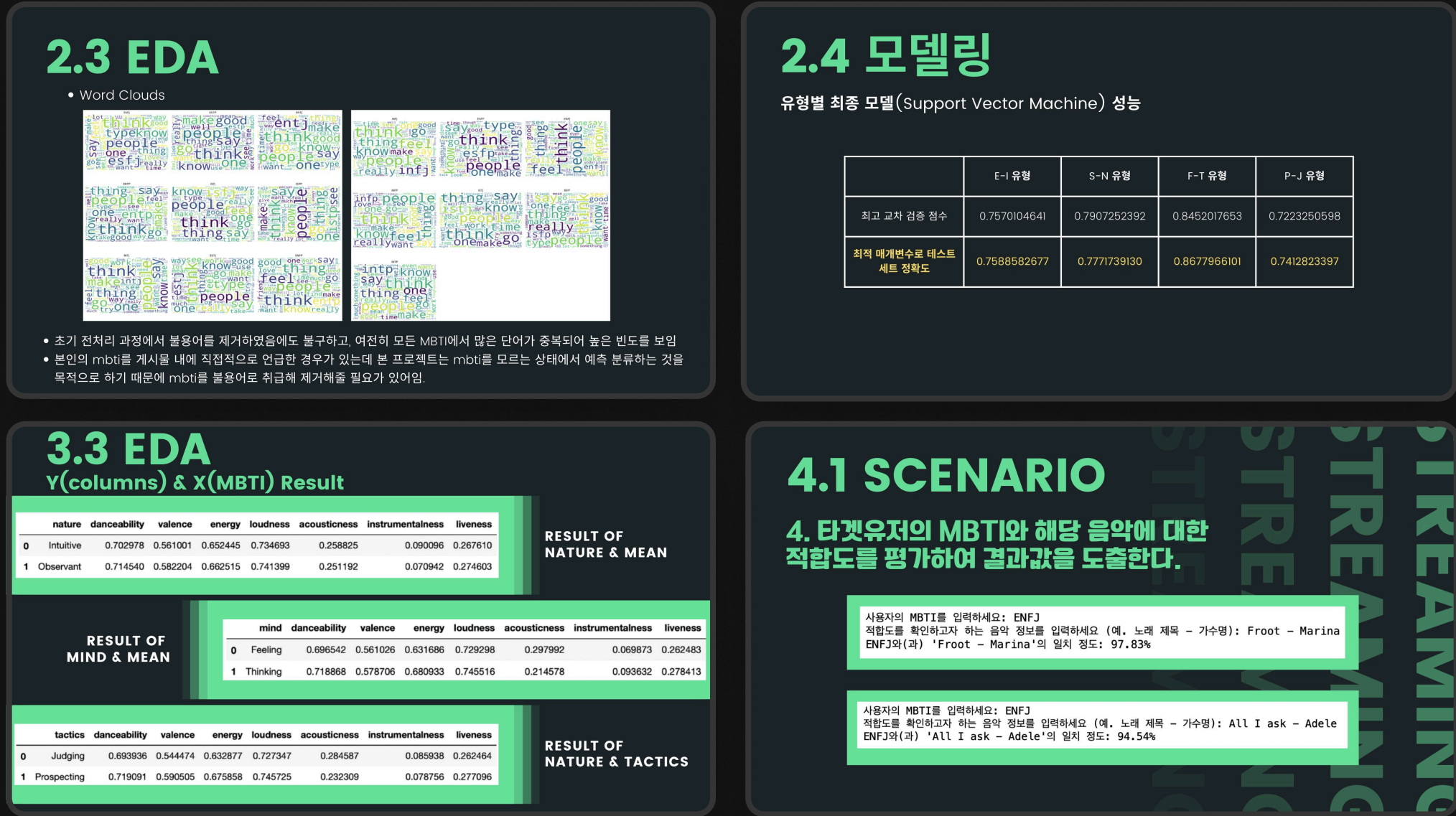


[전공] 스마트 AI 시스템

텍스트마이닝 기반 MBTI별 노래 추천



2023.11-2023.12
3인 프로젝트/ 데이터 수집, 분석, 모델링 / PYTHON
기여도 40%

분석 배경

- 최근 서비스 동향에서 유저의 성향 및 선호 상품을 파악할 수 있는 방법이 앱 내 활동에 대한 분석으로 한정됨.
- 간편가입과 같은 절차 간소화에 의해 과거처럼 온보딩이 상세하게 이루어지지 않으며, 유저 정보 파악이 어려움.
- 로그 데이터 뿐만 아니라 ‘SNS 내 텍스트를 통해 유저의 성향을 확인할 수 있을까?’ 라는 접근을 하게됨.
- 파악된 성향을 바탕으로 기존 상품 추천 서비스의 고도화를 이룰 것이라는 가설을 세우고 프로젝트를 진행함.

수행과정 및 인사이트

MBTI 16유형 전체를 분석 대상으로 삼기에는 데이터 불균형, 해석 혼란 등의 문제가 있었고, 이를 해결하기 위해 E/I, S/N, T/F, P/J라는 4가지 독립 척도로 나누어 분석했습니다. 이 방식은 특히 SNS 텍스트 기반 감성 분석과 결합했을 때, 유형별 언어 사용 패턴과 감정 표현 방식의 차이를 더 명확히 포착할 수 있었습니다.

NLP 기반 감성 분석과 키워드 추출을 통해 데이터를 전처리했고, 이후 다양한 모델을 실험한 결과, 최종적으로 SVM(Support Vector Machine)을 선택했습니다. 딥러닝 기반의 MLP에 비해 학습 데이터가 많지 않고 텍스트 임베딩이 단순한 상황에서, SVM은 오히려 더 안정적인 성능을 보였습니다. 특히 SVM은 고차원 텍스트 데이터에 적합하고, 명확한 분류 경계를 설정할 수 있어 MBTI 척도처럼 기준이 모호한 이진 분류 문제에서도 과적합 없이 분류 기준을 명확히 하는 데 유리했습니다.

이러한 결과는 모든 유형을 일괄적으로 반영하는 기존 성향 분석 접근보다, 서비스의 목적에 맞는 척도만 선택적으로 활용하는 전략이 더 정밀하고 효과적일 수 있다는 인사이트를 제공했습니다.