

7조 Term Project 계획서

- 패션 트렌드 추천 시스템 -

7조

60190312 김동신

60200621 남보라

60202464 원도현

60221313 윤윤아

60221324 홍건우

명지대학교 빅데이터프로그래밍(5999)

목차

1. 개요
 - a. 주제
 - b. 목표
2. 실행 계획
3. 기대 효과
4. 확장 가능성

1. 개요

- a. 주제
무신사 데이터를 기반으로 한 패션 트렌드 추천 시스템 구축
- b. 목표
 - i. 무신사 실시간 상품 데이터를 통해 현재 패션 트렌드에 적합한 추천 상품 리스트 제공
 - ii. 사용자가 특정 키워드(예: "검은색 후드티")를 입력하면, 해당 조건에 부합하는 추천 상품 리스트 제공

2. 실행 계획

- a. 웹 스크래핑을 통해 무신사 랭킹 데이터 수집
 - i. Python의 **request**, **BeautifulSoup** 등을 통해 데이터 추출
 - ii. 추출 데이터: 상품명, 브랜드, 가격, 카테고리, 색상, 조회수, 평점 및 리뷰 등
 - iii. **Cron Job** 등을 통해 데이터를 주기적으로 수집
- b. 수집된 데이터 정제
 - i. 불필요한 데이터 필터링
 - 1. 광고성 데이터 제거
 - 2. 중복 데이터 제거
 - ii. **Csv** 파일로 로컬 및 **HDFS**에 데이터 저장
 - iii. 주기적으로 추가 수집된 데이터 **HDFS**에 저장
- c. 데이터 분석 및 모델 학습
 - i. 데이터 전처리
 - 1. 색상, 브랜드, 카테고리 등 범주형 데이터 **One-Hot encoding**
 - 2. 수치형 데이터 정규화
 - 3. 리뷰 데이터는 형태소 분석을 통해 주요 키워드를 추출하여 **TF-IDF**로 벡터화.
 - ii. 모델 학습
 - 1. 벡터화된 리뷰 데이터 모델(**Random Forest**, **Logistic Regression** 등) 학습에 사용
 - 2. 수치형 데이터는 **Random Forest** 모델로 학습
 - 3. 모델 통합: 개별 모델의 예측 점수를 가중 평균 방식으로 통합
- d. 시스템 구축
 - i. 현재 트렌드에 맞는 상품 리스트 출력
 - ii. 사용자는 구매하고자 하는 상품 정보(특징)를 텍스트로 입력
 - iii. 해당 조건에 부합하는 추천 상품 리스트 제공

3. 기대 효과

소비자는 구매하고자 하는 상품의 특징을 입력하면 현재 트렌드에 적합한 상품을 빠르게 확인하고 구매 결정을 내릴 수 있다. 또한 브랜드와 마케팅 관점에서 트렌드 분석 데이터를 활용할 수 있다.

4. 확장 가능성

무신사 웹 사이트의 랭킹 정보는 실시간부터 최대 1달간 랭킹 정보를 제공한다. 따라서 프로젝트 진행 시점(2024.11 ~ 2024.12)에는 가을부터 겨울옷까지의 랭킹 정보만 수집 가능하다. 하지만 주기적으로 데이터를 수집하여 데이터가 축적되면 장기 트렌드 예측 시스템으로 확장 가능하다.