



MBTI 성격 예측 프로젝트

YBIGTA 박태건, 백상현, 이명진, 이시욱











성격이란?

- 개인의 생각하고, 느끼고, 행동하는 특징적인 패턴, 그리고 이 패턴을 이끄는 심리학적 기제들 (Funder, 2013)
- 심리학을 비롯한 사회과학 연구와 마케팅 등 응용 분야에서 다양한 반응변수를 예측할 수 있음
- 표준화된 성격 검사를 일일이 시행하는 것은 큰 비용이 따름

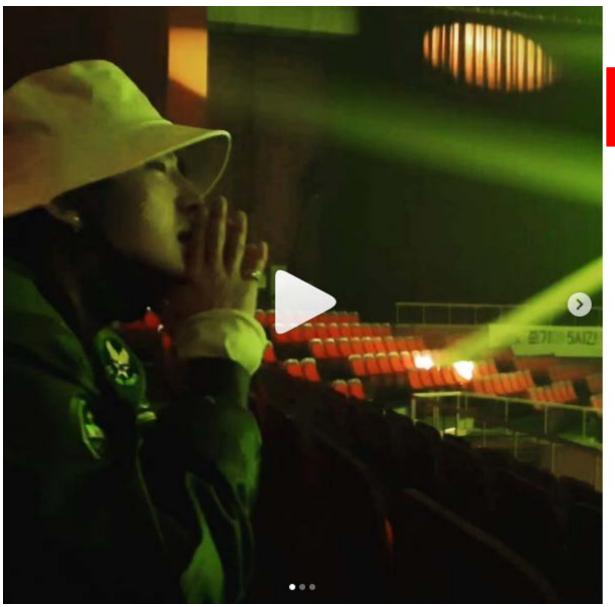
성격이란?

• 개인의 생각하고, 느끼고, 행동하는 특징적인 패턴, 그리고 이 패턴을 이끄는 심리학적 기제들 (Funder, 2013)

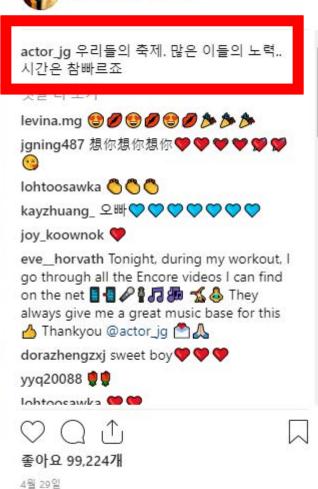
주어진 글만 보고도 쓴 사람의 성격을 예측할 수 있을까?

• 표준화된 성격 검사를 일일이 시행하는 것은 큰 비용이 따르며 자기 보고식 검사의 한계점을 지님









좋아요 또는 댓글을 남기려면 로그인.







godjp 🐡 • 팔로우

godjp 요오우~ 인재 주말쓰는끝 ㅠㅜ 그니 깐 새로운주를위해 얼릉 새수들하구 발꼬락 쓰들 빡빡쓰딱구 굿밤쓰해 매애앤! 빼애앰!!! Yo~ The weekend is coming to a closing sooo prepare for the new week to start aaaand wash up and get to bed maaaan~ BBBAAAMMM!!!

유브이 #JoonPark #ParkJoonHyung #UV # 미세초 #FaceMask #FineMistFaceMask #나두아라매애앤일요일밤쓰돼믄다음날윌 요일쓰가두려운거하지만다음주말이더가꺼 워진다구생각쓰하구

GetSomeGoodSleepMaaaan빼애앰 #왜내맛 크만구멍쓰가크구세윤쓰구멍쓰는요올라뤼 쓰똥꼬오쓰만큼쬬옥꿈하쥐차암나 #SeoulKorea #Kpop

댓글 더 보기

kong.juyeon @leewoongsubb ㅋㅋㅋㅋ









좋아요 12,286개

4월 8일

좋아요 또는 댓글을 남기려면 로그인.

ISTJ 세상의 소금형

한번 시작한 일은 끝까지 해내는 사람들

ISFJ 임금퉛편의 권력형

성실하고 온화하며 협조 를 잘함

INFJ 예인자형

사람과 관련된 뛰어난 통 찰력을 가짐

INTJ 과학자청

전체적인 부분을 조합하 여 비전을 제시하는 사람 들

ISTP 백과사전형

논리적이고 뛰어난 상황 적응력을 가짐

ISFP 성인군자형

따뜻한 감성을 가지고 있 는 겸손한 사람들

INFP 전다르크형

이상적인 세상을 만들어 가는 사람들

INTP 아이디어 뱅크형

비평적인 관심을 가지고 있는 뛰어난 전략가들

ESTP 수완좋은활동가형

친구, 운동, 음식등 다양함 을 선호

ESFP 사교적인유형

분위기를 고조시키는 우 호적인 사람들

ENFP 스파크형

열정적으로 새로운 관계 를 만드는 사람들

ENTP #978

풍부한 상상력을 가지고 새로운 것에 도전하는 사람

ESTJ 사업가형

사무적, 실용적, 현실도모 적인 일을 많이 하는 사람

ESFJ 친선도모형

친절과 현실감을 비탕으로 타인에게 봉사하는 사람

ENFJ 언변능숙형

타인의 성장을 도모하고 현동하는 사람

ENTI 지도자현

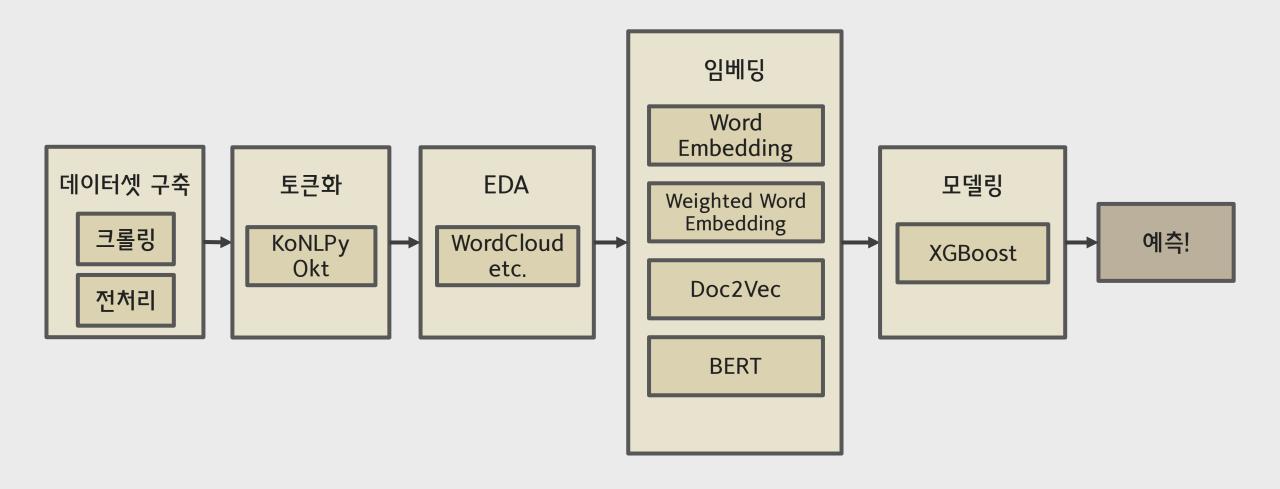
비전을 가지고 사람들을 활력적으로 이끌어 감

MBTI 성격 유형

- 성격심리학에서 표준적으로 인정되는 검사는 아님
- 매우 대중적. 많은 데이터의 축적을 통해 타당화
- 약식 검사를 해보고 싶으시면…

https://www.16personalities.com/ko

워크플로우



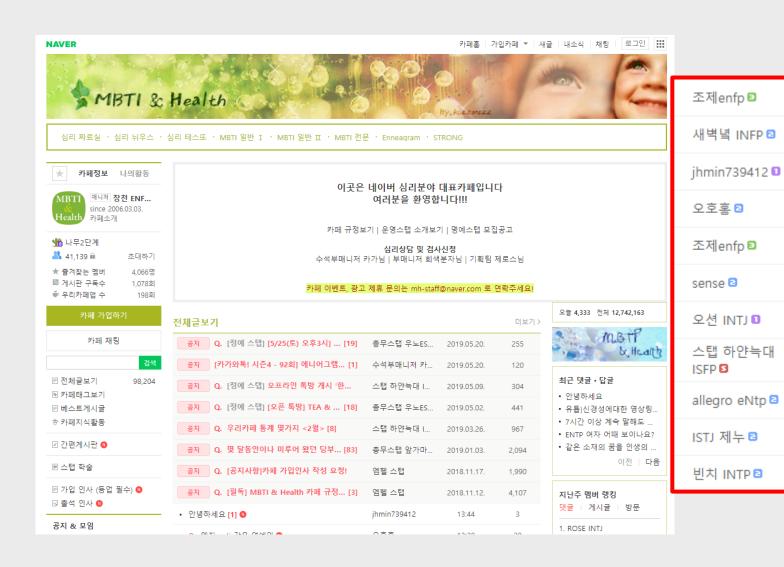
크롤링 및 데이터셋 구축



네이버 카페 MBTI & Health

2006년 설립회원 수 41,914명전체 게시글 수 100,562개

크롤링 및 데이터셋 구축



네이버 카페 MBTI & Health

2006년 설립회원 수 41,914명전체 게시글 수 100,562개

역명성 보장 많은 유저들이 ID에 성격 유형 기재 → 쉽게 라벨링 가능

크롤링 및 데이터셋 구축

Selenium과 BeautifulSoup 이용

'전체글보기' 접근 가능한 **모든 게시물 크롤링** 작성자 ID에 MBTI 성격 유형 포함된 게시글만 사용

가입 인사는 제외

- 최종 데이터셋 -

28,569개의 게시물

6,266,305개의 형태소

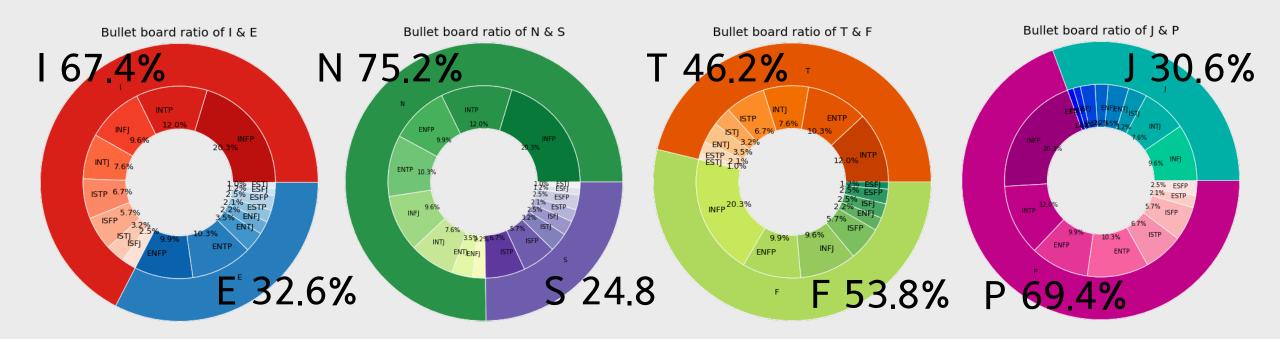
4,458명의 unique 회원

2012년 6월 15일 ~ 2019년 5월 18일 (26,485일)

성격 유형별 회원 수 (비율)

성격 유형	회원 수 (비율)	대한민국 비율		
INFP	827 (18.55%)	3.67%		
INTP	604 (13.55%)	3.3%		
ENFP	467 (10.48%)	3.03%		
ENTP	449 (10.07%)	2.18%		
INFJ	433 (9.71%)	2.4%		
INTJ	340 (7.63%)	5.47%		
ISTP	293 (6.57%)	7,83%		
ISFP	237 (5.32%)	6.5%		
ISTJ	141 (3.16%)	21.5%		
ENTJ	137 (3.07%)	3.44%		
ENFJ	118 (2.65%)	1.84%		
ISFJ	104 (2.33%)	8.18%		
ESTP	100 (2.24%)	5.33%		
ESFP	86 (1.93%)	5.37%		
ESFJ	65 (1.46%)	5.59%		
ESTJ	57 (1.28%)	14.2%		

성격 유형별 게시물 비율



토큰화

KoNLPy Okt

'잠자는 시간이 아깝다는 분? 저는 infj인데 어릴때부터 자는 시간이 너무 아까웠어요ㅋㅋ물론 피곤하면 오래자기도 하지만 필요이상으론 안자요..남 들은 "쉬는날에 잠 자야지" 이렇게 말하는거 저는 공감을 못하네요뭐라도 하는 편이에요 그게 노는 거든 쉬는거든 공부하던 간에ㅋㅋㅋㅋㅋ그래서 어 릴때 아침 일찍 주말에 일어나서 디즈니 만화영화를 매번 봤답니다ㅋㅋㅋ저처럼 잠자는걸아까워하는 다른 유형은 뭘까여??'

> ['잠자는', '시간', 'Ol', '아깝다는', '분', '?', '저', okt.morphs() 'infj', '인데', '어릴', norm=False 'Œ#', stem=False '부터', '자는', '시간', 'Ol', '너무', '아까웠어요', '=='.

['잠자다', '시간', 'Ol', '아깝다', '분', '?', '저', okt.morphs() '늗', 'infj', '인데', norm=True '어리다', 'αΗ', stem=True '부터', '자다', '시간', '0|', '너무', '아깝다', '==',

MBTI 성격 유형 차원 1: E/I



에너지의 방향이 어디를 향하는가?



Extroversion

외부로부터

Introversion

내부로부터

MBTI 성격 유형 차원 1: E/I





Extroversion

외부로부터

Introversion

내부로부터

MBTI 성격 유형 차원 2: S/N



세상을 어떻게 **인식**하는가?



Sensing 현실에서 직접 부딪치면서 iNtuition 관념과 아이디어를 통해

MBTI 성격 유형 차원 2: S/N





Sensing

현실에서 직접 부딪치면서

iNtuition

관념과 아이디어를 통해

MBTI 성격 유형 차원 3: T/F



무엇에 기초하여 **판단**을 내리는가?



Thinking

이성에 의해

Feeling

감성에 의해

MBTI 성격 유형 차원 3: T/F





Thinking

이성에 의해

Feeling

감성에 의해

MBTI 성격 유형 차원 4: J/P



어떤 **생활 양식**을 따르는가?



Judging

규칙과 계획을 따른다

Perceiving

즉흥적인 상황을 즐긴다

MBTI 성격 유형 차원 4: J/P



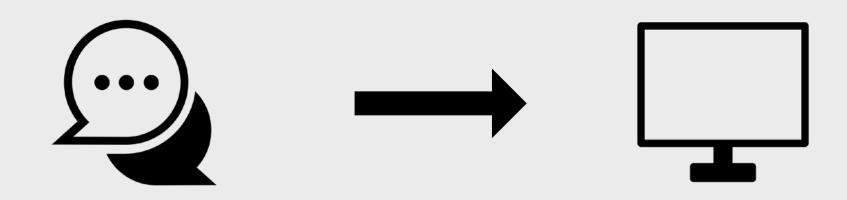
Judging

규칙과 계획을 따른다



Perceiving

즉흥적인 상황을 즐긴다



자연어를 기계가 이해할 수 있게!

Averaged Word Embedding

Word2Vec (skip-gram, CBOW) fastText (skip-gram, CBOW)

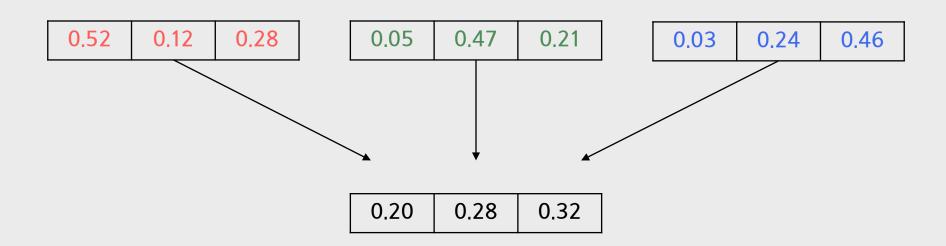
Averaged **tf-idf weighted** Word Embeddings

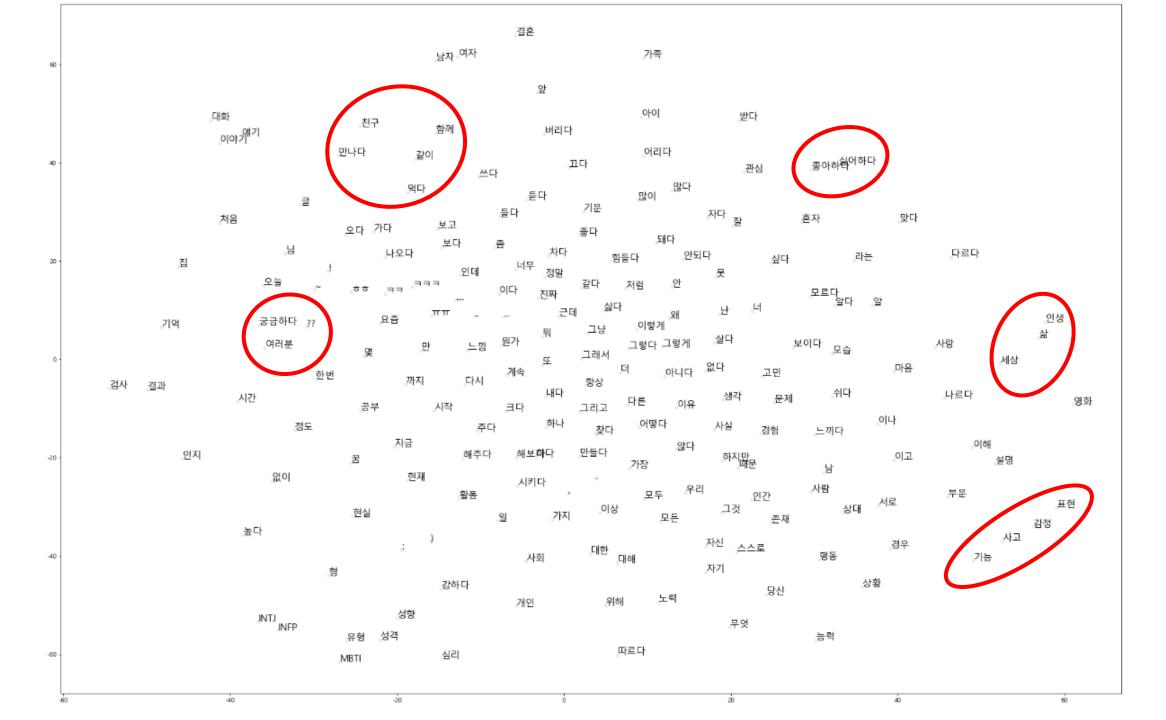
Doc2Vec (PV-DM, PV-DBOW)

BERT

1. Averaged Word Embedding

안녕 내 친구





TF-IDF

단어 문서 행렬(Term-Document Matrix; TDM) 내에 있는 <u>각 단어에 대한 중요도 계산</u>

tf(d,t): 특정 문서 d에서의 특정 단어 t의 <u>등장 횟수</u>.

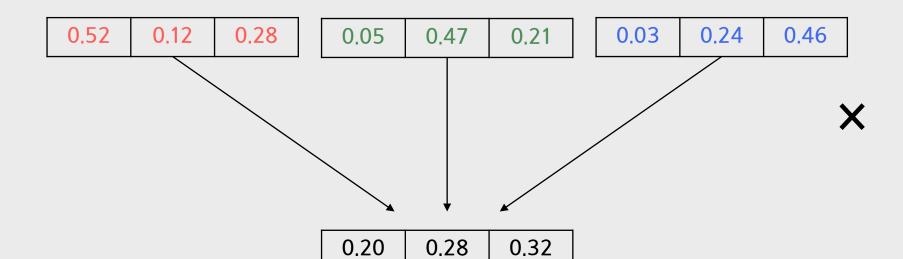
df(t): 특정 단어 t가 <u>등장한 문서의 수</u>.

$$\operatorname{idf}(d,t) = \log \frac{n}{1 + \operatorname{df}(t)}$$

모든 문서에서 자주 등장하는 단어는 중요도가 낮고, 특정 문서에서만 자주 등장하는 단어는 중요도가 높다고 판단

2. Averaged tf-idf weighted Word Embeddings

안녕 내 친구



	안녕	내	친구
Doc1	0.27	0.13	0.09
Doc2	0	0	0.10

3. Document Embedding

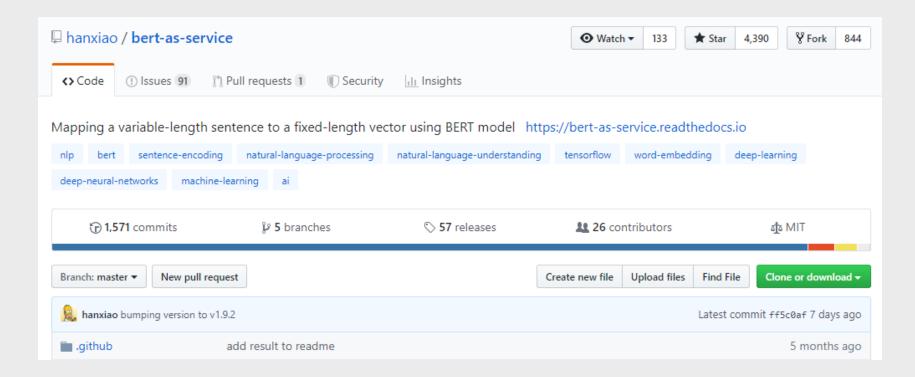
안녕 내 친구

0.52 0.12 0.28

와이빅타 컨퍼런스 재미있다

0.03 0.24 0.46

4. BERT



fine-tuning 없이 pre-trained model 사용

모델링

XGBoost 사용 5-fold CV → F1 score 계산

MBTI 성격 유형의 네 가지 차원에 대한 네 개의 binary classifiers 생성

F1 score	w2v	w2v	ft	ft	doc2vec	doc2vec	tf-idf	tf-idf	tf-idf	tf-idf	Bert
	cbow	sg avg	cbow	sg avg	PV	PV DMO	w2v	w2v sg	ft cbow	ft sg	Бен
I/E	0.75	0.77	0.74	0.76	0.75	0.76	0.76	0.76	0.75	0.76	0.7
N/S	0.37	0.36	0.36	0.36	0.36	0.33	0.37	0.38	0.37	0.38	0.29
T/F	0.58	0.59	0.58	0.59	0.59	0.56	0.58	0.58	0.57	0.57	0.5
J/P	0.77	0.78	0.76	0.78	0.79	0.79	0.78	0.79	0.77	0.79	0.75

모델링

최종 모델 선정 결과

E/I	Averaged Word2Vec (skip-gram)			
N/S	tf-idf weighted Word2Vec (skip-gram)			
F/T	Doc2Vec (PV-DBOW)			
J/P	Doc2Vec (PV-DBOW)			







new_text = '요오무~ 인재 주말쓰는끝 ㅠㅜ 그니깐 새로운주를위해 얼릉 새수들하구 발꼬락쓰들 빡빡쓰딱구 굿밤쓰해 매애앤! 빼애앰!!!'



유진다구생각쓰하구
GetSomeGoodSleepMaaaan빼애앰 #왜내맛 크만구멍쓰가크구세윤쓰구멍쓰는요올라뤼 쓰똥꼬오쓰만큼쬬옥꿈하쥐차암나 #SeoulKorea #Kpop 댓글 더 보기 kong.juyeon @leewoongsubb ㅋㅋㅋㅋ

좋아요 12,286개

동이표 12,200

4월 8일

좋아요 또는 댓글을 남기려면 로그인.





godjp 🌣 • 팔로우

godjp 요오우~ 인재 주말쓰는끝 ㅠㅜ 그니 쓰들 빡빡쓰딱구 굿밤쓰해 매애앤! 빼애앰!!! Yo~ The weekend is coming to a closing

1. 새로운 input 받기

new_text = '요오우~ 인재 주말쓰는끝 ㅠㅜ 그니깐 새로운주를위해 얼릉 새수들하구 발꼬락쓰들 빡빡쓰딱구 굿밤쓰해 매애맨! 빼애앰!!!'



프로드기구너군기에시킨니금구르이니기까 워진다구생각쓰하구

GetSomeGoodSleepMaaaan빼애앰 #왜내맛 크만구멍쓰가크구세윤쓰구멍쓰는요올라뤼 쓰똥꼬오쓰만큼쬬옥꿈하쥐차암나

#SeoulKorea #Kpop

댓글 더 보기

kong.juyeon @leewoongsubb = = = =

Output

array(['ENFP'], dtype=object)



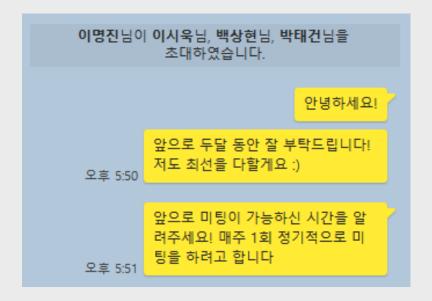
좋아요 12,2867

좋아요 또는 5

4월 8일

ENFP 스파크형

열정적으로 새로운 관계를 만드는 사람들

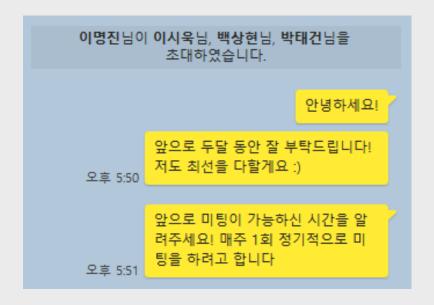


그리고 저번에 선배분들이 저희 프로젝트 칭찬 많이 해주셨어요 살아고 있으니 다들 마지막까지 힘내보아요~ 기말시험 잘보세요ㅎㅎ

이명진: INFJ

INFJ 예언자형

사람과 관련된 뛰어난 통찰력을 가짐



그리고 저번에 선배분들이 저희 프로젝트 칭찬 많이 해주셨어요 실어 가능 하고 있으니 다들 마지막까지 힘내보아요~ 기말시험 잘보세요ㅎㅎ

이명진: INFJ

Output

array(['ENFJ'], dtype=object)

박태건

후 유일하게 학번을 공개해도 내 가 막내내

박태건

어차피 돌리면서 할거 하면 되니 까ㅋㅋㅋ헤헿

오전 12:34

박태건

아니야

형이 살아있다는것 자체로

우리는 설레

박태건: ENTP

ENTP 발명가형

풍부한 상상력을 가지고 새로운 것에 도전하는 사람

박태건

후 유일하게 학번을 공개해도 내 가 막내내

박태건

어차피 돌리면서 할거 하면 되니 까ㅋㅋㅋ헤헿

오전 12:34

박태건

아니야

형이 살아있다는것 자체로

우리는 설레

박태건: ENTP

Output

array(['INTP'], dtype=object)

백상현

우와 짱짱입니다 형!!

오후 3:40

백상현

진짜 진짜 이상한 그래프가 나와 버렸네요;;;

여튼 CV에러 체크 하는거랑 파라 미터 2개 튜닝하는거 진행해봤습 니다~

오전 4:36

백상현

거기서도 먹으면 안대지만

오후 3:32

백상현

눈치봐서

먹자 ㅋㅋㅋㅋㅋㅋ

백상현: INFP

INFP 잔다르크형

이상적인 세상을 만들어가는 사람들

백상현

우와 짱짱입니다 형!!

오후 3:40

백상현

진짜 진짜 이상한 그래프가 나와 버렸네요;;;

여튼 CV에러 체크 하는거랑 파라 미터 2개 튜닝하는거 진행해봤습 니다~

오전 4:36

백상현

거기서도 먹으면 안대지만

오후 3:32

백상현

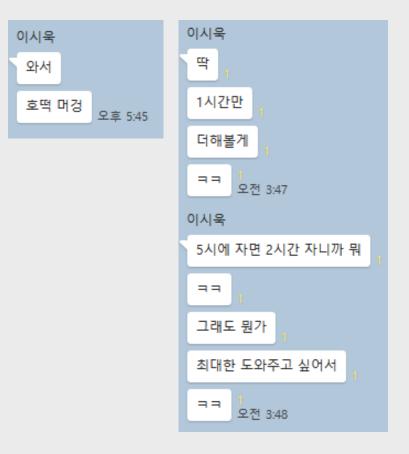
눈치봐서

먹자 ㅋㅋㅋㅋㅋㅋ

백상현: INFP

Output

array(['ENTP'], dtype=object)



이시욱: INFP

INFP 잔다르크형 이상적인 세상을 만들어가는 사람들

```
이시욱
이시욱
와서
                 1시간만
호떡 머겅
       오후 5:45
                 더해볼게
                    오전 3:47
                이시욱
                 5시에 자면 2시간 자니까 뭐
                 ==
                 그래도 뭔가
                 최대한 도와주고 싶어서
                    오전 3:48
```

이시욱: INFP

Output array(['INFP'], dtype=object)

의의

정형화되지 않은 자연어에서

추상적인 성격을 예측

감사합니다

Q & A