

# 「2020 빅콘테스트」 데이터 분석 계획서

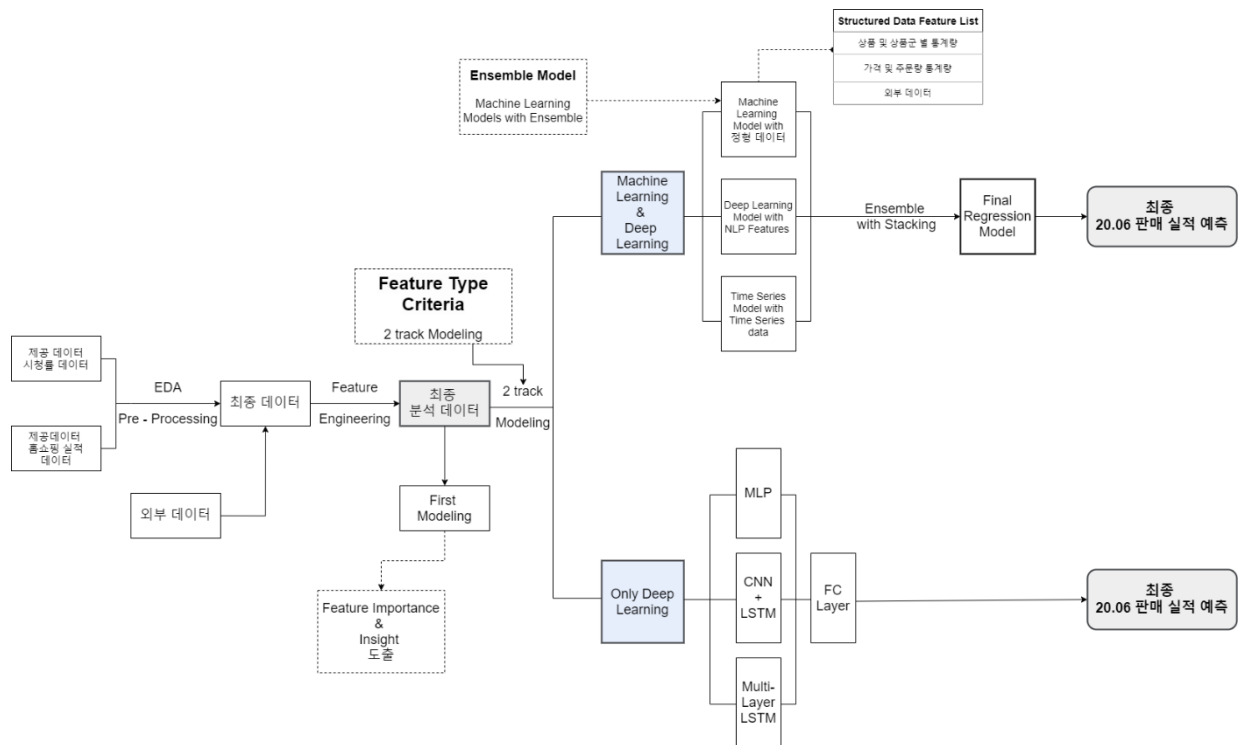
\* 해당란에 체크 표시

참가분야	<input type="checkbox"/> 혁신아이디어분야 <input checked="" type="checkbox"/> 데이터분석분야		
세부분야	<input type="checkbox"/> 퓨처스리그 <input checked="" type="checkbox"/> 챔피언리그 *데이터분석 분야에 한함		
개인/팀여부	<input type="checkbox"/> 개인 <input checked="" type="checkbox"/> 팀(구성원 4 명)	개인/팀명	쇼핑광고등어
대표ID	jinjin960318@gmail.com		

※ 5장 내외로 목차는 준수하여 자유롭게 작성

분석 주제명	NS SHOP+ 판매실적 예측을 통한 편성표 최적화 방안(모형) 도출 <데이터 특성을 고려한 융합 알고리즘 기반 취급액 예측 모형>
분석 배경	<ul style="list-style-type: none"> <li>◆ 현재 TV 시청률 하락, 온라인·모바일 쇼핑의 급성장 등 유통환경 및 소비 패턴이 변화함에 따라, TV홈쇼핑의 성장은 둔화되고 있는 실정</li> <li>◆ 따라서 짧은 시간의 TV 방송 시간을 효율적으로 활용하여, 소비자들의 제품 구매력을 높이는 것이 중요함</li> <li>◆ 내부요인 (NS SHOP+ 편성데이터) 및 외부요인 데이터 분석에 기반하여 소비자의 구매에 대한 패턴을 찾고, 최적의 수익률을 기대할 수 있는 편성모형을 설계함으로써, 방송편성 차별화를 통해 NS SHOP+ 만의 온디맨드 서비스 차별화를 실현시키고자 함</li> </ul>
분석 내용 요약	<p><u>Task1</u> NS Shop+ 방송편성표에 따른 판매실적 예측모델 개발</p> <p><u>Task2</u> 최적 수익을 고려한 편성표 추천 모델 설계</p>

## 분석방법 및 계획



### < Model Flow >

#### A. 분석 데이터

##### 1. 제공 데이터

- 실적데이터: 19년 1월 ~ 19년 12월 프로그램 별 실적 데이터
- 외부데이터: 19년 1월 ~ 19년 12월 요일별/시간대별 분 단위 시청률 데이터

##### 2. 수집 데이터

- 날씨데이터: [기상청] 일별 기온/강수량, [에어코리아] 미세먼지
- 경제지수 데이터

[KOSIS] 온라인쇼핑몰 판매매체별 상품군별거래액,  
 [KOSIS] 전국 소비자동향조사, [KOSIS] 소매업체별 판매액지수,  
 [KOSIS] 소비자/생산자물가지수, [KOSIS] 경기종합지수

- 사회과학적 특성에 민감한 유통 시장의 특성을 고려하여,  
경제 지표에 대한 통계청 지수를 Feature로 반영

---

## C. Modeling

### 1. **Task1:** NS Shop+ 방송편성표에 따른 판매실적 예측모델 개발

#### ✓ **2-track modeling**

#### - **1. Machine Learning & Deep Learning**

##### 1) Machine Learning Model with Structured Data

- ◆ Feature Engineering 과정을 통해 생성된 Train Dataset을 이용하여 모델링
- ◆ Random Forest, XGBoost, CatBoost, LightGBM 등의 다양한 머신러닝 모델을 적용, 앙상블을 통해 최종 결과값 예측
- ◆ **Time Series Split Cross Validation**  
유행 및 이슈에 따라 월별로 판매 경향이 달라지리라 판단하여,  
기준 달을 test set으로 두고, 기준 이전 달 까지를 train set으로 두어 모델 학습

##### 2) Deep Learning Model with NLP Features

- ◆ embedding 된 단어 벡터들을 input으로 하는 Neural Network를 통해 상품명으로 취급액을 예측

##### 3) Time Series Model with Time Series Data

- ◆ 판매액이 시계열적 특성을 따른다고 가정하고, Prophet 모형 적용  
(관련 논문: [Forecasting at Scale] <https://peerj.com/preprints/3190.pdf>)

- 4) 위의 세 가지 모델을 통해 예측된 값을 **stacking**하여,  
Final Regression Model 구축 후 최종 판매액 예측

#### - **2. Deep Learning**

##### 1) Multi-Input Model

가) 데이터 특성에 따른 딥러닝 구조 선택

###### ① NLP feature:

상품명 내의 단어 위치를 순차적으로 고려할 수 있는 CNN-LSTM Layer

###### ② Structured data: Numeric data의 특성을 살릴 수 있는 Fully-connected Layer

###### ③ Time Series data:

시계열 데이터의 긴 기간 의존성 문제를 해결할 수 있는 LSTM Layer

---

---

#### 나) Joint Representation & Regression Network

- ✓ NLP feature, Time Series data, structured data의 **특징을 추출한 vector**들을 하나의 **Joint Representation**(Concatenation/Sorting)으로 구성
  1. 결합적 방식(Concatenation):  
feature vector들을 **연속적**으로 이어 붙여 vector로 사용
  2. 정렬 방식(Sorting):  
각 레이어를 **시간순**으로 정렬한 후 vector로 사용
- ✓ Joint Representation된 vector를 input으로 하는 **Regression Network**를 통해 최종 취급액 예측

#### 2) Auto-Keras

주어진 테스크와 데이터셋에 최적화된 뉴럴 네트워크 아키텍처를 자동으로 구축해주는 NAS(Neural Architecture Search) 사용

#### 3) 월별 Cross Validation Fold

월별 특성을 고려하여 1 Epoch당 12번(1월~12월)의 교차검증 수행

### 2. Task2: 최적 수익을 고려한 **편성표 추천 모델** 설계

#### ✓ Recommendation System

- 1) 상품 별 판매 최적 시간대를 찾아, 판매액이 최대가 되게끔 추천해 주는 방식
- 2) LGBMRanker 모델을 적용하여 편성표를 Re-ordering 하는 방식으로 진행할 예정

#### ✓ Multi Label Model

- 1) 주어진 feature들로 (상품군, 판매단가)를 예측
- 2) 예측된 결과값을 통해 최대 수익을 고려할 수 있는 상품을 고려하여 편성표 구성

### D. 기타 사항

1. python의 matplotlib, seaborn 패키지 및 SHAP, HiPlot, tensorboard를 이용해 분석 결과에 대한 시각화
  2. github과 Notion을 통해 협업 및 버전 관리
-

**분석결과 활용 및  
시사점**

기대효과

- ◆ **취급액 극대화 모형 기반 최적 편성표 구축 활용 가능**
  - 시간대별 상품 및 취급액 예측을 통해 보다 구체적인 편성표 구축 가능
  - 날씨, 경제지표, 사회현상 등의 외부데이터를 기반으로 다양한 변수가 고려된 편성표 구축 가능
- ◆ **NS SHOP+ 고객들에게 개인화 된 맞춤형 서비스 제공**
  - 예측된 모형을 통해 상품 및 상품군의 특성을 파악
  - 추후 소비 패턴에 분석을 기반으로 고객을 세그멘테이션하여 차별화된 고객 경험 제공

ex) 고객 맞춤형 큐레이션 서비스: 소비 패턴에 따라 다른 콘텐츠 송출
- ◆ **높은 판매율 달성을 위한 데이터 기반 마케팅 전략 수립**
  - 판매에 중요한 영향을 끼치는 요인을 파악
  - 요인 기반 제품 소싱 및 판매 활성화 방안 수립

※ 제출자료는 최종 출품작 평가시 활용될 수 있음