

텍스트 스타일로 MBTI 예측

Predicting MBTI Personalities through Text Styles

NLP 팀 중앙대학교 유승욱, 김상렬, 김중훈, 박경빈



팀원 소개

About Our Team



유승욱

중앙대학교 소프트웨어학부
크롤링 코드 구현, 숏페이퍼 및 PPT 작성



김상렬

중앙대학교 컴퓨터공학부
딥러닝 모델 및 웹 프로토타입 구현



김중훈

중앙대학교 응용통계학부
전처리 코드 및 딥러닝 모델 구현



박경빈

중앙대학교 소프트웨어학부
머신러닝 모델 구현 및 PPT 작성

Part 1 | 주제 선정 개요

Topic Selection



지금은 MBTI 열풍!

MBTI



나도 몰랐던 내 성향을 알려 주네... MBTI에 빠진 2030

f t TALK b blog | 0



[요즘 것들의 문화 답사기] 에너지의 방향·인식·판단·생활양식 네 가지 지표를 16개 유형으로 조합

‘재기 발랄한 활동가’ 등 내 모습 발견

나와 다를 수밖에 없는 타인도 이해

MBTI 유형 인쇄된 티셔츠도 나와

“맹신하지 말고 단점 개선에 활용을”



지금은 MBTI 열풍!

MBTI

- 성격유형검사, MBTI별 - 직장인 유형

ISTJ

세상의 소금형

ISFJ

상사 뒤편의 권력형

INFJ

예언자형

INTJ

과학자형

ISTP

백과사전형

ISFP

성인군자형

INFP

잔다르크형

INTP

아이디어뱅크형

ESTP

활동가형

ESFP

사교형

ENFP

스파크형

ENTP

발명가형

ESTJ

사업가형

ESFJ

친선도모형

ENFJ

언변능숙형

ENTJ

지도자형

외향형(E)과 내향형(I)
직관형(N)과 감각형(S)
감정형(F)과 사고형(T)
그리고 판단형(J)과 인식형(P)

4가지 지표에 따라 결정되는 MBTI

텍스트로 MBTI를 예측할 수 있다면?

MBTI

1. 작성된 텍스트를 글쓴이 시선에서 바라볼 기회가 생김
2. 상대방의 관점에서 텍스트 맥락을 이해하며 서로의 글에 더욱 공감할 것임
3. 글로 파악된 MBTI가 실제 MBTI와 다르다면
평소 글쓴이의 작문 성향을 파악할 수 있음

Part 2 | 데이터 정의 및 수집

Data Definition and Collection



데이터 정의

Data Definition



George Ng >

Myers Briggs Personality Type Test Dataset

Dataset

(MBTI) Myers-Briggs Personality Type Dataset

Includes a large number of people's MBTI type and content written by them



Mitchell J • updated 4 years ago (Version 1)

영어로 이루어진 MBTI 텍스트 데이터셋 다수 존재



한국어로 이루어진 MBTI 텍스트 데이터셋 부재

데이터 후보 1) MBTI를 밝힌 유명인 SNS

→ 광고성, 홍보성 글이 많음

데이터 후보 2) 디시인사이드 MBTI 갤러리

→ MBTI와 상관 없는 글이 많음

데이터 정의

Data Definition



MBTI & HEALTH 심리 카페

CAFE.NAVER.COM/MBTICAPE

심리 자료실 · 심리 뉴스 · MBTI 일반 · MBTI 전문 · Enneagram · STRONG

카페정보 나의활동

MBTI & Health since 2006.03.03. 카페소개

90,719 명 초대하기

★ 즐겨찾는 멤버 5,459명

게시판 구독수 2,703회

우리카페앱 수 272회

주제 인문/과학 > 심리학

지역 서울특별시 서초구

이곳은 네이버 심리분야 대표카페입니다
여러분을 환영합니다!!!

카페 규정보기

심리상담 및 검사신청
수석 부매니저 카가님 | 부매니저 회색분자님
에니어그램 검사 및 상담 여강현님

이벤트, 광고 문의는 mh-staff@naver.com으로 문의바랍니다

▲ 대문접기 | 지난카페대문

한국어로 이루어진 MBTI 텍스트 데이터셋 선정
MBTI & HEALTH 심리 카페 (네이버 카페)

- 2006년부터 지금까지 운영된 카페
- 2020년에는 네이버카페 대표 인기 카페로 선정
- 최근까지도 글이 실시간으로 많이 올라오며 SNS와 같은 광보성, 홍보성 글이 거의 없음
- 거의 모든 게시판에서 MBTI를 주로 다루며 대화함 디시인사이드와 같은 MBTI에 부적절한 글이 없음

데이터 수집: 게시판 안에서

Data Collection in Menu

- 게시판 카테고리 특성상 MBTI와 연관되지 않은 글이나오는 카테고리는 제외

- 총 13개의 게시판을 크롤링함

(신변잡기, 유형감별, 심리 관련 질문, 심리 유형 질문, INFP/ENFP, INTP/ENTP, ISFP/ESFP, ISTP/ESTP, INFJ/ENFJ, ISFJ/ESFJ, INTJ/ENTJ, ISTJ/ESTJ 게시판)

- 한 게시판당 최소 200 ~ 많게는 1000 페이지를 크롤링함
(한 페이지당 글 15개)

답소 질문 상담

신변 잡기 N

사회 소식 N

긍정 일기 N

유형 감별 N

심리 관련 질문 N

심리 유형 질문 N

심리 테스트 및 설문 N

고민, 연애, 진로 상담 N

사랑방

INFP & ENFP N

INTP & ENTP N

ISFP & ESFP N

ISTP & ESTP N

INFJ & ENFJ N

ISFJ & ESFJ N

INTJ & ENTJ N

ISTJ & ESTJ N

심리기능 8기능 소통 N

에니어그램 소통 N

궁합 & 관계 소통

데이터 수집: 페이지 안에서

Data Collection in Page

524397	좋아하는 사람이 있으면.. N	infp 햄스톤 N	02:14	8	0
524390	반대인데 은근 잘 맞는 조합 [7] N	infp 햄스톤 N	01:50	63	0
524389	제 카페 목록이군요 ☹️ [1] N	herro N	01:50	27	0
524387	현재 엠헬에 대한 개인적인 생각. [13] N	마피 INFP N	01:43	105	1
524379	흙수저의사.. [13] N	intp6 N	01:26	66	1
524374	동물유형 테스트!! [1] N	치킨먹고싶다남 N	01:19	17	0
524373	인륜의 공부 N	infp 햄스톤 N	01:18	12	0
524367	인팁 질문 받아여 [2] N	별이 intp N	01:00	29	1
524363	전 감정이 너무 어려워여 [3] N	positive 미라클 N	00:49	52	0
524362	술먹고 왜..나한테 그러니..ㅋㅋㅋㅋ [2] N	필연적환상 ENTJA 3w4 N	00:42	73	0
524360	재미로적는 돈많은 사람과 결혼하는방법 [2] N	물뱅이infp N	00:40	116	1
524355	사회적 에너지 바닥 상태~~ N	ISTJ 1w9 spso N	00:28	30	0
524352	나이키 [1] N	엣프제 N	00:14	21	0
524351	잠깐만 사랑하실거면 죽을힘을 다해 사랑하세요 [7] N	비엔나 INFJ N	00:09	167	4
524348	우울해있을 분들을 위해 이글을 올립니다 [1] N	샷건 ISFPwINFP N	00:02	123	2

게시글 articleid 크롤링

:페이지를 나눠서 크롤링할때

데이터를 어디까지 모았는지 판단하기 위함

데이터 수집: 페이지 안에서

Data Collection in Page

524397	좋아하는 사람이 있으면.. N	infp 햄스톤 N	02:14	8	0
524390	반대인데 은근 잘 맞는 조합 [7] N	infp 햄스톤 N	01:50	63	0
524389	제 카페 목록이군요 ☹️ [1] N	herro N	01:50	27	0
524387	현재 엠헬에 대한 개인적인 생각. [13] N	마피 INFP N	01:43	105	1
524379	흙수저의사.. [13] N	intp6 N	01:26	66	1
524374	동물유형 테스트!! [1] N	치킨먹고싶다남 N	01:19	17	0
524373	인륜의 공부 N	infp 햄스톤 N	01:18	12	0
524367	인팁 질문 받아여 [2] N	별이 intp N	01:00	29	1
524363	전 감정이 너무 어려워여 [3] N	positive 미라클 N	00:49	52	0
524362	술먹고 왜..나한테 그러니..ㅋㅋㅋㅋ [2] N	필연적환상 ENTJA 3w4 N	00:42	73	0
524360	재미로적는 돈많은 사람과 결혼하는방법 [2] N	물뱀이infp N	00:40	116	1
524355	사회적 에너지 바닥 상태~~ N	ISTJ 1w9 spso N	00:28	30	0
524352	나이키 [1] N	엣프제 N	00:14	21	0
524351	잠깐만 사랑하실거면 죽을힘을 다해 사랑하세요 [7] N	비엔나 INFJ N	00:09	167	4
524348	우울해있을 분들을 위해 이글을 올립니다 [1] N	샷건 ISFPwINFP N	00:02	123	2

게시글 제목, href 주소 크롤링
: 해당 게시글 주소로 접속하여
본문 내용을 긁어오기 위함

데이터 수집: 페이지 안에서

Data Collection in Page

524397	좋아하는 사람이 있으면.. 	infp 햄스톤 	02:14	8	0
524390	반대인데 은근 잘 맞는 조합 [7] 	infp 햄스톤 	01:50	63	0
524389	제 카페 목록이군요 ☹️ [1] 	herro 	01:50	27	0
524387	현재 엠헬에 대한 개인적인 생각. [13] 	마피 INFP 	01:43	105	1
524379	흙수저의사.. [13] 	intp6 	01:26	66	1
524374	동물유형 테스트!! 🐾 [1] 	치킨먹고싶다남 	01:19	17	0
524373	인륜의 공부 	infp 햄스톤 	01:18	12	0
524367	인팁 질문 받아여 [2] 	별이 intp 	01:00	29	1
524363	전 감정이 너무 어려워여 [3] 	positive 미라클 	00:49	52	0
524362	술먹고 왜..나한테 그러니..ㅋㅋㅋㅋ [2] 	필연적환상 ENTJA 3w4 	00:42	73	0
524360	재미로적는 돈많은 사람과 결혼하는방법 [2] 	물뱅이infp 	00:40	116	1
524355	사회적 에너지 바닥 상태~~ 	ISTJ 1w9 spso 	00:28	30	0
524352	나이키 [1] 	엠프제 	00:14	21	0
524351	잠깐만 사랑하실거면 죽을힘을 다해 사랑하세요 [7] 	비엔나 INFJ 	00:09	167	4
524348	우울해있을 분들을 위해 이글을 올립니다 [1] 	샷건 ISFPwINFP 	00:02	123	2

게시글 작성자 닉네임 크롤링
: 게시글에 대한
MBTI 라벨링을 위함

데이터 수집: 페이지 안에서

Data Collection in Page

524397	좋아하는 사람이 있으면.. 	infp 햄스톤 	02:14	8	0
524390	반대인데 은근 잘 맞는 조합 [7] 	infp 햄스톤 	01:50	63	0
524389	제 카페 목록이군요 ☹️ [1] 	herro 	01:50	27	0
524387	현재 엠헬에 대한 개인적인 생각. [13] 	마피 INFP 	01:43	105	1
524379	흙수저의사.. [13] 	intp6 	01:26	66	1
524374	동물유형 테스트!! 🐾 [1] 	치킨먹고싶다남 	01:19	17	0
524373	인륜의 공부 	infp 햄스톤 	01:18	12	0
524367	인팁 질문 받아여 [2] 	별이 intp 	01:00	29	1
524363	전 감정이 너무 어려워여 [3] 	positive 미라클 	00:49	52	0
524362	술먹고 왜..나한테 그러니..ㅋㅋㅋㅋ [2] 	필연적환상 ENTJA 3w4 	00:42	73	0
524360	재미로적는 돈많은 사람과 결혼하는방법 [2] 	물뱅이infp 	00:40	116	1
524355	사회적 에너지 바닥 상태~~ 	ISTJ 1w9 spso 	00:28	30	0
524352	나이키 [1] 	엣프제 	00:14	21	0
524351	잠깐만 사랑하실거면 죽을힘을 다해 사랑하세요 [7] 	비엔나 INFJ 	00:09	167	4
524348	우울해있을 분들을 위해 이글을 올립니다 [1] 	샷건 ISFPwINFP 	00:02	123	2

페이지 내에서 진행하는 모든 크롤링
Selenium 라이브러리 사용

데이터 수집: 게시물 안에서

Data Collection in Article

[신변 잡기 >](#)

좋아하는 사람이 있으면..



infp 햄스톤 하날 1:1 채팅

2021.08.31. 02:14 조회 185



댓글 8

URL 복사



눈을 마주 보는 성격인가요?

눈을 피하는 성격인가요?

저는 서서히 호감 생기는 사람은 눈을 보는 편인데,
빠르게 호감이 생긴 사람은 눈은 무슨 얼굴도 못봐요ㅋㅋㅋ.



infp 햄스톤님의 게시물 더보기 >

게시글 본문 텍스트 크롤링
: BeautifulSoup 라이브러리 이용
추후 MBTI 라벨링이 이루어질 대상

데이터 수집: 전체

Data Collection for All



네이버 카페 'MBTI & HEALTH 심리 카페'의



MBTI & HEALTH 심리 카페

CAFE.NAVER.COM/MBTICAFE



선택된 게시판 13개에서

- 신변 잡기
- 유형 감별
- 심리 관련 질문
- 심리 유형 질문
- INFP & ENFP
- INTP & ENTP
- ISFP & ESFP
- ISTP & ESTP
- INFJ & ENFJ
- ISFJ & ESFJ
- INTJ & ENTJ
- ISTJ & ESTJ

524397	좋아하는 사람이 있으면..	infp 햄스톤
524390	반대인데 은근 잘 맞는 조합	infp 햄스톤
524389	제 카페 목욕이군요	herro
524387	현재 엠헬에 대한 개인적인 생각.	마피 INFP
524379	흙수저의사..	intp6
524374	동물유형 테스트!!	치킨먹고싶다남
524373	인품의 공부	infp 햄스톤
524367	인팁 질문 받아여	별이 intp
524363	전 감정이 너무 어려워여	positive 미라클
524362	술먹고 왜..나한테 그러니..ㅋㅋㅋㅋ	필연적환상 ENTJA 3w4
524360	재미로적는 돈많은 사람과 결혼하는방법	물범이infp
524355	사회적 에너지 바닥 상태~~	ISTJ 1w9 spso
524352	나이키	엠프제
524351	잠깐만 사랑하실거면 죽을힘을 다해 사랑하세요	비엔나 INFJ
524348	우울해있을 분들을 위해 이글을 올립니다	샷건 ISFPwINFP



닉네임에 MBTI가 포함된 게시글 주소, 본문 크롤링

Part 3 | 데이터 전처리 및 라벨링

Data Preprocessing and Labeling



데이터 전처리

Data Preprocessing

기본 텍스트 전처리

\r, \n 제거

email 제거

url 주소 제거

특수기호 제거

(), [] 속 문자 제거

이중 스페이스 제거

숫자 및 특수문자 제거

본문 텍스트 길이가 전체대비 10%이하인 경우 제거

한국어 전처리

맞춤법 검사

띄어쓰기 검사

텍스트 정규화

토큰나이저 활용

MBTI 전처리

텍스트에 MBTI 명칭 유의어가 포함된 경우
원래 MBTI 명칭으로 바꿔줌

각 텍스트에 e_i, n_s, f_t, j_p 변수를
부여해 이진분류 라벨링

데이터 라벨링

Data Labeling

ISFP < ESFP >

ISFP 잇프피 분들 질문이요~



ESFJ 112 새음 1:1 채팅

2021.08.31. 08:24 조회 109

작성자 닉네임으로 미루어보아 아래 텍스트는 ESFJ로 라벨링될 예정, 그러나...

안녕하세요!

저는 잇프피남을 짝사랑하고 있는 ESFJ 여자 입니다~

5번째 약속을 잡기위해! 선뎡을하려고하는데,
항상

"출근은 잘하셨어요~?"

"비가 많이오네요ㅠㅠ 출근 조심하 하세요~ㅠ"

이런 평범한? 카톡들만 했는데ㅠㅠ 매번 똑같은 카톡만 하면 잇프피분들은 관심이 없어한다고 들어서ㅠㅠ

이번에는 어떤식으로 보내야 할까요...??

여러분 도와주세요ㅠㅠㅠ



ESFJ 112님의 게시글 더보기 >

데이터 라벨링

Data Labeling

1) MBTI의 종류가 16가지나 있는데,
전체 텍스트 데이터를 16가지 레이블로 세세한 분류가 가능할지 의문
만약 가능하다고 하더라도 지금보다 훨씬 많은 양의 데이터가 필요할 것이라 예상

2) MBTI가 같더라도 사람마다 글 쓰는 스타일의 차이가 확고할 수 있는데,
단순 MBTI 16가지로만 레이블로 분류한다면 이 차이가 더 도드라지지 않을지 의문



텍스트를 16가지 클래스에 분류하는 일은 몹시 어려울 것으로 판단



데이터 라벨링

Data Labeling

16가지 클래스로의 텍스트 분류 문제를,
이진 분류를 4번 수행하는 문제로 전환

기존 방법:
제시된 본문 텍스트가
다음 16개 MBTI 중 원지 하나로 라벨링

예)
본문1 → intj
본문2 → estp
본문3 → enfp

intj
intp
entj
entp
infj
infp
enfj
enfp
istj
isfj
estj
esfj
istp
isfp
estp
esfp

새로운 방법:
제시된 본문 텍스트가
MBTI 지표 4개 중 원지 각각 이진 분류 라벨링

예)
본문1 → e_i=1, n_s=0, f_t=1, j_p=0
본문2 → e_i=0, n_s=1, f_t=1, j_p=1
본문3 → e_i=0, n_s=0, f_t=0, j_p=1

e_i 변수는 i면 1, e면 0
n_s 변수는 s면 1, n이면 0
f_t 변수는 t면 1, f면 0
j_p 변수는 p면 1, j면 0

데이터 최종 형태

Generated Data Form

	articleid	mbti	content	menu_id	e_i	n_s	f_t	j_p
43518	470632	infp	대 중반부터 대 초반인 지금까지 평범한 사무직일을다녔는데 조금 더 길게 하다가는 평...	18	0	1	1	0
43519	470601	infp	카톡으로 칭찬 계속 들으면 부담스러우신가요 아니면 그냥 기분 좋은가요 저는 enfp...	18	0	1	1	0
43520	470460	isfp	시간이 비어서 심심해요 엠비티아이 요즘 완전 과몰입 중이라enfp에게 궁금한점이 뭘...	18	0	0	1	0
43521	470380	intj	물론 이기적이고 나쁜 생각이라는 거 알면서도 솔직하게 얘기해보면 그래요 근데 그렇다...	18	0	1	0	1
43522	470291	infp	혹시 공지 위반이라면 말씀해주시면 바로 삭제하겠습니다 마땅한 게시판을 찾지 못해 제...	18	0	1	1	0
43523	470269	infp	용건이 아닌 그냥 일상 톡 두달여간 커피며 선물이며 선택이며부담안되게 꾸준히 밀고빠...	18	0	1	1	0
43524	470143	infp	저는 ISTJ고 친구는 ENFP 입니다 먼저 저에 대해 설명하자면 저는 충분히 생각...	18	0	1	1	0
43525	470098	enfp	아무래도 INFP인데 INTJ나 INTP로 오해받는 경우가 있다거나 하는 거 말이죠...	18	1	1	1	0



전처리가 완료된 본문 텍스트



MBTI 네가지 특성의
이진분류로 텍스트 라벨링

Part 4 | 머신러닝 모델로 자연어 분류

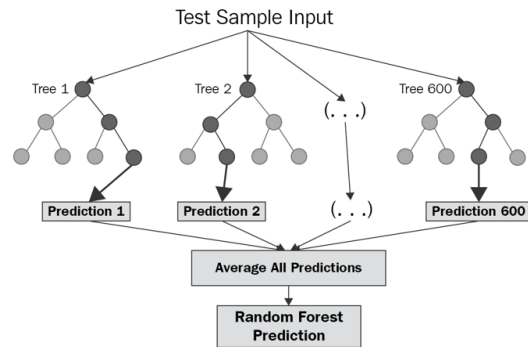
Classifying Texts with ML Models



머신러닝 모델 소개

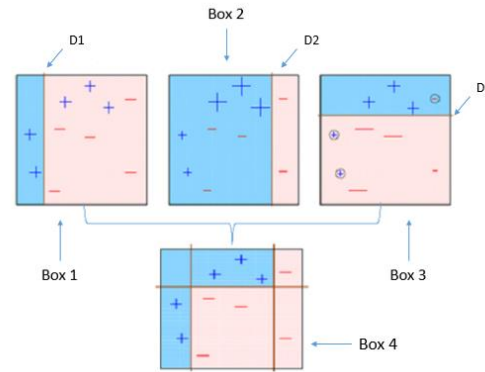
About ML Models

Random Forest



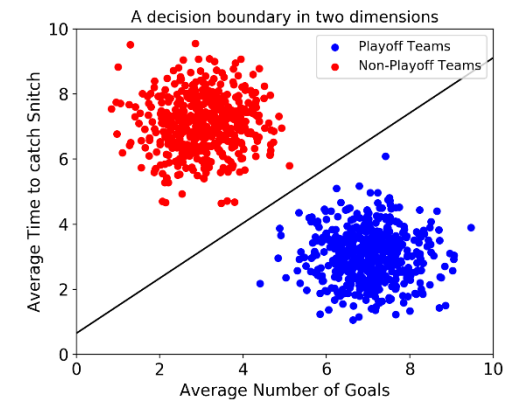
여러 개의 결정 트리 분류기가
개별적으로 학습을 수행한 뒤,
최종적으로 모든 분류기가
Voting을 통해 예측 결정

XGBoost



Random Forest와 동일 원리를
사용하나 Boosting을 적용해
오답에 가중치를 부여해가며
예측 방향 결정

SVM

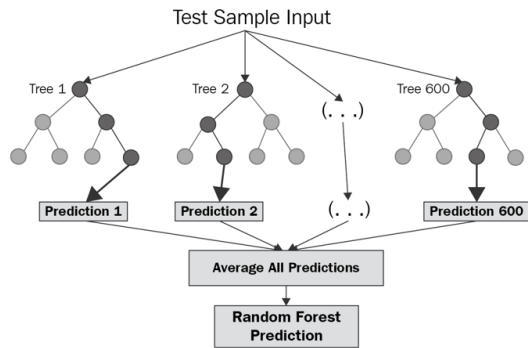


데이터 분류를 위해 분류될
클래스들 사이의 Margin을
최대화하는 초평면을 찾으며,
선형과 비선형으로 나뉨

(1) Random Forest

ML Models-Random Forest

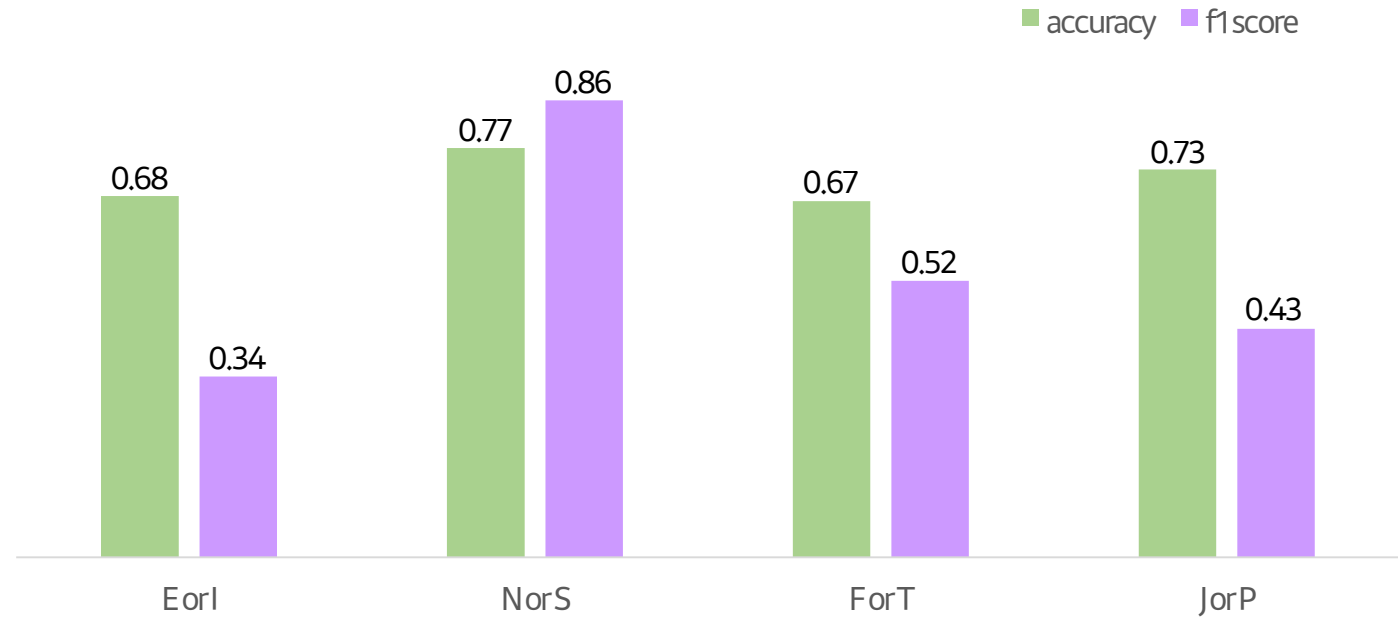
Random Forest



여러 개의 결정 트리 분류기가
개별적으로 학습을 수행한 뒤,
최종적으로 모든 분류기가
Voting을 통해 예측 결정

텍스트
전처리

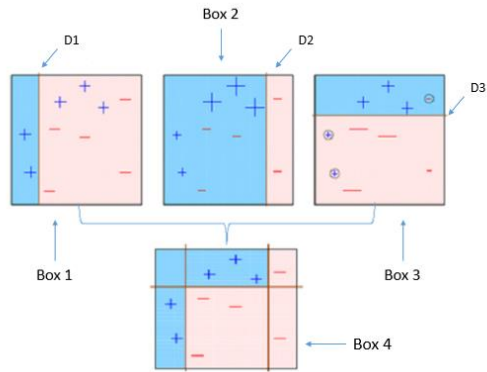
리스트 형태의 텍스트 input을 처리하고자 TF-IDF를 이용,
중요도가 낮으나 여러 텍스트에서 자주 등장하는 단어를 효과적으로 배제함
이 과정에서 Okt 라이브러리를 사용하여 형태소 단위 벡터라이징을 적용함



(2) XGBoost

ML Models - Extreme Gradient Boosting

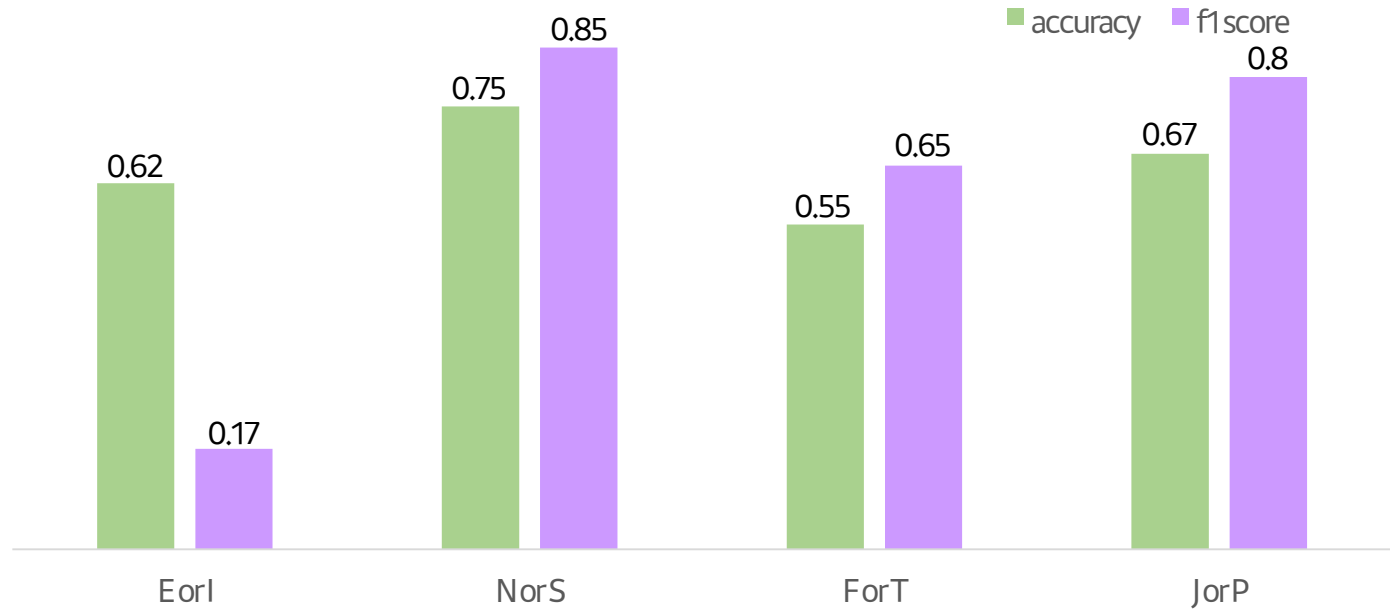
XGBoost



Random Forest와 동일 원리를
사용하나 Boosting을 적용해
오답에 가중치를 부여해가며
예측 방향 결정

텍스트 전처리

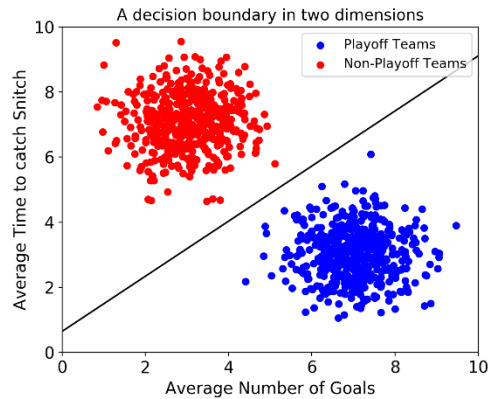
패딩된 numpy input을 처리하고자 Tensorflow Tokenizer를 이용,
Tensorflow 라이브러리에 자체 내장된 클래스를 사용해 토큰화를 진행함
이 과정에서 Okt 라이브러리를 사용하여 형태소 단위 벡터라이징을 적용함



(3) SVM

ML Models - Support Vector Machine

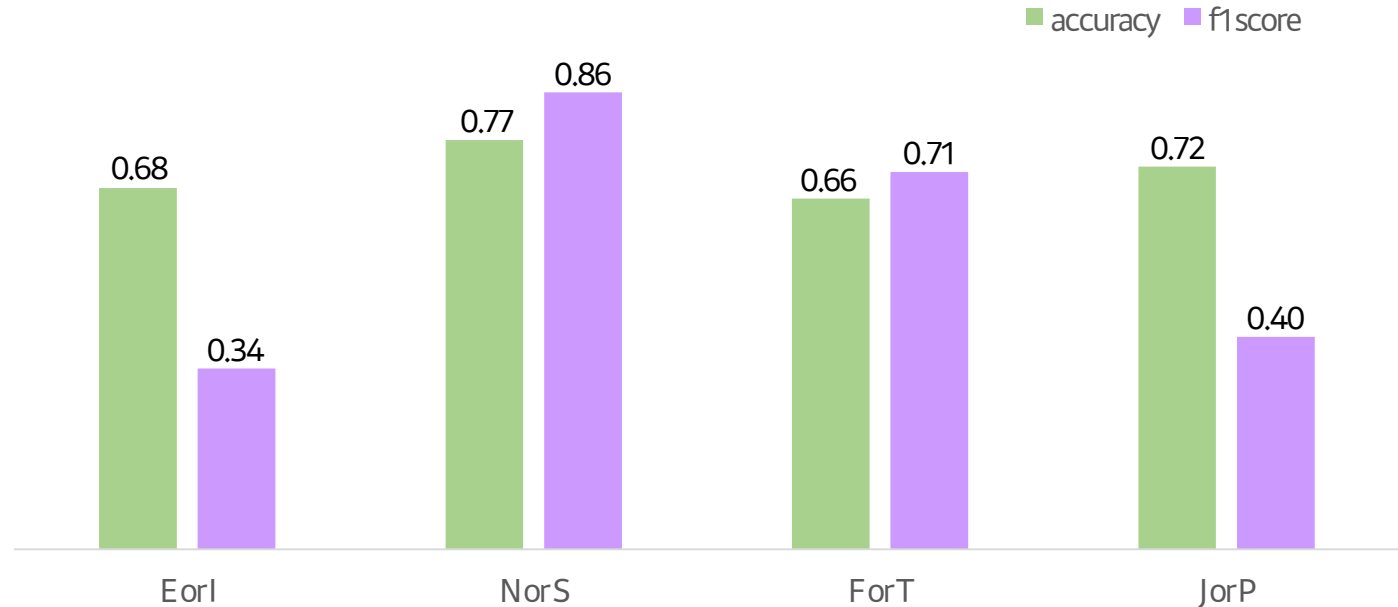
SVM



데이터 분류를 위해 분류될 클래스들 사이의 Margin을 최대화하는 초평면을 찾으며, 선형과 비선형으로 나뉨

텍스트 전처리

리스트 형태의 텍스트 input을 처리하고자 TF-IDF를 이용, 중요도가 낮으나 여러 텍스트에서 자주 등장하는 단어를 효과적으로 배제함 이 과정에서 Okt 라이브러리를 사용하여 형태소 단위 벡터라이징을 적용함



*성능평가 그래프에서 비선형 SVM 결과를 사용

Part 5 | 딥러닝 모델로 자연어 분류

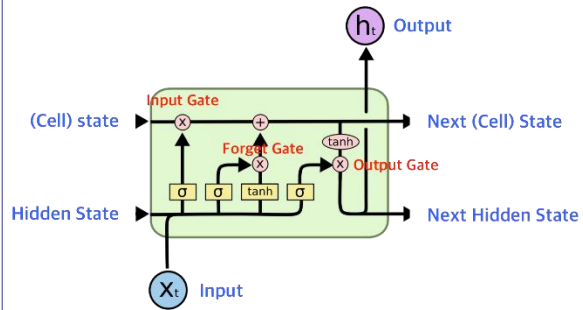
Classifying Texts with DL Models



딥러닝 모델 소개

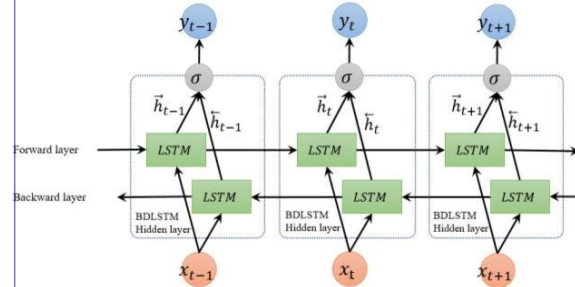
About DL Models

LSTM



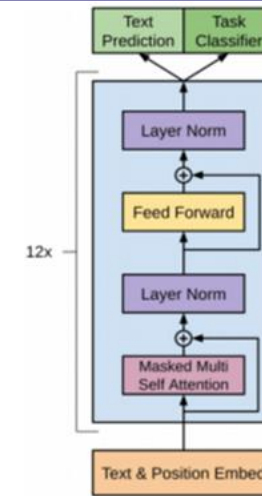
기존의 Simple RNN이 가졌던
장기 기억 손실 문제를 극복하며,
그 과정에는 forget gate가
중요하게 작용하는 모델

Bi-LSTM



LSTM의 양방향 모델로서
정방향과 역방향으로의
두 가지 순서 모두에서
텍스트 파악이 가능함

GPT2

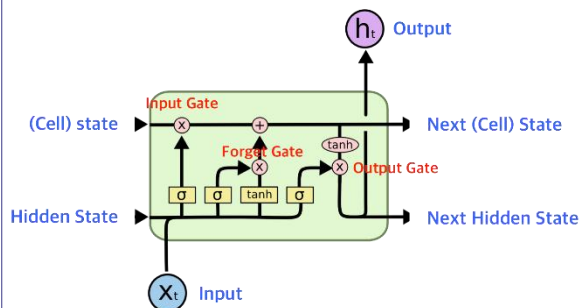


다양한 영역의 텍스트로
사전 학습을 진행하여
모델이 기존보다 다양한 영역의
글을 이해할 수 있음

(1) LSTM

DL Models - Long Short Term Memory

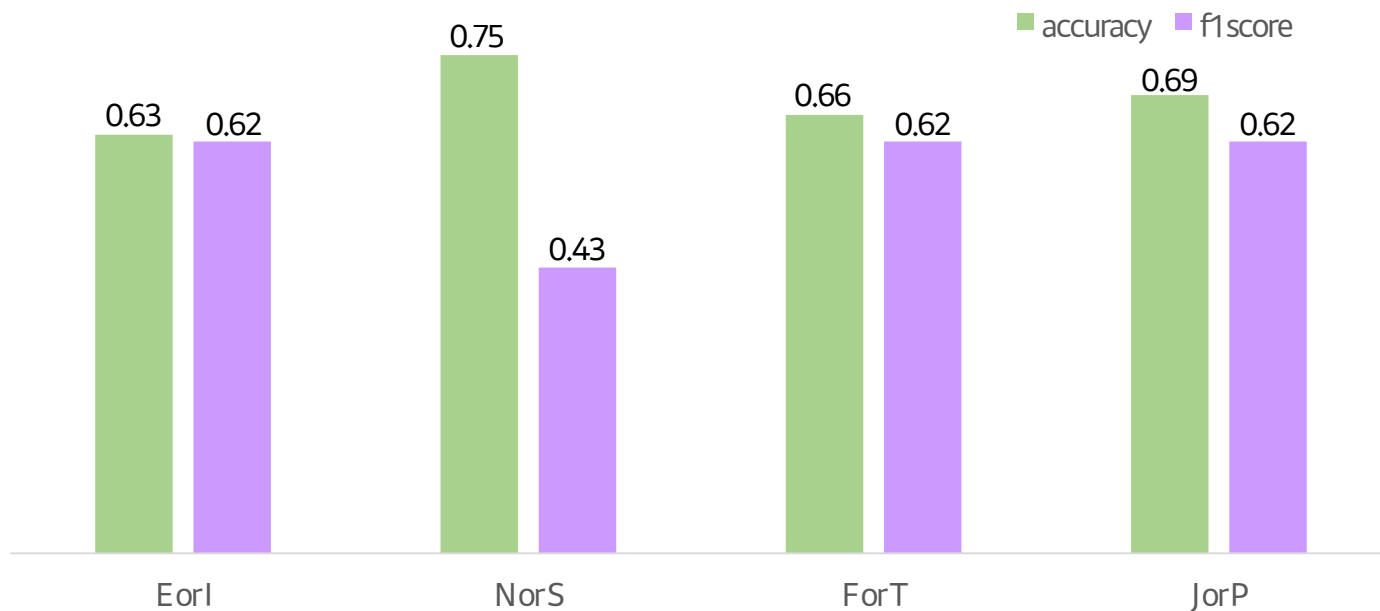
LSTM



기존의 Simple RNN이 가졌던
장기 기억 손실 문제를 극복하며,
그 과정에는 forget gate가
중요하게 작용하는 모델

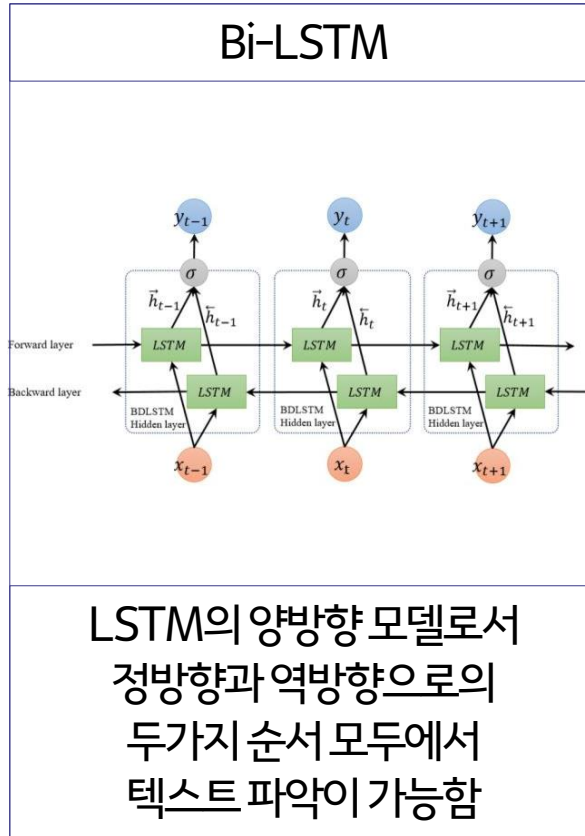
텍스트
전처리

형태소 분석기를 사용하여 본문 내용을 인식하고
분리된 형태소는 정수 인코딩을 거침
그 후 max_len을 256으로 설정해 패딩을 진행함



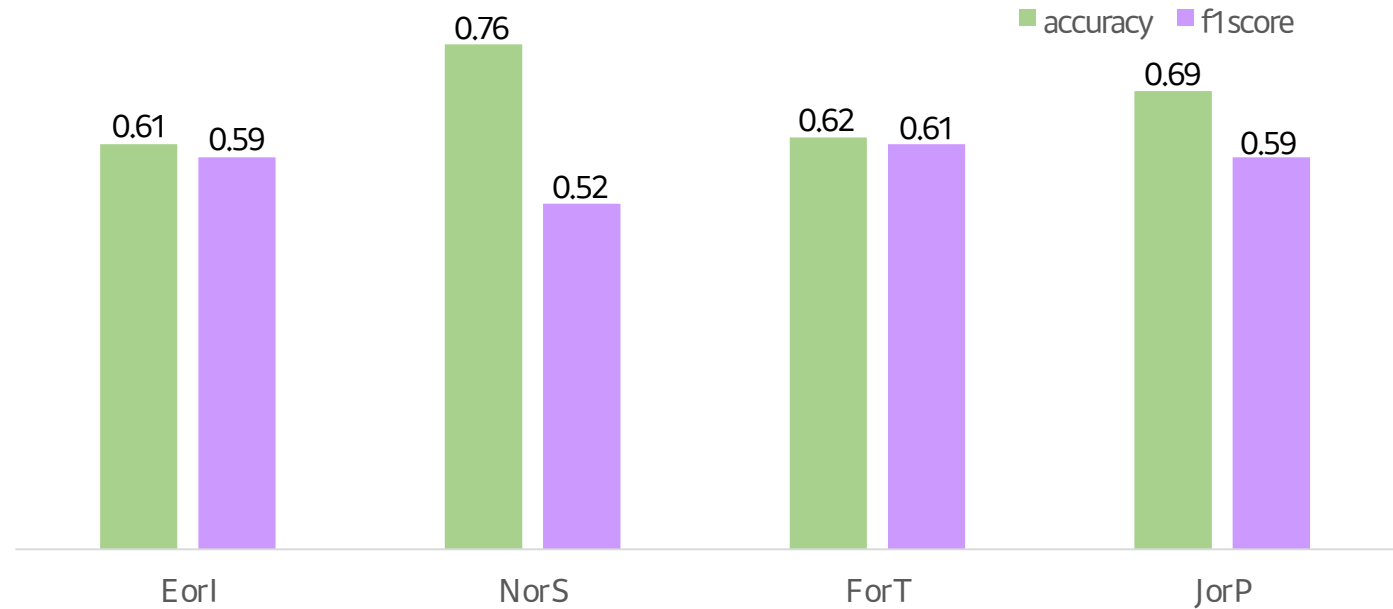
(2) Bi-LSTM

DL Models - Bidirectional Long Short Term Memory



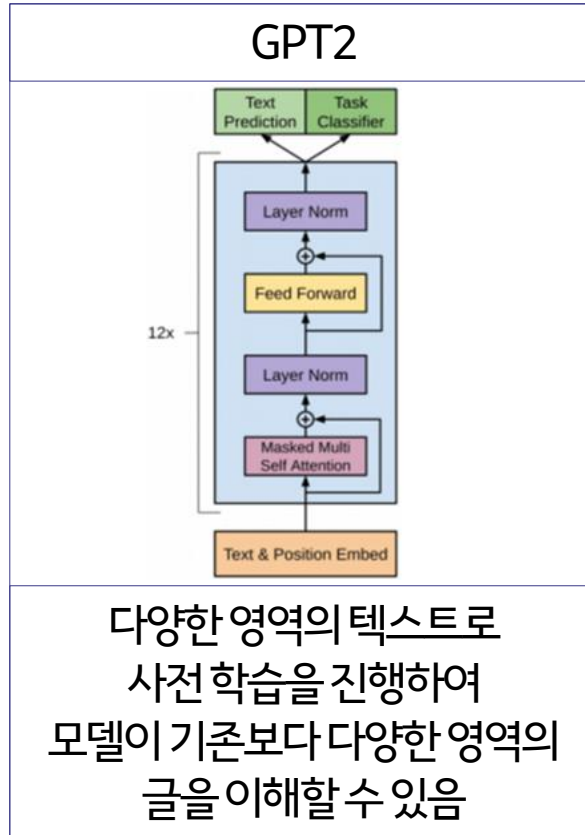
텍스트
전처리

형태소 분석기를 사용하여 본문 내용을 인식하고
분리된 형태소는 정수 인코딩을 거침
그 후 max_len을 256으로 설정해 패딩을 진행함



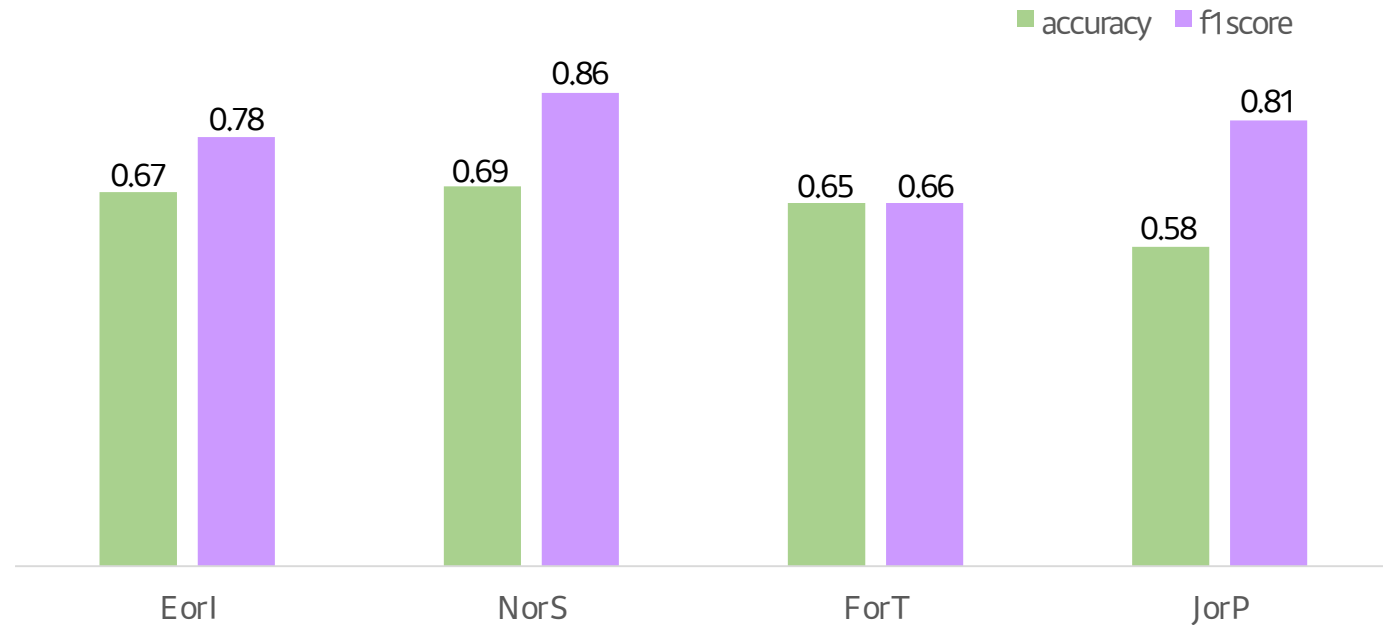
(3) GPT2

DL Models - Generative Pretrained Transformer 2



텍스트
전처리

BPE(Byte Pair Encoding)라는 방식을 사용해
텍스트를 나눠서 입력값으로 넣어줌
이는 글자와 문자 사이의 적절한 단위를 찾아 나누는 방식으로 높은 성능을 보여줌



Part 6 | 웹 프로토타입 및 결론

Web Prototype and Conclusion



웹 프로토타입

WebVersionPrototype

<웹 프로토타입 제작>

1. 학습한 모델을 웹 상에서 실시간으로 적용
2. 사용자로부터 새로운 텍스트를 입력받으면, MBTI지표 4가지에서 이진분류 리턴결과 도출

Front: vue.js

Back: fastapi

Server: aws에 ec2 인스턴스 올려서
웹서버 nginx 열어두고 사용

당신의 MBTI는 무엇일까요?
아래의 질문을 통하여 당신의 mbti를 알아보세요

본인의 성격을 간단하게 알려주세요

다른 사람을 잘 이해하려고 노력하지만 포기가 빨라요.

입력한 텍스트

ACCEPT TERMS

당신의 MBTI 결과는 입니다

51

50

49

E

I

51

50

49

N

S

웹 프로토타입

WebVersionPrototype

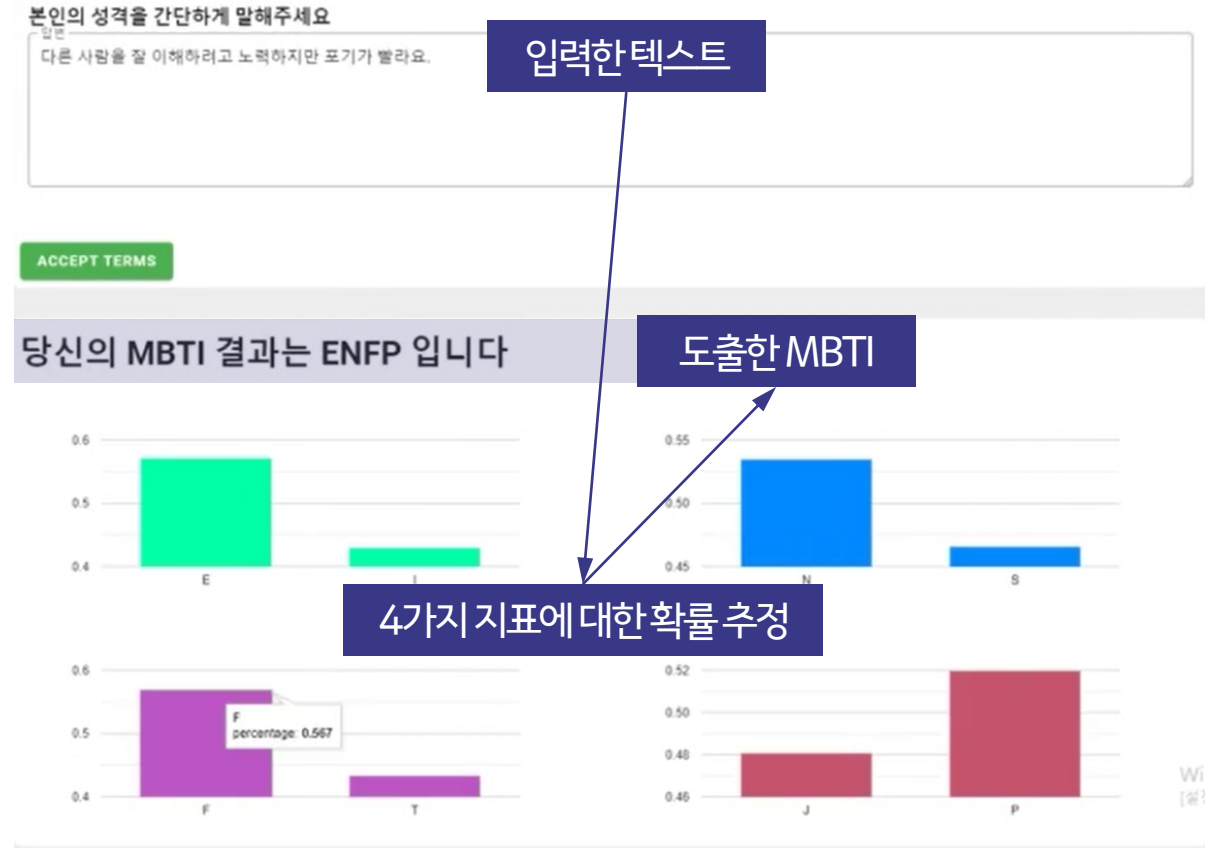
<웹 프로토타입 제작>

1. 학습한 모델을 웹 상에서 실시간으로 적용
2. 사용자로부터 새로운 텍스트를 입력받으면, MBTI지표 4가지에서 이진분류 리턴결과 도출

Front: vue.js

Back: fastapi

Server: aws에 ec2 인스턴스 올려서
웹서버 nginx 열어두고 사용



cuaimbti.site에서 바로 이용 가능!

1. 글 스타일로 본인의 검사 결과를 정확히 예측하는 것은 불가능하다는 한계

오로지 다른 것을 제외하고 ‘글 스타일’만으로 mbti를 예측하는 모델이므로
실제 본인의 mbti와 동떨어질 수 있다는 한계가 존재

2. 카페 게시판의 특성이 데이터셋에 반영됨에 따라 생기는 문제점 및 개선방향

12개 게시판 중 8개가 ‘사랑’과 관련된 주제에 치우쳐져 있어
다른 주제의 경우 명확한 학습이 어려울 수 있음

3. 영어로 된 mbti 분류 데이터셋의 부재

선행 연구는 영어로 된 mbti 데이터셋이나 관련 커뮤니티가 명확히 존재하였으나
한국어의 경우 데이터셋을 확보하는 것에 어려움이 있었음

Thank you!

Q&A



CUAI 4th Summer Conference Oral Presentation
August 31, 2021