|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 주제 | 분석 방법 | 데이터 | 보완할 점 | 비고 |
| 1 | 농산물 예측 | - 농산물 가격에 영향을 주는 다양한 변수들(자연재해 등) 모아서 예측하는 모델 만들어보기!  - 다양한 시계열 기법들 고려  (지난학기 시계열팀의 비트코인 예측 참고하면 좋을 듯) | -데이콘  [2021 농산물 가격예측 AI 경진대회 - DACON](https://dacon.io/competitions/official/235801/data) |  |  |
| 2 | 서울시 산책 루트 | - 사용자의 상황(데이트, 운동 등) 별로 알맞은 산책루트를 추천하는 시스템  - 코로나랑 연계해서 코로나로부터 좀 더 안전한 산책루트 추천도 해볼 만하지 않을까? | -서울시 공공데이터 (두드림길 검색)  <https://data.seoul.go.kr/dataList/datasetList.do#>  : | - 단순히 산책로에 대한 정보만 있어서 추가 데이터를 좀 더 찾아봐야 함  - 이용자에 대한 데이터가 있으면 제일 좋을듯 |  |
| 3 | 애니 추천 | - 평점, 작화 등을 고려해서 사용자가 좋아할 만한 애니 추천!  - 정보는 라프텔 사이트에서 크롤링해서 활용하면 충분할 듯 | -라프텔 크롤링(<https://laftel.net/>)  -캐글 데이터  <https://www.kaggle.com/hernan4444/anime-recommendation-database-2020?select=anime.csv> | - 캐글 데이터는 영어 기반 데이터라 제목을 번역해야함  - 이미지 크롤링을 할 수도 있어서 크롤링 작업이 좀 어려울수도 |  |
| 4 | 재난지원금 효과 | 재난지원금 지급 기한 전후로 지역상권 이 얼마나 활성화됐는지 비교  - 유동인구, 소비금액 등의 지표 확인해서 비교해보면 좋을듯 | -Kt 데이터 플랫폼  <https://bdp.kt.co.kr/invoke/SOKBP2602/> | - 지역매출 데이터 등 경제분야에서의 변화 확인할 데이터가 필요함  - 그냥 단순 비교가 될 수도 있어서 예측이나 지원금액별 효과의 차이 분석 등 내용 구체화가 필요 |  |
| 5 | 인플루언서의 영향력 측정 | -구독자 얼마까지 늘어날까?  - 광고료 얼마일까 | -블링  <https://vling.net/>  - 크롤링 | 크롤링이 페이지 단위가 아니여서 어떤 식으로 해야할 지 아직은 모르겠음  광고료 예측하기엔 블링에서 너무 잘 돼있음 |  |
| 6 | 제주도 음식물쓰레기 배출 예측 및 해결 | 데이터를 활용한 배출량 예측 모델 개발을 통해 제주 읍면동별 배출량을 예측하고, 배출 요인에 따른 배출량 감소 방안 도출 | - 제주데이터허브  <https://www.jejudatahub.net/data/list/> | 데이터 찾아봐야함  (제주 테크노파크 연락 취해보겠음) |  |
| 7 | 넷플릭스 콘텐츠 흥행 여부 예측 | 넷플릭스 컨텐츠 장르, 러닝타임, 나라, 감독, 배우 고려해서 흥행과 별점을 분석 후 모델을 만들어서 개봉할 영화들의 흥행정도를 예측 | 1- Netflix original films & IMDB scores  넷플릭스 오리지널 영화의 평점  <https://www.kaggle.com/luiscorter/netflix-original-films-imdb-scores>  변수이름: genre, premiere(Date the film premiered),  Runtime, IMDB Score, Language,  2- Netflix Movies and TV Shows  넷플릭스 전체 영화와 티비쇼 정보  <https://www.kaggle.com/shivamb/netflix-shows>  변수이름: type , title, director, cast, country, date\_added(Date it was added on Netflix)  Release\_year(Actual Release year of movie), rating, duration  3- Netflix subscribers and revenue by country  Netflix's subscriber figures & revenue growth by region - from 2018 to Q2 2020  나라별 넷플릭스 구독자 수와 수익  <https://www.kaggle.com/pariaagharabi/netflix2020>  변수이름: Area, year, revenue, subscribers,  4- Netflix Originals  Original Movies and Show created by Netflix  <https://www.kaggle.com/abhimanyudasarwar/netflix-originals>  변수이름: title, genre, premiere, seasons (시즌 몇 개고, 에피소드 몇 개인지), length(영상길이), Netflix exclusive(worldwide/ selected territories ), status(끝났는지/계속하고 있는지) |  |  |
| 8 | 넷플릭스 오리지날 주력 장르 예측 |  |  |  |
| 9 | 넷플릭스 콘텐츠 세계적인 흥행 위한 전략 | 우리나라 영화인데 해외에서 아주 흥행한 영화나 시리즈의 공통요소 분석 |  |  |
| 10 | 부담 없이 보기 좋은 괜찮은 영화 장르 분석 | IMDB 평점 평균이 높고, Runtime이 길지 않아 부담없이 보기 좋은 영화가 어떤 장르에 많이 분포해있는지 |  |  |
| 11 | OTT와 영화 관계 |  | -OTT와 극장 동시개봉이 수익감소에 미치는 영향 feat.불법복제  <https://www.kobiz.or.kr/new/kor/03_worldfilm/news/news.jsp>  -영화관입장권통합전산망 <http://www.kobis.or.kr/kobis/business/main/main.do>  <https://rpubs.com/leesohee/689993>  데이터명)  1 연도별 박스오피스  2 역대 박스오피스  변수이름) 순위, 영화명, 개봉일, 매출액, 매출액 점유율, 관객수, 스크린수  **데이터명)**  **3 (재)한국저작권보호원\_콘텐츠별 오프라인 불법복제물 유통량(08년~17년)**  데이터 설명) 한국저작권보호원에서 연도별로 발간하는 저작권 보호 연차보고서 내 불법복제물 시장규모 및 합법저작물 시장 침해규모 조사결과 제공  <https://www.data.go.kr/data/15048842/fileData.do>  **4 (재)한국저작권보호원\_불법복제물 콘텐츠 목록 정보**  **5 (재)한국저작권보호원\_불법복제물 오프라인 단속현황** |  |  |
| 12 | Ai 작곡 | 차트 순위권 안에 있는 노래들의 악보 받은 뒤  R 라이브러리 이용 박자와 음계 등 노래 관련 데이터 생성  각 변수들을 이용해 (Markov chain)모델링을 통하면 다음 멜로디 예측 가능하다고 함. 예측된 멜로디 통해 요새 트렌드에 맞는 노래 만들 수 있을 것 같음. | 참고사이트<https://jaydata.tistory.com/38?category=846387> |  |  |
| 13 | 시청률과 OST의 상관관계 | 시청률과 ost의 상관관계, -> 관련 논문 조사 결과 시청률과 ost간의 상관관계가 없음이 밝혀짐. 후보 탈락 |  |  |  |
| 14 | 배달사고 예방 | 퀵서비스와 배달대행문화의 증가로 인한 배달사고 급증. 특히 빗길 오토바이 사망 사고 증가. 사고다발지점 및 사고발생시간대 파악 이후 위험 예측 및 경고 시스템 구축. 사고다발지점은 이미 확인이 가능하므로 이 주제로 한다면 ,, 기존 데이터 + 시간대 + 기후 데이터를 고려해서 경고/예측시스템을 구축하는 방향으로 잡아야 할 듯 | 관련 뉴스기사: <http://www.newstapa.org/article/lT3VM>  (데이터) 최근 5개년 이륜차 사고 현황  <http://taas.koroad.or.kr/sta/acs/gus/selectTfcacdOccrrncSttusAnnual.do?menuId=WEB_KMP_OVT_MVT_TAC_TOA>  (데이터) 사고유형별 가해운전자 차종별 교통사고 데이터  <http://taas.koroad.or.kr/sta/acs/exs/typical.do?menuId=WEB_KMP_OVT_UAS_ASA>  이륜차 배달 플랫폼 재해예방 시스템(시스템 구축이 이미 되어있음, 데이터도 다운로드 가능)  <http://service.kosha.or.kr/web/oai/afi/acdntFreqocZoneMap.do;jsessionid=6E5F9761078DCA27F4879A4B5637B0B4>  <http://service.kosha.or.kr/web/oai/afi/acdntFreqocZoneApilntrcn.do;jsessionid=4B7622EAA766C96FFDC07DEBDA2BD488> |  |  |
| 15 | 대선 예측 | 언론의 노출도를 크롤링으로 구한 후, 선거율, 선거지구의 인구 수, 선거 여론조사 결과를 변수로 사용하여 선거 예측 시스템을 만든다.  20대 대통령 선거 결과 예측이 목적. | 역대 대선 기간 후보의 언론 노출도(기사 수),  <http://info.nec.go.kr/> - 투표 정보(선거율, 인구 수)  <http://www.realmeter.net/category/politics/page/43/> - 선거 여론 조사 | 작년에 했다고….. |  |
| 16 | 물류대란 | 물류의 무게와 종류에 따라 운임 방식 선정. 운임비 예측을 통해 가장 적합한 운임 시점 추천.  물류대란 속에서 기업들이 빠르게 물류를 운송할 수 있도록 돕는 것이 목적. | [**https://www.bigdata-sea.kr/datasearch/issue/view.do?prodId=PROD\_000046**](https://www.bigdata-sea.kr/datasearch/issue/view.do?prodId=PROD_000046)  **-해양수산데이터거래소(선박스케줄 정보)**  [**http://www.kcla.kr/web/inc/html/4-1\_3.asp**](http://www.kcla.kr/web/inc/html/4-1_3.asp)  **-한국관세물류협회 (운임지수, 해상운임비, 항공운임비)**  [**https://www.tradlinx.com/container-terminal-schedule**](https://www.tradlinx.com/container-terminal-schedule)  **-TRADW LINKS(컨테이너 터미널 정보)** | 운임 구조가 복잡해서 모델을 만들기 위해서 물류 운임구조에 대해 자세히 공부할 필요가 있을 거 같음.  물류 견적을 내주는 사이트가 이미 존재함. |  |
| 17 | 정치인 비리 예측 | ‘비리’ 검색 결과로 나오는 정치인들 이름 수집. 해당 정치인들 이력 조사 및 비리 종류 분류.  겹치는 이력을 찾아냄. 나온 데이터들로 모델 만든 후, 정치인의 이력을 입력하면 비리 확률을 예측하여 반환함. | <https://jaesan.newstapa.org/>  -고위공직자 재산 정보 공개  -네이버 등에서 ‘비리’ 검색어와 연관도가 높은 정치인 색출 및 해당 정치인에 대한 데이터 수집.  <https://www.rokps.or.kr/profile/profile_number.asp> - 역대 국회의원 경력 | 어떤 정치인이 비리를 저질렀다고 해서 같은 집단에 있는 모든 정치인들이 비리를 저지르는 것은 아님. 수집한 데이터와 예측결과 간의 연관성이 떨어질 확률이 큼. |  |