

# # 어쩌다 한국어 텍스트 분석 with 파이썬

경복고 김선경



# 어쩌다? 어쩐다?

# # 모두의 텍스트 분석 with 파이썬, 우리가 공부할 내용은

## 모두의 한국어 텍스트 분석

with 파이썬

기초부터 챗GPT까지,  
누구나 쉽게 시작하는 자연어 처리

박조은, 송영숙 지음



1. 파이썬(python) 언어로 **프로그래밍**을 할 수 있어야 한다.
2. 판다스(pandas), 넘파이(numpy)를 이용해 **데이터를 다룰** 수 있어야 한다.
3. matplotlib, seaborn을 이용해 **데이터를 시각화** 할 수 있어야 한다.
4. **텍스트 분석** 방법을 이해하고 적용할 수 있어야 한다(게다가 한국어).
5. **머신러닝**을 이해하고 적용할 수 있어야 한다.

**#해야지, 뭐**



# # 집에 있는 책을 모두 꺼내보기



26권, 10,920쪽

- 열혈강의 파이썬 - 개정판
- 게임으로 배우는 파이썬
- 창의적 프로그래밍을 위한 파...
- 모두의 파이썬 - 20일 만에 배...
- 파이썬 Jupyter Notebook ...
- 빠르게 활용하는 파이썬 3.6 ...
- 초보자를 위한 파이썬(Pytho...
- 파이썬으로 풀어보는 수학 - ...
- 파이썬 코딩 수학 with 딥러...
- 만들면서 배우는 파이썬과 40...
- 데이터 분석을 위한 파이썬 철...
- 파이썬으로 데이터 주무르기 ...
- 파이썬을 활용한 데이터 길들...
- 파이썬 기초 문법 : 파이썬을 ...
- 판다스 인 액션 - 인기 영화, ...
- 모두의 데이터 과학 with 파...
- 모두의 데이터 분석 with 파...
- Do it! 데이터 분석을 위한 판...
- 혼자 공부하는 데이터 분석 w...
- fastai와 파이토치가 만나 꽃...
- 케라스 창시자에게 배우는 딥...
- 밑바닥부터 시작하는 딥러닝 ...
- 인공지능 전문가가 알려 주는 ...
- 모두의 딥러닝 - 누구나 쉽게 ...
- 헨즈온 머신러닝 - 사이킷런, ...
- 혼자 공부하는 머신러닝 + 딥...

# # I Think.....?!



## 파이썬 프로그래밍 (1년 과정)

## 데이터분석과 시각화 (1년 과정)

## 머신러닝 (1년 과정)

오텔강의 파이썬 - 개정판

게임으로 배우는 파이썬

창의적 프로그래밍을 위한 파...

모두의 파이썬 - 20일 만에 배...

파이썬 Jupyter Notebook ...

빠르게 활용하는 파이썬 3.6 ...

초보자를 위한 파이썬(Pytho...

파이썬으로 풀어보는 수학 - ...

파이썬 코딩 수학 with 딥러...

만들면서 배우는 파이썬과 40...

데이터 분석을 위한 파이썬 철...

파이썬으로 데이터 주무르기 ...

파이썬을 활용한 데이터 길들...

파이썬 기초 문법 : 파이썬을 ...

판다스 인 액션 - 인기 영화, ...

모두의 데이터 과학 with 파...

모두의 데이터 분석 with 파...

Do It! 데이터 분석을 위한 판...

혼자 공부하는 데이터 분석 w...

fastai와 파이토치가 만나 꽃...

케라스 창시자에게 배우는 딥...

딥바닥부터 시작하는 딥러닝 ...

인공지능 전문가가 알려 주는 ...

모두의 딥러닝 - 누구나 쉽게 ...

핸즈온 머신러닝 - 사이킷런, ...

혼자 공부하는 머신러닝 + 딥...

**#해야지, 뭐22**

# # For Example, 연세대 인공지능학과의 커리큘럼

출처 : [https://ai.yonsei.ac.kr:58135/sub3\\_1\\_a.php](https://ai.yonsei.ac.kr:58135/sub3_1_a.php)

## 1학년

	학기	학정번호-분반-실습	과목명	학점	분류
▼	1	CAC1100	컴퓨터프로그래밍	3	전공선택
▼	2	CSI2102	객체지향프로그래밍	3	전공필수

## 2학년

	학기	학정번호-분반-실습	과목명	학점	분류
▼	1	CSI2103	자료구조	3	전공필수
▼	1	CSI2101	이산구조	3	전공기초
▼	1	MAT2013	확률통계	3	전공기초
▼	2	-	인공지능개론	3	전공필수



# # For Example, 연세대 인공지능학과의 커리큘럼

출처 : [https://ai.yonsei.ac.kr:58135/sub3\\_1\\_a.php](https://ai.yonsei.ac.kr:58135/sub3_1_a.php)

## 3학년

	학기	학점번호-분반-실습	과목명	학점	분류
▼	1	-	기계학습	3	전공필수
▼	1	-	분산학습시스템	3	전공선택
▼	1	-	빅데이터분석및모델링	3	전공선택
▼	1	-	멀티모달데이터처리	3	전공선택
▼	2	-	텍스트마이닝	3	전공선택
▼	2	-	데이터마이닝	3	전공선택
▼	2	-	딥러닝	3	전공선택
▼	2	-	GPGPU프로그래밍	3	전공선택

# # For Example, 연세대 인공지능학과의 커리큘럼

출처 : [https://ai.yonsei.ac.kr:58135/sub3\\_1\\_a.php](https://ai.yonsei.ac.kr:58135/sub3_1_a.php)

## 4학년

	학기	학점번호-분반-실습	과목명	학점	분류
▼	1	-	정보검색및추천시스템	3	전공선택
▼	1	-	로봇인공지능	3	전공선택
▼	1	-	컴퓨터비전	3	전공선택
▼	1	-	강화학습	3	전공선택
▼	1	-	인공지능종합설계(1)	3	전공필수
▼	2	-	HCI&AI	3	전공선택
▼	2	-	예측및의사결정시스템	3	전공선택
▼	2	-	자연어처리	3	전공선택
▼	2	-	AI보안	3	전공선택
▼	2	-	데이터모델및시각화	3	전공선택
▼	2	-	AI윤리	3	전공선택
▼	2	-	인공지능종합설계(2)	3	전공필수

**#해야지, 뭐333**

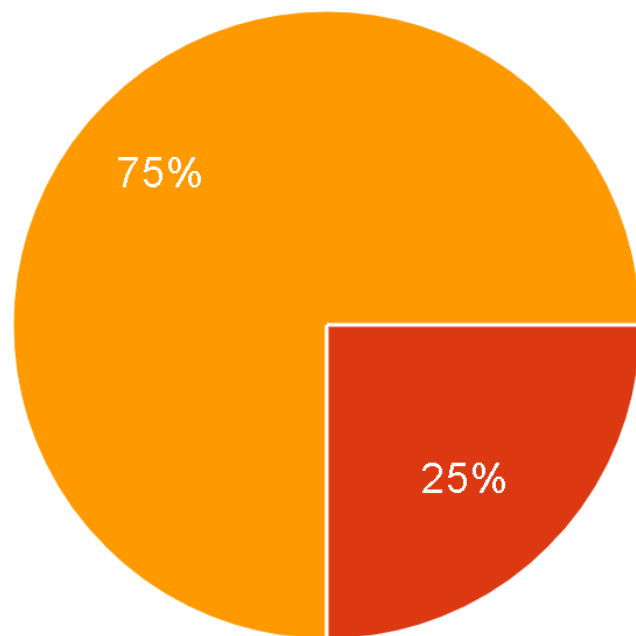


# 그런데,

# # 그런데,

## 1. 파이썬 문법을 잘 알고 있나요?

응답 4개

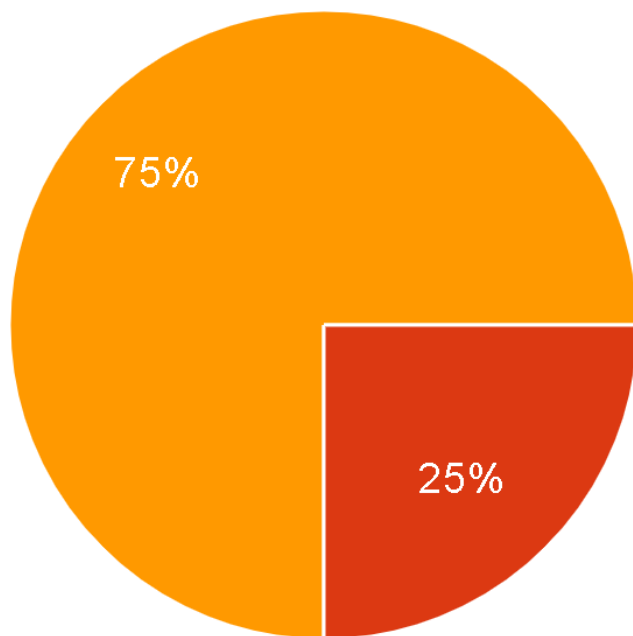


- 상 (클래스, 함수 등의 사용이 자유로워요!)
- 중(기본 문법 조건문, 반복문 정도를 쓸 줄 알아요!)
- 하(파이썬은 잘 모르지만, 잘 알고 싶어요!)

# 그런데,

## 2. 데이터 분석을 해 본 경험이 있나요?

응답 4개



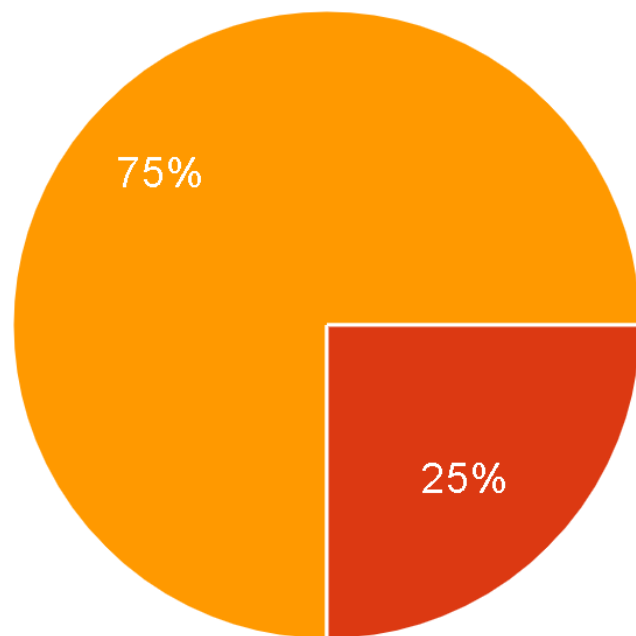
- 상(판다스, 넘파이 라이브러리를 사용해서 데이터 분석을 해 본 적이 있어요!)
- 중(엑셀을 사용해서 데이터 분석을 해 본 적이 있어요!)
- 하(데이터분석은 처음이지만 열심히 해 보고 싶어요!)



# 그런데,

### 3. 데이터를 시각화 해 본적이 있나요?

응답 4개

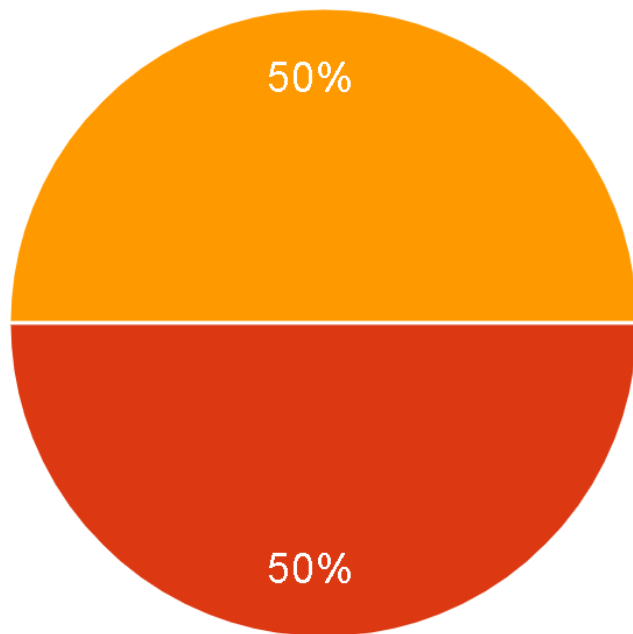


- 상(matplotlib, seaborn 등의 라이브러리를 써서 그래프를 그려 본 적이 있어요!)
- 중(엑셀의 그래프를 그려본 적이 있어요!)
- 하(데이터 시각화를 해 본 적은 없지만 관심이 있어요!)

# 그런데,

#### 4. 머신 러닝에 대해 잘 알고 있나요?

응답 4개



- 상(머신 러닝의 카테고리를 분류하고 여러가지 모델을 사용해 본 적이 있어요!)
- 중(코드를 이용하여 머신 러닝을 해 본 적은 없지만 개념을 이해하고 간단한 툴을 사용해 본적이 있어요!)
- 하(머신 러닝, 진짜 관심 있는데 접해 볼 기회가 없었어요!)

**#해야지, 뭐4444**

## # 그런데,

5. '모두의 한국어 텍스트 분석 with 파이썬'을 함께 공부하며 나에게 남기고 싶은 것은 무엇인가요?



# 우리는 함께



파이팅 해야지!

## # Time Table

8월 13일	오전 (09:00~12:00)	오리엔테이션 5장 연합뉴스 타이틀 주제 분류
	오후 (13:00~16:00)	6장 국민청원 데이터 시각화와 분류
8월 14일	@home	개별 프로젝트 구상 및 발표준비 (생성형 인공지능을 많이 많이 활용해 보세요! ChatGPT, Bing, wrtn, bard 중 하나를 선택하여 활용하기)
8월 15일	오전 (09:00~12:00)	7장 '120다산콜재단' 토픽 모델링과 RNN, LSTM 8장 인프런 이벤트 댓글분석 개별 프로젝트 발표 준비 시간
	오후 (13:00~16:00)	<개별 프로젝트 발표>



# # 개별프로젝트 안내

1. 주제 선정 및 주제 선정의 이유

2. 데이터 가져오기

- 데이터를 어디에서 가져올 것인가?

3. 데이터 분석 및 시각화

- 데이터를 어떻게 정제할 것인가?
- 데이터가 의미하는 것은 무엇인가?
- 데이터가 의미하는 것을 시각화 하여 드러내 보기

4. 텍스트 분석하여 학습시키기(머신 러닝)

5. 결론 도출 및 결과 분석하기

6. 프로젝트 성찰 : 나의 발전기록 / 나의 실패 기록 / 나의 가능성 기록

# 아니 사칙연산을 알려주고, 논문을 쓰라구요?



# # 개별프로젝트 안내 : 사례 1

## 1. 기똥찬 주제를 선정해 보기!

### 2. 데이터 가져오기

- 데이터를 어디에서 가져올 수 있나 탐색해 보기.

### 3. 데이터 분석 및 시각화

- 데이터를 어떻게 정제할 것인가?
- 데이터가 의미하는 것은 무엇인가?
- 데이터가 의미하는 것을 시각화 하여 드러내 보기

### 4. 텍스트 분석하여 이렇게 학습시키면 어떨까, 제안해보기(머신 러닝)

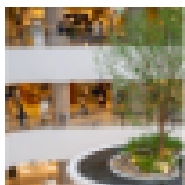
### 5. 미래를 위한 제안

### 6. 프로젝트 성찰 : 나의 발전기록 / 나의 실패 기록 / 나의 가능성 기록

## # 개별프로젝트 안내 : 사례 2

### 1. 텍스트 분석과 관련된 인공지능 분석 사례 따라해 보기

데이콘 : <https://dacon.io/search?tab=total&q=nlp>



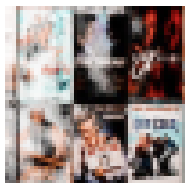
#### 데이콘 Basic 쇼핑물 리뷰 평점 분류 경진대회

알고리즘 | NLP | 분류 | 리뷰 | Accuracy



#### 데이콘 Basic 뉴스 그룹 분류 경진대회

데이콘 베이직 Basic | NLP | Accuracy



#### 데이콘 Basic 영화 리뷰 감성분석 경진대회

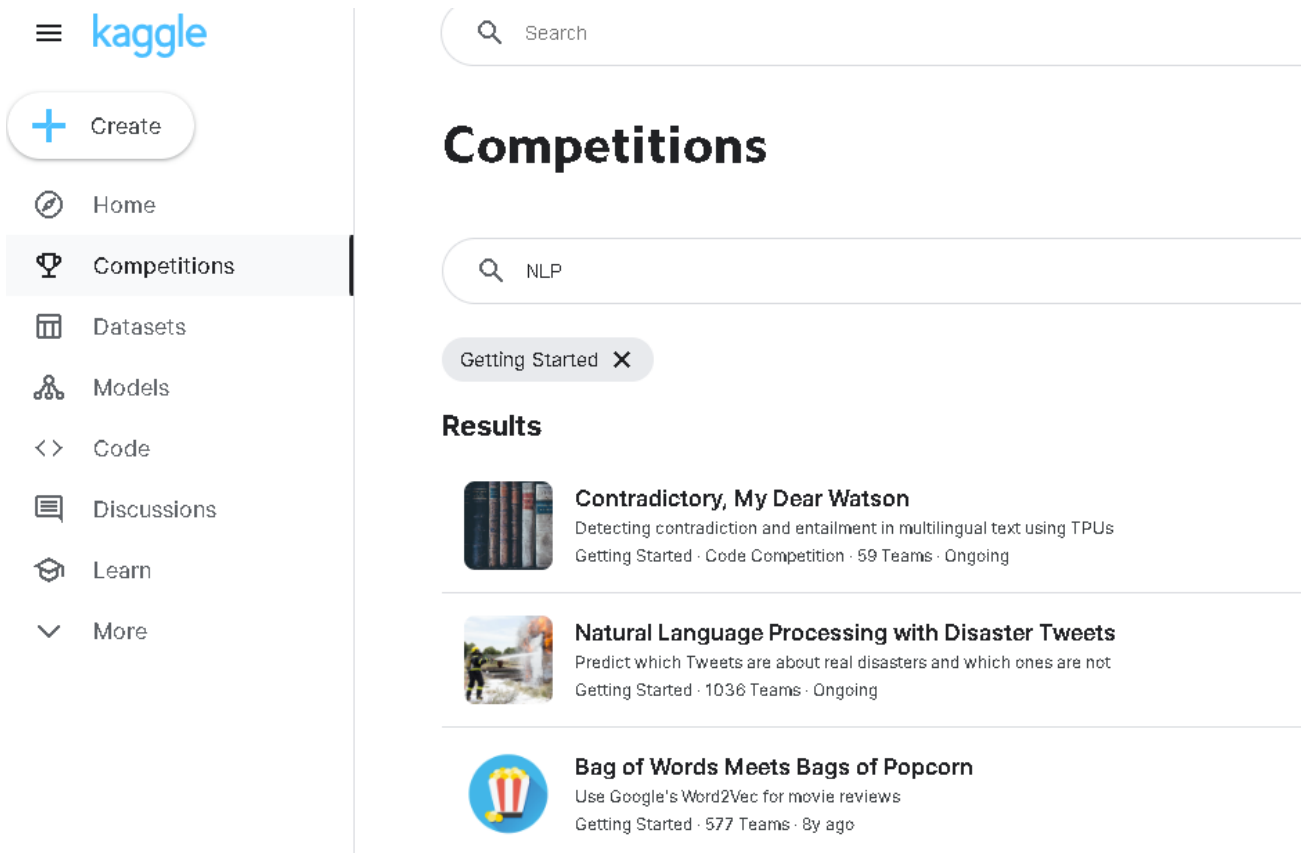
데이콘 베이직 Basic | NLP | Accuracy

6. 프로젝트 성찰 : 나의 발전기록 / 나의 실패 기록 / 나의 가능성 기록

# # 개별프로젝트 안내 : 사례 2

## 1. 텍스트 분석과 관련된 인공지능 분석 사례 따라해 보기

캐글 : <https://www.kaggle.com/competitions?hostSegmentIdFilter=5&searchQuery=NLP>



The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation links: Home, Competitions (highlighted), Datasets, Models, Code, Discussions, Learn, and More. The main content area is titled 'Competitions' and features a search bar with 'NLP' entered. Below the search bar is a 'Getting Started' button. The 'Results' section lists three competitions:

- Contradictory, My Dear Watson**  
Detecting contradiction and entailment in multilingual text using TPUs  
Getting Started · Code Competition · 59 Teams · Ongoing
- Natural Language Processing with Disaster Tweets**  
Predict which Tweets are about real disasters and which ones are not  
Getting Started · 1036 Teams · Ongoing
- Bag of Words Meets Bags of Popcorn**  
Use Google's Word2Vec for movie reviews  
Getting Started · 577 Teams · 8y ago

6. 프로젝트 성찰 : 나의 발전기록 / 나의 실패 기록 / 나의 가능성 기록

# 우리의 목표는 : **머신 러닝을 이용한 문제 해결 위한 사이클에 대해 알게 된다.**

# 모두의 한국어 텍스트 분석

with 파이썬 기초부터 챗GPT까지,  
누구나 쉽게 시작하는 자연어 처리

박조은, 송영숙 지음



1. 파이썬(python) 언어로 쓰여진 **프로그래밍**을 읽을 수 있다.
2. 판다스(pandas), 넘파이(numpy)를 이용해 **데이터를 다룬다**는 것을 안다.
3. matplotlib, seaborn을 이용해 **데이터를 시각화** 한다는 것을 안다
4. **텍스트 분석** 방법을 대충 안다.
5. **머신러닝이 이런 거구나** 대충 안다.





**# 이틀 안에 이런 것 할 수 있는 사람 없다**

**경복고 김선경**