

2022 빅데이터 붐업 워크숍

# 데이터톤

오리엔테이션

주최 | 숙명여자대학교 IT공학전공 학생회

주관 | 숙명여자대학교 빅데이터 혁신공유사업단

## 슬랙 안내

- 개인 메일로 '참가자 안내 사항' 발송

- 슬랙 워크스페이스 참가

- 표시 이름 변경

**\*\* 슬랙 알림 활성화 \*\***

(공지 미확인으로 인한 불이익은 진행 위원회가 책임지지 않습니다.)

# 슬랙 안내

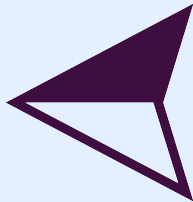
▼ 채널

# 전체공지

# 정확도-순위-공개

# 질의응답

team01



UD

2B

+

2022 BigData

모든 DM

@ 멘션 및 반응

더 보기

▼ 채널

# 전체공지

# 정확도-순위-공개

# 질의응답

team01

team02

team03

team04

team05

team06

team07

team08

team09

team10

team11

team12

team13

team14

team15

team16

team17

Slackbot

+ 채널피 추가

안녕하세요! Slackbot이 여기 있습니다.

저는 한두 가지 일(주로 알림 및 리마인더)을 할 수 있는 간단한 봇입니다. 도움이 필요하시다면 [고객지원센터](#)를 방문해 보세요.

B I

Slackbot에게 메시지 보내기

+ @ Aa

▶ ▼

# 목차

01

---

## 대회 소개

---

- 대회 개요
- 팀빌딩 결과 안내

02

---

## 세부 일정

---

- 중간 정확도 순위 공개
- 최종 코드 제출 및 본선 팀 발표
- 발표 자료 제출
- 발표 및 심사

03

---

## 데이터셋

---

- 데이터셋 소개
- 채점 방식
- 제출 방법

04


---

## 심사 및 시상


---

- 심사위원 소개
- 시상 내역
- 수상 자격

# 대회 개요




빅데이터 컨소시엄  
7개 대학 학부생



최대 900만원의 상금





2박 3일  
전면 비대면 진행



예선 / 본선 진행

# 팀 빌딩 결과 안내

팀명	이름
1팀	김지원
	백지영
	정시은
2팀	윤진섭
	최지혜
	한수정
3팀	유시모
	윤정희
	김성수
4팀	남유지
	신진영
	정은정
5팀	임유나
	김수빈
	박종한
6팀	김수현
	손승해
	김가현
7팀	백범수
	성민기
	조선영

팀명	이름
8팀 폐쇄	
9팀	김동화
	이가은
	권은혁
10팀	최윤서
	정성문
	김영준
11팀	이지수
	유정섭
	이주성
12팀	전채린
	황인성
	조인후
13팀	장혜지
	최민동
	손아영
14팀	김세희
	조수민

팀명	이름
15팀	고나경
	고경수
	최진광
16팀	최재성
	김가영
17팀	조상민
	최승연
18팀	박윤서
	김시현
19팀	김진형
	한정아
20팀	문유진
	송우제
21팀	신혜리
	강하림
	장나리
22팀	양희찬
	최규진

# 세부 일정 안내

	시간	내용	비고
1일차 (2/23 수)	~	워크숍	
	16:30 - 17:30	데이터톤 OT 및 질의응답	
	17:30 ~	개발	
2일차 (2/24 목)	00:00	1차 정확도 순위 공개	슬랙 제출
	12:00	2차 정확도 순위 공개	
	18:00	3차 정확도 순위 공개	
3일차 (2/25 금)	00:00	4차 정확도 순위 공개	슬랙 제출
	08:00	최종 코드 제출	
	10:00	본선 진출 팀 발표	슬랙 공지
	12:00	발표 자료 제출	본선 진출 팀만 슬랙 제출
	13:00 - 15:00	발표 및 심사	발표 10분 + 질의응답 5분
	15:30 - 16:00	시상식 및 폐회	

# 중간 정확도 순위 공개

2일차 (2/24 목)	00:00	1차 정확도 순위 공개
	12:00	2차 정확도 순위 공개
	18:00	3차 정확도 순위 공개
3일차 (2/25 금)	00:00	4차 정확도 순위 공개

순위 공개 1시간 전까지 슬랙 팀 채널에 제출한 파일에 대해 정확도 측정  
( 업로드 시간 확인 예정이오니, 시간 준수 )

전체 결과는 슬랙 '정확도-순위-공개' 채널에 이미지 형태로 업로드

\* 최종 정확도 순위는 시상식 이후 공개



## 최종 코드 제출 및 본선 팀 발표

3일차 (2/25 금)	08:00	최종 코드 제출
-----------------	-------	----------

최종 예측 파일과 소스 코드 등  
팀 채널에 업로드

3일차 (2/25 금)	10:00	본선 진출 팀 발표
-----------------	-------	------------

본선 진출 상위 6개 팀 발표

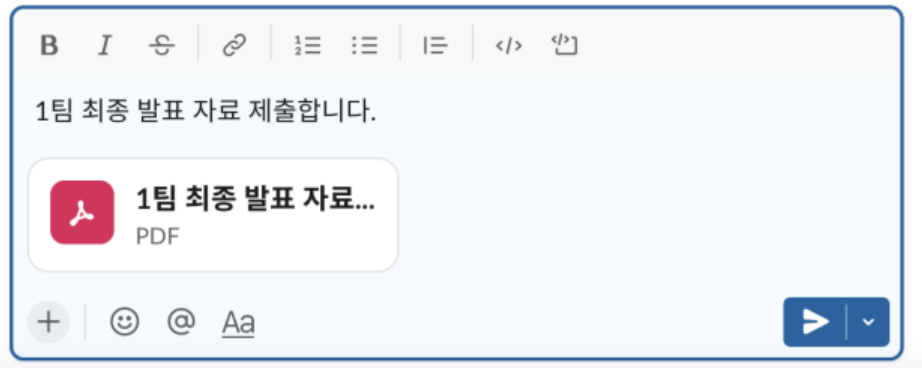
\* 전체공지 채널에 공개

\*\* 팀별 순위 공개 X

# 발표 자료 제출

3일차 (2/25 금)	12:00	발표 자료 제출 (본선 진출 팀만)
-----------------	-------	------------------------

슬랙 팀 채널에 발표 자료 업로드  
파일 이름 포맷 : @팀 최종 발표 자료.pptx 또는 .pdf



제출 이후 자료 수정은 불가

## 발표 및 심사

3일차 (2/25 금)	13:00 - 15:00	발표 및 심사 * 발표 10분 + 질의응답 5분
-----------------	---------------	-------------------------------

### 온라인 진행

\* zoom 링크는 당일 슬랙 전체공지 채널에 업로드

\*\* 모든 발표 및 심사 세션은 녹화될 예정

**활용 아이디어 + 코드 실행 시연 필수**

**모든 팀 전원 참석 필수**

\* 심사 대상 - 본선 진출 팀

# 발표 및 심사

## <예선 심사 기준>

심사 항목	평가 지표	배점
정확도	예측이 얼마나 정확한가?	100

## <본선 심사 기준>

심사 항목	평가 지표	배점
정확도	예측이 얼마나 정확한가?	50
모델 및 코드 이해도	모델과 코드를 어느 정도 이해하는가?	20
데이터 및 모델 활용 아이디어	활용 아이디어가 얼마나 창의적인가? 실생활에 적용 가능한가?	30

# 데이터셋 소개

## 주제

인터넷 사용자의 MBTI 예측하기

## 모델링

인터넷 사용자가 작성한 게시글에서  
랜덤 추출한 400개의 단어를 통해  
사용자의 MBTI 예측

INTP	say process model list like subscriber channel
INFJ	upon much manipulate retail finish like sacrifici
INFJ	fit yes certain bff social feel goal go know nor
INTJ	complete love within someone ideal joke solv
ENTJ	public strictly thing person x question person
INFJ	opinion mean dim nuclear one like upset deci
INFJ	thing zebra fool aw co word take day talk giv
INTP	hard lazy yet technique convince many hard l
INTP	incite use market especially market late try ne
ENTP	east relate probability thinker live know simple
INFJ	understand seem hard one use environment b
INTJ	let true need nice u nice even like fi life fi mal
INFJ	rot discipline live get overall house black te di
INTP	build beyond polynomial perceive unexpected
ENTP	idea piece moral usually try f shit enfp value a

# 데이터셋 소개

## 데이터셋

트레이닝 데이터 (MBTI\_train.csv)

74,357건  
MBTI, 400개 단어 -> 총 2 컬럼 구성

테스트 데이터 (MBTI\_test.csv)

9,337건  
400개 단어 -> 총 1 컬럼 구성

## 데이터셋 예시

MBTI	400 words
INTP	preferably p hd low except wew lad video p mind good reason agree statement relationship become diff...
ENTJ	eye close bitch half get sure cringeworthy beat topic like sjw introduce break wall bring mind e bre...

## 주의사항

- \* training data 읽을 때 ISO 8859-1 로 인코딩해야 오류 없음
- \*\* 데이터셋 내 MBTI 분포는 균등하지 않음
- \*\*\* 참고 자료: <https://www.kaggle.com/zeyadkhalid/mbti-personality-types-500-dataset>

## 채점 방식

### 부분 정답 인정

한 사람의 MBTI 예측 시, 4가지의 선호 지표 조합 중 맞춘 지표 수에 따라 정확도 판별

ex) 정답 : ESTJ / 결과 : ISTJ

-> 3개 정답 인정

ex) 정답 : ESFP / 결과 : ISTP

-> 2개 정답 인정

## 제출 방법

중간 제출

VS

최종 제출

1차 ~ 4차 정확도 측정을 위한 제출

( 예측 파일 제출 )

최종 정확도 측정을 위한 제출

( 예측 파일 & 소스 코드 & requirements.txt 제출 )



# 제출 방법 - 중간 제출

슬랙 팀 채널에 예측 파일 업로드



예측 파일 형식 : .txt 또는 .csv

파일 이름 포맷 : R#\_team@ (#: 정확도 예측 파일 제출 차수 / @: 팀 번호)

## \* 주의사항

- 코드를 통해 도출된 MBTI 4가지 지표의 조합 10,000개를 예측한 파일을 제출하는 것으로 개별 타이핑 금지!
- 띄어쓰기 없이 문자 4개 붙이기
- 대문자 형식 지키기
- 포맷이 맞지 않는 결과물을 제출할 경우 정확도 0점 처리

INFP  
INTJ  
INTP  
INFJ  
ESTP  
INTJ  
INFJ  
INTP  
INTP

< 예측 파일 예시 >

## 제출 방법 \_ 최종 제출

슬랙 팀 채널에 예측 파일 & 소스 코드 & requirements.txt 업로드

예측 파일 형식 : .txt 또는 .csv

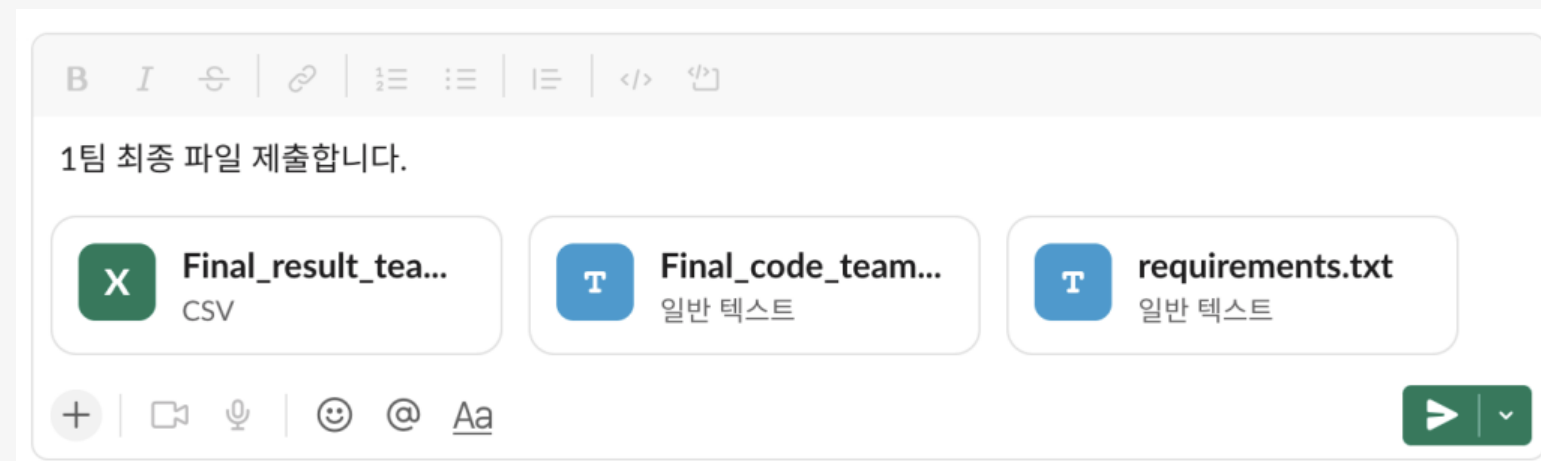
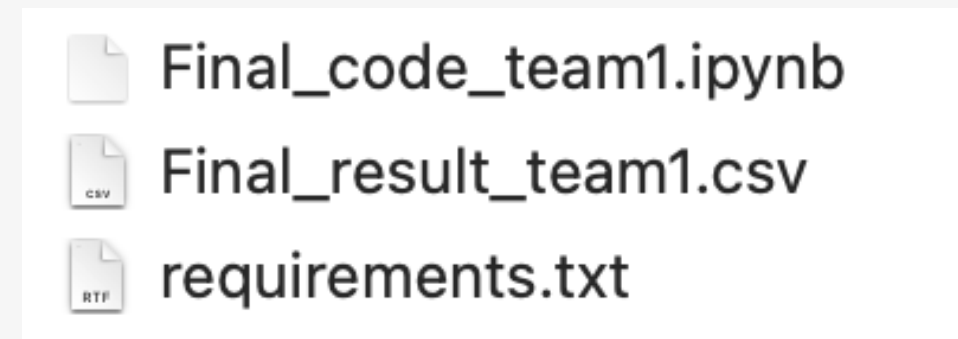
파일 이름 포맷 : Final\_result\_team@ (@: 팀 번호)

소스 코드 형식 : .ipynb

파일 이름 포맷 : Final\_code\_team@ (@: 팀 번호)

\* 소스 코드 제출시, 데이터를 포함시키지 않음

\*\* 소스 코드에는 학습 과정이 포함되어야 함



INFP  
INTJ  
INTP  
INFJ  
ESTP  
INTJ  
INFJ  
INTP  
INTP

< 예측 파일 예시 >

# 심사위원 소개

소속	부서	직위	이름
경기과학기술대학교	스마트 경영학과	교수	김용준
경상국립대학교	빅데이터 혁신공유사업단	단장	한관희
서울대학교	e러닝 지원부	e러닝 지원부장	안용희
서울시립대학교	빅데이터 혁신공유사업단	단장	이용희
숙명여자대학교	IT공학전공	교수	김철연
전북대학교	통계학과	교수	최규빈
한동대학교	AI 융합교육원	교수	김경외

# 시상 내역

상격	상금	수상 팀
대상	150만원	1팀
최우수상	100만원	2팀
우수상	50만원	3팀
장려상	20만원	10팀
발전상	10만원	다수의 팀

\* 최종 제출 결과물에 따라 수상 팀의 수는 조정될 수 있음

## 수상 자격

1

소스 코드 제출

2

최종 예측 파일 제출

3

발표 및 심사, 시상식 참여

\* 팀당 최소 1명은  
캠을 키고 있어야 함

\* 상기 3가지 조건 미충족시, 수상이 취소됩니다.



# Q & A

