|  |
| --- |
| **주제: 패션관련 제품에 대한 평가 종합적 평가 및 추천 사이트 구축**  **나반, 7팀, 20192884 강영민 나반, 7팀, 20192932 최원석** |

|  |  |
| --- | --- |
| **2. 요약**  기존 제품을 구매하는 과정에서 겪는 복잡성과 불편함을 해결하기 위해 이 프로젝트를 시작을 했다. 여러 사이트에 들어가 제품에 대한 정보를 추함하고 추합한 정보들에 대한 신뢰성을 판단하는데 우리는 많은 시간을 낭비하고 있다. 우리는 신뢰성을 판단 해 사용자가 원하는 제품에 대한 정보를 자동적으로 추합을 해준후 제품에 대한 중요한 내용을 알려주며 또한 해당 제품과 비슷한 제품들을 추천도에 기반해 추천을 해주는 사이트를 만들고자 한다. 이를 통해 앞으로 우리가 제품을 구매를 할때 여러 사이트(유튜브, 다양한 판매 사이트들)를 탐색하며 정보를 판단하고 추합하지 않고 오직 한 사이트(Full-review)를 통해 복잡하지 않고 간단하게 정확한 정보를 얻어 합리적으로 내가 원하는 제품을 판단해 구매할 수 있도록 한다. | **3. 대표 그림**  그림 1. 구축한 Full-review 사이트  개발 배경: 기존 제품을 구매할때 이 제품이나에게 필요한가, 구매할만한가에 대한 것을 평가하는 방식이 복잡하고 불편하다. 즉, 다른 유저 반응 탐색, 제품 스펙 확인, 유튜버 평가 확인 등을 하는 시간이 오래걸린다.  예상 결과: 다양한 유저 정보와 유튜버의리뷰 정보를 확보한 Full-review를 통해 합리적인 구매를 신속하고 정확하게 할 수 있다. |

|  |
| --- |
| **4. 서론**  많은 사람들이 제품을 구매할 때 다양한 정보 소스를 활용한다. 특히 요즘에는 두 가지 경로 온라인 쇼핑몰 사이트와 유튜브 영상으로 정보를 얻는다. 이들 경로를 통해 소비자들은 제품에 대한 자세한 정보, 사용 후기, 평가 등을 얻을 수 있으며 이러한 정보는 구매 결정 과정에서 중요한 역할을 하고있다.  이러한 정보들은 때때로 제품 사용 시 발생할 수 있는 불편함이나 문제점에 대해 충분히 설명하지 않을 수 있으며 또한 잘못된 정보와 리뷰의 신뢰성등의 문제로 소비자들은 리뷰와 평가를 비판적으로 분석하고, 가능하다면 다양한 소스에서 정보를 수집하여 종합적인 평가를 하는 것이 중요해지고 있다.  제품을 구매할 때, 다양한 소스에서 정보를 수집하는 것이 불가피 하지만 다양한 온라인 사이트를 방문하며 정보를 수집하는 과정에서 불필요한 복잡성과 혼란을 겪고 있으며 각 사이트마다 제공되는 정보의 양과 품질이 상이하다, 이로 인해 어떤 정보가 신뢰할만하고 어떤 것이 그렇지 않은지를 판단하는 데 어려움을 겪고 있다.  또한, 각 사이트는 긍정적인 리뷰를 강조하려는 경향이 있어 과장된 정보나 편향된 의견을 식별하는 것이 힘들어지고 있다. 이로 인해 소비자로서 정확하고텍스트, 폰트, 라인, 번호이(가) 표시된 사진  자동 생성된 설명텍스트, 웹사이트, 웹 페이지, 온라인 광고이(가) 표시된 사진  자동 생성된 설명  그림2. 무신사 리뷰 사례  신뢰할만한 정보를 얻기 위해서는 더 많은 노력과 시간을 투자해야 하는 상황이다. 좀 더 나은 정보를 찾기 위해 어떤 사이트를 신뢰할 수 있는지, 어떤 평가 기준을 활용해야 하는지에 대한 명확한 지침이 부족하면서, 이로 인해 제품 선택에 대한 불안이 더해지고 있다. 이러한 어려움을 극복하고자 제품에 대한 정보를 탐색할 수 있는 여러 사이트들을 통해서 정보를 취합해 소비자들의 불필요한 복잡성과 혼란을 겪지 않을 수 있는 제품에 대해 신뢰할 수 있는 정확한 정보 제공해주는 서비스를 만들어보자고 생각을 했다. |

|  |
| --- |
| **5. 본론**  텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진  자동 생성된 설명  먼저 핵심 기능 설명이다. 주 기능은 제품에 대한 총 요약, 유사도, 추천도를 제공하는 사이트이다. 먼저 요약에 대한 설명을 하자면, 사용자가 제품명을 입력하고 추천 제품 및 유사 제품을 받기 전에 모든 제품군에 대하여 DB에 크롤링해서 얻은 정보를 저장한다. 해당 정보들은 다음과 같다. [제품명, 카테고리, 주 소재, 리뷰 수, 이미지\_URL, 리뷰 별점, 리뷰 Comment] 제품명을 특수문자나 모델명과 같은 의미없는 글자는 text cleaning을 거치고 정제된 제품명을 통해 Google에서 제공하는 Youtube Data api V3[[1]](#footnote-1)의 쿼리로서 입력된다. Response로 제품에 대한 Video\_id를 저장한다. 저장된 Video\_id는 다시 youtube-transcript-api 0.6.1[[2]](#footnote-2)를 사용하여 비디오와 관련된 자막 데이터(사용자 입력, Youtube 자동자막)을 추출한다. 추출된 자막을 Open AI사의 ChatGPT api gpt-3.5-turbo[[3]](#footnote-3)를 사용하여 세가지 지표로 요약한다. 세가지 지표는 다음과 같다.[사이즈 추천, 스타일링 제안, 제품 고유 특징] 이러한 정보를 최대 10개의 영상에서 추출하여 공통된 내용만을 다시 요약하여 정보에 대한 신뢰도를 상승시키는 효과를 가져온다.  다음은 유사도에 관한 설명이다. 유사도는 다음 4가지 지표를 통해 평가가 진행된다. [제품 고유 핵심 키워드, 이미지 유사도, 소재적 유사도, 카테고리적 유사도] 제품 고유 핵심 키워드는 앞선 요약의 제품 고유 특징에서 재활용하여 제품에 관한 특징을 gpt-3.5-turbo3를 통해 추출했다. 해당 키워드, 소재, 카테고리는 다시 hugging face의 KoSimCSE-roberta[[4]](#footnote-4)라는 문장 유사도에 관해 사전 학습된 모델을 통해 각각 따로 다른 제품들과의 유사도를 평가한다. 소재와 카테고리는 일단 단어적 유사도만 측정했으나, 추후 패션과 관련된 도메인 전문과와의 협업을 통해 실질적인 의미에서의 유사도를 벡터화시키고 이를 코사인 유사도로 측정할 계획이다. 이미지적 유사도는 미리 크롤링된 이미지\_url을 가지고 VGG16 모델로 imagenet을 pre-train한 모델을 통해 각 이미지의 Feature를 Extract한다. 그렇게 미리 각 제품당 고유 feature파일을 만들어 넣고 DB에 저장해 두고 유저가 제품명을 입력하게 되면 각 유사도지표에 설정된 가중치를 활용해 총체적 유사도를 출력한다. 100제품을 유사도 기준으로 정렬을 하고 상위 30 제품에서 다시 추천도를 기준으로 정렬을 한 뒤 유사 제품으로서 출력된다.  마지막으로 추천도이다. 추천도는 앞서 크롤링과 가져온 자막을 통해 진행된다. [리뷰 별점, 리뷰 comment, 리뷰 개수, 영상 자막] 영상 자막과 리뷰 comment에 대해서는 jaehyeong의 koelectra-base-v3-generalized-sentiment-analysis[[5]](#footnote-5)모델을 통해 감성평가를 진행했다. 나머지 특징들 또 한 각 점수를 0과 1 사이로 정규화를 시켰다. 총체적으로 모두 더해서 0~1 값이 나오도록 설정했으며 각각의 특징들에 대해 가중치를 다르게 가져가 추천도를 계산했다.  위의 알고리즘은 이렇게 마무리가 되었고 밑은 시스템 개요도 이다. 사용자의 입력은 검색창 입력을 통해 Nginx 웹 서버로 전달된다. Nginx는 정적 파일에 대한 요청을 바로 응답하거나, 동적 콘텐츠에 대한 요청을 Gunicorn으로 전달한다. Gunicorn은 요청을 받으면 이를 Flask 웹 애플리케이션에 전달하며, Flask는 요청을 처리하기 위해 필요한 로직을 실행하고, YouTube Data API v3에서 동영상 관련 데이터를 가져온 후 앞서 설명했던 방식으로 gpt api를 통해 추천과 요약을 진행한다. 그리고 이에 관한 것들을 MySQL 데이터베이스에서 정보를 조회하거나 저장된다. 처리가 완료된 후, Flask는 결과를 Gunicorn에게 반환하고, Gunicorn은 이를 다시 Nginx에게 전달하며 마지막으로, Nginx는 처리된 결과를 사용자에게 보내준다. 이 모든 시스템은 AWS Lightsail에서 호스팅되며, 안정적인 인터넷 연결과 리소스를 제공받아 운영된다.  텍스트, 스크린샷, 폰트, 도표이(가) 표시된 사진  자동 생성된 설명 |

|  |
| --- |
| **6. 결론**  저희는 제품을 구매하기 위해 특히 패션과 관련된 제품을 구매하기 위해서 많은 사전 정보를 필요로 한다. 사용자는 제품에 대한 스펙을 보고 다른 유저의 평가와 도메인적 지식이 꽤 있는 유튜버의 평가 그리고 제품의 고유 특징들을 알아보고 다른 사이트를 들려가며 비슷한 제품이 있는가를 탐색한다. 이러한 피곤함을 없애기 위해 저희는 요약, 추천, 유사제품을 제공하는 사이트를 만들었고 계속해서 이를 업그레이드할 계획이다.  다음은 향후 추가 및 향상시킬 기능들에 대해서 설명하겠다.   * **입력**: 현재 입력은 신발 제품명으로만 받고 있지만 추후에는 모든 제품군에 대해서 입력을 받고 사진으로도 입력을 받을 수 있게 할 계획이다. * **유사도**: 소재와 카테고리적 유사도를 측정함에 있어서 그 소재의 특성과 용도적 특성을 잘 구별해야하기에 도메인 전문가와의 협업을 통해 해당 특징에 대한 유사도를 측정할 것이다. * **추천도**: 현재는 영상에서 음성에 대한 자막에 관한 감성평가만 진행했지만 추후 OpenAI 사의 Whisper를 사용하는 방향으로 음성적 지표(어조적지표)를 활용하는 방향으로 향상시킬 예정이다. * **유저반응**: 추천된 결과에 대해서는 유저의 [좋아요, 싫어요] 반응을 통해 피드백을 받고 특성에 대한 가중치를 조절해 나갈 생각이다. * **DB 및 서버**: MySQL에서 확장성과 성능을 고려하여 Oracle Database을 사용할 계획이고 서버는 높은 내구성, 가용성, 무제한의 스토리지 공간을 고려하여 Amazon S3를 사용할 계획이다. |

**출처 – 각주 참고**

1. Google Youtube Data API : <https://developers.google.com/youtube/v3/getting-started?hl=ko> [↑](#footnote-ref-1)
2. youtube-transcript-api 0.6.1 : <https://pypi.org/project/youtube-transcript-api/> [↑](#footnote-ref-2)
3. Chat gpt api : <https://platform.openai.com/docs/api-reference/authentication> [↑](#footnote-ref-3)
4. BM-K/KoSimCSE-roberta : <https://huggingface.co/BM-K/KoSimCSE-roberta> [↑](#footnote-ref-4)
5. 감성평가 모델: <https://huggingface.co/Copycats/koelectra-base-v3-generalized-sentiment-analysis> [↑](#footnote-ref-5)