

# AI Project

# MBTI Personality

# Predictor

Team#01



# Team

Boseong Kwon

Minjae Kim

Jungsik Kim

Hyunsoo Kim

Jubong Park

Youngseok Yoo



# CONTENTS

01

Project Background including related work

02

Goal and what is new?

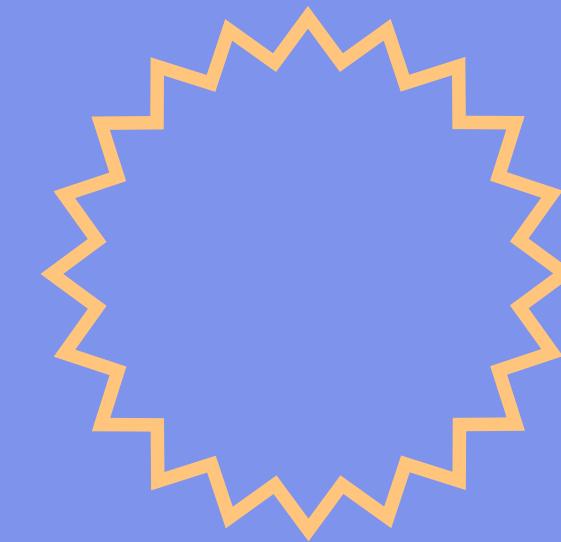
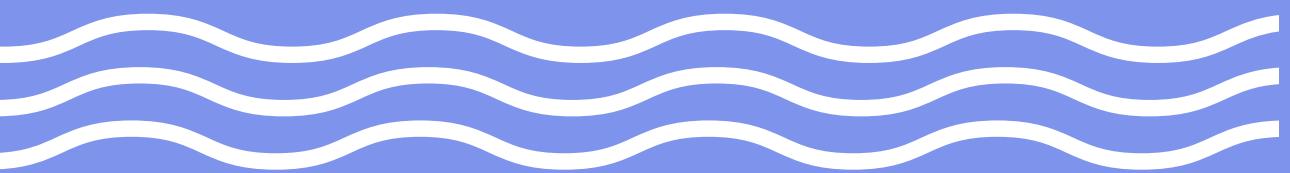
03

Dataset

04

Evaluation methods and baselines for comparison.

# 01



Project Background including  
related work

# MBTI

personality assessment which categorizes people into 16 different types

E vs I

Extroversion  
Introversion

T vs F

Thinking  
Feeling

S vs N

Sensing  
Intuition

J vs P

Judging  
Perceiving

# How?



## Common Way

Answering test questions.

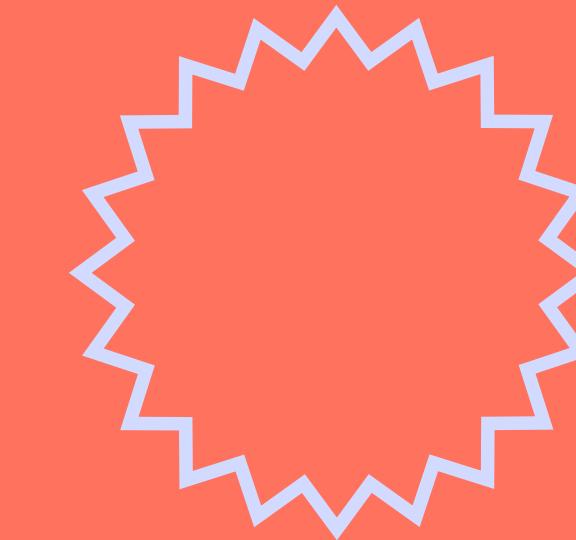
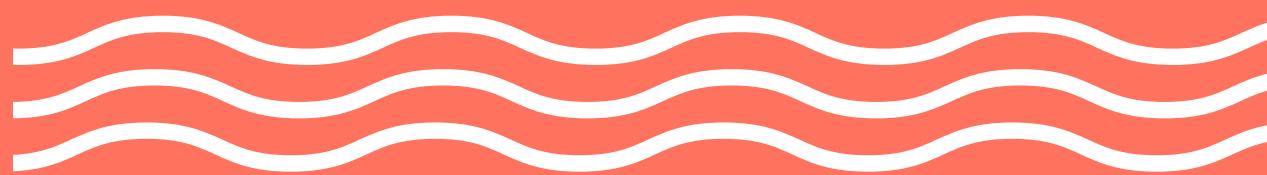
Relies on individual's subjective measurement.

So,

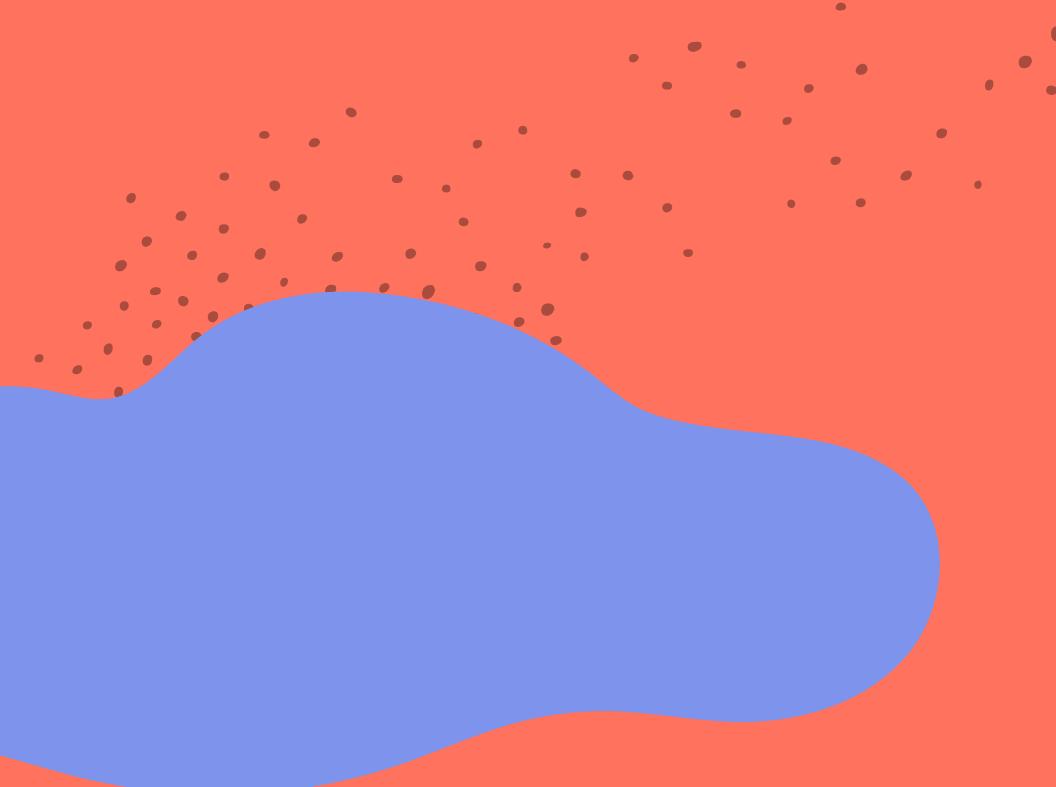
In SNS post, people are more natural and reflective.

There have been some tries to classify person's MBTI using what they post on SNS.

# 02



Goal and What is new?



# Goal

Build a model which achieves much higher accuracy than any other previous models.

# What is new?



## BERT

BERT is a recent pre-trained natural language processing model developed by Google



## Datasets

Generate our own datasets by collecting posts from our acquaintances with their agreements

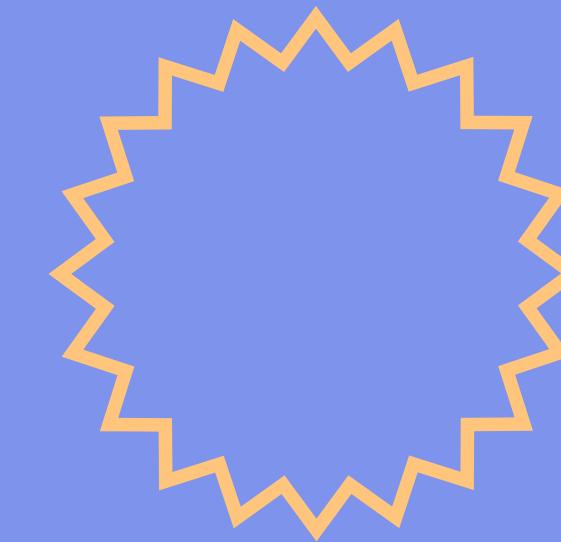


## Contrastive Learning

Use contrastive learning for learning representations by similarity of each MBTI pairs



# 03



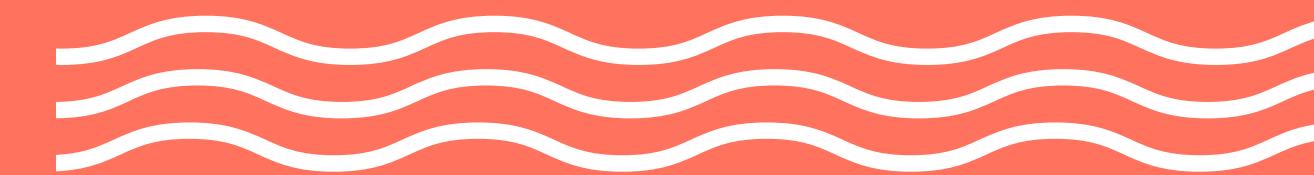
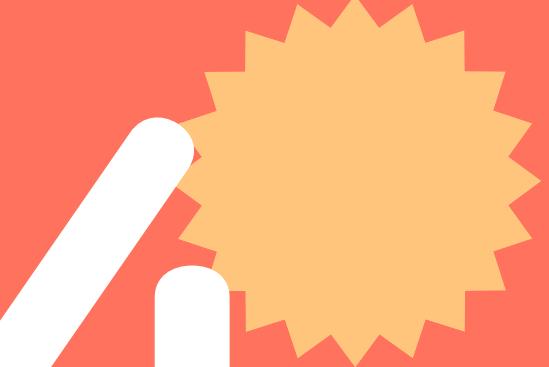
Dataset

# Dataset

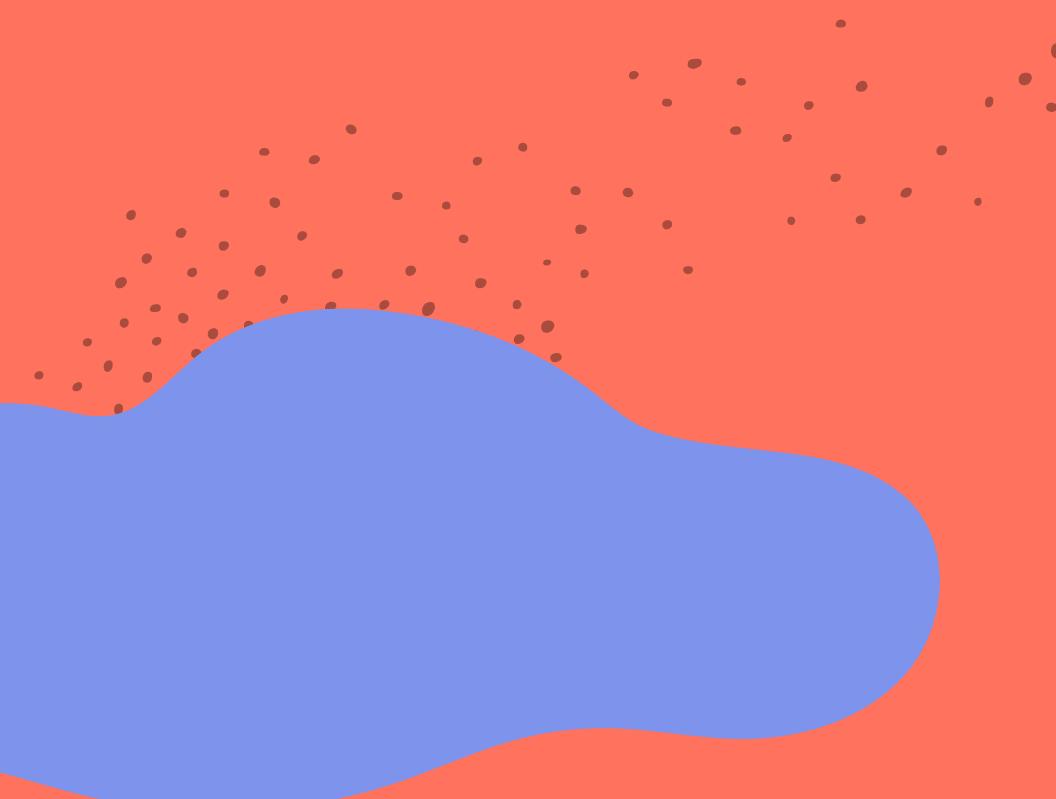
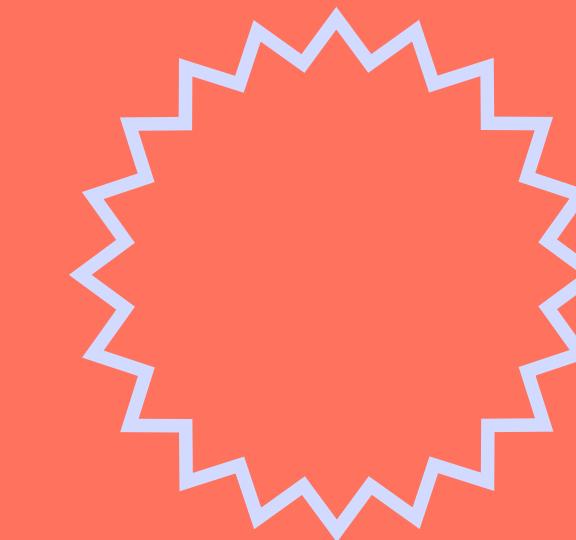
- Dataset contains over 8600 rows of data
- Each row is a person's MBTI and a section of last 50 things they have posted
- Collected through Personalitycafe forum

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw  ...  '
1	ENTP	'I'm finding the lack of me in these posts ver...  '
2	INTP	'Good one ____ https://www.youtube.com/wat...  '
3	INTJ	'Dear INTP, I enjoyed our conversation the o...  '
4	ENTJ	'You're fired.   That's another silly misconce...  '

# 04



Evaluation methods and  
baselines for comparison



# Evaluation Method

- Model will be scored on mean column-wise log loss across each of the four categories
- Lower the score, the higher accuracy is.

# Comparison Baseline

- Baseline for comparison includes all previous models submitted to Personality Profile Prediction competition on Kaggle.

# THANK YOU

