



텍스트 데이터 기반 성격 유형 분류 모델 서비스 기획서

Executive Summary

본 기획안은 데이터 기계학습을 통한 인공지능 사업으로서
사용자 **성향 분석** 서비스 및 **활용 방안**을 제안한다.

사용자의 다양한 **텍스트 데이터**를 기반으로
성격 유형을 분류하는 **Multi Label Classification** 구현

Bi_GRU, CNN, HAN 모델 앙상블을 통해
Accuracy : **87.6%** , F1_score : **91.2%** 달성

고객 자체 성향 분석 서비스, 추천시스템 성능 개선 등
다양한 파생 서비스 활용 가능

CONTENTS

1. Business Introduction

2. Data Processing

2-1) Total Architecture

2-2) Crawling

2-3) Preprocessing

3. Data Modeling

3-1) Algorithm

3-2) Comparison and Analysis

4. Prototype

5. Service

(1) 성격 검사 현황

성격 검사의 간단한 역사

- 성격 검사는 제1차 세계대전 중 ‘셀 쇼크’에 가장 취약한 군인을 예측하기 위해 미 육군이 최초로 사용했다. (셀 쇼크: 전투 상황에서 신체적·정신적으로 견딜 수 없을 만큼 불안감이 극에 달하는 현상)
- 해마다 직장인 수백 만명이 인사 선발, 협업 및 팀워크 향상, 진로/적성 탐색을 목적으로 성격 검사를 받고 있으며, 매년 10~15%씩 성장하는 5억 달러 규모의 산업이 됐다.
- 성격검사 산업의 활발한 성장은 이러한 검사를 노동력 최적화의 도구로 신뢰하는 관리자들이 늘고 있다는 사실을 암시한다.

출처: Harvard Business Review [A Brief History Of Personality Tests]
Eben Harrel (Senior Editor)

성격 검사의 목적 및 국내 현황

- 개인은 자신의 성향을 파악하고 주로 그에 맞는 진로 선택을 위해 성격 검사를 이용하고 있다.
- 국내 기업들은 성격 검사를 통해 그들의 성향을 파악하고 조직에 잘 적응할 수 있는 사람인지 판단하여 자신들이 원하는 스타일의 인재를 선별한다.
- 기업 입사 과정 중 시험 또는 면접의 형태로 성격/인성 검사가 실시되고 있다.



1. Business Introduction

(2) 성격 검사의 문제점과 User Text Data

기존 성격 검사의 문제점

- **비용 발생** - 성격 검사를 실시하는 데에 많은 자원이 소요된다. (시간, 돈)
- **객관성 부족** - 개인(지원자)은 자신의 목적이나 지원한 기업의 인재상에 따라 자신의 답변을 준비해가거나 문제의 답안을 선택할 수 있다.
(실제로 인성 검사/면접의 가이드라인을 준비해주는 강의/코칭도 존재)

이러한 문제점들을 사용자 텍스트 데이터를 활용하여 해결하고자 한다.

사용자 텍스트 데이터

- 개인 사용자가 온라인 공간에서 활동을 하며 남긴 데이터는 물론 **개인이 작성한 텍스트 형태의 모든 데이터** (커뮤니티 게시판, 상품/영화 리뷰 등)
- 화자의 의도적/비의도적 맥락을 모두 포함하고 있으나 자연어 형태로 작성되어 있는 "**비정형**" 데이터이기 때문에 이를 분석하기 위한 **방법론이 필요하다**.

→ 사용자의 텍스트 데이터를 활용한 **성격 유형 분류 모델**을 구현하였고, 성격 분류 유형(Label)은 대중적으로 널리 알려진 **MBTI**를 사용하였다.

1. Business Introduction

(3) MBTI 설명

MBTI란?

- C. G. Jung의 심리유형론을 근거로 일상생활에 유용하게 활용할 수 있도록 고안한 자기 보고서 성격유형 지표
- 현재 세계에서 가장 널리 사용되는 심리검사 중 하나

MBTI 척도 4가지

외 향(E) Extraversion	에너지의 방향(주의초점)	내 향(I) Introversion
감 각(S) Sensing	인식기능(정보수집)	직 관(N) Intuition
사 고(T) Thinking	판단기능(결정, 선택)	감 정(F) Feeling
판 단(J) Judging	행동양식	인 식(P) Perceiving

MBTI 성격 유형 16가지

ISTJ Responsible, sincere, analytical, reserved, realistic, systematic. Hardworking and trustworthy with sound practical judgment.	ISFJ Warm, considerate, gentle, responsible, pragmatic, thorough. Devoted caretakers who enjoy being helpful to others.	INFJ Idealistic, organized, insightful, dependable, compassionate, gentle. Seek harmony and cooperation, enjoy intellectual stimulation.	INTJ Innovative, independent, strategic, logical, reserved, insightful. Driven by their own original ideas to achieve improvements.
ISTP Action-oriented, logical, analytical, spontaneous, reserved, independent. Enjoy adventure, skilled at understanding how mechanical things work.	ISFP Gentle, sensitive, nurturing, helpful, flexible, realistic. Seek to create a personal environment that is both beautiful and practical.	INFP Sensitive, creative, idealistic, perceptive, caring, loyal. Value inner harmony and personal growth, focus on dreams and possibilities.	INTP Intellectual, logical, precise, reserved, flexible, imaginative. Original thinkers who enjoy speculation and creative problem solving.
ESTP Outgoing, realistic, action-oriented, curious, versatile, spontaneous. Pragmatic problem solvers and skillful negotiators.	ESFP Playful, enthusiastic, friendly, spontaneous, tactful, flexible. Have strong common sense, enjoy helping people in tangible ways.	ENFP Enthusiastic, creative, spontaneous, optimistic, supportive, playful. Value inspiration, enjoy starting new projects, see potential in others.	ENTP Inventive, enthusiastic, strategic, enterprising, inquisitive, versatile. Enjoy new ideas and challenges, value inspiration.
ESTJ Efficient, outgoing, analytical, systematic, dependable, realistic. Like to run the show and get things done in an orderly fashion.	ESFJ Friendly, outgoing, reliable, conscientious, organized, practical. Seek to be helpful and please others, enjoy being active and productive.	ENFJ Caring, enthusiastic, idealistic, organized, diplomatic, responsible. Skilled communicators who value connection with people.	ENTJ Strategic, logical, efficient, outgoing, ambitious, independent. Effective organizers of people and long-range planners.

사람들의 성격을 4가지 척도를 기준으로 16가지 유형으로 분류

1. Business Introduction

(4) 텍스트 데이터 기반 모델

	기존 성격 검사 서비스 (설문, 인성 검사)	텍스트 데이터 기반 성격 유형 분류 모델
소요 시간	수 십 분에서 수 시간	수 십 초에서 몇 분
서비스 형태	설문, 검사지 작성	텍스트 입력
서비스 비용	수 만원	-
객관성	응답자가 상황에 맞추어 조작 가능	사용자 텍스트 데이터의 맥락을 통한 검사 방식으로 조작 불가 (정확도 90% 이상)

텍스트 데이터 기반 성격 유형 분류 모델은
기존의 검사 방식에 비해 **월등한 성능과 이점들**을 가지고 있다. (위 표)



탈락 자소서

[무한한 도전을 함께할 동반자]

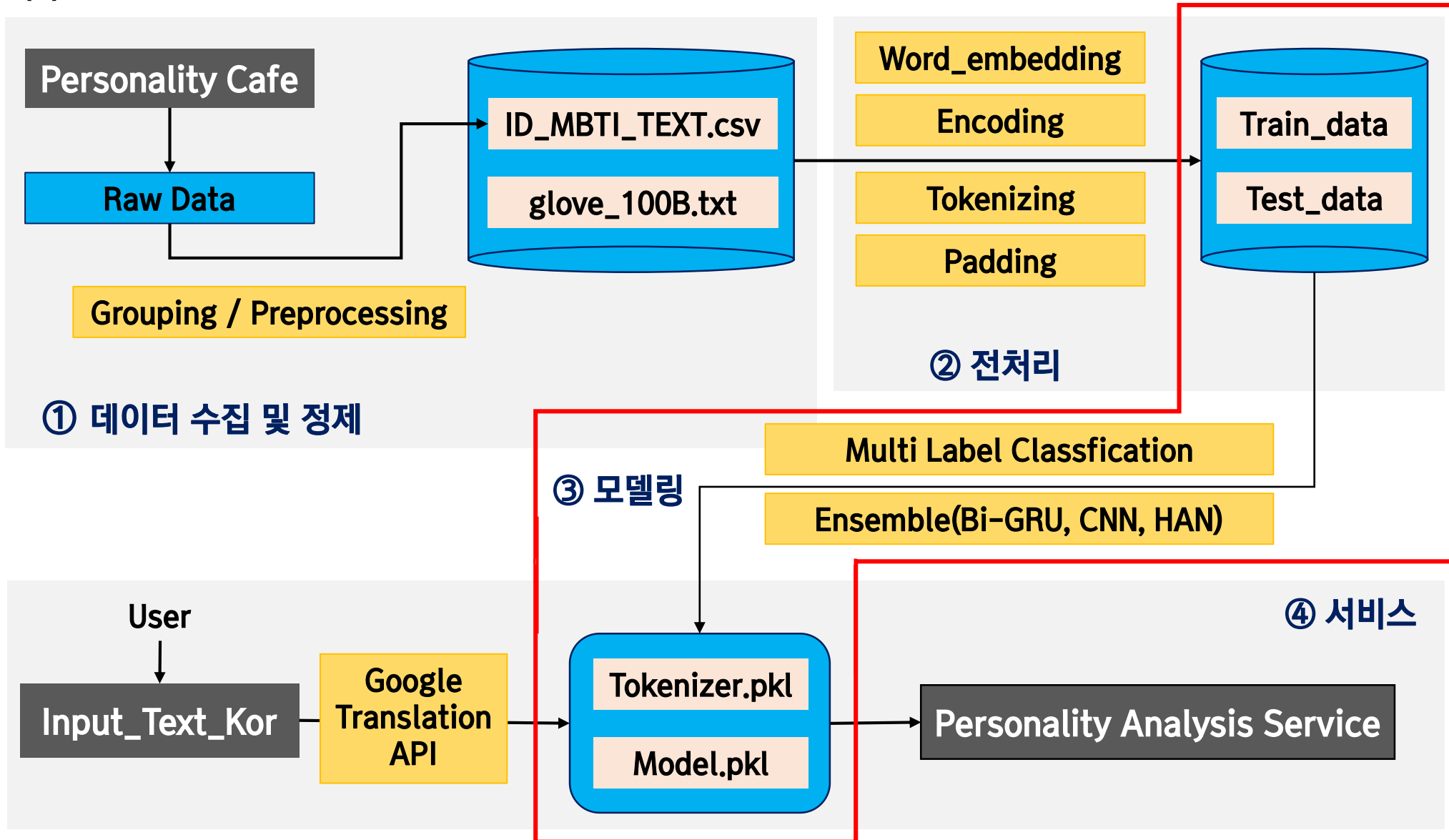
끊임없는 도전을 함께할 수 있는 회사를 찾고 싶습니다. 저는 고등학교 시절 공부에 흥미도 없고 잘하지도 못하는 학생이었습니다. 그러던 중에 고등학교 2학년 시절 학교 선생님이 "20년 안에 모든 자동차의 엔진이 다 없어질 것이고 엔진 관련 회사는 모두 망할 것이다"라는 말에 굉장한 흥미가 생겼고 관련 대학 학과에 진학하고 싶은 욕심이 생겼습니다. 그래서 굳은 결심을 하고 공부라고는 해본 적 없는 제가 1년 남은 기간 학원에 다니면서 공부를 시작했고 주위 친구들과 선생님들조차 무정적으로 바라봤습니다. 저는 공고에 다니고 있었기 때문에 목표로 하고 있는 서울 4년제 대학은 누구도 진학한 적이 없습니다. (하락)



다양한 데이터(리뷰 데이터, 자소서 등)를 Input으로 사용하여
추천 시스템 고도화, 자기소개서 검토용 등의 목적으로도 활용이 가능하다.

2. Data Processing

(1) Total Architecture



2. Data Processing

(2) 데이터 수집

크롤링 사이트 소개


















































Personality Cafe

- **MBTI 성격 유형으로 분류된 사용자**들이 여러 주제에 관한 게시글을 올리는 커뮤니티
- 특정 주제의 글에 대하여 자신의 의견을 답글로 달며 토론이 활발히 진행됨

↳ 16가지 성격 유형으로 Labeling된 텍스트 데이터(게시글) 수집 가능

Topics of Interest	
Title	Last Post
General Psychology Discuss psychology related issues here. Psychology News Forum	Can people with PDs be properly typed? by series0 Yesterday 10:21 PM
Trends Forum This is the place to discuss/predict the latest trends or past trends. Yes, let's hope Pikachu and mullets come back! Better yet, let's BRING it back!	How Pippi has inspired Enigma.... by Pippi Yesterday 10:10 PM
The Art Museum Post your artistic visuals for all to see.	Explain Your Avatar by Suntide Yesterday 10:04 PM
Book, Music, & Movie Reviews Review or discuss some of your favorite books, movies, and music in this forum. Video Games	What Song Are You Listening To Now? by Kaznos Today 01:28 AM
The Debate Forum The place for intelligent discussions.	No Permissions To View Last Post Info In This Forum
Current Events Talk about current events that are happening around the world.	No Permissions To View Last Post Info In This Forum
Education & Career Talk A forum for students and the working.	Is your current job one that is recommended for... by ImminentThunder Yesterday 04:20 AM
Science and Technology A forum for discussing science and technology breakthroughs.	Can you recommend Wordpress tutorials? by Lucan1010 08-01-2019 12:02 PM
Health and Fitness Let's get physical! Share tips on exercising and how to maintain a healthy lifestyle.	1000 calories A DAY by Duo Today 01:04 AM
Sex and Relationships The forum for sex and relationship advice and talk. *Contains Explicit Talk, Viewer Discretion Is Advised.*	Romantic Confession Time by Skeletalz Today 01:22 AM

Community: General Chat		Forum Tools		Search Forum
Title / Thread Starter		Replies / Views		Last Post By ▼
Sticky Threads				
	 Sticky: How are you today? Started by WickedQueen, 07-05-2009 08:01 PM 1 2 3 ... 1555	★★★★★ 	Replies: 15,547 Views: 541,533	Firelily Yesterday, 11:30 PM 
	 Sticky: What are you thinking about? Started by Tad Cooper, 03-08-2014 02:43 PM 1 2 3 ... 7989	★★★★★ 	Replies: 79,880 Views: 1,736,595	Suntide Yesterday, 08:34 PM 
	 Sticky: What do you currently desire? Started by NeonBomb, 06-03-2011 02:14 PM 1 2 3 ... 1135	★★★★★ 	Replies: 11,340 Views: 385,065	Dr Robotnik Yesterday, 08:19 PM 
	 Sticky: Post a random Youtube video (interesting, amusing - whatever) Started by heartturnedtoporcelain, 05-03-2010 06:36 PM 1 2 3 ... 1099	★★★★★  	Replies: 10,989 Views: 286,022	Forest Nymph Yesterday, 09:31 AM 
	 Sticky: What did you last eat? Started by TheWaffle, 11-30-2010 04:01 PM 1 2 3 ... 1649	★★★★★  	Replies: 16,482 Views: 621,691	Dr Robotnik 07-30-2019, 06:48 PM 
	 Sticky: What are you doing right now? Started by Happy, 06-20-2009 01:24 AM 1 2 3 ... 2179	★★★★★ 	Replies: 21,788 Views: 871,639	Xanthus Primus 07-30-2019, 03:58 PM 
	 Sticky: What did you google? Started by Promethea, 04-01-2012 05:58 PM 1 2 3 ... 190	★★★★★ 	Replies: 1,895 Views: 130,949	PoikaMies 07-24-2019, 04:59 AM 
	 Sticky: The Happy Birthday Thread :D Started by Cerulean, 04-06-2009 08:43 PM 1 2 3 ... 33	★★★★★  	Replies: 329 Views: 51,507	PoikaMies 07-24-2019, 04:41 AM 
	 Sticky: The "Help me find my Avatar (or other image)" thread Started by Maybe, 06-03-2019 05:29 PM		Replies: 7 Views: 214	Kazos 07-04-2019, 01:32 PM 
	 Sticky: The all-MENTIONS-enabled Compliment a Member thread Started by Cheshire Tower, 08-22-2014 05:37 PM 1 2 3 ... 80	★★★★★ 	Replies: 794 Views: 72,242	Froody Blue Gem 06-16-2019, 09:11 AM 
	 Sticky: The "Help me find this Thread" Thread. Started by starri, 04-13-2014 12:38 PM 1 2 3	★★★★★  	Replies: 20 Views: 8,306	Mmmm 12-06-2018, 09:42 AM 

↳ 특정 분야에 관한 의견부터 흥미 위주의 주제, 일상적인 대화 소재에 대한 데이터 추가 수집이 가능

2. Data Processing

(2) 데이터 수집

1. 03-20-2019 11:05 AM #3

2. SummerRoads
3. ESFJ
Progress 33%

4. I don't know that I can point to one or two specific people that I consider role models as much as I notice small things people do that I admire. I will notice a gesture of kindness or an act of selflessness someone does and then I'll try to incorporate that into the way I treat others as well. I admire people whose actions and words align and who are consistently a good person and aren't arrogant or pompous. I appreciate other people who work on being inclusive and considerate. In a leadership setting, I appreciate those who are organized and give clear directions of their expectations and deadlines and know how to say thank you to their employees from time to time.

Honesty, loyalty, kindness, empathy, selflessness, organizational skills, and humor are all character traits I admire.

Gender: ♀
Posts: 332
Join Date: Feb 2018
Post Rank: Enthusiast
Thanked: 398 times

수집 내용

- 1. 글 게시 시각
- 2. 유저 ID
- 3. 유저 MBTI 유형
- 4. 게시글 내용(텍스트)

SimplyEnigmatic and Suntime thanked this post.

You need 15 posts to view signatures.

Report Post | Reply | Reply With Quote

↳ 총 42,081 명의 유저가 게시한
3,714,918 개의 텍스트 데이터(게시글) 수집

[Python] BeautifulSoup, requests 패키지 사용

2. Data Processing

(3) 데이터 전처리 - grouping

[표 1] Raw data
3,714,918개의 게시글

name_list	mbti_list	text_list	datetime_list
SummerRoads	ESFJ	Like the title says. Seems like most other typ...	02-21-2019 05:22 PM
SimplyEnigmatic	INTP	I'll get this thread started :PWhat traits do ...	03-20-2019 08:40 AM
SummerRoads	ESFJ	I don't know that I can point to one or two sp...	03-20-2019 12:05 PM
orion83uk	ESFJ	For me, its people who are not frightened to S...	03-23-2019 02:22 PM
SimplyEnigmatic	INTP	How often are you approached by someone new?Wh...	03-27-2019 11:37 PM
SummerRoads	ESFJ	These questions are confusing to me. Maybe tha...	03-28-2019 10:39 AM
SimplyEnigmatic	INTP	I was trying to be quite general. I'm sorry fo...	03-29-2019 12:05 AM
HIX	ESFP	Why is the sky blue ???	03-29-2019 05:49 PM
Katie Tran	ENFP	Yes! This makes me so happy. More ESFJ represe...	03-29-2019 06:01 PM
SummerRoads	ESFJ	I would pick the one that never changed. I wan...	03-29-2019 06:36 PM



[표 2] ID_Grouping data
42,081명의 게시글

ID	MBTI	TEXT
0	INTP	Is this your natural eye color, or do you wear...
1	INTJ	It occurs to me that if I didn't have Ni playi...
2	ENTJ	Name: Soured Lie is an anagram of my real name...
3	INFP	"I don't understand why you are scared to driv...
4	INTP	The ringing in my ear is almost drowning out t...
5	INFJ	Oh, I'm not ISTJ, but I really like you guys s...
6	ISFP	Currently on my pandora I got Viva La Vida by ...
7	INFJ	You know you're an INFJ when someone tells you...
8	ISFP	From what I have been reading, it seems like m...
9	ENTJ	k Indifferent, but I agree with the governme...
10	INTJ	My tenant, while introducing her new cat to me...

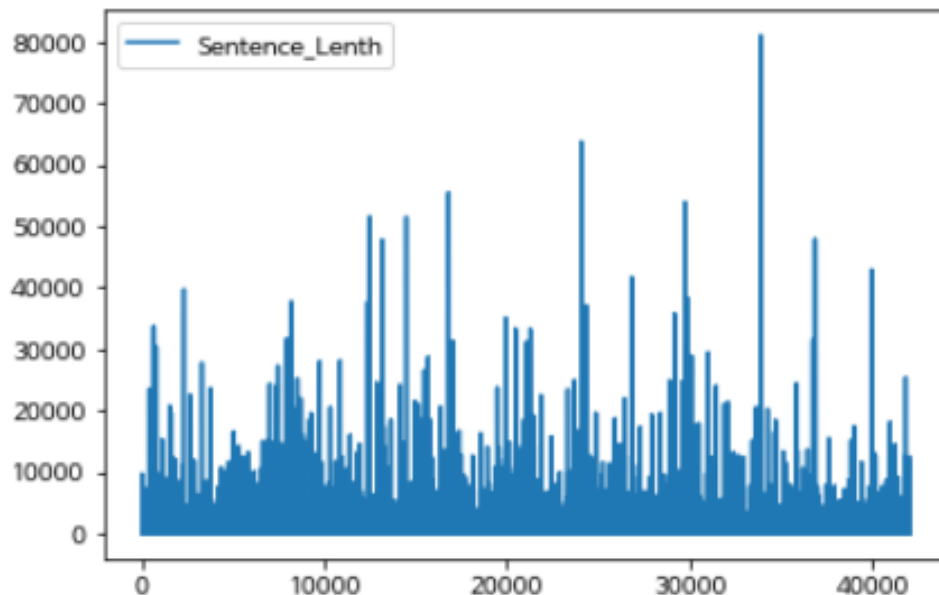
유저 ID를 기준으로 수집된 텍스트 데이터 그룹핑

2. Data Processing

(3) 데이터 전처리 – 문장 길이

유저별 게시물 및 문장 길이가 10개미만에서 40,000개 이상 차이가 있음

=> 문장 개수가 100개 미만인 유저 제거, 문장 개수가 200개 이상인 유저 분할



[그림 1] 유저별 게시물 길이 분포

MBTI	TEXT
INTP	[natural eye color wear contact lenses, me hig...
INTP	[concerning photo sky tried couple time make s...
INTP	[may post later, yes know thread posted well,...
INTP	[i get angry quite rarely safer surrounding pe...
INTP	[think darwin theory year old theory, thing ...
INTP	[again smile silly carefree want to, pleasure...
INTJ	[i like drawing picture head imagining people ...
INTJ	[dear everywhere nondescript social commentary...
INTJ	[i ponder big questions, gonna grab corporate...
INTJ	[so would say kinda something or, it joke pred...
INTJ	[i mind cute post is, sister yard sale moving...

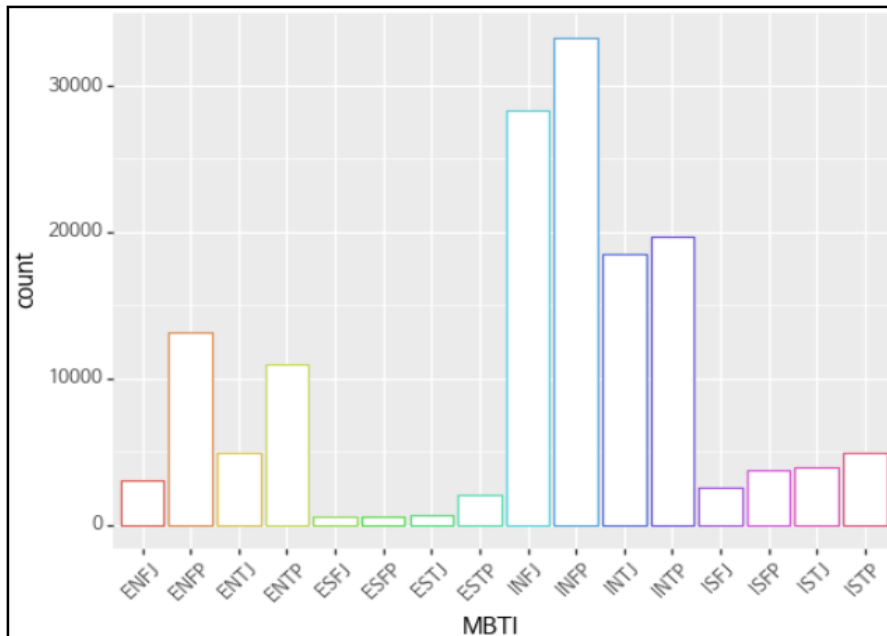
[표 3] 게시물 길이에 따라 데이터 재정립
총 **150,668**개 데이터

2. Data Processing

(3) 데이터 전처리 – TASK 정의 [1/2]

총 16개의 MBTI를 분류하는 Multi Class Classification으로 볼 경우
다중 클래스 분류 예측의 어려움과 **극심한 데이터 불균형** 문제 발생

⇒ Mbti는 I 또는 E, N 또는 S, F 또는 T, J 또는 P
네가지 변수의 이진분류이기 때문에 **Multi Label Classification**으로 문제 재정의



[그림 2] MBTI별 분포

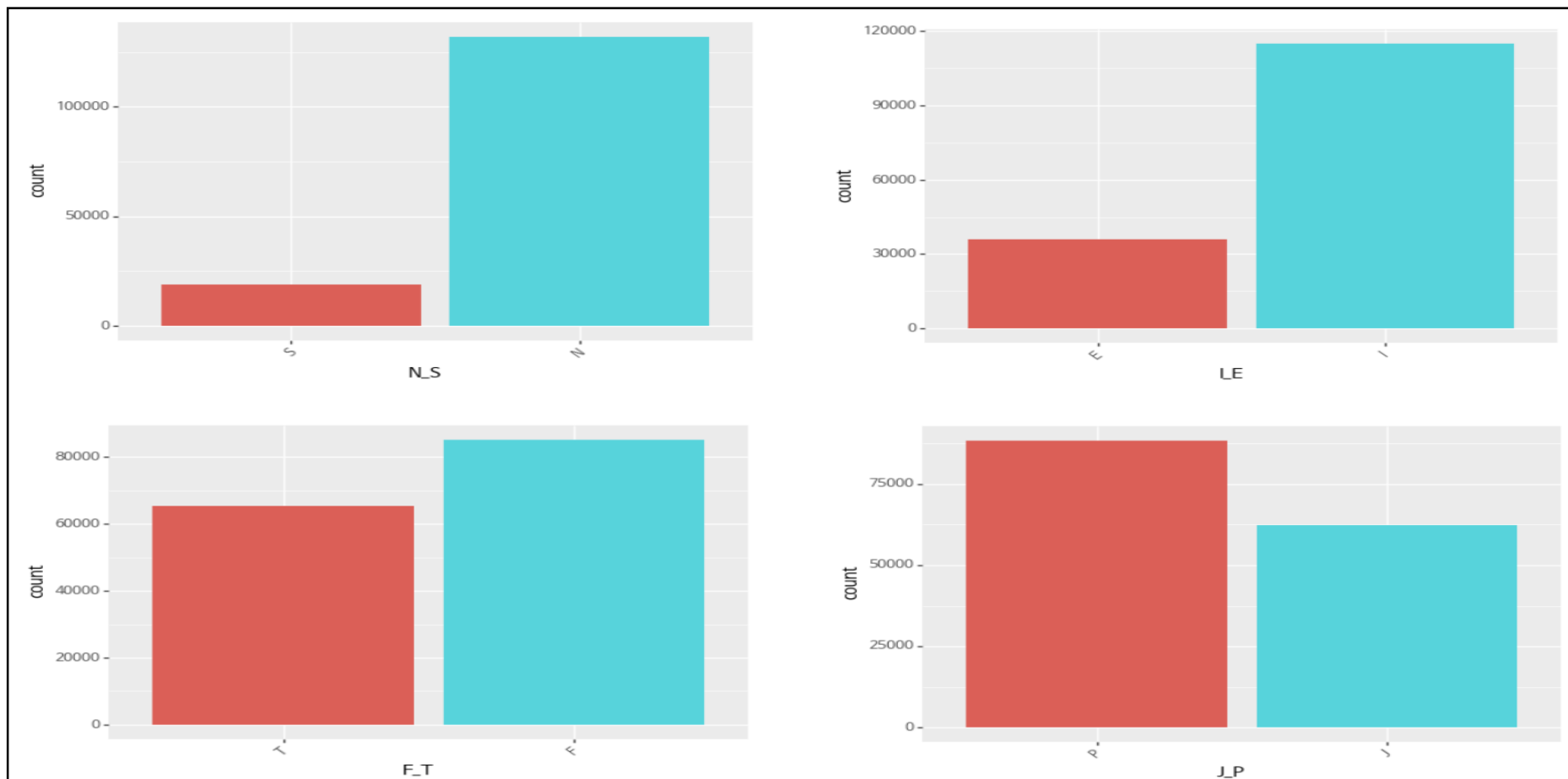
TEXT	I_E	N_S	F_T	J_P
Is this your natural eye color, or do you wear...	True	True	False	False
It occurs to me that if I didn't have Ni playi...	True	True	False	True
Name: Soured Lie is an anagram of my real name...	False	True	False	True
"I don't understand why you are scared to driv...	True	True	True	False
The ringing in my ear is almost drowning out t...	True	True	False	False
Oh, I'm not ISTJ, but I really like you guys s...	True	True	True	True
Currently on my pandora I got Viva La Vida by ...	True	False	True	False
You know you're an INFJ when someone tells you...	True	True	True	True
From what I have been reading, it seems like m...	True	False	True	False
k Indifferent, but I agree with the governme...	False	True	False	True
My tenant, while introducing her new cat to me...	True	True	False	True

[표 4] Multi Label Classification

2. Data Processing

(3) 데이터 전처리 – TASK 정의 [2/2]

N_S, I_E 의 경우 여전히 imbalance 문제가 있지만
평균적으로 다중 클래스 불균형 문제보다 완화 된 것을 확인



[그림 3] 각 Label 별 분포

2. Data Processing

(3) 데이터 전처리 – encoding

[“Temporarily helpful, but not to the perpetual hater.”,
“Like - seriously - one morning I hated an individual in a coffee shop because she was unkind.”,
“I expected her to behave better.”
“Hate exists, in my case, because I'm not ready or am unwilling to move forward with shedding certain beliefs.”]

Normalizing & Tokenizing

불용어 및 특수문자 제거 / 어간 및 표제어 추출 / max_features = 100,000

[["temporarily", "helpful", "perpetual", "hater"],
["like", "seriously", "one", "morning", "hated", "individual", "coffee", "shop", "unkind"],
["expected", "behave", "better"],
["hate", "exists", "case", "ready", "unwilling", "move", "forward", "shedding", "certain", "beliefs"]]

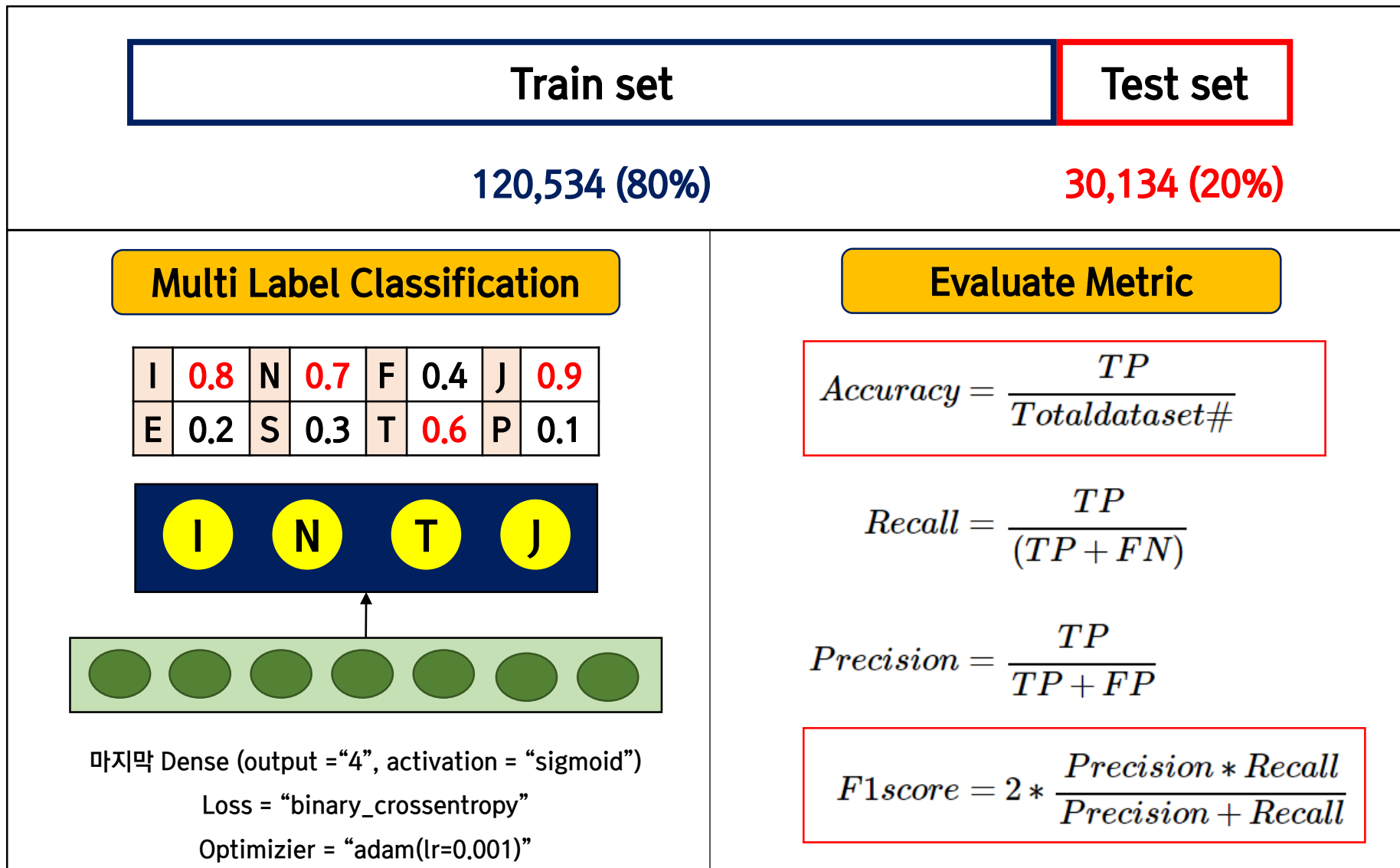
Pretraining

glove.6B.100d / padding / 최대 문장 길이 = 20 / 최대 문장 수 = 200

[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4539, 948, 6689, 7579],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 468, 6, 805, 1497, 414, 1076, 1869, 9087],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1208, 1778, 61],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 183, 1438, 228, 881, 4755, 311, 849, 12817, 220, 3241]]

3. Data Modeling

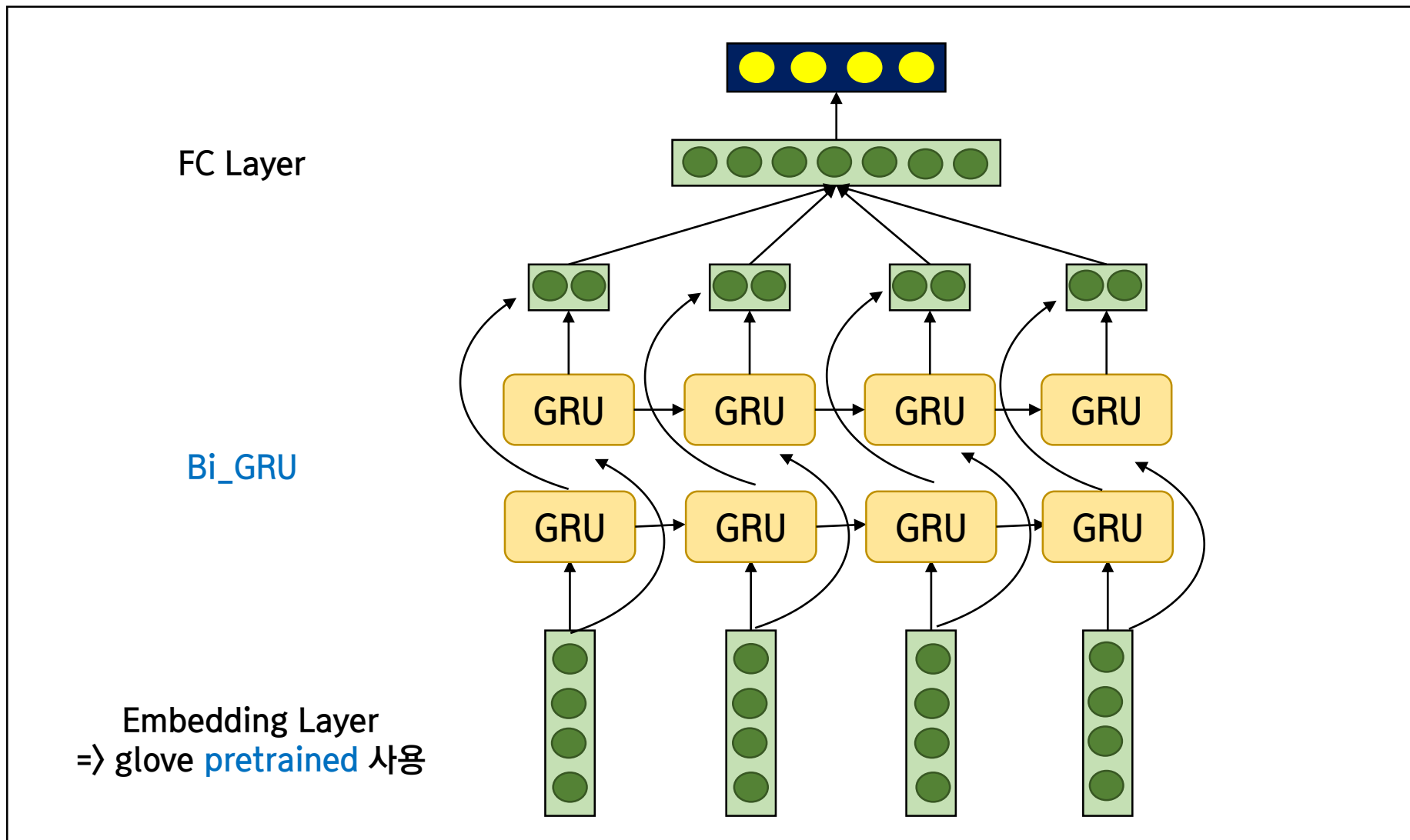
(1) 전체 OVERVIEW



3. Data Modeling

(2) 모델 [1/3] – *Bi_Directional_GRU

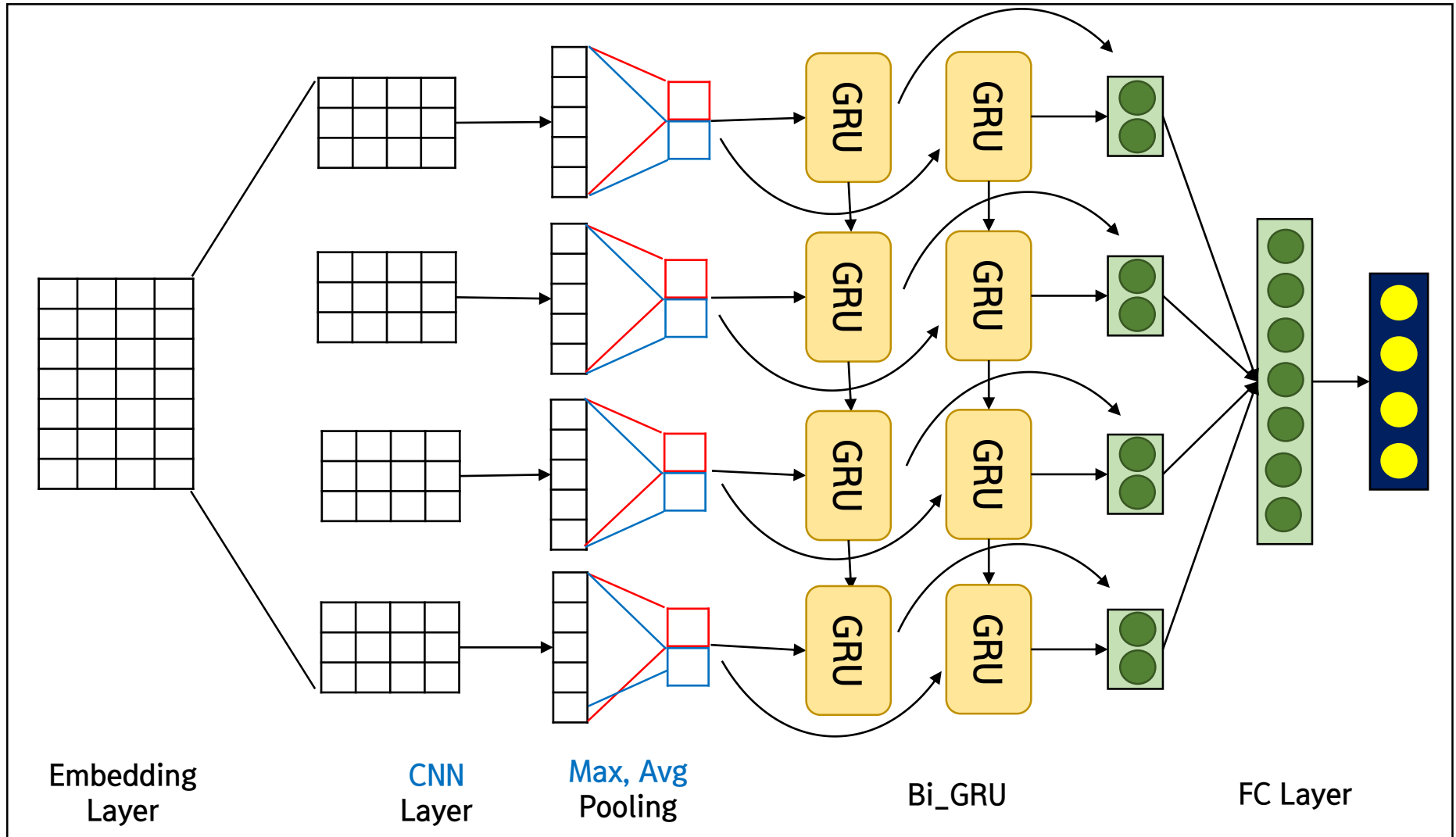
*TEXT 데이터에 양방향 순차적 신경망 적용



3. Data Modeling

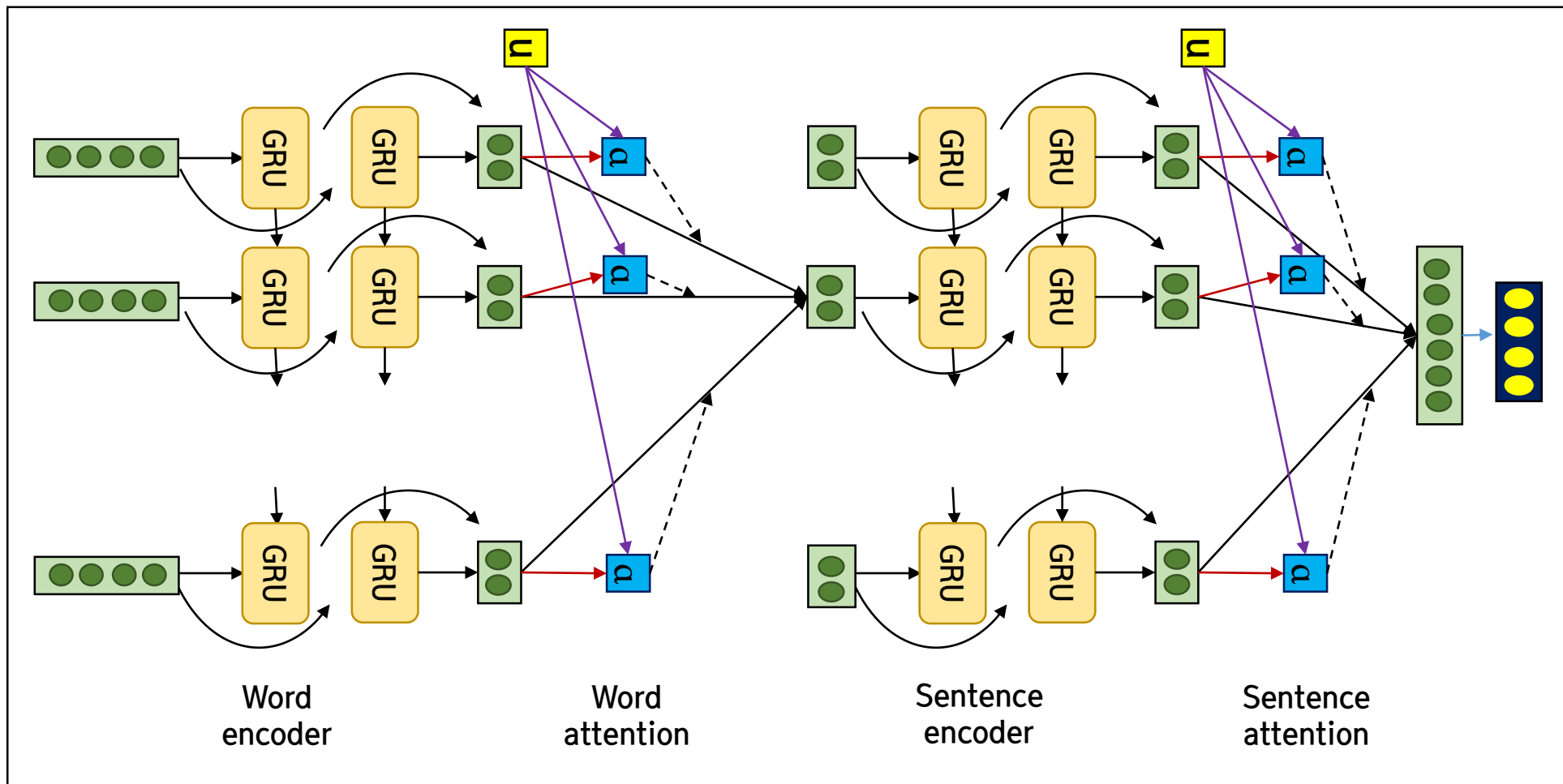
(2) 모델 [2/3] – *CNN + Bi_Directional_GRU

*기존 Bi_GRU 앞에 CNN, Pooling 추가
=> TEXT 특징 추출



3. Data Modeling

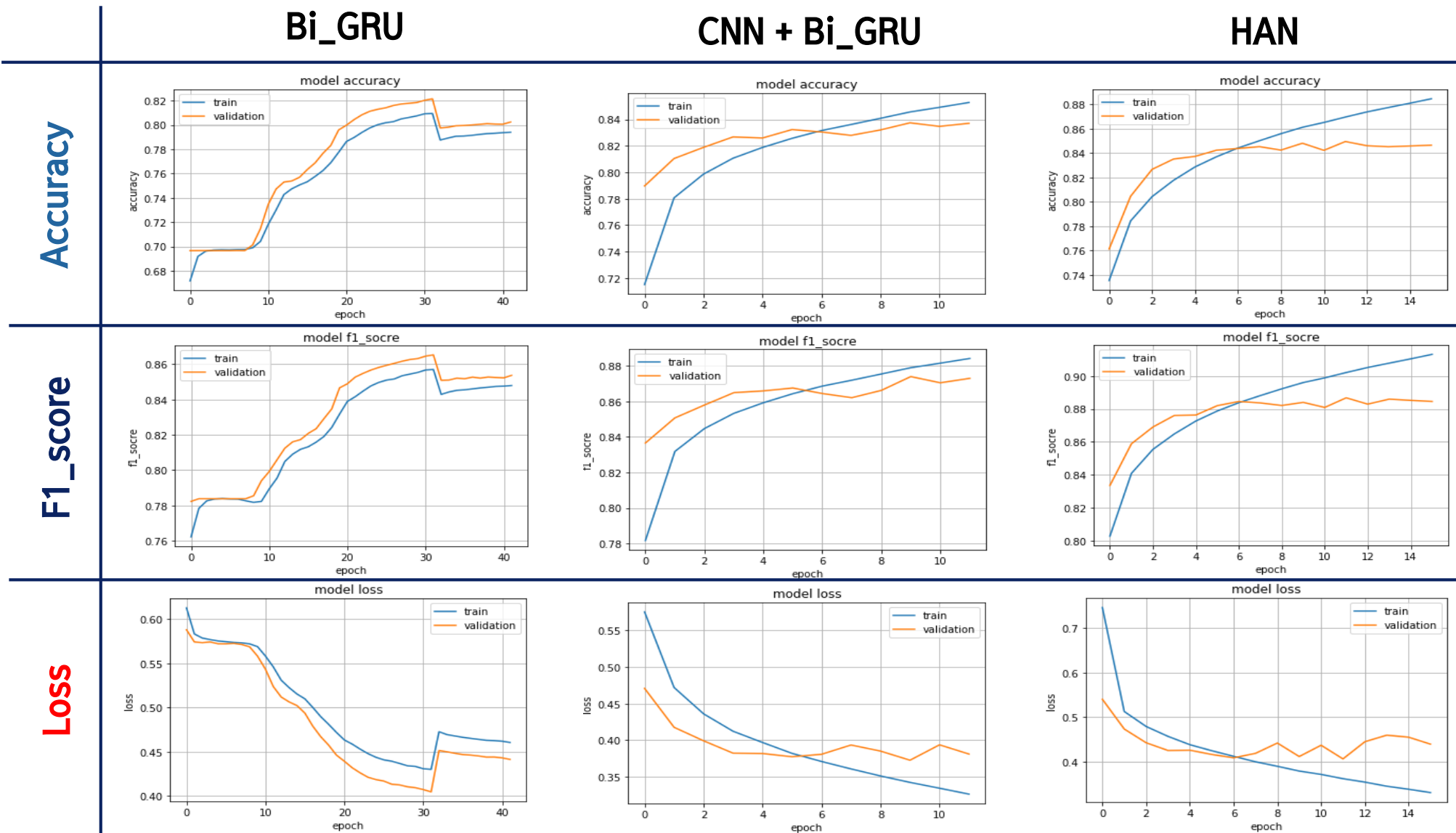
(2) 모델 [3/3] – *Hierarchical Attention networks (+ Bi_Directional_GRU)



*word attention을 통해 문장내 중요 word 가중치 적용 하여 sentence 벡터 생성 후,
sentence attention을 통해 중요 sentence 가중치 적용 하여 최종 document 벡터 생성

3. Data Modeling

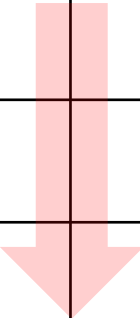
(3) 결과 [1/2]



3. Data Modeling

(3) 결과 [2/2]

	Accuracy	F1 score
Bi_GRU	82.12	86.47
CNN + Bi_GRU	83.91	87.56
HAN	85.13	88.86
Ensemble	87.64	91.21



텍스트 데이터에서 양방향 순환 신경망 이전 Layer에서

중요 특징을 적절히 추출할수록 좋은 성능을 보임

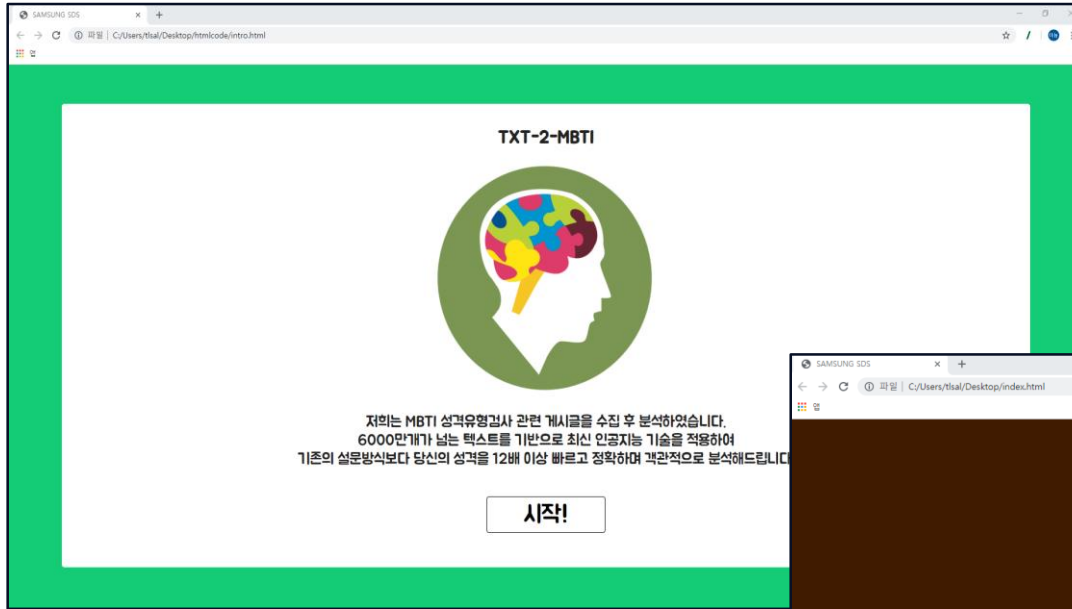
=> 가중평균을 통해 세개의 모델을 Ensemble한 결과

최종 Accuracy : 87.6%, F1_score : 91.2% 달성

4. Prototype

(1) 시연 화면 [1/3] - 초기 페이지

서비스에 대한 간략한 소개와 분석 모델에 필요한 데이터를 입력 받는 메인 페이지



← [서비스 첫 화면]
서비스에 대한 소개

[유저 입력 화면]
이름, 성별,
대화내용/텍스트 파일 입력



TXT-2-MBTI

작성한 글 혹은 대화내용을 통해 당신의 성격 유형을 분석합니다.
분석에 소요되는 시간은 약 1분입니다

성명:

성별:
☒ 남
☐ 여
☐ 불명

대화내용:
내용을 입력해주세요.

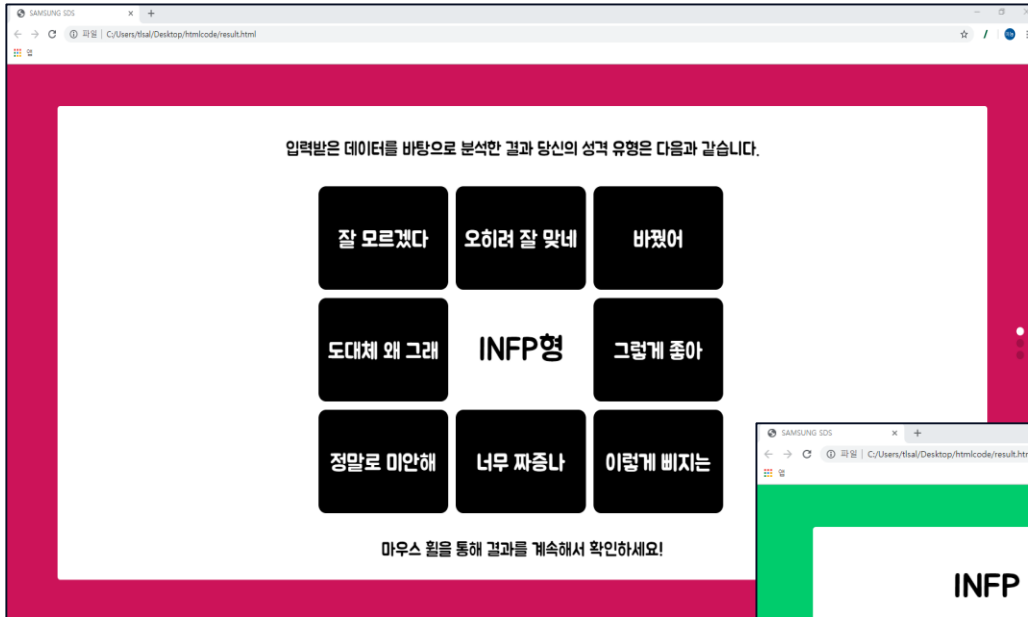
파일 선택 선택된 파일 없음

확인

4. Prototype

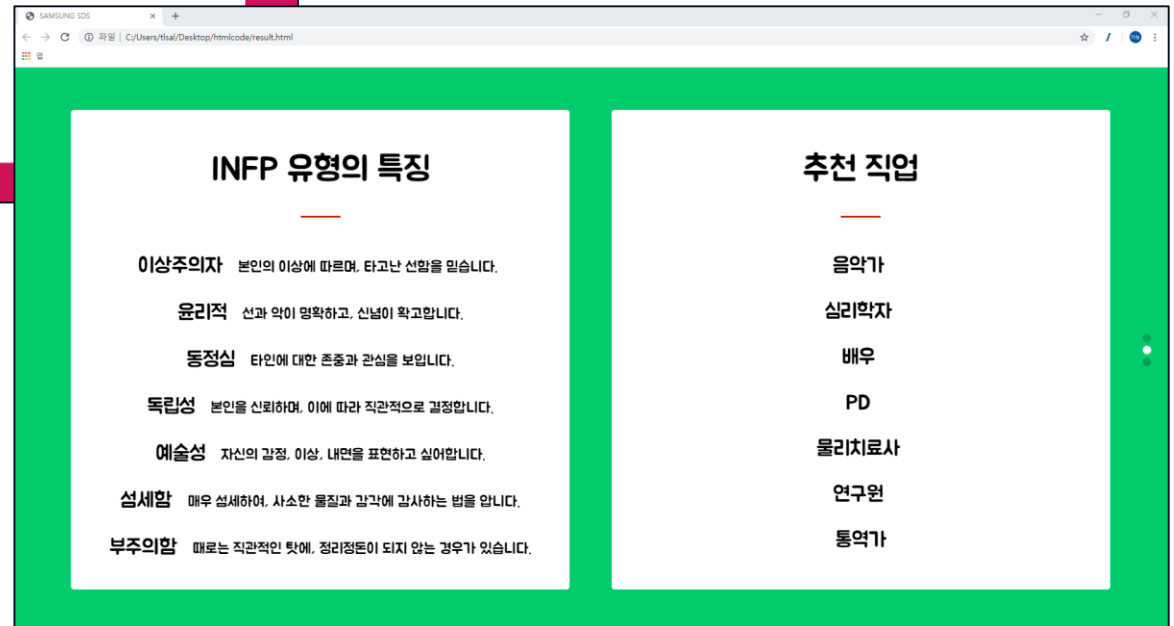
(1) 시연 화면 [2/3] – 분석 결과 페이지

유저 입력화면에서 Django로 데이터를 받아 생성한 모델을 통한 분석 결과를 결과 페이지로 전송



← [분석 결과 화면 1]
유저의 성격유형 제공
결과에 큰 영향을 준 TEXT

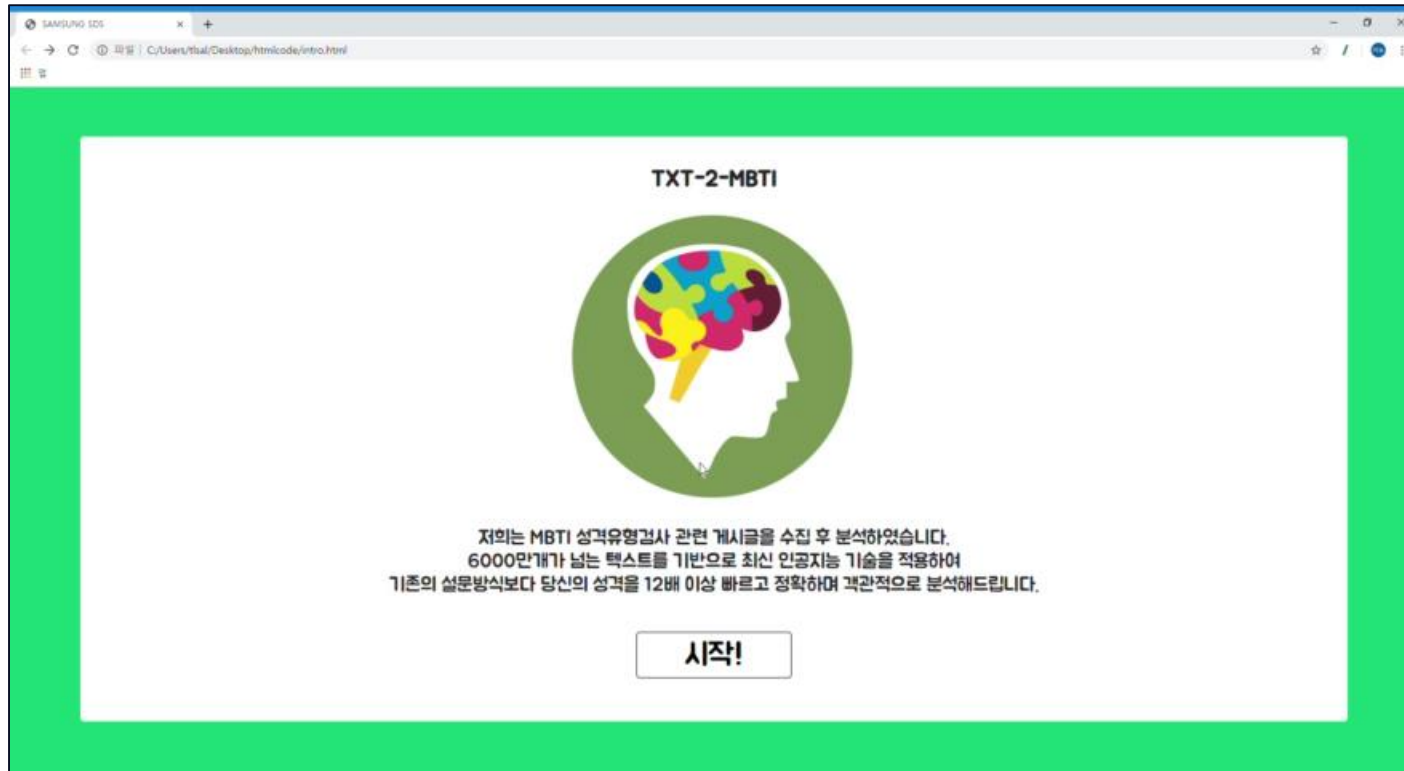
[분석 결과 화면 2]
성격 유형에 대한
설명과 직업 추천



[분석 결과 화면 3]
유저에게 제공하는 분석 시각화 차트

4. Prototype

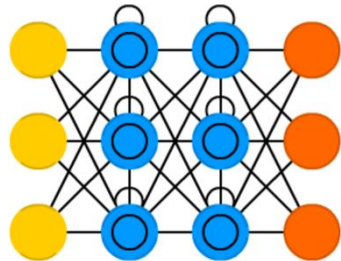
(2) 시연영상



[서비스 시연 영상]
Prototype_Video.mp4 확인

(1) Service Preview

기존 성격 검사 방식과 달리 텍스트 기반 성격 유형 분류 모델은
텍스트 데이터만 있다면 어디에든 적용할 수 있다.
이러한 범용성을 기반으로 다양한 서비스로의 확장이 가능하다.



텍스트 기반 성격
유형 분류 모델

간편 진로 탐색 서비스

- 개인을 대상으로 하여 자신이 작성한 텍스트 데이터를 업로드 하는 것만으로 손쉽게 자신의 성향을 파악하고 그에 따라 자신에게 알맞은 직무/직업을 추천 받을 수 있도록 한다

입사 지원자/ 임직원 대상 인성 검사 서비스

- 입사 지원자/기존 직원에게 특정 주제에 대한 자신의 의견을 피력하게 하여 이를 모델에 넣어 성향을 파악한다.
- 기존의 성격 검사를 이것으로 대체하거나 기존 검사의 결과와 모델의 결과를 대조하여 이가 일치하는지 파악함으로써 지원자 또는 직원이 성격 검사에 진실되게 응답하였는지 판단한다.

리뷰 데이터 적용 추천 시스템 고도화

- 여행/맛집/쇼핑 사이트의 리뷰 데이터를 모델에 입력하여 고객의 성격 유형을 분류한다.
- 분류한 성격 유형을 추천 시스템의 Feature로 사용하여 성능을 향상시킬 수 있다.

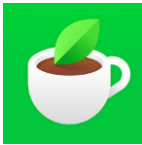
AI 비서의 개인화

- AI 챗봇에 텍스트 기반 성격 유형 분류 모델을 적용하여 개인화된 맞춤 대화를 제공한다.
- 사용자들이 AI비서에 친근감을 느끼게 하고 체류 시간을 증가시킨다.

5. Service

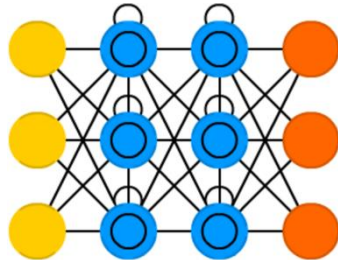
(2) 간편 진로 탐색 서비스

Input



다양한 형태의
텍스트 데이터

Model



텍스트 기반 성격
유형 분류 모델

Output

ISTJ Responsible, sincere, analytical, reserved, realistic, systematic. Handworking and trustworthy with sound practical judgment.	ISFJ Warm, considerate, gentle, responsible, pragmatic, thorough. Devoted caretakers who enjoy being helpful to others.	INFJ Idealistic, organized, insightful, dependable, compassionate, gentle. Seek harmony and cooperation, enjoy intellectual stimulation.	INTJ Innovative, independent, strategic, logical, reserved, insightful. Driven by their own original ideas to achieve improvements.
ISTP Action-oriented, logical, analytical, spontaneous, reserved, independent. Enjoy adventure, skilled at understanding how mechanical things work.	ISFP Gentle, sensitive, nurturing, helpful, flexible, realistic. Seek to create a personal environment that is both beautiful and practical.	INFP Sensitive, creative, idealistic, perceptive, caring, loyal. Value inner harmony and personal growth, focus on dreams and possibilities.	INTP Intellectual, logical, precise, reserved, flexible, imaginative. Original thinkers who enjoy speculation and creative problem solving.
ESTP Outgoing, realistic, action-oriented, curious, versatile, spontaneous. Pragmatic problem solvers and skillful negotiators.	ESFP Playful, enthusiastic, friendly, spontaneous, tactful, flexible. Have strong common sense, enjoy helping people in tangible ways.	ENFP Enthusiastic, creative, spontaneous, optimistic, supportive, playful. Value inspiration, enjoy starting new projects, see potential in others.	ENTP Inventive, enthusiastic, strategic, enterprising, inquisitive, versatile. Enjoy new ideas and challenges, value inspiration.
ESTJ Efficient, outgoing, analytical, systematic, dependable, realistic. Like to run the show and get things done in an orderly fashion.	ESFJ Friendly, outgoing, reliable, conscientious, organized, practical. Seek to be helpful and please others, enjoy being active and productive.	ENFJ Caring, enthusiastic, idealistic, organized, diplomatic, responsible. Skilled communicators who value connection with people.	ENTJ Strategic, logical, efficient, outgoing, ambitious, independent. Effective organizers of people and long-range planners.

성격 유형

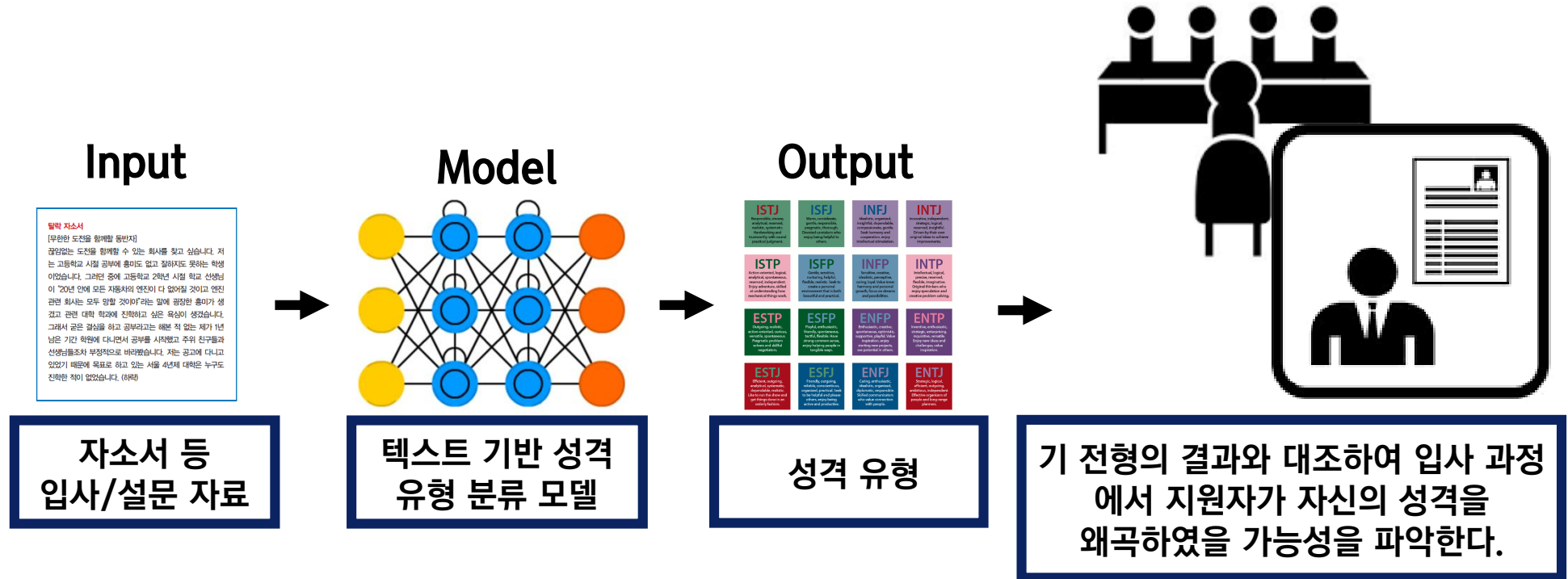


성격 유형별 직업 추천

- 기존 성격 검사를 개인이 실시할 경우 주로 진로 탐색 및 적성 검사의 목적으로 활용되어 왔다.
- 텍스트 기반 성격 유형 분류 모델을 사용하면 복잡한 검사와 절차를 없애고 손쉽게 개인의 성격을 진단해주고 이를 기반으로 적성에 맞는 직업을 추천해 줄 수 있다.

5. Service

(3) 입사 지원자/임직원 대상 인성 검사 서비스



- 개인(지원자)은 자신의 목적이나 지원한 기업의 인재상에 따라 자신의 답변을 준비해가거나 문제의 답안을 선택한다는 문제점이 있다.
- 따라서 기업 입장에서는 이러한 거짓 답변을 분간해낼 장치가 필요하며 텍스트 기반 성격 유형 분류 모델이 이를 수행할 수 있다.
- 기존 검사의 결과와 모델의 결과를 대조하여 이가 일치하는지 파악함으로써 지원자 또는 직원이 성격 검사에 진실되게 응답하였는지 판단한다.

(4) 추천 시스템 현황

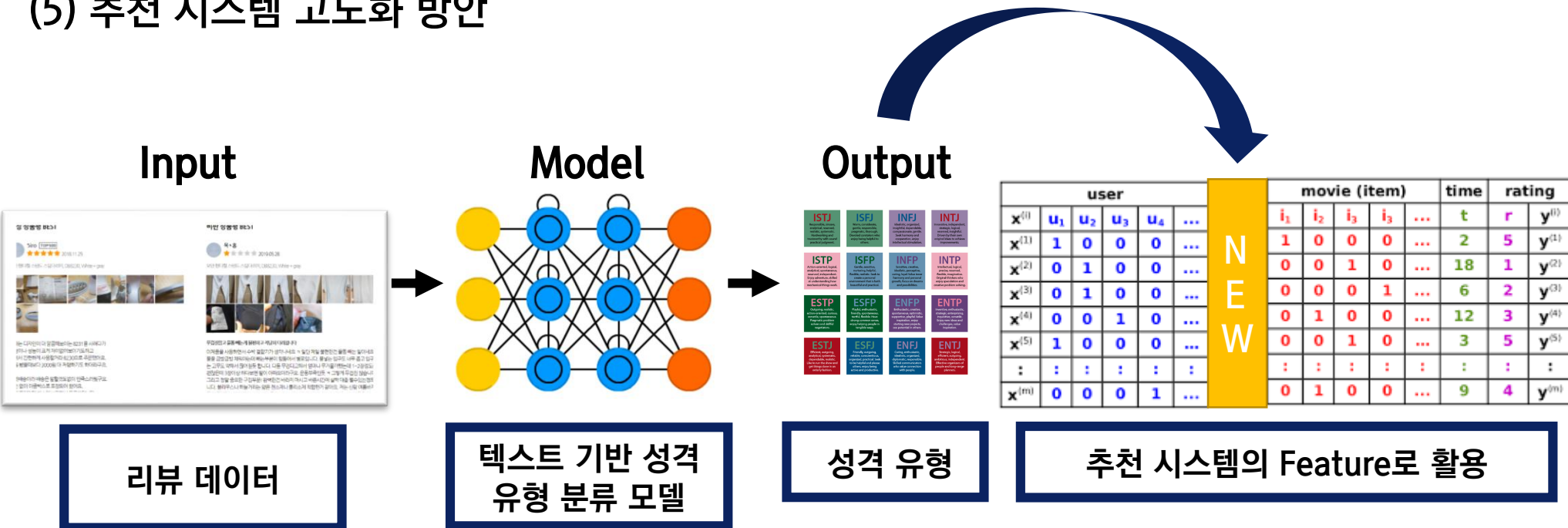
추천 시스템 현황

- 초기 추천 시스템은 collaborative filtering(CF) 기술을 기반으로 발전하였으며 사용자에게 중심을 두고 추천을 수행하게 된다.
- 핵심 아이디어는 사용자와 비슷한 다른 사용자가 존재할 경우 해당 정보를 활용하여 추천을 하는 것이다. 그 중에서 Model-based 방식은 Memory-based 방식의 cold start problem을 해결하기 위하여 고안되었다.
- Model-based 방식에 서는 사용자 혹은 아이템 사이의 유사성을 찾기 보다는 데이터에 내재된 hidden feature를 활용한다. Hidden feature를 계산하기 위한 방식으로 matrix factorization (MF) 방법이 주로 사용된다.
- 1~2년 전부터 **개인화 추천**이 뜨면서 **Factorization Machine**이 각광받기 시작했다.
- MF는 유저-상품 데이터만 가지고, 숨겨진 정보들(side-features, latent feature -> 상품카테고리, 검색어, 이전에 본 상품 등)을 표현(학습)하는 알고리즘이라, 입력 데이터로 유저-상품(클릭 여부)만 사용 가능하나 FM에서는 이러한 **side-feature들을 직접 입력**으로 넣어서 학습이 가능하다.
- FM은 MF보다 더 일반적이고 확장된 모델이어서, 추천 뿐만 아니라 회귀나 이진분류와 같은 다양한 ML에서도 사용 가능하다.(MF는 불가능)
- FM의 가장 큰 장점은 기존 **user,item 외의 feature들을 실제로 사용**할 수 있다는 부분이다

→ 텍스트 기반 성격 유형 분류 모델을 사용하여 **Factorization Machine**에 사용될 **Feature**를 생성하여 추천시스템을 고도화시키고자 한다.

5. Service

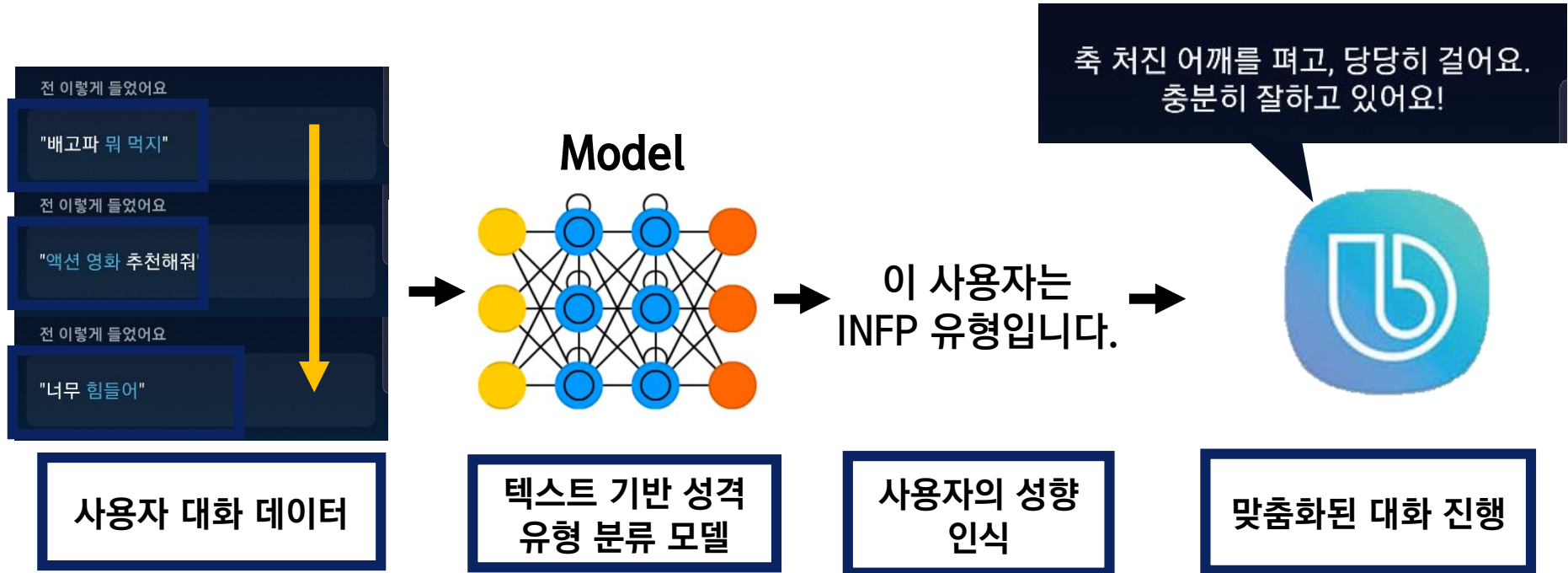
(5) 추천 시스템 고도화 방안



- 장소/음식/상품에 관련된 추천 시스템 이슈에서 리뷰 데이터를 활용하고자 하는 노력이 이어져왔으나 뚜렷한 성과를 보이지는 못하고 있었다.
- 그러나 텍스트 기반 성격 유형 분류 모델의 경우 유저를 중심으로 리뷰 데이터를 분석하여 군집화 시킬 수 있다.
- 이렇게 분류된 성격 유형을 FM의 Feature로 사용하여 추천 시스템의 고도화가 가능하다.

5. Service

(6) AI 비서의 개인화



- 현재 AI 비서는 주로 정보 검색 및 알람 설정/통화/문자 전송 등의 목적으로 사용되고 있다.
- 점차 대화형 AI의 핵심 기술로 일상 대화를 반영하고자 하는 시도가 일어나고 있으며 제품에 일상 대화를 적용하면 사람들이 친근감을 느끼고 사용 시간이 길어질 것이다.
- 텍스트 기반 성격 유형 분류 모델로 사용자와의 대화를 분석한다면 성격 유형에 맞는 일상 대화를 진행할 수 있을 것이다.

(7) 발전 방향

발전 방향

- Computing Power 문제로 인하여 제한적으로 알고리즘을 사용하였다. 이러한 문제가 해결되면 Pre-Train에 사용될 Embedding 차원을 증가시키거나, BERT 또는 XLNET등 최신 Embedding 방법론을 적용하여 **정확도를 더욱 향상시킬 수 있다.**
- 메인 아이디어인 텍스트 기반 성격 유형 분류 모델은 높은 범용성을 가지기 때문에 상기의 활용방안외에도 **다양한 분야에 적용을 시도**해볼 수 있다.
- 영어 데이터로 학습이 되었기 때문에 한글 데이터를 입력 시 번역 Api를 거쳐야하는 **번거로움**이 있고 번역기의 성능에 따라 **결과에 차이**가 날 수 있다.
(반면, 번역기를 통하여 단어가 **평이해진다**는 장점도 존재한다.)



Projected by
TEAM 609