 최종 결과 도출

결과 (데이터)

원본데이터를 사용한 성능 결과 : 재현율 0.1404, 정밀도 0.1559 F-1 Score 0.1353,
최종 수행 후 성능 결과 : 재현율 0.1873, 정밀도 0.2543 F-1 Score 0.2078

기간

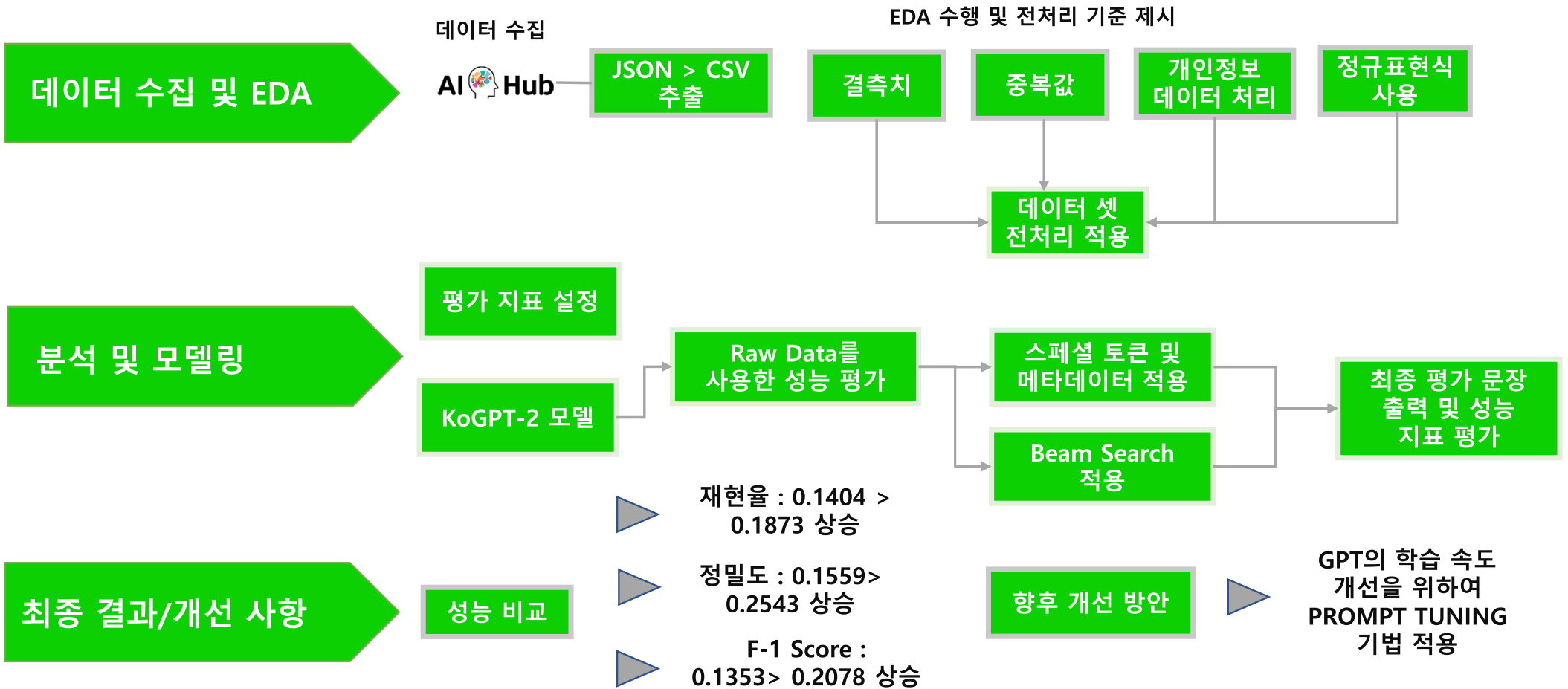
2022.11 ~ 2022.12



요약문	텍스트
실제 요약문	2019년 연말 결산을 해본 결과 영화를 116편 보았다.
모델 생성 요약문	2019년 연말결산 해봤는데 영화 116편을 봤다고 한다.
실제 요약문	월드 스쿨 신곡 127이 좋고 뮤직비디오는 엔시티 유의 체리 밍이다.
모델 생성 요약문	월드스쿨의 신곡에 대해 이야기하고 있다.
실제 요약문	시험기간이라 할게 너무 많아서 밤을 새야 할 거 같다.
모델 생성 요약문	할 일이 너무 많아서 밤을 새울 것 같다.

문자 텍스트 데이터를 사용한 요약문 생성 AI 개발

AI허브 문자 데이터를 사용한 대화 상황요약 AI 서비스입니다.
문자 메시지에서 각 화자의 대화의 상황을 요약하여 빠르게 결론을 낼 수 있습니다.
각종 문자 텍스트를 요약할 수 있어 정보의 습득을 효율적으로 할 수 있습니다.



영화관 입장권 데이터 시각화

프로젝트 개요

영화 배급사의 입장에서 영화관 관객 수 데이터 셋을 활용하여 데이터 속에서 유의미한 관계를 도출하고자 하였습니다..
13~22년 영화관 입장권 통합 전산망 데이터를 사용했습니다.

동기 및 문제

분기별 국내에 배급되는 영화에 트렌드가 있었는지 확인하고 배급사 입장에서 각 분기 별 어떤 영화의 장르들을 개봉하는지 파악함.

솔루션

분석 과정에서 발생한 문제점 및 해소 방법입니다.

1. 특정 장르 배제 기준 마련 필요

- 성인물, 독립 영화 등 실제 분석에서 사용하지 않기 위하여 특정 장르들의 관객수 분포를 추출하여 400명 미만의 영화는 사용하지 않기로 결정함.

과정 및 프로세스

조사 과정은 다음과 같은 순서로 진행 되었고 전 과정을 참여하였다.

1. 데이터 셋 수집
2. 데이터 EDA 수행 (결측, 중복, 전처리 기준 파악)
3. 데이터 전처리 (결측치 제거, 일부 장르 삭제, 전국 관객수 400명 미만 삭제)
4. 데이터 시각화 진행 (라인 차트, 워드 클라우드, 코사인 유사도)
5. 시각화 결과를 바탕으로 트렌드 분석

결과 (데이터)

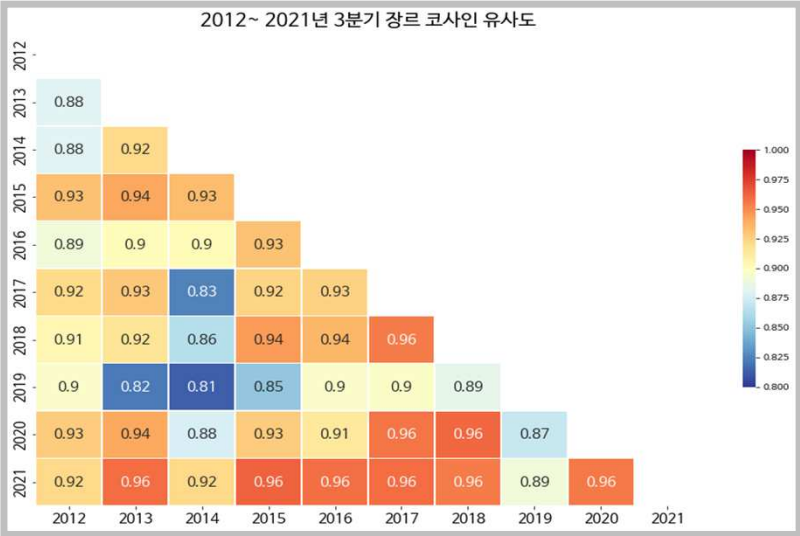
전국 상영 영화 수가 가장 적을 때 전국 관객수가 가장 적은 현상을 발견

매년 워드 클라우드는 일정한 키워드가 상위권이었고 분기별 코사인 유사도 비교에서는 대부분 유사한 관계를 지니고 있음.

배급의 영향을 주는 다른 요소를 찾아 고려하여 분석에 활용이 필요함.

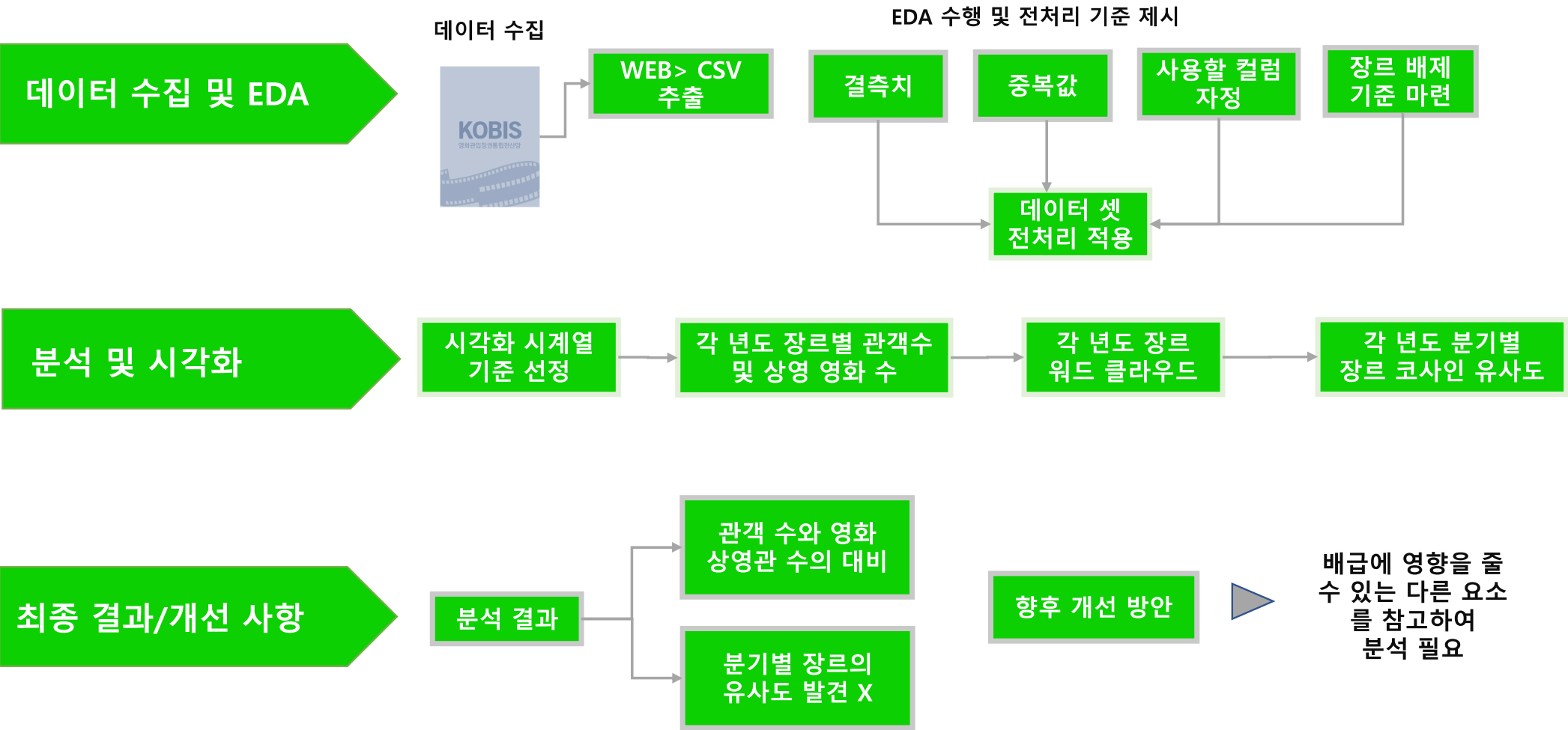
기간

2022.07 ~ 2022.07



영화관 입장권 데이터 시각화

영화 배급사의 입장에서 영화관 관객 수 데이터 셋을 활용하여
데이터 속에서 유의미한 관계를 도출하고자 하였습니다..
13~22년 영화관 입장권 통합 전산망 데이터를 사용했습니다.



예천군 불법주정차 완화 데이터 분석

프로젝트 개요

예천군 불법주정차 완화를 위한 주차장 최적입지를 선정하는 프로젝트입니다.
불법 주정차가 많이 발생하는 지역을 도출하고 건축물 용도에 따라 가점을 부여했습니다.
이후 이 둘의 상관관계를 보고 주차장이 신설되어야 할 최적입지를 도출했습니다.

동기 및 문제

19년, 기준 예천군 불법주정차 4,130건 중 2,827건 이 예천읍에서 적발되었고, 고정형 CCTV의 적발 건수가 3,625건으로 전체 건수에서 약 87%를 차지하고 있음

집중된 불법주정차 원인을 파악하고 군민의 불법주정차 문제를 해소하기 위함.

솔루션

분석 과정에서 발생한 문제점 및 해소 의견입니다.
1. 이동형 CCTV 단속 데이터의 부족 및 데이터 표준화
- 이동형 단속 데이터는 더욱 정확한 좌표 데이터를 확인할 수 있기에 고정형 CCTV가 잡아내지 못하는 음영지역의 주민 및 보행자의 불평을 해소할 가능성
- 단속대장 데이터 형식이 각 지자체마다 달라 표준화가 요구됨.

과정 및 프로세스

조사 과정은 다음과 같은 순서로 진행 되었고 전 과정을 참여하였다.

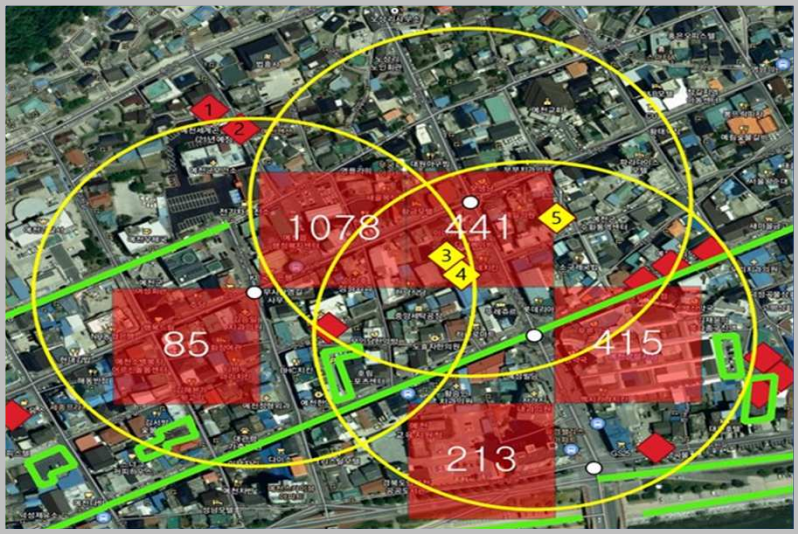
1. 예천군 불법주정차 데이터 요청
2. 불법주정차 현황 데이터 분석 (상위 적발 구역, 월, 일, 시간별 적발 건수, 전체 주차면수 , 건축물 사용 용도 확인)
3. 불법주정차 해소를 위한 가설 설정
- 건축물 사용용도와 적발 건수를 합하여 상위 지역을 주차관리 핵심지역이라고 설정
4. 데이터 전처리 수행 (결측치 제거, 불필요한 컬럼 삭제, 주건축물 가점 부여)
5. 주건축물 가점과 적발건수 간 상관관계 분석
6. 불법주정차 해소를 위한 주차장 최적입지 선정

결과 (데이터)

예천군 식당 61면, 멕시칸 치킨 66면, 노하리 마을문화센터 10면 총 137면을 추가적으로 신설할 수 있다는 결론을 제시함.

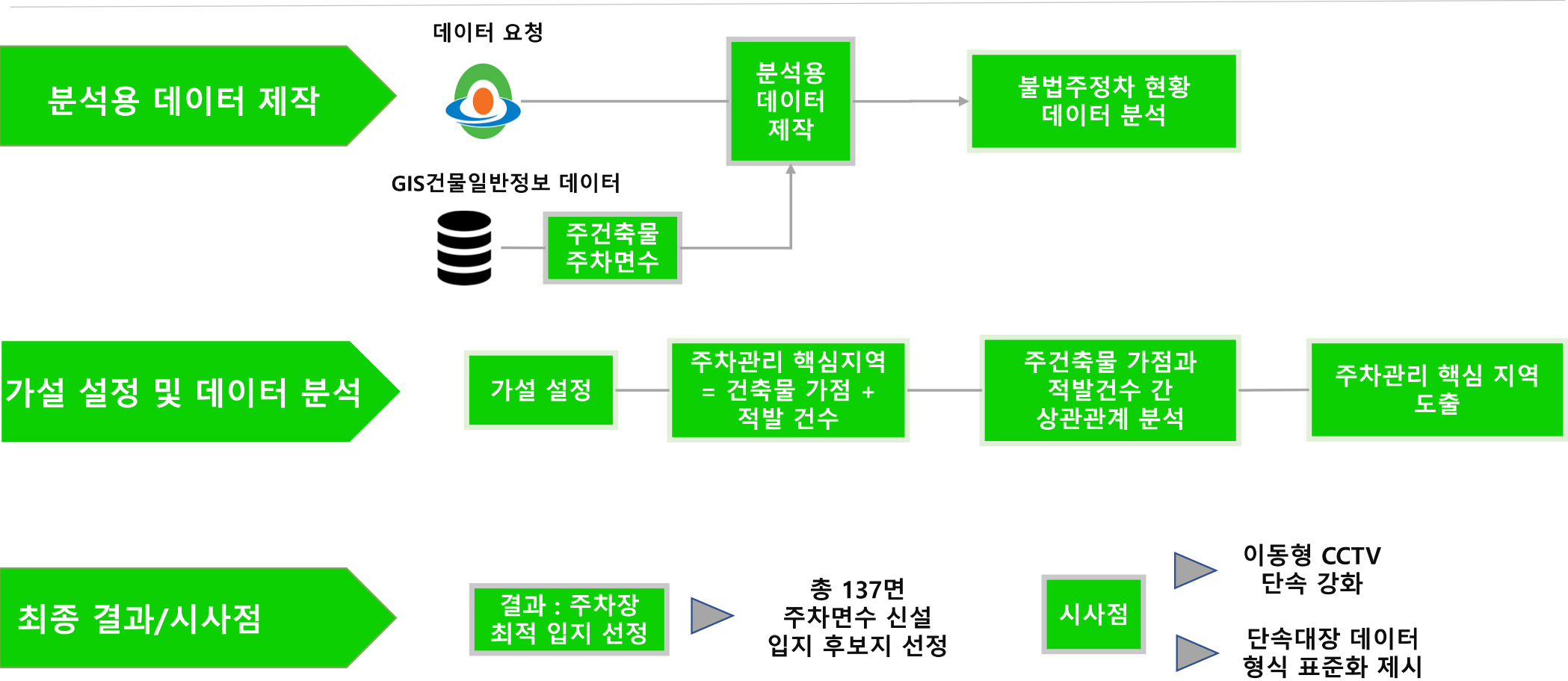
기간

2020.10 ~ 2020.12

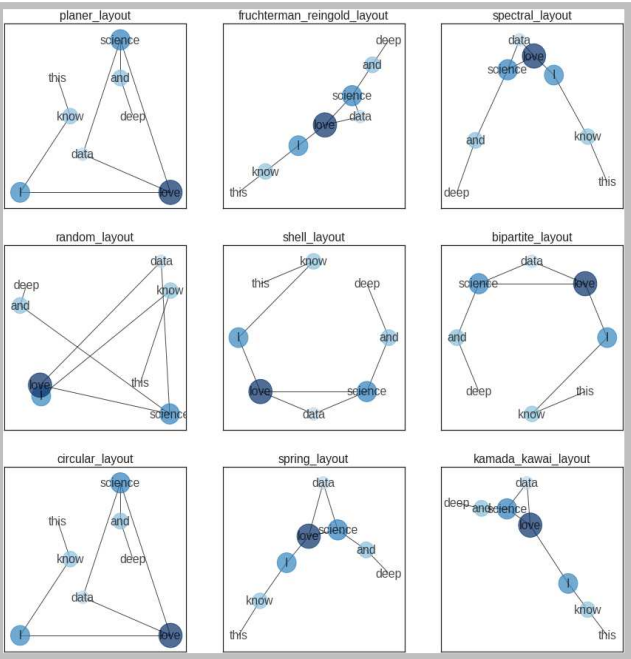


예천군 불법주정차 완화 데이터 분석

예천군 불법주정차 완화를 위한 주차장 최적입지를 선정하는 프로젝트입니다.
불법 주정차가 많이 발생하는 지역을 도출하고 건축물 용도에 따라 가점을 부여했습니다.
이후 이 둘의 상관관계를 보고 주차장이 신설되어야 할 최적입지를 도출했습니다.



한걸음 씩 더 나아가기 위한
저의 모습입니다.



다양한 NLP 시각화 능력

NLP에서 좋은 모델을 선택하는 것도 중요하지만 전처리를 어떻게 할 것인지 결정하는 것도 중요하다고 생각하여 다양한 전처리 방법을 학습하고 실습함.

다음에 진행할 프로젝트에서는 이런 방법을 사용하여 모델의 성능 개선을 이루도록 할 것임.



개발자로 나아가기 위한 행사 참여

다른 개발자들의 관점을 더욱 많이 알고 싶어서 오프라인 세미나를 참여함. 이런 행사를 방문함으로써 저의 부족함 부분을 이해하고 다양한 관점을 알아가는 계기였음.



sgr118 Delete Chapter9_심층신경망1.ipynb	08f7514 5 days ago	42 commits
Chapter15_PyFile	Add files via upload	2 weeks ago
Chapter19_PyFile	Add files via upload	2 weeks ago
Chapter03_파이썬치_튜토리얼.ipynp...	complete	5 days ago
Chapter04_선형 회귀.ipynb	complete	5 days ago
Chapter05_손실 함수.ipynb	complete	5 days ago
Chapter06_경사하강법.ipynb	complete	5 days ago
Chapter07_선형 회귀.ipynb	complete	5 days ago
Chapter08_로지스틱 회귀.ipynb	complete	5 days ago
Chapter10_심층신경망1.ipynb	complete	5 days ago
Chapter10_확률적 경사하강법.ipynb	complete	5 days ago
Chapter11_최적화.ipynb	Colaboratory를 통해 생성됨	3 weeks ago
Chapter12_오버피팅을 방지하는 방...	Colaboratory를 통해 생성됨	3 weeks ago
Chapter13_심층신경망2.ipynb	Colaboratory를 통해 생성됨	3 weeks ago
Chapter14_장규화.ipynb	Colaboratory를 통해 생성됨	3 weeks ago
Chapter15_실무 환경에서의 프로젝...	Colaboratory를 통해 생성됨	3 weeks ago
Chapter16_표현학습.ipynb	Colaboratory를 통해 생성됨	2 weeks ago
Chapter17_확률론적 관점.ipynb	Colaboratory를 통해 생성됨	2 weeks ago
Chapter19_RNN.ipynb	Colaboratory를 통해 생성됨	2 weeks ago
PyTorch_Tutorial.ipynb	Colaboratory를 통해 생성됨	5 months ago
README.md	Update README.md	last week

프레임워크 학습

부트캠프에서 배우지 못했던 Pytorch 프레임워크 및 py를 사용한 실무 환경을 이해하기 위하여 Pytoch 입문서를 활용하여 학습함. 이외에도 한국어 NLP 처리 능력을 위해 꾸준히 GitHub에 학습한 내용을 업로드 중임.



직면하는 문제를 해소하고
보다 더 나아가기 위한 손가락입니다.



2015 -
2023

WORK PORTFOLIO.

● THANK YOU +

읽어주셔서 감사합니다.
한 걸음씩 성장하여 미래에
NLP 개발자가 되고 싶은 지원자
손기락입니다 !

+ CONTACT

CALL. +82 10. 6669. 2361
MAIL. rlfkr1234@naver.com
GIT. <https://github.com/sgr1118>
NOTION. <https://www.notion.so/AI-NLP-59314840165748878d1550f481c7e1b1>