

Project. 3

TEAM 3조 벤쿠버

이현지, 김응진, 유한솔, 김나현, 김준철

목차

1. 선정 배경
2. 대상 고객
3. 데이터 수집 / 활용
4. 모델
5. 검증 및 결과
6. 활용
7. 데모 시연
8. 자체 평가 및 보완

1. 선정 배경

- 팀 공동의 목표를 달성하기 위한 **커뮤니케이션**
- 상호(interdependence)와 신뢰(Dependability)
- 팀구성원의 **성향에 따라** 달라지는 성과
- 성격유형 선호지표 MBTI 활용 ➡ 팀빌딩



IDEA

MBTI 원지알죠?

- Myers-Briggs Type Indicator
- 스위스 정신의학자 칼 구스타브 융의 성격 유형 이론 근거
- 이사벨 마이어스 & 캐서린 브릭스가 만든 지표 (1975)
- 한국에서 특히 큰 유행, 다양한 밈(meme) 및 콘텐츠

MBTI 성격유형

ISTJ 소금형 한번 시작한 일은 끝까지 해내는 성격	ISFJ 권력형 성실하고 온화하며 협조를 잘하는 사람	INFJ 예언자형 사람에 관한 뛰어난 통찰력을 가진 사람	INTJ 과학자형 전체를 조망해 비전을 제시하는 사람
ISTP 백과사전형 논리적이고 뛰어난 상황 적응력	ISFP 성인군자형 따뜻한 감성을 가진 겸손한 사람	INFP 잔다르크형 이상적인 세상을 만들어가는 사람들	INTP 아이디어형 비평적인 관점을 가진 뛰어난 전략가
ESTP 활동가형 친구, 운동, 음식 등 다양함을 선호	ESFP 사교형 분위기를 고조시키는 우호적인 성격	ENFP 스파크형 열정적으로 새 관계를 만드는 사람	ENTP 발명가형 풍부한 상상력으로 새로운 것에 도전
ESTJ 사업가형 사무적, 실용적, 현실적인 스타일	ESFJ 천선도모형 친절, 현실감을 바탕으로 타인에게 봉사	ENFJ 연변농숙형 타인의 성장을 도모하고 협동하는 사람	ENTJ 지도자형 비전을 갖고 타인을 강력적으로 인도

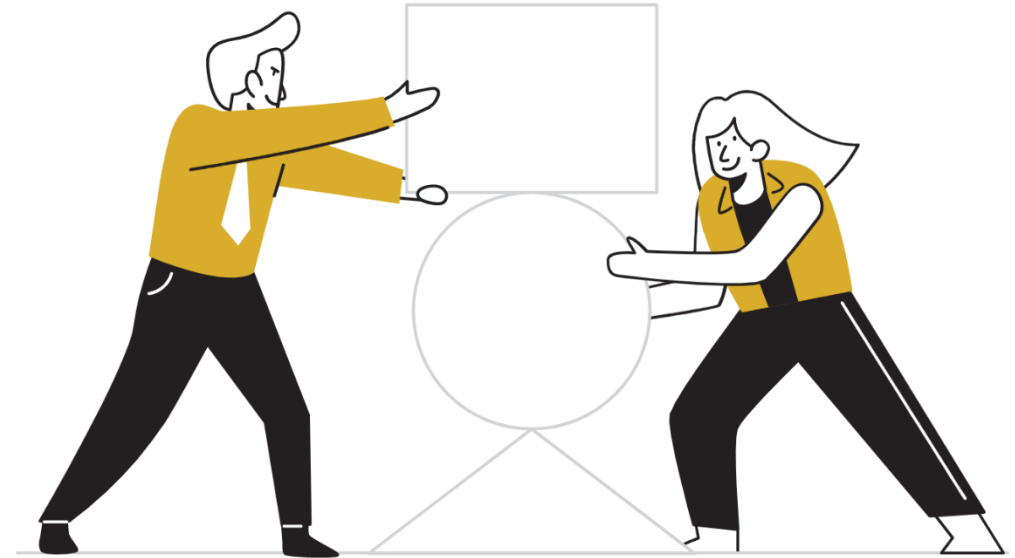
1. 선정 배경

MBTI profile 및 개인-조직 일치여부

선호성향		개인성격	조직성격	개인-조직 일치비율
주의초점	외향 (E)	321 (59.0%)	298 (54.8%)	57.2%
	내향 (I)	223 (41.0%)	246 (45.2%)	
정보수집	감각 (S)	409 (75.2%)	424 (77.9%)	71.5%
	직관 (N)	135 (24.8%)	120 (22.1%)	
의사결정	사고 (T)	360 (66.2%)	428 (78.7%)	68.4%
	감정 (F)	184 (33.8%)	116 (21.3%)	
대응양식	판단 (J)	411 (75.6%)	387 (71.1%)	68.0%
	인식 (P)	133 (24.4%)	157 (28.9%)	

출처 : MBTI에 의한 개인-조직 적합성과 조직효과성의 관계분석 박영미(2016)

* 대상 : 행정/경찰 공무원 544명



- 직무 적성 / 팀원, 상사 등을 고려한 인력 배치로 업무 효율성 증대
 - 개인-조직 일치비율이 높을수록 직무만족/조직에 대한 몰입이 높음
 - 하나의 management tool로서의 인재 유출 방지 기대
- (* 채용 시 지표가 아닌, 채용 후 팀 management을 통한 조직효과성 증진을 위함)

2. 대상 고객

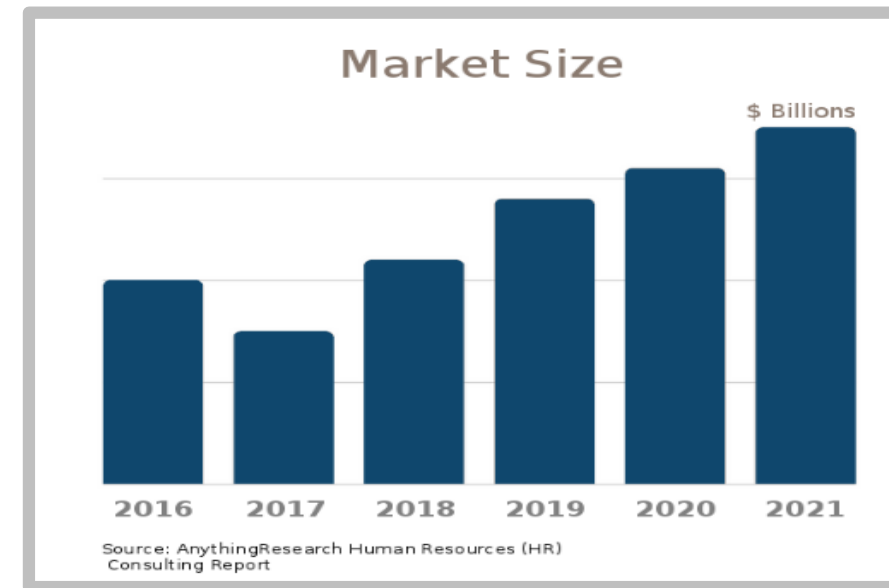
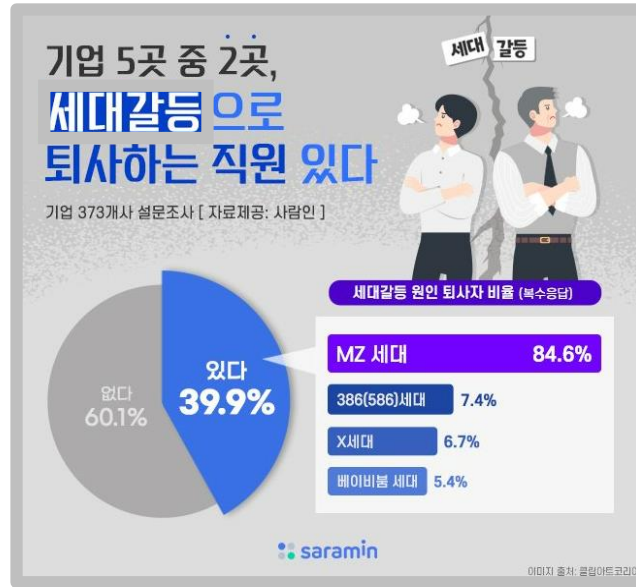
직장인 89.5%, 퇴사고민...이유 1위는 '상사 잔소리'

조사대상 : 직장인 665명

조사기간 : 20년 10월 19일 ~ 20일

Q. '입사 연차별' 퇴사를 유발하는 빌런(악당)은?

신입직원	상사 잔소리 (15.3%)
입사 3~4년차	대인관계 스트레스 (각 14.8%, 16.7%)
입사 5~6년차	연봉 (각 13.7%, 22.8%)
입사 25년차 이상	사내정치 (15.8%)



- 인간관계는 퇴사 주요한 사유
- 기업의 미래인 MZ세대의 경우 더욱 민감함
- 인력부서(HR)에서의 다양한 대응 컨설팅의 중요도가 높아짐
- 일상 언어 습관을 바탕으로 검사하면, 선택지 검사 등의 시간 절약 가능

3. 데이터 수집 / 활용

공정 일기>
아파도 쉬는 시간은 있으니까

Uni INFP 카톡스텝
1:1 채팅
2020.06.07. 00:42 조회 80

아프고 아픈 순간들이
연속인 듯 해도,
아름에서 헤어나오지 못할 것 같아도

오늘도 타이밍만 재다가 끝났네요...

ISFJ 불날 새음
1:1 채팅
2022.10.24. 19:03 조회 124

말걸고 싶고 커피 한잔 마시자고 하고 싶고
언제 말할까하다가 오늘도 끝났네요...

후다닥 퇴근하려 나가는 뒷모습을 보니까
이러는 사이에 연인이 생긴건 아닐까 혼자 상상하게 되는 저 자신을 보

전 자존감 높다고 생각하는데 그사람만 생각하면
마음이 작아지는 것 같아요.

그래도 힘내봐야죠!

```

import pandas as pd
file_name = "our_test.csv"
dataset = pd.read_csv(file_name, index_col=0)
dataset.head()

[68] ✓ 0.3s

...

```

	text	EN	mbti
0	저는 감정에 일상이 흔들리는 편이에요~ 썬 감정형이기도 하고, 존재로 지지받음이 적...	I tend to get distracted by my emotions.It's a...	INFP
1	아프고 아픈 순간들이 연속인 듯 해도, 아픔에서 헤어나오지 못할 것 같아도 자전거 ...	Even if the painful and painful moments are co...	INFP
2	한창 관계에서 받는 상처로 자존감이 바닥을 치면 조음, 아침의 장면이에요. 제 내면...	It's the scene in the morning when my sense of...	INFP
3	대화하는데 갑자기 토 할것 같은 느낌인데 오토바이를 보니까 속이 편해졌네요 근데 주...	I feel like I'm suddenly puking but I feel rel...	ESTJ
4	계단 올라갈때 뛰어가는게 편한 사람 있나요 계단은 뛰어가는게 속편이 편해요	Is it convenient to run when climbing the stai...	ESTJ

```

text = dataset.loc[:, "EN"].to_list()
mbti = dataset.loc[:, "mbti"].to_list()

text[11], mbti[11]

[70] ✓ 0.3s

...
("I want to say hi, and I want to ask you to have a cup of coffeeSeeing the back of my back office after my work was such a bless
I'll have to cheer you up!",
'ISFJ')

```

text	mbti	
안녕하세요 저는 istp 남자입니다 몇번을 검사해도 ist	ISTP	4
인팁 친구는 생각이 창의적이고 새로운 편이라 토론이	ISTJ	1
오늘 제 생일이랍니다 ~ 나이가 들면서 축하받는것도	ISFP	4
나는 1로 살아도 괜찮다 나를 1랑한다 나한테 고맙다	ISFJ	4
공황장애를 앓고 있거나 경험해본 분들은 아시겠지만..ISTJ	INTP	0
좋아해서 그런듯 해보입니다 ㅋㅋ	INFP	9
 만상 만나며 안겨두 무하느 부득 으크 만아유!	INFJ	1
멜로망스- 사랑인가봐(사내맞선 ost) 수많은 사람들이	INTJ	2
인터넷을 보다가 인''을 포기하는 한국인이 진짜 많다	ISTJ	9
말걸고 싶고 커피 한잔 마시자고 하고 싶고 언제말할까	ESFP	4
문든, 요즘 새로운 사람(동성,이성) 만나고 싶어서 직장	ESFJ	2
처음엔 애정표현도 그닥 없고 자주 보자고 하지도 않고	ESTP	4
사소한 질투같은거 하지않고 상대방이 누구 만나든, 뭘	INTP	2
이별한지 20일지났구 여자는 인스타로 여행도가고 운	FNFP	1
싸움... 까진 아니었지만 어제 여기에 올렸던 카톡내용	ENFJ	5
이틀 전 월요일 아침 새 현장에서 담당자 접선 대기하는	ENTJ	3
서로 호감 있는 것 같은데, 둘 다 연령대가 어느정도는		
여성 분들이 사내 임시직 남자 분께 이성적 관심을 표		
잔소리에 반박할 답변이 드디어 생각났네요		63

- MBTI 카페 데이터 크롤링
- 한국어 데이터셋 부족
- 영어 데이터를 살리기 위해 “번역” 진행
 - + 데이터500학습
 - + 구어체 추가 학습

4. 모델

한-영 번역기

연토님이 제공해준 best_model을 가지고 AI-hub Dataset인 25. 일상생활 및 구어체 한-영 번역을 가지고 fine-tune.

추가 Dataset: 1,200,306개 ko-en pair임.

```
!pip install transformers wandb sentencepiece
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: transformers in /usr/local/lib/python3.7/dist-packages (4.23.1)
Requirement already satisfied: wandb in /usr/local/lib/python3.7/dist-packages (0.13.4)
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.7/dist-packages (0.1.97)
Requirement already satisfied: tqdm<4.27 in /usr/local/lib/python3.7/dist-packages (from transformers) (4.64.1)
Requirement already satisfied: regex<2019.12.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (2022.6.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers) (3.8.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages (from transformers) (1.21.6)
Requirement already satisfied: huggingface-hub<1.0,>=0.10.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.10.1)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (21.3)
Requirement already satisfied: tokenizers<0.11.3,>=0.11.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.13.1)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from transformers) (2.23.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (6.0)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages (from transformers) (4.13.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from huggingface-hub<1.0,>=0.10.0->transformers) (4.1.1)
Requirement already satisfied: pyparsing<3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=20.0->transformers) (3.0.9)
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (5.4.8)
Requirement already satisfied: sentry-sdk<=1.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (1.9.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from wandb) (57.4.0)
Requirement already satisfied: BitPython<=1.0.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (3.1.29)
Requirement already satisfied: promise<3,>=2.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (2.3)
Requirement already satisfied: shortuuid<0.5.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (1.0.9)
Requirement already satisfied: protobuf<4.0,>=3.12.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (3.17.3)
Requirement already satisfied: docker-pycreds<0.4.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (0.4.0)
Requirement already satisfied: six<=1.13.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (1.15.0)
Requirement already satisfied: pathtools in /usr/local/lib/python3.7/dist-packages (from wandb) (0.1.2)
Requirement already satisfied: setproctitle in /usr/local/lib/python3.7/dist-packages (from wandb) (1.3.2)
Requirement already satisfied: Click<8.0.0,>=7.0 in /usr/local/lib/python3.7/dist-packages (from wandb) (7.1.2)
```

MBTI 분류기

```
[ ] # train 하게 할 데이터 셋 미개 첫번째 데이터셋
dt = pd.read_csv('drive/MyDrive/Colab Notebooks/MBTI_500.csv')

# validation 하게 할 데이터 셋 두번째 데이터셋 + 전처리
df_test = pd.read_csv('drive/MyDrive/Colab Notebooks/nbti_1.csv')

# 첫번째 시도
df_test = pd.read_csv('drive/MyDrive/Colab Notebooks/our_test.csv')

# 네이버를 포함한 두번째 시도
df_test = pd.read_csv('drive/MyDrive/Colab Notebooks/add_naver.csv')

# 슬랙 데이터
df_test = pd.read_csv('drive/MyDrive/Colab Notebooks/our_slack.csv')

# df_test = pd.read_csv('drive/MyDrive/Colab Notebooks/tean3_SLACK.xlsx')
```

```
[ ] for i in range(len(df_test['mbti'])):
    if not (df_test['mbti'][i] == "enfj" or df_test['mbti'][i] == "enfp" or df_test['mbti'][i] == "infj" or df_test['mbti'][i] == "isfp"):
        df_test['mbti'][i] = "entj"
```

/usr/local/lib/python3.7/dist-packages/pykernel_launcher.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
[ ] # 데이터 확인
df.head()
df_test.head()
```

Unnamed: 0	text	EN	NAVER	mbti
0	안녕하세요! 앞으로 함께 프로젝트를 진행하게 될 3월입니다! 반가워요!	Good day to you too! We are the three teams th...	Hello! We are the 3 teams who will work on the...	enfj
1	안지오남이랑 병창선님은 개인 사정으로 수업을 그만 들으신다고 합니다! 그리고 오늘 ...	I heard that the director and the Hwajangseon ...	Angjo and Bingchangsun are going to stop takin...	enfj
2	아 그리고 프로젝트 1,2는 팀별로 구분되다고 하니까 팀워크는 더 중요하겠네요!	Oh, and the 1/2 project is called a team P...	And project 1 and 2 are divided into team P...	enfj

- GPT Base-model 기반
- AI-hub 내 구어체 한-영 데이터셋
- Fine-tuned

- ML Supervised 모델
- Classification : Support Vector Machine
- (분류 과제를 위한 scikit-learn 참고)
- Cross validation시 CountVectorizer 및 SVM Linear 모델을 pipeline 으로 사용

5. 검증 및 결과

```
f1 score for Extraversion (E) - Introversion (I):  
0.8329686360320934  
f1 score for Sensation (S) - INtuition (N):  
0.878594249201278  
f1 score for Thinking (T) - Feeling (F):  
0.8406466512702079  
f1 score for Judgement (J) - Perception (P):  
0.8470124753775443
```

```
↳ f1 score for Extraversion (E) - Introversion (I):  
0.9290515309932785  
f1 score for Sensation (S) - INtuition (N):  
0.9251934651762683  
f1 score for Thinking (T) - Feeling (F):  
0.9407236335642801  
f1 score for Judgement (J) - Perception (P):  
0.9451137884872824
```

AdaBoost

```
▶ list_score_per_class = np.array(list_score_per_class)  
print('Mean score per classes: %s' % np.around(list_score_per_class.mean(axis=0),2))
```

↳ Mean score per classes: [0.74 0.75 0.76 0.78]

+ 코드

+ 텍스트

XGBoost

```
[134] list_score_per_class = np.array(list_score_per_class)  
print('Mean score per classes: %s' % np.around(list_score_per_class.mean(axis=0), 2))
```

Mean score per classes: [0.74 0.75 0.76 0.78]

AdaBoost :

- 과적합의 영향 축소
- 분류기 실수를 반복 수정

XGBoost:

- 과적합 규제
- Classification and Regression Tree
양상블 모델로 분류 및 회귀의 예측

6. 활용 : slack 대화 학습을 통한 MBTI 유형 파악으로 팀 규칙 생성

슬랙 대화 +

일정이너무 타이트 10월 24일 월요일 발표하고 피드백 받아서 진행하는게 좋을지..

개인 SNS (insta, facebook) 분석해서 mbti와 같이 성향 비슷한 사람 찾아주는? 서비스 이런쪽이 좀더 설득력 있어보이네요

👍 1 😊

k 김응진 (자연어처리 7회차) 오후 10:59

지금 그냥 속 지나가는 아이디어는 회사들이 보통 sns 주소도 받기도 하는데 오히려 그 sns 를 분석하는걸 도와주는 툴로 보면 좋겠네요

Y 유한솔(자연어처리 7회차) 오후 11:00

좋은 생각이신데요! (편집됨)

k 김응진 (자연어처리 7회차) 오후 11:01

특히 미국같은 경우는 무조건 선택사항으로 개인 sns 받는데 그 영향으로 우리나라도 조금씩 받는 것 같더라고요 그래서 장기적으로 보면 트렌드에 맞게 더 효율적으로 분석할 수 있게 도와주는걸로 하면 어떨까요

↓ 최근 메시지

B I S | | | | | </> ↻

김응진 (자연어처리 7회차), 김준철 (자연어처리 6회차), 유한솔(자...

+ | 📎 🗑️ | 😊 @ Aa ▶ | ▾

<TEAM3의 PROJECT3 규칙>

1. 내성적인 사람들에게 의견을 한 번 더 물어본다.(I를 위한 규칙)
2. 새로운 아이디어에 대해 직감적인 선호도를 확인한다.(FP를 위한 규칙)
3. 회의시간은 최소 전날 계획하여 공유하고 늦지 않는다.(J를 위한 규칙)
4. 작은 아이디어와 비현실적인 상상력에도 쓸데없다고 하지 말고 존중해준다.(N을 위한 규칙)
5. 프로젝트에서 의견이 일치되지 않거나 반대되는 생각일 때 감정을 정리하고 의견 공유한다.(F를 위한 규칙)
6. 비유와 은유보다 상대가 이해하기 쉬운 아이디어를 전달하려고 노력한다.(NP를 위한 규칙)
7. 프로젝트 진행 시 다양한 변수가 발생할 수 있으므로 이전 계획과 다르게 진행되어도 팀원들과 충분히 상의해서 극복하려고 노력한다. (J를 위한 규칙)
8. 외향적인 성향의 자유를 보장한다. (E를 위한 규칙)
9. 다른 의견을 갖더라도 공감하며 팀원들과 대화하여 더 좋은 방향으로 한다. (T를 위한 규칙)
10. 우리 팀에 기여된 주도적 성향을 채우기 위해 서로 회의 전에 충분히 대화하며 분위기를 풀고 자유롭게 이야기할 수 있는 분위기를 형성한다.

(슬랙 채팅 크롤링)

7. 데모 시연

- <https://acb95ff06ad63926.gradio.app>

8. 한계 및 추후 진행 방향

1. 데이터 수집

- 카페 데이터 내, 질문글, 홍보글, 데이터 삭제, 데이터 접근 시 일정 멤버 등급 이상 필요
- 기업이 보유한 자소서 등에 대한 자신을 나타내는 자료가 있으면..

2. 특정 MBTI에 대한 자료가 적음

→ 모델 개선을 위한 추가 데이터 수집 및 특정 MBTI의 추가 학습자료 확보

Q & A

감사합니다.