

Data Analysis

[전공] 스마트 AI 시스템 텍스트마이닝 기반 MBTI별 노래 추천

3인 팀 프로젝트/ 텍스트 기반 MBTI 분류모델 개발 / ML(파이썬)

인사이트 및 결론

MBTI의 16가지 유형을 4가지 독립적 척도(E/I, S/N, T/F, P/J)로 간소화함으로써 분석의 정확도를 높게 됨. 종합적인 척도에서 정확도의 한계를 인정하고, 독립적인 척도로서의 활용성에 대해 탐구함. 그 결과 특정 서비스에 적합한 MBTI 척도를 선별적으로 활용하는 것이 더 효과적일 수 있다는 점을 발견하고, 목적에 맞는 정밀한 타겟팅 방법으로 활용 가능성을 확인함.

배운점

프로젝트를 진행했을 당시 모델 성능과 같은 정량적 목표는 세웠지만 추천 방식에 있어서 이용자에게 어떤 경험을 제공할 것인지에 대한 목표 설정을 생략하고 기능적인 부분에만 집중하였습니다. 결국 서비스 요소 간 일관성을 잃고, 각자 담당한 기능을 자랑하는 듯한 결과물이 나오며 대규모의 수정이 불가피했습니다. 이 경험을 통해 타겟 설정과 상세한 정성적, 정략적 목표를 세우고 지속적으로 상기하는 것이 프로젝트의 방향성 유지에 중요함을 배울 수 있었습니다.

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

2.3 EDA

- Word Clouds



- 초기 전처리 과정에서 불용어를 제거하였음에도 불구하고, 여전히 모든 MBTI에서 많은 단어가 중복되어 높은 빈도를 보임
- 본인의 mbti를 게시물 내에 직접적으로 언급한 경우가 있는데 본 프로젝트는 mbti를 모르는 상태에서 예측 분류하는 것을 목적으로 하기 때문에 mbti를 불용어로 취급해 제거해줄 필요가 있음.

2.4 모델링

유형별 최종 모델(Support Vector Machine) 성능

	E-I 유형	S-N 유형	F-T 유형	P-J 유형
최고 교사 검증 점수	0.7570104641	0.7907252392	0.8452017653	0.7223250598
최적 매개변수로 테스트 세트 정확도	0.75889582677	0.7771739130	0.8677966101	0.7412823397

3.3 EDA

Y(columns) & X(MBTI) Result

	nature	danceability	valence	energy	loudness	acousticness	instrumentalness	liveness
0	Intuitive	0.702978	0.561001	0.652445	0.734693	0.258825	0.090096	0.267610
1	Observant	0.714540	0.582204	0.662515	0.741399	0.251192	0.070942	0.274603

RESULT OF NATURE & MEAN

RESULT OF MIND & MEAN

	mind	danceability	valence	energy	loudness	acousticness	instrumentalness	liveness
0	Feeling	0.696542	0.561026	0.631686	0.729298	0.297992	0.069873	0.262488
1	Thinking	0.718868	0.578706	0.680933	0.745516	0.214578	0.093632	0.278411

	tactics	danceability	valence	energy	loudness	acousticness	instrumentalness	liveness
0	Judging	0.693936	0.544474	0.632877	0.727347	0.284587	0.085938	0.262464
1	Prospecting	0.719091	0.590505	0.675858	0.745725	0.232309	0.078756	0.277096

RESULT OF NATURE & TACTICS

4.1 SCENARIO

4. 타겟유저의 MBTI와 해당 음악에 대한 적합도를 평가하여 결과값을 도출한다.

사용자의 MBTI를 입력하세요: ENFJ
적합도를 확인하고자 하는 음악 정보를 입력하세요 (예. 노래 제목 - 가수명): Froot - Marina
ENFJ와(과) 'Froot - Marina'의 일치 정도: 97.83%

사용자의 MBTI를 입력하세요: ENFJ
적합도를 확인하고자 하는 음악 정보를 입력하세요 (예. 노래 제목 - 가수명): All I ask - Adele
ENFJ와(과) 'All I ask - Adele'의 일치 정도: 94.54%