|  |
| --- |
| **「2020 빅콘테스트」데이터 분석 계획서** |

|  |  |  |  |
| --- | --- | --- | --- |
|  | \* 해당란에 체크 표시 | | |
| **참가분야** | □ 혁신아이디어분야 ■ 데이터분석분야 | | |
| **세부분야** | □ 퓨처스리그 ■ 챔피언리그 | | |
| **개인/팀여부** | □ 개인 ■ 팀(구성원 4 명) | **개인/팀명** | 엄덕구 |
| **대표ID** | zhddhkdn@naver.com | | |

|  |  |
| --- | --- |
| **분석 주제명** | **Tree Based Ensemble과 Candidate-Rank-Formation 모델을 통한**  **NS홈쇼핑 판매실적 예측 및 편성 최적화 방안 제시** |
| **분석 배경** | * **시장 배경**   : 최근 마이크로 미디어(micro-media)와 오픈마켓의 성장으로, 홈쇼핑의 성장세가 둔화되는 한편[[1]](#footnote-1), 4년간 홈쇼핑 송출 수수료 인상률은 302%[[2]](#footnote-2)에 달하고 있는 상황이며, 이를 상쇄시키기 위한 매출 성장이 반드시 필요함.  : 또한 포스트 코로나19 시대와 같이 급변하는 시장 상황에 맞는 적절한 상품 기획 및 편성이 필요하며, 이러한 편성 개편 및 외부요인 분석을 통해 실적 개선한 사례가 발생하고 있음.[[3]](#footnote-3) 이와 같이 한정된 자원을 효율적으로 활용할 수 있는 상품 기획과 편성 최적화를 이루는 방안을 모색해야함   * **분석 목표**   : 이에 최신 데이터 분석 기법을 활용하여 시간대별 상품 판매실적을 예측하는 예측 모델과 이를 통한 편성 최적화 모델을 도출하고자 함. |
| **분석 내용**  **요약** | - 예측 모델링 프로세스는 CRISP-DM에 따름  - 예측 모델에 Tree Based Ensemble Model을 사용하여 최대한 다양한 변수를 수용하며, 과적합을 방지  - 2-Stage-Model에 Nelder-Mead Simplex Algorithm을 적용하여 적은 연산량으로 최적화 가능 |
| **분석방법 및**  **계획** | * **분석 방법론**  1. **예측모델: CRISP-DM 적용**     [Figure ] CRISP-DM Process Diagram**[[4]](#footnote-4)**   1. **Business Understanding:**  * 홈쇼핑 방송 편성에 대한 중요성이 점차 높아지고 있음 * 코로나19의 영향력으로 인해 고객의 소비 물품 및 패턴이 변화하고 있는 상황 * 내/외부 데이터를 활용한 매출품목 선정 및 방송 편성 최적화 방안이 필요  1. **Data Understanding:**  * 기술 통계량 확인 * 데이터 타입 및 구조 확인 * 독립변수 간의 관계 확인 * 독립-종속 변수 간 연관성 확인  1. **Data Preparation**  * 데이터 타입 변환 * 데이터 분포 변환 * 내부 데이터 활용 변수 생성 1: 시간 관련 변수 * 내부 데이터 활용 변수 생성 2: 제품 관련 변수 * 외부 데이터 활용 변수 생성 1: 날씨 관련 변수 * 외부 데이터 활용 변수 생성 2: 외부 쇼핑몰 관련 변수  1. **Modeling**  * Train Data Set / Test Data Set 변화 실험 * Tree Based 모델 / Ensemble 모델 적용 * 각 모델의 최적 Hyperparameter 도출  1. **Evaluation**  * 시각화를 통한 실제 값 / 예측 값 비교 * Cross-Validation 기반 최종 모델 평가 (MAPE 활용) * 최적 모델 기반 Test Data Set 예측  1. **편성 최적화 모델: two-stage 모델 + 최적화 모델 적용**     [Figure 2] Candidate-model, Rank-model, Formation-model   * Candidate-model, Rank-model, Formation-model 최적화 모델을 직렬로 배치하여 최소의 연산으로 최적의 편성표 출력  1. Candidate-model: MF(Matrix Factorization) 또는 Nelder-Mead Simplex Algorithm을 이용한 시간대별 n개의 카테고리 후보 추출 2. Rank-model: Candidate-model을 통해 추출된 카테고리 후보에 속한 상품 후보를 ‘예측모델’을 활용하여 상품별 취급액 예측 3. Formation-model: two-stage (Candidate-model~Rank-model)을 통해 추출된 상품별 취급액 총합이 최대가 되는 편성 출력  * **활용 외부 데이터**   1) 기상청 기상자료 개방포털 – 기온, 습도, 기압, 강수량 등  : https://data.kma.go.kr/cmmn  2) 빅카인즈 – ‘코로나’ 및 ‘재난 지원금’ 관련 일별 뉴스기사 건수  : <https://www.bigkinds.or.kr/>  3) 네이버 Open API |
| **분석결과 활용 및 시사점** | * 활용 방안   : 편성 상품에 대한 취급액 예측을 통해 편성 및 상품 기획에 대한 참고 지표로 활용  : 한정된 방송시간에 최적의 효율성을 기대할 수 있는 최적화된 편성표를 추출, 이를 편성에 대한 참고 자료로 활용   * 시사점   : 최신 데이터 분석 기법을 활용하여 예상 취급액을 예측 가능한 영역으로 끌어올림  : 예측 및 편성 최적화 모델을 통해 내∙외부 시장 변화에 적절히 대응할 수 있는 상품 기획 및 편성 가능 |

1. [ZD Net KOREA] '갑질'서 '백수오'까지…위기의 홈쇼핑 (2015.05.11) [↑](#footnote-ref-1)
2. (비즈니스 WATCH) IPTV 송출료 '수직상승'…홈쇼핑 '죽쒀서 남준다' (2020.02.20) [↑](#footnote-ref-2)
3. [뉴데일리경제] ‘코로나19 전략’에 희비 엇갈린 홈쇼핑…‘MD가 갈랐다’ (2020.05.11)

   [머니투데이] 마트·홈쇼핑이 '날씨' 분석하자, 매출이 뛰었다 (2020.08.05) [↑](#footnote-ref-3)
4. [그림1] CRISP-DM Process Diagram Image(Source: Kenneth Jensen) [↑](#footnote-ref-4)