

# NS SHOP+ 판매 실적 예측을 통한 편성 최적화 방안 모형 도출

팀명 : InsightOut

박민형 pminhyung12@naver.com

김우용 zxc1843zzz@naver.com

박서희 seohuipark95@gmail.com

문찬호 buddy6274@naver.com

# Contents

- 예측 모델 개발
- 최적화 모델 개발

01

## 분석 주제 소개

- Nsshop +
- 분석 목표 설정

02

## 데이터 취합 및 분석

- 데이터 전처리
- EDA
- 유용한 변수 생성

03

## 예측, 최적화 모델 개발

- 예측 모델 개발
  - 모델 학습
  - 하이퍼파라미터 최적화
  - Train Error- based Feature Engineering
- 최적화 모델 개발

04

## 결과 및 기대효과

- 기대효과



01

## 분석 주제 소개

- Nsshop +
- 분석 목표 설정

## “온디맨드 서비스에서 딥러닝 기반 상품별 매출 예측, 자동 편성 시도”



홈쇼핑 시장 점유율 5위, 매출액 4858억원 (2019년 기준)



“온디맨드 서비스 제공을 통한 TV홈쇼핑과 차별화” 전략 추구

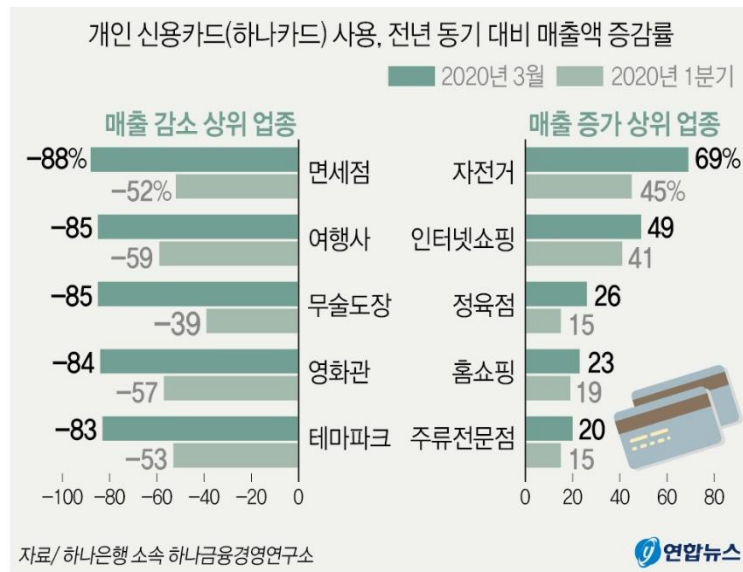
On-demand 기반 모든 채널이 융합되어

고객에 대한 맞춤형 서비스를 제공

하트, 리모콘 구매 등 온디맨드 서비스들을 제공함

최근 빅데이터 상품 추천, 딥러닝 기반 매출 예측, 자동 편성 시도

## “편성 예측 모델 코로나 19와 같은 환경적 요소 반영 필요”



이재윤 기자 / 20200521 트위터 @yonhap\_graphics 페이스북 tuneey.kr/LeYN1

### # 코로나19로 인한 2020년도 소비행태 변화

작년 대비.홈쇼핑 온라인 쇼핑 소비량, 매출액 증가  
작년 대비 상품 품목별 소비 행태 및 소비량 변화

### # 코로나19로 인해 생활 패턴 변화

사회적 거리두기로 인해 “집콕” 족 등장

이에 따라 프로야구, 뉴스 등 각종 프로그램 시청률 상승

내부 요인을 고려한 규칙 기반 편성

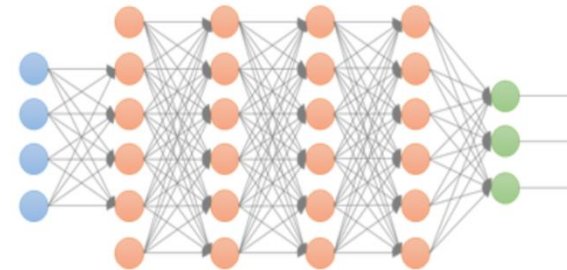


머신러닝/ 딥러닝 알고리즘 기반 편성

NS홈쇼핑

시간	07.07 화	07.08 수	07.09 목
11시	35 주말 HD	35 액세서리 HD	35 화장품 HD
12시	40 현타 HD	40 주말 HD	40 건강보조식품 HD
13시	40 건강기능식품 HD	40 생활잡화 HD	40 화장품 HD
14시	35 계절가전 HD	35 건강보조식품 HD	40 집구 HD
15시	30 상조상품 HD	30 상조상품 HD	40 화장품 HD
16시	30 농산물 HD	30 수산물 HD	40 건강기능식품 HD
17시	10 축산물 HD	10 농산물 HD	50 축산물 HD
18시	35 농산물 HD	35 건강기능식품 HD	35 건강기능식품 HD

NS홈쇼핑  
롯데홈쇼핑  
현대홈쇼핑  
CJ홈쇼핑  
GS SHOP  
CJ오쇼핑플러스  
K쇼핑  
공영쇼핑  
쇼핑엔디  
홈8쇼핑



최대 수익 방송 편성표 도출

시각	07.07 (화)	07.08 (수)	07.09 (목)	07.10 (금)	07.11 (토)
09:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
10:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
11:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
12:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
13:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
14:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
15:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
16:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
17:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)
18:00	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)	현대홈쇼핑 (20)

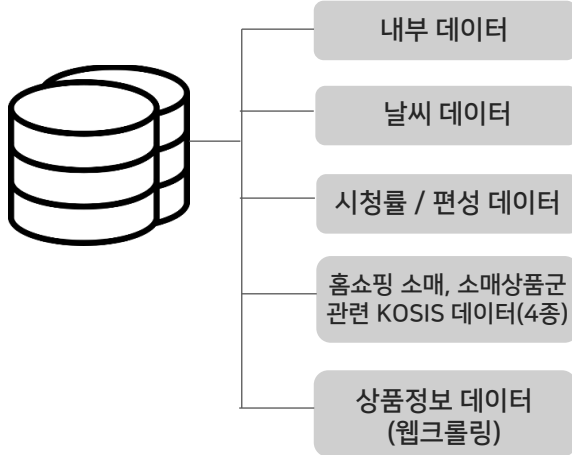
고객 니즈 상품력, 시즌별 상품 변화 고려  
날씨, 행사, 타 방송사 현황 자료조사 하여 반영  
기존 편성 규칙들을 기반으로 향후 편성 추진

상품 메타데이터 수집 및 활용, 코로나 19로 인한 소비 추세 학습  
외부 요인 환경, 날씨, 경제 지수 데이터 반영  
학습 데이터를 기반으로 매출액을 극대화하는 편성 제시

## NS SHOP+ 판매실적 예측을 통한 편성 최적화 방안(모형) 도출

- 2019년, 2020년 6월의 NS SHOP+ 편성/판매 데이터, 기상 데이터, 상품정보 데이터, 해당 기간의 소매 및 홈쇼핑 관련 통계자료 데이터를 추가하여 분석 및 판매 실적 예측 모델 개발.
- 예측 모델을 활용한 2020년 하반기 편성 최적화 알고리즘 개발

## 1. 데이터 전처리 및 통합 데이터 생성



## 2. 탐색적 데이터 분석 및 시각화(EDA)

## 3. 유용한 변수 생성(Feature Engineering)

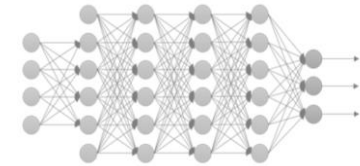
방송일시  
상품명  
노출시간

요일, 휴일, 주당단위시간,  
연중 단위 수

결제방식, 성별, 프로모션  
세트여부, 상품 종류

중간시간, Zapping

## 4. 상품 판매 실적 예측 모델 개발



## 5. 편성 배치 알고리즘 제시

최대 수익 방송 편성표 도출									
시각	18:00-18:30	18:30-19:00	19:00-19:30	19:30-20:00	20:00-20:30	20:30-21:00	21:00-21:30	21:30-22:00	22:00-22:30
18:00	18:00-18:30 (KBS) 뉴스 편성: 18:00-18:30 (KBS) 수익: 18000000 평가: 4.5	18:30-19:00 (KBS) 뉴스 편성: 18:30-19:00 (KBS) 수익: 18000000 평가: 4.5	19:00-19:30 (KBS) 뉴스 편성: 19:00-19:30 (KBS) 수익: 18000000 평가: 4.5	19:30-20:00 (KBS) 뉴스 편성: 19:30-20:00 (KBS) 수익: 18000000 평가: 4.5	20:00-20:30 (KBS) 뉴스 편성: 20:00-20:30 (KBS) 수익: 18000000 평가: 4.5	20:30-21:00 (KBS) 뉴스 편성: 20:30-21:00 (KBS) 수익: 18000000 평가: 4.5	21:00-21:30 (KBS) 뉴스 편성: 21:00-21:30 (KBS) 수익: 18000000 평가: 4.5	21:30-22:00 (KBS) 뉴스 편성: 21:30-22:00 (KBS) 수익: 18000000 평가: 4.5	22:00-22:30 (KBS) 뉴스 편성: 22:00-22:30 (KBS) 수익: 18000000 평가: 4.5
18:30	18:30-19:00 (KBS) 뉴스 편성: 18:30-19:00 (KBS) 수익: 18000000 평가: 4.5	19:00-19:30 (KBS) 뉴스 편성: 19:00-19:30 (KBS) 수익: 18000000 평가: 4.5	19:30-20:00 (KBS) 뉴스 편성: 19:30-20:00 (KBS) 수익: 18000000 평가: 4.5	20:00-20:30 (KBS) 뉴스 편성: 20:00-20:30 (KBS) 수익: 18000000 평가: 4.5	20:30-21:00 (KBS) 뉴스 편성: 20:30-21:00 (KBS) 수익: 18000000 평가: 4.5	21:00-21:30 (KBS) 뉴스 편성: 21:00-21:30 (KBS) 수익: 18000000 평가: 4.5	21:30-22:00 (KBS) 뉴스 편성: 21:30-22:00 (KBS) 수익: 18000000 평가: 4.5	22:00-22:30 (KBS) 뉴스 편성: 22:00-22:30 (KBS) 수익: 18000000 평가: 4.5	22:30-23:00 (KBS) 뉴스 편성: 22:30-23:00 (KBS) 수익: 18000000 평가: 4.5



# 02

## 데이터 취합 및 분석

- 데이터 전처리
- EDA
- 유용한 변수 생성

# 1. 탐색적 데이터 분석 및 시각화(EDA)

## ● NS SHOP+ 제공 데이터

프로그램별 매출실적 데이터

2019년 (2019.01.01 ~ 2019.12.31)

NS Shop+ 2019.01.01~12.31

방송일시	노출(분)	마더코드	상품코드	상품명	상품군	판매단가	취급액
2019.1.1 6:00	20	100346	201072	테이트 남성 셀린나트3중	의류	39,900	2,099,000
2019.1.1 6:00		100346	201079	테이트 여성 셀린나트3중	의류	39,900	4,371,000
2019.1.1 6:20	20	100346	201072	테이트 남성 셀린나트3중	의류	39,900	3,262,000
2019.1.1 6:20		100346	201079	테이트 여성 셀린나트3중	의류	39,900	6,955,000
2019.1.1 6:40	20	100346	201072	테이트 남성 셀린나트3중	의류	39,900	6,672,000
2019.1.1 6:40		100346	201079	테이트 여성 셀린나트3중	의류	39,900	9,337,000
2019.1.1 7:00	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	6,819,000
2019.1.1 7:20	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	15,689,000
2019.1.1 7:40	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	25,370,000
2019.1.1 8:00	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	16,133,000
2019.1.1 8:20	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	30,061,000
2019.1.1 8:40	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	53,542,000
2019.1.1 9:00	20	100816	202404	보코 리버시블 무스탕	의류	79,000	16,075,000
2019.1.1 9:20	20	100816	202404	보코 리버시블 무스탕	의류	79,000	25,628,000
2019.1.1 9:40	20	100816	202404	보코 리버시블 무스탕	의류	79,000	49,270,000
2019.1.1 10:00	20	100809	202395	CERINI by PAT 남성 풀패키지 기모니트 3중	의류	79,900	48,799,000
2019.1.1 10:20	20	100809	202395	CERINI by PAT 남성 풀패키지 기모니트 3중	의류	79,900	97,907,000
2019.1.1 10:40	20	100809	202395	CERINI by PAT 남성 풀패키지 기모니트 3중	의류	79,900	112,972,000
2019.1.1 11:00	20	100664	201999	크로커다일 The 편안한 코튼 브라팬티 4세트	속옷	69,900	24,531,000
2019.1.1 11:20	20	100664	201999	크로커다일 The 편안한 코튼 브라팬티 4세트	속옷	69,900	49,829,000
2019.1.1 12:00	20	100088	200236	에코라믹 통주물 스톤 냄비세트	주방	60,900	33,823,000
2019.1.1 12:20	20	100088	200236	에코라믹 통주물 스톤 냄비세트	주방	60,900	59,083,000
2019.1.1 12:40	20	100088	200236	에코라믹 통주물 스톤 냄비세트	주방	60,900	99,235,000

요일별/시간대별 분단위 시청률 데이터 (단위 %)

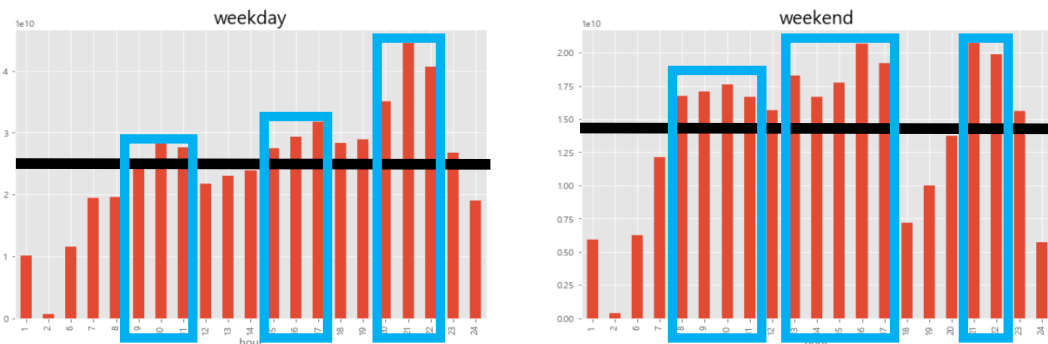
2019년 (2019.01.01 ~ 2020.01.01)

NS SHOP+ 일자별, 시간대별 시청률 (2019년)

시간대	2019-01-01	2019-01-02	2019-01-03	2019-01-04	2019-01-05	2019-01-06	2019-01-07
02:00	0	0	0	0	0	0	0
02:01	0	0	0	0	0	0	0
02:02	0	0	0	0	0	0	0
02:03	0	0	0.014	0	0	0	0
02:04	0	0	0	0	0	0	0
02:05	0	0	0	0	0	0	0
02:06	0	0	0	0	0	0	0
02:07	0	0	0	0	0	0	0
02:08	0	0	0	0	0	0	0
02:09	0.014	0	0	0	0	0	0
02:10	0	0	0	0	0	0	0
02:11	0	0	0	0	0	0	0
02:12	0	0	0	0	0	0	0
02:13	0	0	0	0	0	0	0
02:14	0	0	0	0	0	0	0
02:15	0	0	0	0	0	0	0
02:16	0	0	0	0	0	0	0
02:17	0	0	0	0	0	0	0
02:18	0	0	0	0.014	0	0	0
02:19	0	0	0	0.014	0	0	0
02:20	0	0	0	0	0	0	0
02:21	0	0	0	0	0	0	0
02:22	0	0	0	0	0	0	0

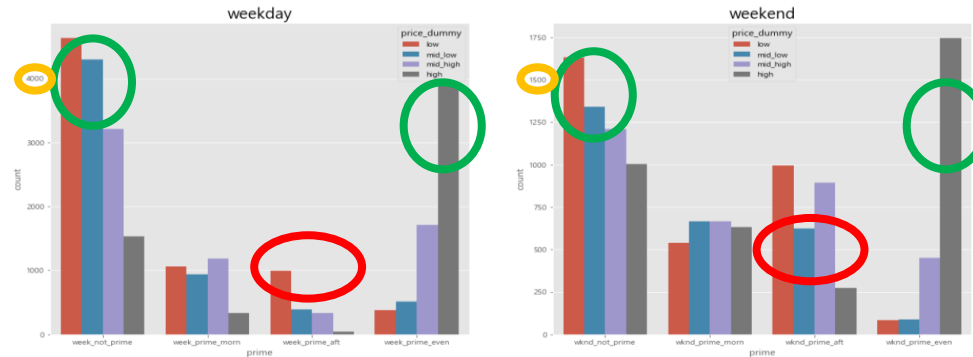
# 1. 탐색적 데이터 분석 및 시각화(EDA)

- 시청률, 매출액에 따른 프라임 타임



[평일] 오전 : 9~11시 / 오후 : 16~17시 / 저녁 : 20~22시,  
[주말] 오전 : 8~11시 / 오후 : 13~17시 / 저녁 : 21~22시

- 판매 단가

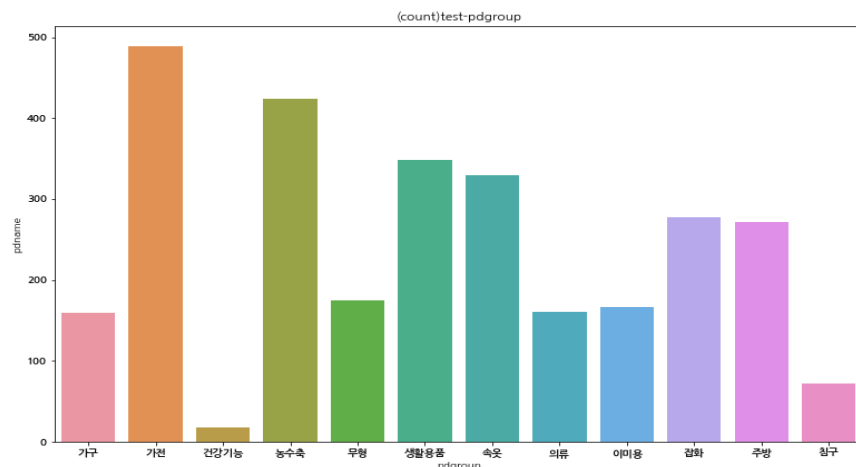


➔ 평일과 휴일(주말+공휴일) 방송상품의 판매단가에 큰 차이를 보여 이를 반영한 명목형 변수를 생성.

➔ 평일, 저녁 내에서 오전, 오후, 저녁의 프라임타임 시각 그룹, 나머지 시간 그룹으로 총 4개의 명목형 범주 생성

# 1. 탐색적 데이터 분석 및 시각화(EDA)

## ● 2020.06 예측 데이터 탐색



hour	1	2	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
pdgroup																						
가구	0	0	6	3	0	12	12	3	0	0	0	0	0	0	0	16	50	27	9	15	6	
가전	0	0	0	0	0	2	5	24	21	13	32	24	7	1	16	26	59	109	114	36	0	
건강기능	2	0	0	0	0	0	0	0	0	0	0	0	0	4	2	0	0	0	0	4	6	
농수축	4	0	4	2	2	43	36	29	16	3	10	15	51	84	48	22	14	14	15	7	5	
무형	31	7	0	0	0	0	0	0	14	19	12	6	1	4	14	5	0	0	0	24	38	
생활용품	12	2	16	20	14	4	16	30	35	44	30	33	17	4	7	7	10	4	10	13	20	
속옷	23	8	14	18	11	9	17	15	6	11	9	10	8	4	10	27	33	17	29	28	22	
의류	21	7	16	22	17	9	8	5	7	3	4	6	2	0	6	5	1	0	0	8	14	
이미용	11	2	6	11	16	10	12	7	9	12	18	11	4	2	0	8	8	2	2	7	9	
잡화	10	5	18	31	24	6	4	15	20	72	13	18	2	1	4	18	16	0	0	0	0	
주방	4	2	4	14	29	32	20	28	20	8	8	12	11	2	8	26	17	19	8	0	0	
침구	0	0	0	0	0	0	0	0	0	0	16	16	20	8	0	0	0	0	0	8	4	

- 단위시간(20분) 편성개수 기준으로 가전, 농수축이 상위권, 생활용품과 속옷, 잡화, 주방이 중간을 차지.
- 특히 가전 부문은 이른 더위에 맞게 냉장고와 에어컨을 집중적으로 개수를 늘려  
평일의 경우 오후, 저녁 프라임 타임, 휴일의 경우 오전과 저녁 프라임 타임에 적극 편성한 것을 파악.
- 단위가격이 높은 가전부문, 다수 편성된 농수축, 생활용품 부문 상품의 정확한 예측이 가능한 모델이 필요

## 1. 탐색적 데이터 분석 및 시각화(EDA)

- 해당 분석 문제에 대한 결론과 접근 전략
  - 2020년의 경우 코로나 바이러스로 인해 상품군에 따른 편성비율과 주문량, 인기 상품군 등이 매우 상이하다.
  - 2019년에 상품 매출과 주문량과 상관관계가 높은 통계자료 이용 시, 연속성을 이용한다면 테스트 데이터에 대한 예측 가능
  - 2019년 데이터에 대한 과적합 방지

## 2. 데이터 전처리 및 통합 데이터 생성

내부 데이터

프로그램별 매출 실적 데이터



외부 데이터

1. 상품 분류, 검색량 세부 데이터
2. 공중파 채널 편성, 시청률 데이터 수집
3. 소매 업체별, 상품군별 판매액 지수(불변지수)

## 1) 상품 분류, 검색량 세부 데이터

출처 : 네이버 쇼핑, 아이템 스카우트

NAVER 네이버쇼핑

일시불쿠팡압력밥솥 6인용

전처리 되지 않은 텍스트를 그대로 검색에 입력

'쿠팡' 브랜드스토어 바로가기 →

카테고리: 디지털/가전 > 생활/건강 > 가구/인테리어

가격: 20만원 이하, 20만원 ~ 40만원, 40만원 ~ 80만원, 80만원 이상

배송/배달/색상: 무료배송, 오늘출발, 카드할인, 쿠팡, 적립, 추가할인

전체 5,619 | 가격비교 112 | 네이버페이 1,532 | 백화점/소점 715 | 핫딜 0 | 소량인도 3 | 해외직구 2

1. 상품명 (띄어쓰기 적용된 상품명)

쿠팡 6인용 열판압력밥솥 CJ-S-FA0602V

최저 112,480원 판매처 85

2. 가격정보

3. 카테고리 정보

디지털/가전 > 주방/가전 > 전기밥솥 > 압력밥솥

용량: 6인용 | 방식: 일반압력 | 조작방식: 단순화면 | 뚜껑: 분리형내출커버 | 내솥코팅: 다이렉트코팅 |  
 패킹: 3중파워패킹 | 요리종류: 백미, 잡곡, 현미, 누룽지, 영양점 | 레시피서: 잡곡(29분) |

리뷰 ★★★★★ 1,100 · 등록일 2015.12 · ❤️ 찜하기 173 · 📄 정보 수정요청

소량물별 최저가

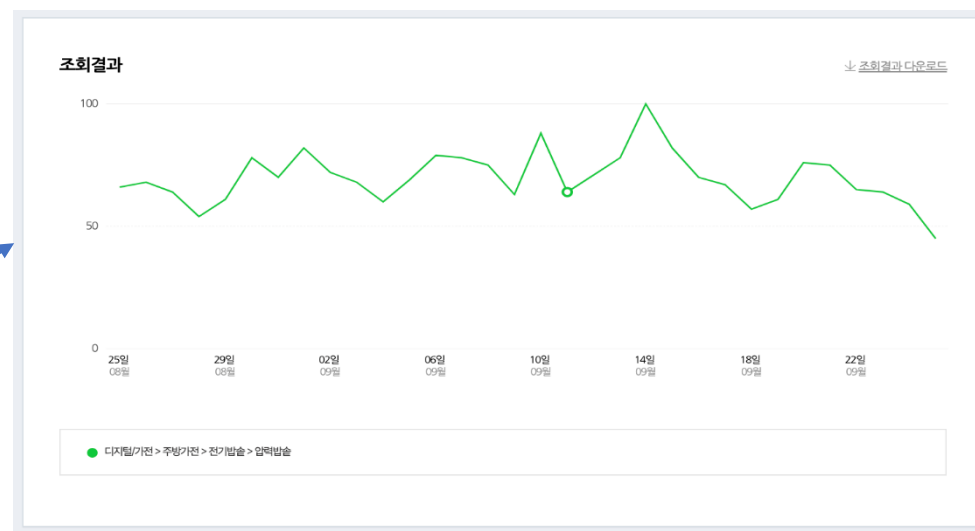
옵션

G마켓 112,480

티몬 112,500

신세계TV쇼핑 112,500

G9 115,520



- 네이버 쇼핑, 아이템 스카우트 사이트에서 상품명 검색 결과 크롤링
- 상품 세부 품목 정보, 상품 브랜드명, 단위별 가격 정보 추출
- 상품 세부 품목 정보를 기반으로 해당 품목 검색량 추이 확인

## 2. 데이터 전처리 및 통합 데이터 생성

## 2) 주요 도시 날씨 데이터 출처 : 기상청

2019년 서울 기상(일단위)

날짜	기온	최고기온	최저기온	일교차	강수량	적설량	풍속	습도	운량	일조시간	날씨	년	월	일	요일	코드(서울)
2019-01-01	-5	-0.6	-8.2	7.6	0	0	2.2	50	3.4	7.5	구름조금	2019	1	1	2	108
2019-01-02	-4.9	0.2	-8.8	9	0	0	1.7	43	0	8.7	맑음	2019	1	2	3	108
2019-01-03	-3.5	3.2	-8.4	11.6	0	0	1.3	39	0.1	8.7	맑음	2019	1	3	4	108
2019-01-04	-1.1	4.1	-6.2	10.3	0	0	1.2	56	5.5	3.9	구름맑음/안개	2019	1	4	5	108
2019-01-05	-2.8	1.1	-5.5	6.6	0	0	2.2	40	0.5	8.6	맑음/안개	2019	1	5	6	108
2019-01-06	-2.8	2.7	-6.3	9	0	0	1.2	35	3.1	7.7	구름조금	2019	1	6	7	108
2019-01-07	-1.9	3.1	-6.2	9.3	0	0	1.5	46	4	7	구름조금/안개	2019	1	7	1	108
2019-01-08	-3.5	0.5	-7.2	7.7	0	0	2.5	31	0	8.7	맑음	2019	1	8	2	108
2019-01-09	-4.7	1.3	-9.4	10.7	0	0	1.3	29	4.9	7.4	구름조금	2019	1	9	3	108
2019-01-10	-0.7	3	-4.5	7.5	0	0	1.4	49	8.1	3.1	흐림/안개	2019	1	10	4	108
2019-01-11	2.4	7.2	-0.4	7.6	0	0	1	59	6.1	6.9	구름맑음/안개	2019	1	11	5	108
2019-01-12	2.8	8	-0.6	8.6	0	0	1.3	48	3.9	6	구름조금/안개	2019	1	12	6	108
2019-01-13	1.2	7.6	-3	10.6	0	0	1.3	54	0	8.6	맑음/안개	2019	1	13	7	108
2019-01-14	1.4	5.3	-2.4	7.7	0	0	0.9	69	4.5	0.4	구름조금/안개	2019	1	14	1	108
2019-01-15	-1.7	2.6	-7.2	9.8	0	0	2.8	59	3.8	4.7	구름조금/눈/안개	2019	1	15	2	108
2019-01-16	-5.3	-1.1	-10.1	9	0	0	2.3	46	4.1	8.6	구름조금/눈	2019	1	16	3	108
2019-01-17	-0.3	4	-3.2	7.2	0	0	1.8	46	1.8	8.7	맑음/눈	2019	1	17	4	108
2019-01-18	0.5	6.7	-4.6	11.3	0	0	1.3	44	3	8.5	구름조금/안개	2019	1	18	5	108
2019-01-19	3.6	8.5	0	8.5	0	0	1.3	60	7.8	2.9	구름맑음/안개	2019	1	19	6	108
2019-01-20	-1.4	4	-5.5	9.5	0	0	3.2	39	1.1	8.8	맑음/안개	2019	1	20	7	108
2019-01-21	-1.2	4.2	-6.7	10.9	0	0	1.6	58	4.1	1	구름조금/눈/안개	2019	1	21	1	108
2019-01-22	1.7	7.1	-3.3	10.4	0	0	1.8	62	1.1	8.5	맑음/안개	2019	1	22	2	108
2019-01-23	2	7	-0.8	7.8	0	0	2.2	49	0.9	8.6	맑음/안개	2019	1	23	3	108

- 기간 : 2019.01.01~ 2019.12.31
- 매출에 영향을 미치는 주요 도시의 날씨 데이터 수집
- 서울, 부산, 대전, 광주, 인천, 대구 -> 최종적으로 서울 데이터 사용
- 최저/최고온도, 미세먼지, 강수량, 습도, 전운량 정보 반영



## 2. 데이터 전처리 및 통합 데이터 생성

### 3) 소매 업태별, 상품군별 판매액 지수(불변 지수) 출처 : KOSIS

- 기간 : 2019.01 ~ 2020.06 (코로나 영향 기간 포함)

업태별	2019. 01	2019. 02	2019. 03	2019. 04	2019. 05	2019. 06	2019. 07	2019. 08	2019. 09	2019. 10	2019. 11	2019. 12	2020. 06
인터넷 쇼핑	201.7	175.8	210.7	200.1	213.4	195.8	205.2	200.9	203.8	215.3	241.3	231.0	261.7
홈쇼핑	139.3	113.3	134.0	130.1	133.8	123.4	124.7	120.3	124.3	139.0	138.8	138.0	139.2

업종별	2019. 01	2019. 02	2019. 03	2019. 04	2019. 05	2019. 06	2019. 07	2019. 08	2019. 09	2019. 10	2019. 11	2019. 12	2020. 06
소매판매액지수	109.8	99.8	116.0	111.4	116.6	110.0	110.5	111.4	113.3	115.8	120.5	121.6	116.9
음식점 및 주점업	97.1	86.4	95.1	95.2	99.9	94.8	98.5	99.7	90.2	96.2	95.7	109.2	86.5

- 경상지수를 디플레이터(소비자물가지수)로 나눈 불변지수 데이터 획득
- 2019년 대비 2020년의 소매 업계 현황에 대한 비교 가능 한 지수를 획득 소매 업태별 데이터에서  
홈쇼핑 업태의 판매액 지수 데이터 추출 및 분석

## 2. 데이터 전처리 및 통합 데이터 생성

### 3) 소매 업태별, 상품군별 판매액 지수(불변 지수) 출처 : KOSIS

- 기간 : 2019.01 ~ 2020.06 (코로나 영향 기간 포함)

업태별상품군	2019. 01	2019. 02	2019. 03	2019. 04	2019. 05	2019. 06	2019. 07	2019. 08	2019. 09	2019. 10	2019. 11	2019. 12	2020. 06
무점포 소매 총지수	166.5	142.0	168.9	162.7	171.4	157.1	163.7	159.5	162.2	172.6	186.7	180.8	197.6
가전제품	180.7	168.1	217.3	188.5	216.6	208.3	211.4	209.7	187.9	203.5	228.9	207.7	278.8
통신기기 및 컴퓨터	184.0	178.9	172.2	144.3	147.0	130.4	151.7	147.5	142.0	145.2	176.5	171.5	168.8
가구	137.2	135.4	154.1	145.1	158.4	144.3	130.5	142.2	144.8	161.1	166.6	160.6	210.8
의복	144.5	127.0	163.4	161.4	170.8	149.2	151.0	124.3	141.8	173.8	200.3	181.1	173.4
오락, 취미, 경기용품	151.2	138.7	179.9	192.7	204.1	189.6	201.1	178.8	185.1	218.0	230.5	227.6	243.6
음식료품	198.0	140.3	168.2	175.1	182.2	170.7	179.2	193.2	196.5	189.4	193.1	196.1	225.4
화장품	150.3	120.2	128.6	136.9	134.5	119.9	122.9	121.8	127.4	131.5	135.8	134.1	125.1
서적, 문구	151.2	132.1	156.7	130.9	118.0	106.7	135.2	126.2	121.9	124.5	137.6	164.0	143.1
기타상품	162.9	143.2	172.1	161.9	171.7	158.2	163.2	160.8	166.3	173.6	182.2	178.3	198.4

- 코로나 영향 기간을 포함하여 각 업태별 상품군 10개에 해당하는 판매액 지수 데이터 확보
- NS shop+ 에서 제공한 편성 데이터의 상품 군과 비교하며 해당 판매 상품에 대한 매출 추이 확인

### 3. 유용한 변수 생성(Feature Engineering)

- 프로그램별 매출 실적 데이터 칼럼 기반 피쳐 생성

방송일시

상품명

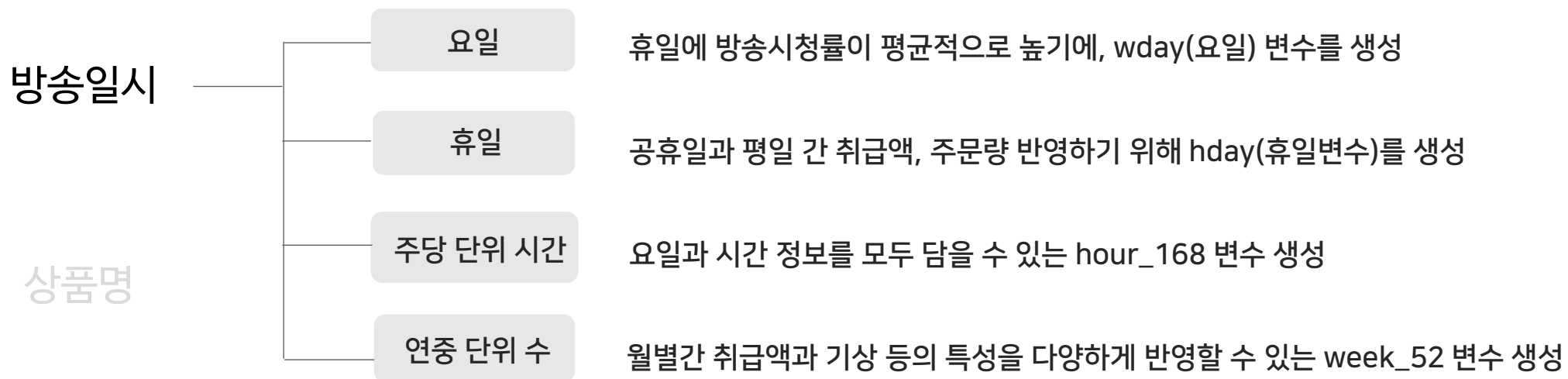
노출(분)



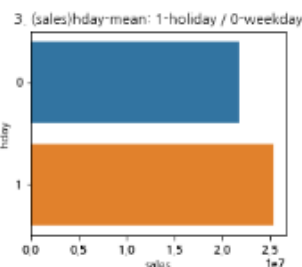
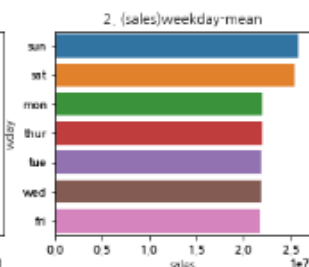
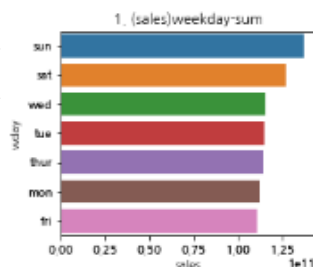
방송일시	노출(분)	마더코드	상품코드	상품명	상품군	판매 단가	취급액
2019.1.1 6:00	20	100346	201072	테이트 남성 셀린니트3중	의류	39,900	2,099,000
2019.1.1 6:00		100346	201079	테이트 여성 셀린니트3중	의류	39,900	4,371,000
2019.1.1 6:20	20	100346	201072	테이트 남성 셀린니트3중	의류	39,900	3,262,000
2019.1.1 6:20		100346	201079	테이트 여성 셀린니트3중	의류	39,900	6,955,000
2019.1.1 6:40	20	100346	201072	테이트 남성 셀린니트3중	의류	39,900	6,672,000
2019.1.1 6:40		100346	201079	테이트 여성 셀린니트3중	의류	39,900	9,337,000
2019.1.1 7:00	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	6,819,000
2019.1.1 7:20	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	15,689,000
2019.1.1 7:40	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	25,370,000
2019.1.1 8:00	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	16,133,000
2019.1.1 8:20	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	30,061,000
2019.1.1 8:40	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	53,542,000
2019.1.1 9:00	20	100816	202404	보코 리버시블 무스탕	의류	79,000	16,075,000
2019.1.1 9:20	20	100816	202404	보코 리버시블 무스탕	의류	79,000	25,628,000
2019.1.1 9:40	20	100816	202404	보코 리버시블 무스탕	의류	79,000	49,270,000

### 3. 유용한 변수 생성(Feature Engineering)

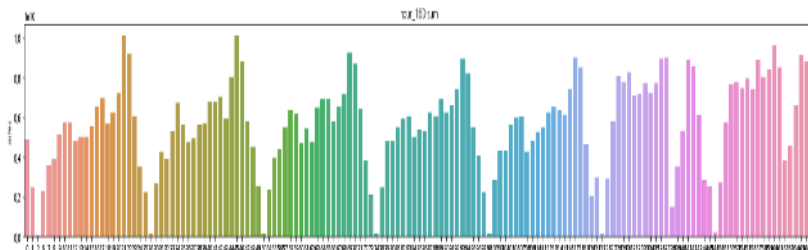
- 프로그램별 매출 실적 데이터 칼럼 기반 피쳐 생성



노출(분)



wday, hday 간 매출액



hour\_168 변수 내 매출액 분포

## 3. 유용한 변수 생성(Feature Engineering)

- 프로그램별 매출 실적 데이터 칼럼 기반 피쳐 생성

방송일시

상품명

노출(분)

결제 방식

군마다 결제방식에 따라 평균 취급액 규모 차이가 있음을 확인

- 0(해당 없음), 1(무이자), 2(일시불)

성별 - 남/녀

남성, 여성 상품 간 취급액 규모 차이에 따라 변수를 생성 - 0(해당 없음), 1(남성), 2(여성)

프로모션 여부

'1+1', '파격가', '특가', '단하루', '역시즌' 키워드 포함 시 1, 그렇지 않을 때는 0.

세트 여부

'세트', '패키지' 키워드 포함 시 1, 그렇지 않을 때는 0.

상품 종류

상품 취급 종류 개수에 따라 평균주문량과 취급액 규모가 상이함

- '3종' 과 같이 '종' 키워드 앞에 있는 숫자를 파싱하여 값을 부여

- '종' 이외에도 정규식을 이용하여 "팩/구/미/인용/박스" 앞의 숫자 값을 부여

### 3. 유용한 변수 생성(Feature Engineering)

- 프로그램별 매출 실적 데이터 칼럼 기반 피쳐 생성

방송일시

상품명

노출(분)

중간 시간 (시)

중간 시간 (분)

Zapping

각 단위 프로그램의 중간 시간(midhour, midminute)을 구함

해당 시간에서 반올림을 한 시간(zapping)이 시청자가

정각에 시작할 다른 프로그램 시청을 위해 채널을 전환하는시각이 zapping time

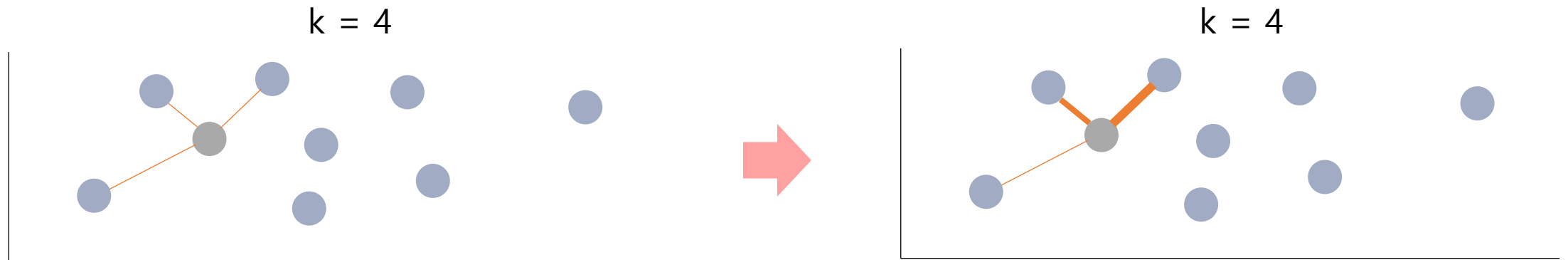
## 4. 결측치 대입 (Null Data Imputation)

- K-Nearest Neighbor를 이용하여 매출액 결측치 처리

장점 : 거리에 따라 값을 추정하는 KNN을 사용하였을 때, Target 인 매출액과 **비선형** 관계에 있는 **많은 명목형 변수**들이 일반 보간법보다 결측치 추정에 좀 더 유용하게 사용될 수 있다.

단점 : 결측치와 k개의 이웃 데이터들과의 거리에 대한 차이를 반영하지 않고, 평균으로 Target 값이 추정된다.

→ k개의 이웃 데이터와의 **거리**에 따라 **가중치(weight)**를 적용하여 정교한 추정이 이루어지도록 함



# 03

## 예측/최적화 모델 개발

- 예측 모델 개발
  - 모델 학습
  - 하이퍼파라미터 최적화
  - Train Error-Based Feature Engineering
- 최적화 모델 개발



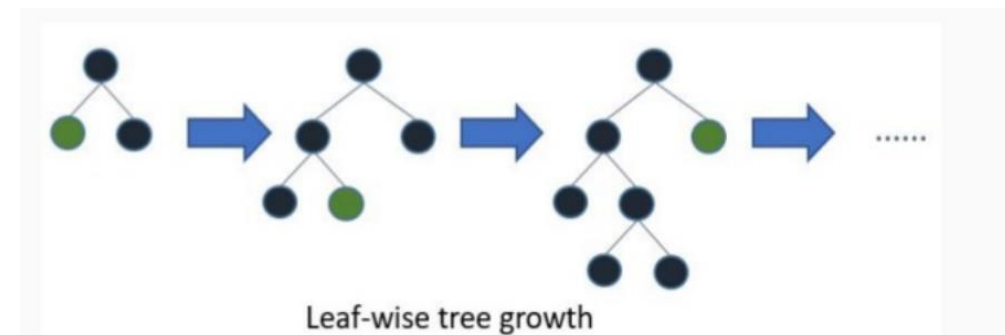
# 1. 통합 데이터 기반 편성 매출액 예측 알고리즘

- 모델학습

## 모델 시도

DNN  
LSTM  
GAN  
Stacking DNN  
Stacking Regression Models

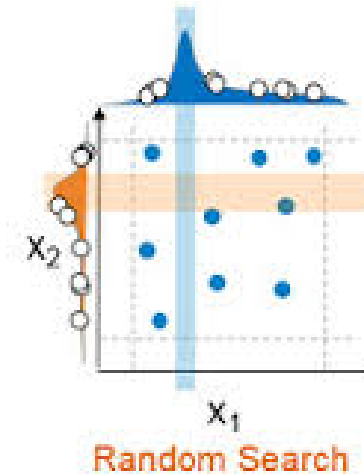
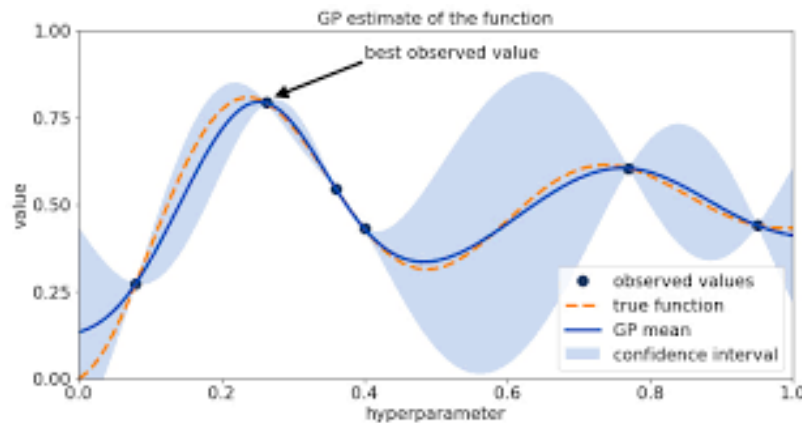
## 최종 모델



# 1. 통합 데이터 기반 편성 매출액 예측 알고리즘

- 하이퍼 파라미터 최적화 : Bayesian Optimization + Random Search

Bayesian Optimization으로 넓은 범위를 대상으로 하이퍼파라미터 결과값을 추정한 후, 성능이 좋았던 하이퍼파라미터 범위에 대하여 Random Search를 진행 결과 정교하게 최적화 수행  
 -> MAPE 7 ~ 8 상승

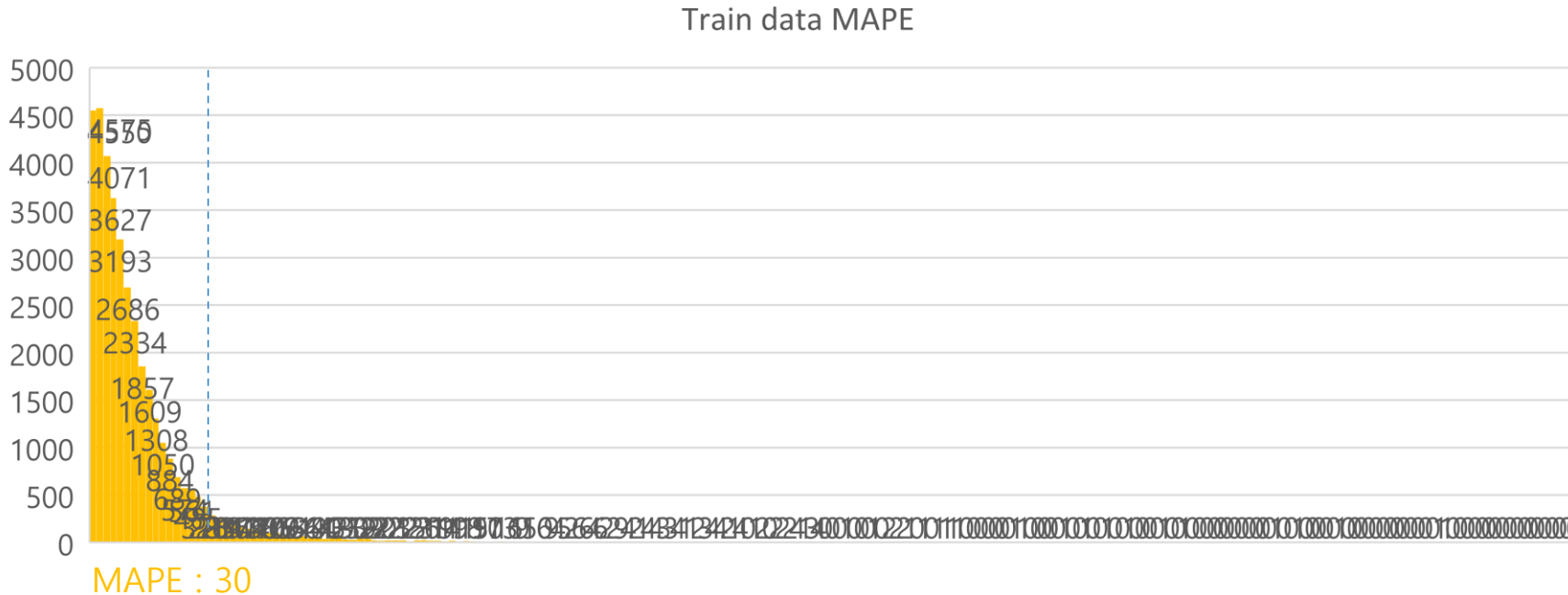


조정 파라미터 :

'num\_leaves', 'max\_depth', 'min\_split\_gain', 'min\_child\_weight', 'learning\_rate',  
 'min\_child\_weight', 'colsample\_bytree', 'subsample', 'n\_estimators'

# 1. 통합 데이터 기반 편성 매출액 예측 알고리즘

- Train error 확인을 통한 모델의 weak part 확인 -> 변수 추가 생성



속옷 과 생활용품 데이터에서 MAPE 수치가 높은 Data 다수 존재 확인  
Test Data 에도 존재하는 관련 속옷 브랜드와 생활용품(마스크) 데이터에 대해 예측정확도를 높이기 위해 **추가변수** 생성

## 1. 통합 데이터 기반 편성 매출액 예측 알고리즘

- Train error 확인을 통한 모델의 weak part 확인 -> 변수 추가 생성

**junior** : train error 이 높은 아동의류와 성인의류 구분

**underwear\_sex** : 주로 여성 속옷에서 train error이 높게 존재, 속옷 내에서 남성, 여성 구분

**puma, dickies, USPA** : train error 이 낮은 USPA, dickies 브랜드 상품과 train error 이 높은 puma 브랜드상품 구별

**Part** :

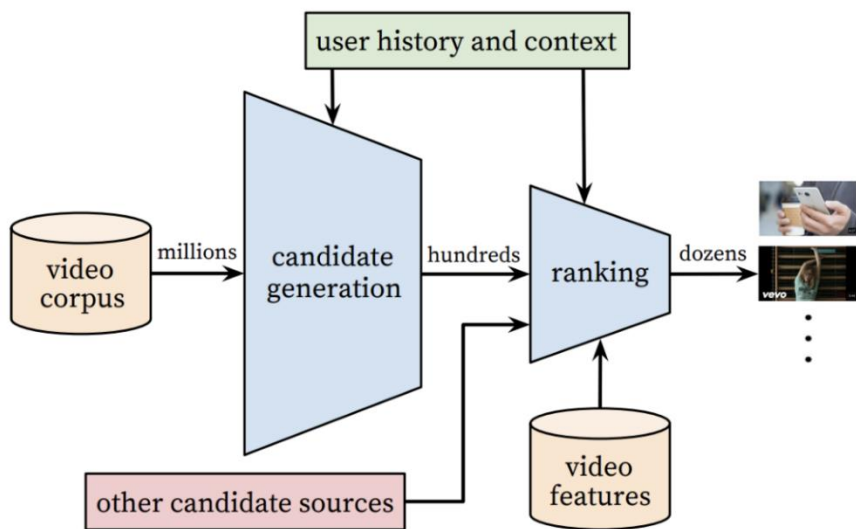
Train error이 프로그램 단위별 마지막구간에 높은 것을 확인하여 동일 프로그램 내 단위별 매출액 분석 결과, 유의미한 매출액 차이가 존재함을 확인, 명목형변수로 생성

다른 프로그램간상이한 단위수에도, 시작단위 구간과 종료 단위 구간에 같은 값이 들어가도록 비율로 생성

**hour\_rank** :

프라임 시간 간에도 train error이 일정하지 않아, 분산이 큰 sales 대신, 주문량을 이용하여, 평일과 공휴일 별 시간당 평균 주문량이 많은 순서대로 가중치를 부여함

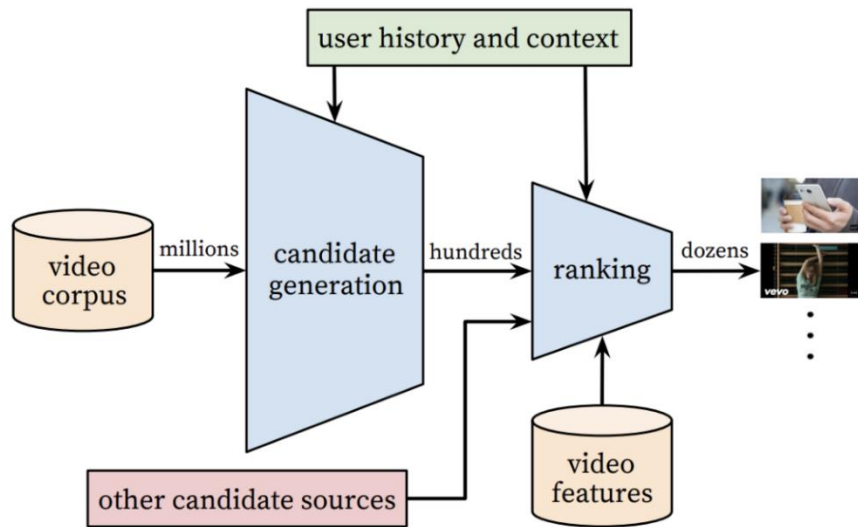
- 딥러닝 추천 알고리즘 기반 편성 최적화 모델



- 유튜브 추천 알고리즘 모델에서 아이디어를 착안해 판매액 예측 모델을 기반으로 한 편성 추천 알고리즘 제시
- 후보 생성 모델을 통해 입력 데이터를 기반으로 학습 데이터 중 유사한 편성 후보군을 제시
- 랭킹 모델을 기반으로 후보군 중에서 매출액을 최대화 하는 방송 편성 추천 및 제안

출처: Deep Neural Networks for YouTube Recommendations 논문

- 딥러닝 추천 알고리즘 기반 편성 최적화 모델



- 현재 제공된 2019년도 이전의 편성 데이터들이 주어진다면 대량의 데이터를 기반으로 최적화 모델 성능 극대화 가능
- 해당 알고리즘을 기반으로 2020년도 매출액을 극대화 하는 편성 테이블, 편성 시나리오 제공할 수 있음

출처: Deep Neural Networks for YouTube Recommendations 논문



04

## 결과 및 기대효과

- 기대효과

## 포스트 코로나 기간의 NS SHOP+ 판매 실적 예측, 편성 계획 수립

- 비선형 회귀 모델을 활용해 2020년도 하반기, 포스트 코로나 기간의 상품 판매 실적 예측
- 포스트 코로나 시즌을 고려한 최적의 NS SHOP+ 판매 편성 계획 수립

## 개인화 맞춤형 T커머스 편성 알고리즘까지 영역 확장 가능

- 편성 별 전체 판매 실적 데이터와 함께 개인 구매 이력 데이터, IPTV 시청 데이터도 추가하여 빅데이터 분석 기반의 개인별 구매 예측 모델 개발
- 개인화된 예측 모델에 본 프로젝트에서 제시하는 편성 배치 알고리즘을 적용하여 개인 맞춤형 홈쇼핑 편성과 큐레이션 서비스 제공 가능