## I AM Inbeom Hwang

2 0 2 1 / P O R T F O L I O



성장하는 AI 개발자 황인범 입니다

### education

16. 12 ~ 17.01 정혜경의 c언어 프로그래밍 (렉토피아)

17.04~17.12 차세대리더육성멘토링 한국장학재단

18. 06 ~ 18.09 R과 Python을 활용한 빅데이터 전문 과정 (멀티캠퍼스)

20. 11 ~ 20.12 데이터 기반 개인화 추천기획 실무 체험 (comento)

20.07~20.12 서비스 산업 데이터를 활용한 빅데이터 분석과정(멀티캠퍼스)

## 황인범 / Inbeom Hwang

1992.12.12/ 서울특별시

**Tel.** 010-7202-829

Email. rydn2004@naver.com

서울특별시 광진구 화양동 132-93

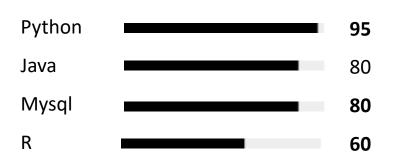
#### **GRADUATION**

2011 산곡고등학교 졸업

2012 성공회대학교 정보통신공학과 입학

2019 성공회대학교 정보통신공학과 졸업

#### **SKILL**



#### **AWARDS**

2018 멀티캠퍼스 - 최우수상(프로젝트)
 2020 KDX - 최우수상 (소비 트렌드 분석)
 2021 국립중앙도서관 - 우수상(서비스 아이디어)

#### **CERTIFICATE**

2018 MOS excel 2010 - Microsoft

**2018** MOS powerpoint 2010 - Microsoft

2019 정보처리기사 - 한국산업인력공단

2020 SQLD - 한국데이터산업진흥원

2020 ADSP - 한국데이터산업진흥원

#### **GITHUB**

https://github.com/hwanginbeom/Portfolio\_ML

#### **BLOG**

https://continuous-development.tistory.com/

## 2018.12 -2019.12

2020.01-2020.06

2021.05 -

## 천재교육

- 부서 : AI 수학 LAB (연구원)

- 직무: 웹 프로그래머,소프트웨어 개발

포맨텍 아이티 서비스

- 부서 : 소프트웨어 개발팀 (대리)

- 직무 : 웹 프로그래머

에프 앤 에프(F&F)

- 부서 : 디지털트랜스포메이션팀 (연구원)

- 직무 : 웹 프로그래머

내전석 웹 페이지 - 백엔드 구현 및 AWS를 통한 운영, 유지 / 보수

데이터 생성 및 삽입 작업 자동화

Restful API 기능 추가 개발

밀크티 중학 웹 페이지 - 사용자들의 학습 데이터를 수집하는 데이터 수집 프로젝트 구현

- Language : Python, SQL

- Framework : Django

- System : Windows, Linux, Ubuntu

- Tool : Pycharm, Heid SQL, AWS, Putty,
Jmeter, git, postman, bitbucket

쇼핑몰의 백오피스 부분인 운영 관리 부분, 시스템 관리 부분 화면 및 기능 구현

캐시소멸, 임직원 인증만료,자동탈퇴 등 회원 관련 배치 개발 마케팅 api 관련 운영 및 API 수집 부분 python으로 구현-

graph api를 통한 데이터 수집 및 대시보드 구현

- Language : Java, PostgreSql, JSP, JavaScript

- Framework : Spring framework

- System : Windows

- Tool : Eclipse, Heid SQL, DBeaver, MobaXtam,
RedMine

- Language : Java, PostgreSql, JSP, JavaScript

- Framework : Spring framework

- System : Windows

Tool : Eclipse, Heid SQL, DBeaver, MobaXtam,
 RedMine

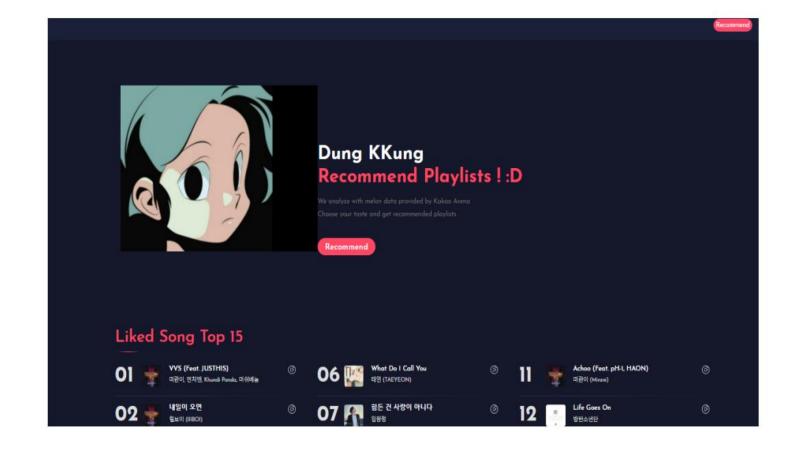
## a presentation About

- 1. 음악 플레이리스트 추천 서비스
- 2. 2020 소비 트렌드 분석
- 3. 오산시 어린이 교통사고 위험 지역 도출
- 4. 영화추천 시스템
- 5. 판매 실적 예측을 통한 편성 최적화 방안
- 6. 수요 예측을 통한 물류 최적화 시스템
- 7. 공공 도서관 입지선정

## 음악 플레이리스트 추천 서비스

## 등꿍등꿍 플레이리스트 추천 서비스

음악 플레이리스트 추천 서비스



멜론 데이터를 통해 사용자가 노래와 태그를 선택하는 것에 따라 새로운 플레이리스트를 생성 후 추천 하는 서비스를 구현하였습니다. 이것을 웹으로 구현 하였습니다.

기간/인원수 2020.11.16 - 2020.12.24 / 4명

Python(Numpy, Pandas, Gensim), Django, Mysql, Slack, GitHub, 기술스택

Pycharm, Html, Css, JavaScript

paly list data, song meta data, melon 앨범 이미지 크롤링, 데이터

youtube URL 주소 크롤링

알고리즘 Word2Vec / Word2Vec + FastText / Neighbor / KNN

시각화 **WEB** 

웹 프론트/백엔드 개발, DB 구조 설계 및 구현 구현 부분

알고리즘 구현

결과 노래와 태그 선택에 따라 맞춤 플레이 리스트 추천

컨텐츠 기반 추천 시스템 구현

알고리즘 별 성능차이 분석

활용방안으로서 매장음악 서비스와 킬러 컨텐츠로서의 역할 수행

3. 프로젝트 수행 결과

 3. 프로젝트 수행 결과
 Service Flow
 배경 준비과정 수웽결과 발전병안 1

 좋아하는 곡/태그 선택
 백엔드에서 모델링 적용

 Recommend 30%
 Recommend 50%

 의 페이지에 곡 리스트 출력

 Recommend 100%

| Recommend Playlist | Singer | Release Date | Play | Share | 1 영향원동 | 포마 | June 19, 2016 | 4 영 | 5 숙 | 3 숙 | 3 불맛 | 경미에 | Aug 19, 2019 | 6 숙 | 6 表로 (Feat 감타영) | 경기 | Aug 31, 2012 | 6 숙 | 6 ★ 로로 (Feat 감타영) | 경기 | Aug 31, 2012 | 6 숙 | 6 ★ 로로 (Feat 감타영) | 공마 | Dec 47, 2007 | 6 숙 | 6 | Aug 14, 2019 | 8 수 육 | 5 | MOM (겨울나무) | 포마 | Dec 47, 2007 | 6 수 육 | 5 | MOM (겨울나무) | 포마 | Dec 47, 2007 | 6 수 육 | 5 | MOM (겨울나무) | 포마 | Dec 47, 2007 | 6 수 육 | 5 | STEP 5.

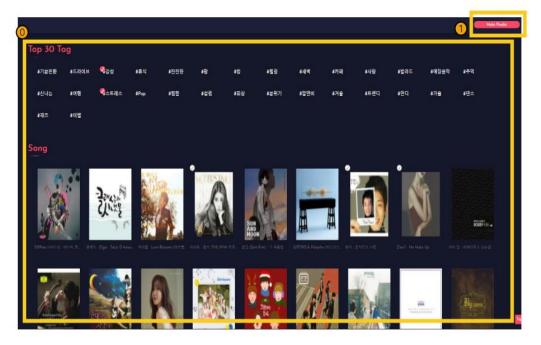
5 All Share

Share 버튼 클릭 시 카카오톡으로 공유

Service 설명

3.프로젝트수행결과 Service 설명





STEP 0.

선호하는 노래/태그 선택

STEP 1.

플레이 리스트 생성 버튼 클릭

4. 활용 및 발전 방안

활용 방안



## Solution

태그와 곡을 선택하여 나만의 맞춤 플레이리스트 생성

## **>>>**

## Improvement

1. 매장음악 서비스

태그를 통해 매장에서 트는 음악 플레이리스트를 제공

2. 킬러 컨텐츠로서의 역할

반복되는 노래 추천이 아닌 내가 선택한 것에 따라 차별 화 된 음악 추천을 받을 수 있음

소비 트렌드 분석

## 소비 트렌드 코리아 2020

코로나는 어떻게 실버세대의 소비를 바꾸었나?

소비 트렌드 코리아 2020

UDT 박세진 이경현 최우진 황인범

한국데이터 거래소에서 주최한 대회로 KDX의 다양한 데이터와 외부 데이터를 활용해 한국의 소비 트렌드를 분석해 인사이트를 도출하는 경진대회였습니다. 이 경진대회에서 좋은 성과를 거둬 최우수상을 수상하였습니다. **기간/인원수** 2020.09.23 - 2020.10.25 / 4명

기술스택 Python(Numpy, Pandas), Slack, GitHub, Pycharm, Colab

**데이터** 통계청 데이터(인구구조,경제활동인구, 인터넷 이용률, 온라인 쇼핑몰 취급액) 오프라인 소비 데이터(삼성카드,신한카드), 온라인 소비 데이터(Mcorporation)

알고리즘 Prophet 모듈을 이용한 forecast 예측 (시계열 예측)

시각화 python 에서 제공하는 plotly 라이브러리 사용

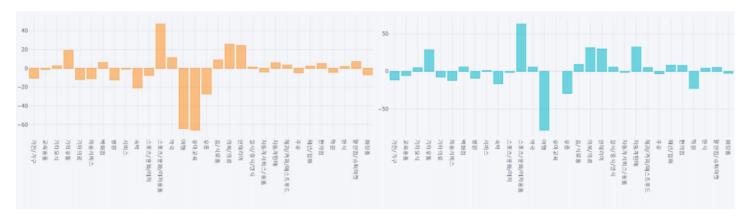
**구현 부분** 기획, 데이터 분석, 데이터 시각화

**결과** 미래의 핵심 소비층인 실버세대를 위한 온라인 시장 공략 코로나로 인한 변한 소비 트렌드 분석

03 사회적거리두기 전/후 분석

### 강도 높은 사회적 거리두기 전후 오프라인 판매량에 따른 카테고리 (신한카드)





상승 : 스포츠/문화/레저용품, 의복/의류, 인테리어, 기타유통

하락: 여행, 숙박, 유아교육, 유흥, 학원

현황 요약

코로나 이전: 온라인 구매건수가 오프라인의 1.5% 수준

사회적 거리두기 이후 : 아웃도어, 레저용품, 스포츠 등의 구매건수가 30% 이상 증가 코로나 이후 : 온라인 구매건수 평균 22% 상승

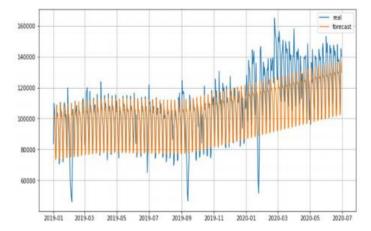
오프라인 시계열 예측 대비 실제 구매수: + 1.6만건

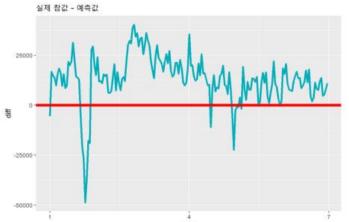
온라인 시계열 예측 대비 실제 구매수: + 178만건

▶ 코로나19로 인해 온라인 시장의 성장 가속화

#### 04 온/오프라인 시계열 분석

### 2020년 온라인 구매수 예측 및 실제값과의 비교





▶ trend는 상승하는 추세이고,실제값은 예측값보다 더 높은 추세를 보임(5060대의 온라인 소비의 보편화로 추정)

온라인	FALSE	TRUE	Total	실제값 > 예측값 인 비율
전체	15	167	182	92%

#### 결론

#### 미래 핵심 소비층인 실버 세대를 위한 온라인 시장 공략

- 고령사회 진입을 앞둔 한국: 빠른 속도로 고령화 되고 있음
- 시니어 세대의 경제 활동 인구 증가는 향후 고령사회에서 핵심 소비 주체로서의 역할
- 온라인 시장의 지속적인 성장
- 오프라인에서는 아직 할인점, 마트와 같은 눈으로 보고 직접 구매하는 경우가 우세하다. 하지만 이번 코로나19로 인해 식료품등 온라인 시장의 성장이 가속화 됨



## 어린이 교통사고위 점지역 도출

## 어린이 교통 사고 제로화

오산시 어린이 교통사고 위험 지역 도출

## 오산시 어린이 교통사고 위험 지역 도출

[ 제로화 ]

국토도시 데이터 분석과제로 오산시 어린이 교통 사고 근절을 위해 두지 과제를 해결하는 문제였습니다.

1. 어린이 보호구역 외 어린이 교통사고 위험 지역 20개로 제시 2.기존 어린이 보호구역중 교통안전시설물 우선 설치지역 20개소 제시 **기간/인원수** 2021.01.16 - 2021.01.28 / 4명

기술스택 Python(Numpy, Pandas), Slack, GitHub, Pycharm, Colab

데이터 17개의 오산시 지역 데이터,, 교통약자 다발지점 사고 지표

**알고리즘** 기본 회귀 알고리즘(Linear, DecisonTree, RF, LassoCV, LinearSVR, Kneighbors), 앙상블 (XGB, GB, Stacking, Voting, Bagging)

시각화 python 에서 제공하는 plotly 라이브러리 사용

**구현 부분** 기획, 데이터 전처리, 알고리즘 개발 및 구현

결과 1. 어린이 보호구역 외 어린이 교통사고 위험지역 20개소 제시 2.기존 어린이 보호구역중 교통안전시설물 우선 설치지역 20개소 제시

1. 프로젝트 배경

## 프로젝트 계획





#### 해결 과제 1

어린이 보호구역 외 어린이 교통사고 위험지역 20개소 제시



#### 해결책 1

현황분석 + 지도학습 통한 예측



#### 해결 과제 2

기존 어린이 보호구역 중 교통안전시설물 우선 <u>설치지역</u> 20개소 제시



#### 배결책 2

지도학습 통한 예측

2. 데이터 분석

## 최적 모델 선정





#### 사용한 모델과 최종 모델로 선택되는 기준

• 모델 선택 기준

• 각 모델 별 MAE 결과 값

#### < MAE (Mean Absolute Error) >

독립변수와 종속변수의 관계를 제대로 설명하는지 정량적인 지표를 이용 하여 평가하는 방법

$$MAE := rac{1}{N} \sum_{i=1}^N |H(X_i) - y_i|$$

Linear	0.359	DecisionTree	0.365
RF	0.326	LGBM	0.343
LassoCV	0.315	XGB	0.346
ElastioNetCV	0.316	GradientBoosting	0.320
LinearSVR	0.236	Stacking	0.233
BayesianRidge	0.339	Voting	0.289
KNeiahbors	0.300	Bagging	0.228

데이터 탐색
 데이터 탐색 및 전처리

## 데이터 수집



① Multipolygon 타입 데이터 합치기

=> intersects 활용



⇒ Multipolygon이 어린이 교통사고 격자와 겹치는 부분을 추가방식으로 데이터 추출 ② Point 타입 데이터 합치기

=> within 활용



⇒ Point가 어린이 교통사고 격자에 포함되는 부분을 추가방식으로 데이터 추출

## 발전방안

Q. 분석결과가 실제로 어디에 활용될까?

보호지역 內 안전시설물 추가 설치 지역



보호지역 外 어린이 교통사고 위험 지역

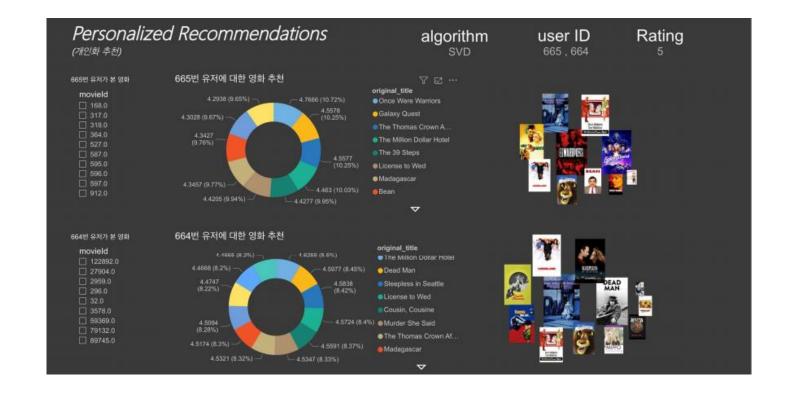
오산시 '스쿨존 어린이교통사고 제로화'정책 추진

'민식이법 '통과 따라 어린이보호구역 교통시설 개선 착수

## 영화 추천 시스템

## 영화 추천 시스템

영화 데이터를 이용한 Hybrid 추천 시스템 구현



사용자 데이터와 영화 데이터를 이용해 영화를 추천하는 시스템을 구현 하였습니다. 추천 방식으로는 Content based filtering / collaborative filtering / Hybrid filtering을 사용하였다. 또한, 유저가 영화를 보는 것에 따라 개인화 추천을 구현하였다. **기간/인원수** 2020.08.21-2020.09.10/4명

기술스택 Python(Numpy, Pandas, sklearn), Slack, GitHub, Pycharm, Power bi

**데이터** Movie meta data(영화 관련 메타 데이터) / review data(사용자 리뷰 데이터)

알고리즘 SVD/Slope/NMF/ALS/KNN

시각화 Power bi 를 통한 시각화

**구현 부분** 기획, 데이터 시각화 및 전처리, 알고리즘 및 추천 시스템 구현

결과 각각의 알고리즘에 따른 영화 추천

개인이 평점을 매긴 영화에 따른 개인화 추천

알고리즘 별 성능차이 분석(RMSE,MAE)

사용자 기반 협업 / 컨텐츠 기반 협업 추천 시스템 구현

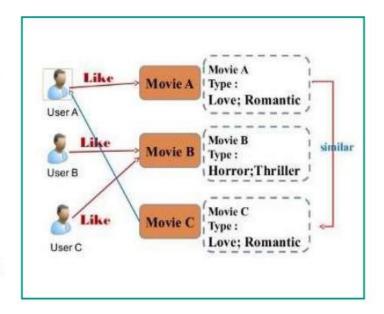
Hybrid 추천 모델 구현

## 4. 프로젝트 수행 결과\_ 기획 및 사전조사

#### 대안

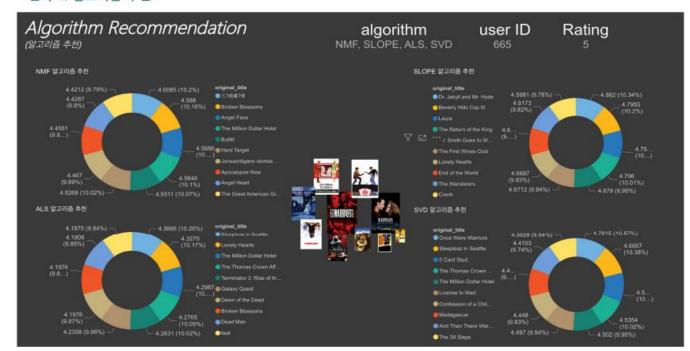
#### 하이브리드 (Hybrid) 필터링 모델

- 위 두 모델의 단점들을 최소화 하기 위해서 고안 된 방법이 바로 Hybrid 필터링이다.
- 사용자 평점 기반 데이터에 콘텐츠 기반 데이터 (ex. 특징들)의 가중치를 주고 두개를 합산하여 좀 더 정확한 결과를 낼 수 있는 모델이다.
- 콜드 스타트를 방지하기 위해 신규 아이템에 대해서는 콘텐츠 기반 가중치가 보완 역할을 하고, 반대로 콘텐츠 기반 필터링의 단점은 협업 필터링 모델로 상쇄하는 상호보완적인 알고리즘 모델이라고 할 수 있다.



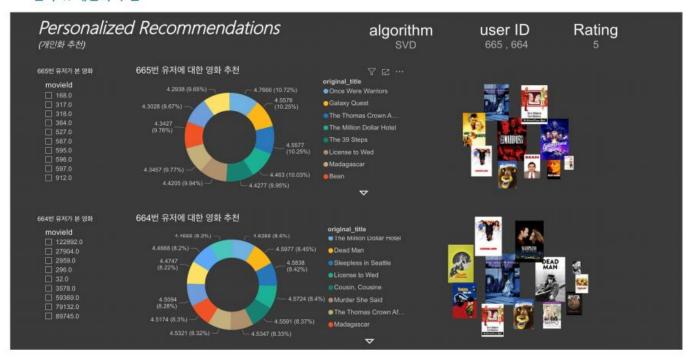
### 4. 프로젝트 수행 결과\_ 시각화

결과 4. 알고리즘 추천



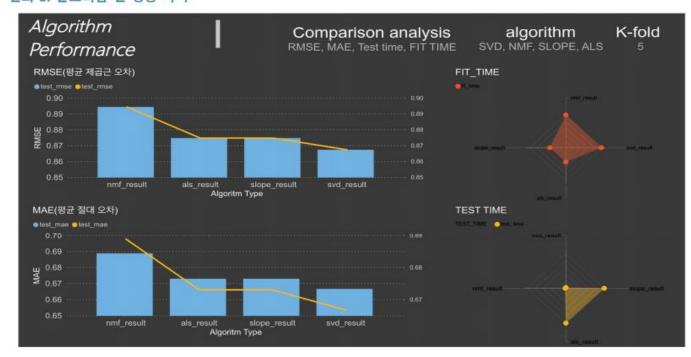
### 4. 프로젝트 수행 결과\_ 시각화

결과 1. 개인화 추천



### 4. 프로젝트 수행 결과\_시각화

결과 5. 알고리즘 별 성능 차이



## 판매실적 예측을 통한 편성 최적화

## 판매 실적 예측을 통한 편성 최적화

NS SHOP 판매 실적 예측과 이에 따른 편성 최적화 방안





박 콘테스트에서 진행한 대회로 NS SHOP 판매 실적을 예측하고 이에 따라 편성을 어떻게 하는 것이 최적화 된 방안 인지를 도출하는 대회였습니다.

- 1.판매 실적 데이터를 통해 상품의 취급액(매출) 예측
- 2.편성표 최적화 방안 도출

**기간/인원수** 2020.08.03 – 2020.09.08/4명

기술스택 Python(Numpy, Pandas, sklearn), R, Slack, GitHub, Pycharm, Colab

**데이터** NS SHOP 판매 실적 데이터, NS SHOP 쇼핑몰 카테고리 분류 데이터(크롤링),

시청률 데이터, 기상청 데이터

알고리즘 DL: DNN / 1D CNN

ML: XGB, LGBM, Random Forest, HGB, Voting, Stacking

시각화 Python - ggplot , R - ggplot

**구현 부분** Python 관련 데이터 전처리 및 시각화, ML 관련 알고리즘 구현,

데이터크롤링

**결과** 판매실적 데이터를 통한 취급액 예측

편성표 최적화 방안 도출

알고리즘에 따른 취급액 예측(MAPE)

#### Feature Engineering

## 코로나 관련 변수



취급	·액 (억원)	19년 2Q	20년 2Q	증감량	증감율
Elec	tronics	458.595	439.229	-19.366	-4.2%
Hea	lth Food	954.557	1070.39	115.833	12.1%
Food	d	760.928	1040.862	279.934	36.8%
Hou	sehold goods	197.026	265.752	68.726	34.9%
Cost	metics/Hair	407.64	372.791	-34.849	-8.5%
App	arel	523.138	398.628	-124.51	-23.8%
Oth	ers	95.116	103.348	8.232	8.7%
총합	계	3397	3691		

2020 2분기 NS 홈쇼핑 IR 자료에서

NS SHOP+ 취급액과 6월 취급액만 특정할 수 없다는 문제 발생

코로나19로 인한 취급액 영향을 알아보고자 IR 자료의 품목별 실적 비중을 전년 동기와 비교

### Television Schedules

#### 편성 최적화 방안 도출: 최적 수익을 고려한 방송일시에 따른 상품 소분류 추천 과정

#### STEP 1

특정 방송일자(ex, 2019년 1월 1일)에 대한 데이터를 Train으로 설정

#### STEP 3

STEP 5 \* 상품별 취급액을 기준으로 \* 추출된 상품 데이터들의 소분류별로 개수를 세어 내림차순 정렬

가장 많은 빈도의 소분류를 선택

• 추천할 최종 소분류 결정

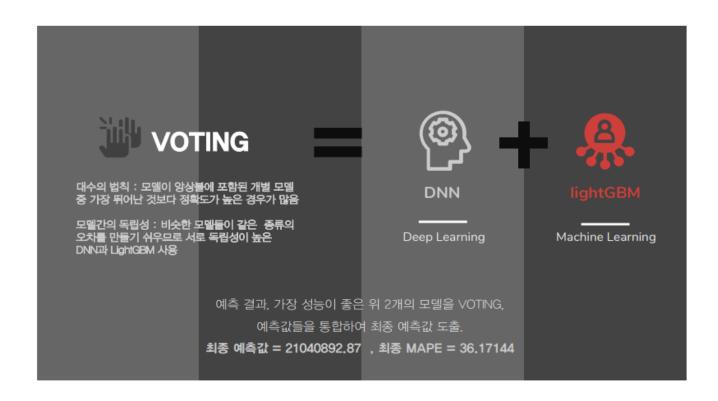
STEP 7

• 특정 방송일자의 방송시간대별로 그룹화해 상품별 취급액 합을 구함

4분위수로 데이터를 나눠 상위 25%의 상품들을 추출

#### • 소분류의 개수가 같다면 소분류들의 취급액 평균을 구해 더 큰 소분류를 선택

\*가장 큰 값을 선택하지 않은 이유는 하나의 상품군에서 이상치에 가까운 취급액이 나온 데이터를 취하는 것을 막기 위함



5,272,283.41

3.089.597.07

#### Optimized **Television Schedules**

## 추천 결과

하루 기준

해스때벨 웜 업속옷 보정언더우 5.420.365 16,727,188 11,073,776 6,702,776.27 트용 극세사 의류 잠옷,이지( 4,843,139 26,078,100 15,460,620 12,198,619.71 위트용 극세사 의류 잠못,이지도 9,084,517 14,430,469 11,757,493 4,802,492.85 히트용 극세사 의류 잠옷,이지되 7.671,401 24,582,228 16,126,814 9,454,814.28 히트용 극세사 의류 잠옷,이지( 13,898,215 28,960,406 21,429,310 12,092,310.38 히트용 극세사 의류 장우.이지도 13.017.999 7.279.565 10.148.782 3.329.781.53 전치 엠마 사철잡화 캐주얼가녕 21,938,219 13,358,487 17,648,353 1,959,353.20 전치 엠마 사철잡화 캐주얼가본 31,893,833 15,779,014 23,836,423 1,533,576.60 드라다 투웨이'잡화 캐주얼가분 22,419,720 25,605,294 24,012,507 7,879,506.76 일시불 쿠쿠전 주방 냄비,압력: 41.662,308 47,235,052 44,448,680 14,387,679.83 일시불 쿠쿠전 주방 냄비,압력; 57.337.332 60,262,120 58,799,726 5,257,726.02 일시불 쿠쿠전 주방 냄비,압력1 15.979.824 18.942.940 17.461.382 1.386.382.20

프라다 투웨이'잡화 캐주얼가성 29.470.160 27,965,034 28,717,597

2.734.501 12.008.066 7.371.283

취급액 차이

해당 날짜에 대한 프로그램 개수

## 수요 예측을 통한 물류 최적화

## 수요 예측을 통한 물류 최적화 시스템

수요 예측을 통한 물류 최적화 + 위치 데이터 기반 배송 위치 최적화

# 수요 예측을 통한 물류 최적화 시스템

#황인범 #박세진 #이경현

한국 농수산 식품 유통공사에서 진행한 해커톤으로서 농식품 관련 서비스 개발을 하고 제안하는 빅데이터 해커톤 이였습니다. 이 주제를 수요 예측을 통한 물류 최적화 시스템으로 잡고 진행하였고 본선까지 진출하였습니다.

**기간/인원수** 2020.10.26 - 2020.11.16/3명

기술스택 Python(Numpy, Pandas), Django, Slack, GitHub,

Pycharm, Html, Css, JavaScript

데이터 가락시장 반입량 데이터, 나이스 지니 데이터, 물류창고 업체 인허가

정보 데이터, 소상공인시장 진흥공단 전국 전통시장 현황 데이터

**알고리즘** Prophet 모듈을 통한 시계열 예측, Haversine 을 통한 위경도 거리 계산

시각화 WEB, Plotly를 통한 시각화

**구현 부분** 웹 개발, 데이터 전치리 및 시각화, 알고리즘 개발 및 구현

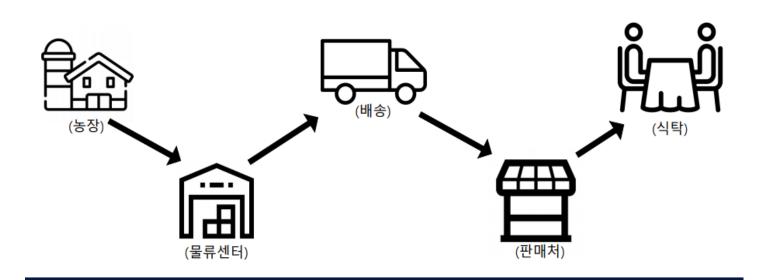
결과 시계열 분석을 통한 농식품 수요 예측

수요 예측을 통한 물류 창고 최적화

지역시장 경쟁력 강화

Part 2

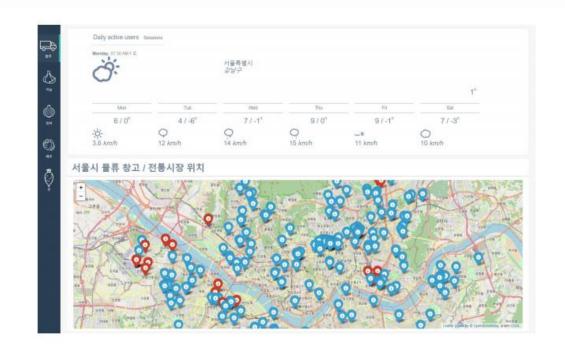
## 서비스 모델 제안



▶ 물류 데이터와 판매처 데이터를 통해 물류시스템을 최적화

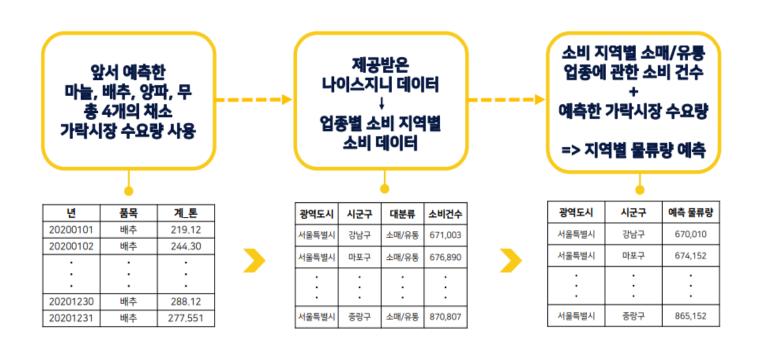
Part 5

## 서비스 시연



Part 4

## ② 물류 예측 – 전통 시장별 물류량 예측



Part 5 결론

## 물류 최적화 서비스를 통한 기대 효과



# 공공 도서관 입지 선정

## 데이터를 통한 공공 도서관 입지 선정

데이터 활용을 통해 공공 도서관의 위치 선정에 관한 의사결정 지원

Part 4 아이디어 실현 방안 Preview



국립중앙도서관에서 진행한 공모전으로서 서비스 아이디어 부문으로 참여하였습니다. 데이터를 통해 각 행정동의 도서관 대출 수요를 예측을 하였고 사후처리를 통해 좀 더 구체적인 입지선정을 하였습니다. 공모전 결과로는 관장상인 우수상을 수상하였습니다. **기간/인원수** 2020.07.10 - 2020.08.16 / 4명

기술스택 Python(Pandas,geopandas),

Pycharm, Html, Css, JavaScript

데이터 서울시공공 도서관별대출데이터, 센서스용행정구역 경계, 행정동 성별 인구수, 행정동

별 연령별 인구 수 서울 지하철 위치 데이터,지하철 승/하차 정보,버스정류장 위치

정보,버스정류장 승/하차 정보

알고리즘 Xgboost, lgbm, randomforest

시각화 Plotly를 통한 시각화

**구현 부분** 기획, 데이터 전처리, 시각화

결과 지역별 편차해소

데이터 기반의 의사결정 도움

독서율 향상 및 공공 도서관 활성화



<sup>Part 4</sup> 분석 결과 사후 처리

794

#### 사후 처리

작은 도서관의 유무



#### 효과

작은 도서관은 공공 도서관의 대체제로서의 역할 가능

#### 사후 처리

기존 도서간과의 거리



#### 立立

지역 간의 정보 불균형 완화

## Part 4 공공도서관 최적 입지 선정



A

데이터 기반의 의사결정 도움

위치 선정에 있어서 데이터를 통해 좀 더 정확하고 효율적인 의사결정을 가능케함 B

지역별 편차 해소

누구나 집 근처 도서관에서 지식·정보· 문화를 누리도록 한다는 목표를 이룰 수 있음

<u>독서율</u> 향상 & 공공 도서관 활성화

접근성 및 편의성 증대를 통한 지역 주민들의 공공 도서관 이용률 증가

# THANK YOU