**데이터 시각화 프로젝트 기획안**

기획안 작성일자 : 2021. 7. 7

|  |  |
| --- | --- |
| 팀 명 | 6조 : 그것이 알고싶다 |
| 프로젝트 주제 및 개요 | 국내 ott 기업 콘텐츠 유통 전략 수립 제안  – Netflix의 여러 독립변수들을 통해 viewpoint(시청자수)를 예측 |
| 프로젝트 수행 방향 | **Business Understanding** :   * Ott 플랫폼 중 점유율 1위에 위치한 Netflix의 컨텐츠를 분석 * 데이터셋의 빈도수 시각화를 통해 Netflix의 유통전략을 대략적으로 파악   **Data Understanding**  1. Netflix 관련 데이터 수집 후 필요한 데이터 셋 선별  - 선행 분석을 참고하여, 필요한 데이터셋과 필요 없는 데이터셋을 분류  2. Netflix 데이터 다각도 분석  - 장르, 감독, 연령 등 시청자수와 영향이 있을 것 같은 것을 다각도로 분석  - 분석 결과 시각화 및 상관관계 파악  **Data preparation**  - Kaggle\_netflix 데이터 컬럼 정리  - (DataSet) netflix-rotten-tomatoes-metacritic-imdb  - filxpatrol의 랭킹 크롤링  - Netflix 분기별 보고서 확인(선행 연구자료)  **Modeling**  **- viewPoint를 타겟변수로 하여 유요한 변수들을 통해 타겟변수를 예측하는 모델을 작성**  - Skill  - python 3.9.  - R  - mySQL  **Evaluation**   * ott산업의 발전으로 기존의 다양한 선행연구 자료 활용 : * 기존 기업 분석 보고서 및 캐글 분석 자료를 통해 본 프로젝트와 오차 범위 확인예정 * flixPtrol의 랭킹 자료의 viewPoint를 통해 모델의 정확성을 평가   **deployment**   * **한국의 ott플랫폼 기업이 콘텐츠를 제작거나 유통할 때에 해당 모델을 통해 viewPoint를 예측하여 유통전략에 사용** |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  팀장 : 채길호  - 프로젝트 총괄, sns 태그들과 상영시간의 viewPoint간의 상관관계 분석  팀원1 : 유화영  – 사용 언어와 tv 품질의 viewPoint간의 상관관계 분석  팀원2 : 천상희  – imdb 점수와 장르의 viewPoint간의 상관관계 분석  팀원3 : 김영성  – 넷플릭스 출시 일자를 통해 월과 년도로 분할하여 viewPoint간의 상관관계 분석 |
| 프로젝트 추진 일정 | -● 일정  6/18~ 7/2: 주제 후보 선정 및 주제 확정, 일정 조정  7/2 ~ 7/7: 선행 연구자료 및 분석 보고서 참조하여 프로젝트 분석 방향 확정  7/7 ~ 7/11: 각자 선택한 변수들과 타겟의 상관관계를 분석하여 결과를 취합  7/12 ~ 7/16: 상관분석 시각화와 유효한 변수 추출 및 모델링 작업 7/17 ~ 7/19: 보고서 작성 및 발표자료 준비  7/20 : 발표 |