| 프로젝트 기획서 |
| --- |

Deep Learning 활용하여 자연어 처리를 통한 VOD 추천 서비스

2022.10.31

광주인공지능사관학교 언어지능반 코람페 팀

1. **요약**

| **팀명** | 코람페 ( 코딩해서 람보르기니 페라리 사자 ) |
| --- | --- |
| **팀원 및 역할** | * 팀장: 이재혁 (프로젝트 총괄, 일정계획) * 팀원: 손재형, 이민흠, 임철성, 주성민, 정우현 |
| **아이디어 주제** | Deep Learning 활용하여 자연어 처리를 통한 VOD 추천 서비스 |
| **제안 배경 및 필요성** | * 제안 배경 * 스트리밍 서비스의 영상 개수가 많아져서 영상들의 정보를 찾아가며 본인의 취향의 영상을 찾기 힘들어졌다. 장르나 ‘평점이 높다’와 같은 키워드로 본인의 취향에 맞는 영상을 찾도록 도와준다. * 필요성 및 차별성 * 일상의 제한된 시간 속에서 내 취향에 맞는 VOD를 빨리 찾는 것은 쉬운 일이 아니다. 키워드를 통한 VOD 추천을 해주고, 선택한 VOD와 다른 VOD의 유사성을 찾아 비슷한 장르의 VOD를 추천해준다. * 추가로 구글 캘린더 정보와 위치 정보를 얻어 상황에 따른 VOD 추천 |
| **선행 기술 및 논문 현황** | * 유비 쿼터스 환경에서의 상황인식 기반 VOD 컨텐츠 추천시스템 * <https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE08993580> * 영상의 유사도 측정을 위한 키워드 추출 * <https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE10490474> |
| **제안 내용** | * 개발 목표 * 키워드를 통한 VOD 추천 * 개발 내용 * 리뷰 데이터를 통한 영화 선호도 키워드 추출 * 영화 기본 정보 데이터를 통한 장르 키워드 추출 |
| **수행 방법** | * 데이터 확보 * 네이버 영화 리뷰 크롤링 * KOFIC 영화관 입장권 통합전산망 |
| **기대효과 및 활용방안** | * 기대효과 * VOD 선택에 있어 정보를 제공해주어 고르는데 시간을 줄여주고, 만족도 증가를 통한 재사용률 증가 * 활용방안 * VOD 스트리밍 어플리케이션(ott 서비스)에 기능 추가 |

## 

### 2. 제안배경 - 외부 환경 분석(PEST / STEEP)

| **Political**  **(정책적 배경)** | * ott 서비스 성장을 위한 국가적 지원 * 한국건강증진개발원의 음주, 흡연 장면에 대한 규제 권고 * 해당 장면들이 들어간 영상에 대해 키워드로 분류하여, 악영향을 미칠 수 있는 대상에게 추천하지 않는 등의 관리 가능 |
| --- | --- |
| **Economic**  **(경제적 배경)** | * 코로나19 장기화로 ott 유료 이용률 대폭 증가 * 기존 서비스 뿐만 아니라 새로운 서비스 출시 * 다른 서비스와는 차별화된 기술 필요 |
| **Social**  **(사회적 배경)** | * 코로나19로 인해 실내활동 선호 (영상 콘텐츠 시청) * 동영상 서비스 플랫폼이 개인의 관심사를 인정함 |
| **Social**  **(트렌드 배경)** | * 영상 시청에 있어서 제약이 없어진 현재에 짧은 시간 안에 원하는 영상을 보기 원 |
| **Technological**  **(기술적 배경)** | * 자연어처리 기술 ‘워드 임베딩’을 통해 중복 단어 키워드 추출 |

### 

### 3. 제안배경 - 내부환경 분석(3C)

| **Capacity**  **(제안자 능력)** | * 데이터 수집 및 분석 가능 * 많은 양의 데이터를 수집하고 분석하며 원하는 정보를 추출할 수 있음 * 분류 모델 구축 가능 * 머신러닝, 딥러닝, 오픈 API 등을 활용해 데이터를 분류하는 모델 구축 가능 |
| --- | --- |
| **Competition**  **(경쟁제품/기술/**  **특허분석)** | * 넷플릭스 * 컨텐츠 소비 기록, 영상 되감기 기록 등으로 소비하는 컨텐츠의 75%가 추천 알고리즘에 의한 컨텐츠 * 유튜브 * 영상에 대한 조회수와 시청 시간에 따른 알고리즘 생성 |
| **Customer**  **(고객 분석)** | * 영상을 찾는데 많은 시간을 쏟지 않길 원함 * 내가 원하는 영상을 빠른 시간 안에 찾길 원함 |

### 4. 수요분석

| **필요성** | * 영상 시청에 있어서 트렌드가 과거와 달라짐 * 원하는 콘텐츠 소비를 위해 방영 채널과 시간을 기다리는 과거에서 * 원하는 콘텐츠를 원하는 시간과 장소에서 소비 |
| --- | --- |
| **차별성** | * 콘텐츠 소비 기록이나 조회수, 시청 시간 등은 소비자가 콘텐츠를 사용해야 남는 것이다. * 소비자 정보를 사용하는 기존의 방법과는 다르게 영화에 대한 정보를 활용하여 키워드 추출 |
| **현재까지 준비사항** | * 데이터 수집을 위한 사이트 정리 * 자연어 처리 기술 * 활용할 분석기법 : 딥러닝, 워드 임베딩, 웹크롤링 |
| **기대효과** | * 이미 기존의 고객이 많은 플랫폼에 새 고객 유치 가능 |
| **활용방안** | * VOD 스트리밍 서비스나 영상 플랫폼의 검색 기능 활성화 |

### 

### 5. 개발내용

| **개발 목표** | * 유사도 비교를 통한 VOD 추천 |
| --- | --- |
| **개발 내용** | * 선택한 VOD에 대해 높은 유사도를 가진 VOD 추천 * VOD 기본 정보를 통한 키워드 추출 * 사용자의 위치 정보를 통한 VOD 추천 |

### 6. 달성 목표 및 달성 전략

| **달성 목표** | * 서비스 소비자가 선택한 VOD에 대해 유사성이 높은 키워드를 추출하여 같은 키워드를 가진 VOD를 추천해준다. |
| --- | --- |
| **달성 전략** | * 데이터 구축 * 웹크롤링을 통해 네이버 영화에 있는 영화 리뷰 가져오기 * KOFIC에 있는 영화 기본 정보 데이터 가져오기 * 데이터 전처리 및 인공지능 모델 개발 * 워드 임베딩을 통해 영화에 대한 리뷰에서 높은 빈도의 단어 추출 * 영화 장르와 줄거리 정보를 통해 영화 추천 기준 설정 * 인공지능 모델 성능 향상 * 인공지능 모델 예측값의 여러가지 지표 향상 (정확도, 평균제곱오차 등) * 인공지능 모델의 목표 수준 달성도 측정 |

### 

### 

### 7. 개발 일정

| 추진내용 | | **제1주차**  (1차 스프린트:  기준 모형 구축) | | | | | | | **제2주차**  (2차 스프린트:  개선 모형 구축) | | | | | | | **제3주차**  (3차 스프린트:  최종 모형 구축) | | | | | | **제4주차**  (발표준비) | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | 14 | 15 | 16 | 17 | 18 | 21 | 22 | 23 | 24 | 25 | 24 | 28 | 29 | 30 | 1 | 2 | 5 | 6 | 7 | 8 | 9 | 12 | 13 |
| 사용자  스토리 작성 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 백로그 작성 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 프로그램  설계 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 데이터 수집 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 데이터 전처리 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| AI 모델링  및 리뷰 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 스프린트 회고 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 프로젝트 종결 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 발표 준비 및 발표 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

❍ 백로그 작성: 프로덕트 백로그(제품 요건 분석서) 작성 + 스프린트별 백로그(중간품 요건 분석서) 작성

❍ 일일 스크럼 회의 및 회의록 작성 - 일정 진행 전 5분에서 10분

❍ 일일 리뷰 회의(검토회의) 및 회의록 작성 - 당일 작업 검토 및 다음 날 목표 설정 20분

### 8. 참여 인원

| **이름** | **역할 및 능력** |
| --- | --- |
| **이재혁** | * 프로젝트 총괄, 데이터 추가 수집, 필요 문서 작성 |
| **손재형** | * 데이터 추가 수집, 데이터 분석(머신러닝, 딥러닝) |
| **이민흠** | * 데이터 추가 수집, 데이터 분석(머신러닝, 딥러닝) |
| **임철성** | * 데이터 추가 수집, 데이터 분석(머신러닝, 딥러닝) |
| **주성민** | * 데이터 추가 수집, 데이터 분석(머신러닝, 딥러닝) |
| **정우현** | * 데이터 추가 수집, 데이터 분석(머신러닝, 딥러닝) |

### 

### 9. 차별성 분석

| **기존 시장 방식과 차별점** | * 기존 시장 방식 * 내용기반 필터링 모델과 콘텐츠 기반 필터링 모델을 이용해 시청자들이 선호할 가능성이 높은 비슷한 범주 데이터들을 엮어, 분석 대상 콘텐츠들 마다 고유의 특성파일을 구성한 다음, 기존 시청자의 콘텐츠 선호 정보를 분석해 타깃 사용자와 유사한 성향과 선호도를 가진 것으로 추정되는 기존 사용자 그룹을 가졍하여 VOD를 추천 * 차별점 * 기존 방식과 달리 실시간 날씨 정보 및 사용자의 위치정보 데이터를 분석하여 그에 맞는 VOD를 추천 * 사용자의 성향과 선호도를 베이스로 하여 그 날의 날씨 또는 사용자의 위치 정보를 토대로 추천을 해준다면 기존 방식의 한계점을 보완할 수 있음 |
| --- | --- |

### 10. SWOT 분석

| **S** | **W** |
| --- | --- |
| * 실시간 날씨 정보를 수집하여 날씨와 어울리는 VOD 추천 * 시청자의 위치정보를 이용한 VOD 추천 | * 영화 정보 및 리뷰 데이터 크롤링에 많은 시간 소요 예상 * 사용자의 성향과의 유사도 또는 연관성을 계량화(지표)를 명확하게 보여주기 어려울 수 있음 |
| **O** | **T** |
| * 기존 방식은 명확하지만 조금은 아쉬운 추천 * 항상 사용자의 성향에만 맞는 VOD 추천시 가끔은 새로운 VOD를 원하는 사용자들에게 신선함을 줄 수 있음. | * 기존의 방식을 잘 사용하고 있는 업체 또는 회사들을 새 고객으로 유치하기 쉽지 않을 수 있음 |

### 

### 11. 마케팅 믹스(4P 전략)

| **Product**  **(제품)** | * 워드임베딩을 통한 유사단어 인식, 위치, 날씨 정보를 이용한 VOD 추천 서비스 |
| --- | --- |
| **Price**  **(가격)** | * 추후 협의 필요 * 정기적으로 데이터를 업데이트하여 시간이 지날수록 추천의 정확도가 올라간다면 구독을 유지 시킬 수 있을 것으로 사료됨 |
| **Place**  **(판매 방법,**  **유통 경로)** | * 진입기 * 홈페이지 개설 * 사용자들의 성향을 알기 위한 설문조사 실시 * 안정기 * 기존 방식의 서비스에 날씨 및 위치정보를 고려해 추천해주는 제품의 특성을 바탕으로 온라인 홍보 * 기 사용자들의 추천을 통해 신규 사용자 확보 * 확산기 * OTT 서비스를 이용중인 사용자들의 유입을 통한 서비스 확대 * 국내 OTT 서비스를 운영중인 회사에게도 판매 |
| **Promotion**  **(홍보 방안)** | * 홈페이지 * 앱 * 길거리 판촉물 * 할인 프로모션 진행 (첫 구독시, 1년 10% 할인 등) |

### 12. 향후 발전 방향

| **구분** | **발전 방향** |
| --- | --- |
| **기술** | * 구글 캘린더 연동을 통한 사용자의 일정을 고려한 VOD 추천 |
| * 서비스 시작 일정기간 후, 사용자 설문을 실시하여 불편사항 개선 |
| * 워드임베딩을 이용한 추천 알고리즘 보다 추천 서비스의 질을 개선할 수 있는 새로운 딥러닝 모델이 나온다면 새로 적용 |
| **사업** | * 일반 사용자를 목표로 시작하여 추후에는 국내 OTT서비스 회사들에게도 판매 |

### 

### 

### 

### 

### 13. 참고문헌

* SWOT:

<https://asana.com/ko/resources/swot-analysis>

* 국내 OTT 서비스의 추천 시스템 현황 논문:

<http://koreascience.kr/article/CFKO201920461759142.pdf>

* 인공지능 맞춤 추천서비스 기반 온라인 동영상(OTT) 콘텐츠 제작 기술 비교 논문::

<http://koreascience.or.kr/article/JAKO202119060144548.pdf>

* 코로나19 장기화로 OTT 유료 이용률 50% 이상 넘었다:

<http://www.mediatoday.co.kr/news/articleView.html?idxno=304427>

* ‘오징어 게임’, ‘우영우’ 이을 OTT 콘텐츠 키운다

<https://gonggam.korea.kr/newsContentView.es;jsessionid=M3q-HO3PcloI1bYpWpG1a1s2.weekly0901?mid=a10201000000&section_id=NCCD_POLICY&content=NC002&news_id=EBC6D4014DAC4203E0540021F662AC5F>

* 유튜브 알고리즘의 선택을 받으려면?

<https://scent.kisti.re.kr/site/main/archive/article/%EC%9C%A0%ED%8A%9C%EB%B8%8C-%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98%EC%9D%98-%EC%84%A0%ED%83%9D%EC%9D%84-%EB%B0%9B%EC%9C%BC%EB%A0%A4%EB%A9%B4>

* OTT 영화, 청소년 관람가에도 음주·흡연 너무 많아

<https://www.healthinnews.co.kr/news/articleView.html?idxno=33293>

* 코로나 19 이후 일상생활 변화

<http://www.kcti.re.kr/webzine2/webzineView.action?issue_count=111&menu_seq=7&board_seq=1>

* OTT와 위드 코로나시대, 방송 제작 현장의 변화

<https://www.kocca.kr/trend/vol28/sub/s13.html>'

* 급변하는 국내 OTT 시장 동향

<https://spri.kr/posts/view/22907?code=industry_trend>

* [Netflix 서비스의 핵심, 추천 알고리즘과 이를 둘러싼 특허 분쟁 가능성](http://blog.gooddaylab.com/2016/02/netflix.html)

<http://blog.gooddaylab.com/2016/02/netflix.html>

* 'Z세대는 왜 유튜브와 OTT를 선호할까?', Z세대를 사로잡은 콘텐츠

<http://www.casenews.co.kr/news/articleView.html?idxno=11127>