

v2022

AI Researcher

Hojun Eric Choi



<https://github.com/hchoi256>



<https://hchoi256.github.io/>



<https://www.linkedin.com/in/hojun-choi-2b10b11a0/>

# *PORTFOLIO*



# 미래를 끌어오는 차기 AI 연구원 최호준입니다.



안녕하십니까. University of Wisconsin-Madison에서 Computer Science와 Data Science를 복수 전공하고 있는 최호준이라고 합니다.

저는 자율주행 로봇과 검색 엔진에 탑재된 언어 모델에 깊이 매료되어 활용 가치가 무궁무진한 인공지능을 연구하는 꿈을 가지게 되었습니다.

저는 학부에서 통계학적 데이터 분석 및 활용 기술을 연마하고, AI 관련 전공 수업들에서 도메인 지식 및 프로젝트 역량을 쌓아왔습니다. 그간 방학 기간을 활용해서 AI 부트캠프, NLP 개발 과정, LG Research AI 교육 과정을 마친 이력이 있습니다.

소수의 해커톤에 참가하여 본선 진출을 이룬 바 있습니다. 현재는 자율주행 센서의 안테나 성능을 예측하는 해커톤에 힘쓰고 있습니다.

저는 초거대 AI 연구 가능성에 큰 기대를 안고 있습니다. 높은 컴퓨팅 환경이 필요해서 현실적으로 당장 연구 기회가 오지 않을 수 있으나, 항상 최신 정보에 눈 뜨고 있습니다.

저의 관심 연구 방향으로 한국어 이해 및 추론 능력을 갖춘 성능 좋은 대용량 AI 시스템 연구 개발과 CV 분야에서 객체 인식 및 XAI 후속 연구를 추구하고 싶습니다. 감사합니다.

## PROGRAMMING LANGUAGES

|        |     |            |
|--------|-----|------------|
| Python | C   | JavaScript |
| Java   | C++ | SQL        |
| R      | C#  |            |

## EDUCATION/EXPERIENCE

|              |               |             |
|--------------|---------------|-------------|
| UW-Madison   | CS & DS 복전 학사 | 2020 - 2022 |
| MAPAL HiTEC. | 산업체           | 2018 - 2020 |
| SUNY Albany  | 편입            | 2016 - 2017 |

## AWARDS

|             |                       |      |
|-------------|-----------------------|------|
| 진행중         | LG Research Hackathon | 2022 |
| 본선 진출       | Hatathon 경진대회         | 2022 |
| Dean's List | 전학기 성적 우수 표창장         | 2022 |

## CONTACT

|         |   |
|---------|---|
| Phone   | 010-7185-1250   |
| E-mail  | <a href="mailto:hchoi256@wisc.edu">hchoi256@wisc.edu</a>              |
| Website | <a href="https://hchoi256.github.io/">https://hchoi256.github.io/</a> |

# Index

주제 클릭 시 해당 페이지 이동

## 기본

### 영어 가능자

어학 관련 demonstrations

4

### 기타

통계, OS, HCI, Unity, Web, Chrome Extn., ...

7 - 8

### ETL & EDA

ETL과 EDA 활용 개발 및 분석 사례

5 - 6

## AI 개발

### 자연어 처리(NLP) Tasks

문장 생성, 개체명 및 음성 인식, 감성분석, ...

9 - 10

### 다양한 ML/DL Tasks

회귀, 분류, 군집화, PCA, 시계열, 전이학습, ...

11

# 영어 가능자:

2016 - current

최호준

미국 4년제 대학교 22년도 졸업 예정자  
미국 4년제 공인대학 어학과정 최고레벨 수료  
미국 학부생 학업 멘토링  
네이버 지식iN 플랫폼 영어 eXpert 활동

## CERTIFICATE OF COMPLETION

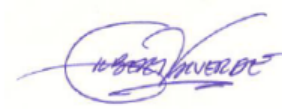
This certificate is presented to

**Hojun Choi**

This certifies that Hojun Choi has successfully completed the

**Advanced 2 Level**

October 24-December 19, 2016



Gilbert A. Valverde, Ph.D.  
Dean and Vice Provost, International Education  
Center for International Education & Global Strategy




Cathleen C. McCarthy  
Assistant Director, Intensive English Language Program  
Center for International Education & Global Strategy

## 지식iN 활동증명서

쥬스 (ch75\*\*\*\*) 님이  
아래와 같이 지식iN에서 활동하고 있음을 증명합니다.

- 등급 식물신
- 시작일 2016.07.09.
- 순위 상위 5만명 (0.0206%) 이내
- 주요활동분야 영어문법, 영어작문, 번역, 통역
- 답변수 2013개 (답변 채택률 71.1%)

- 주요경력 교육기부  
2022 지식iN 교육기부 참여증  
지식 iN eXpert  
(2022.05~) 영어 상담 전문가 활동 중



It may concern,

I confirm our job offer to Hojun (Eric) Choi for the undergraduate student position of Academic Mentor at the Center for Academic Excellence (CAE). Please see the contact information and let me know if there is anything you need.

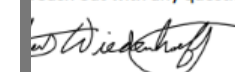
**F-1 student employee name:** Hojun (Eric) Choi  
**Description of employment:** Academic Mentor / Tutor. Hojun (Eric) Choi will work with students in courses in computer sciences, supporting their learning in the course and academic success overall.

**Start date:** 09/07/2021  
**Expected number of hours:** 5-10 hours per week  
**Employer identification number:** 396006492  
**Employer contact information:** Robert Wiedenhoef (supervisor); 608-890-0458; [rwiedenhoef@wisc.edu](mailto:rwiedenhoef@wisc.edu)

**Verification of employment status:** Yes, Hojun (Eric) Choi has been hired and accepted the offer with the start date of 09/07/2021.

**Signature:** See below.

Please reach out with any questions!



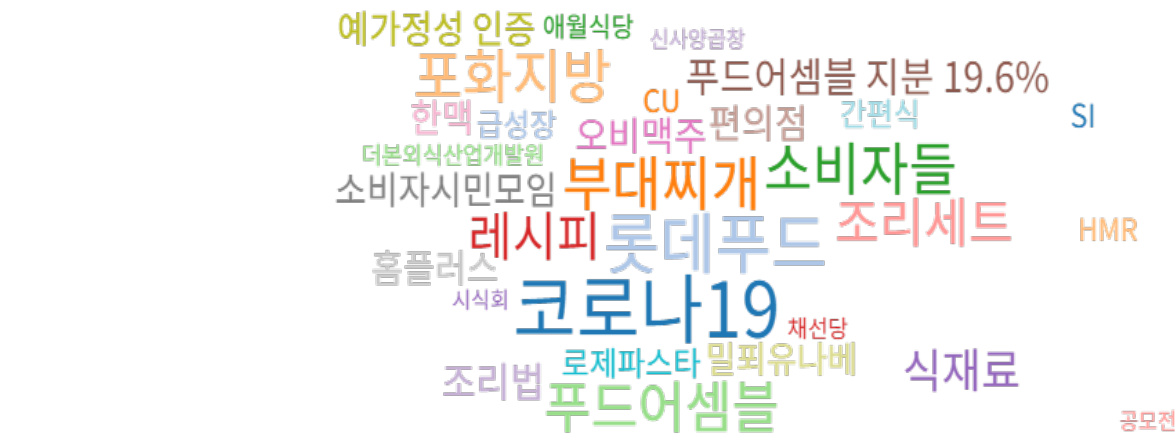
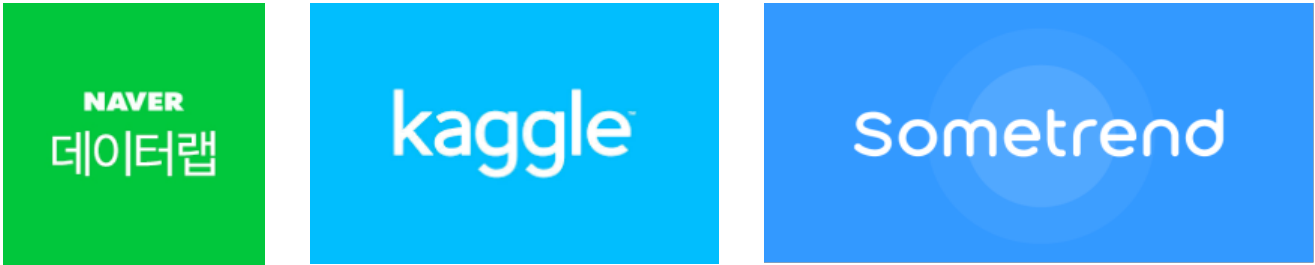
Wiedenhoef

Academic Program Manager: Learning Support Coordinator and Academic Advisor

# ETL & EDA:

2022  
최호준

'밀키트' 연관어 분석 프로젝트  
아이템 기반 협업 필터링 영화 추천 시스템



: 밀피유나베, 코로나19, 간편식 + α

- 신제품 컨셉을 한 마디로 표현하면?  
: **안전 제일!** 혼자서도 즐기는 **초간단 밀피유나베**
- 신제품 고객 타겟층  
: 자취생, 자가격리중인 사람, 1인가구



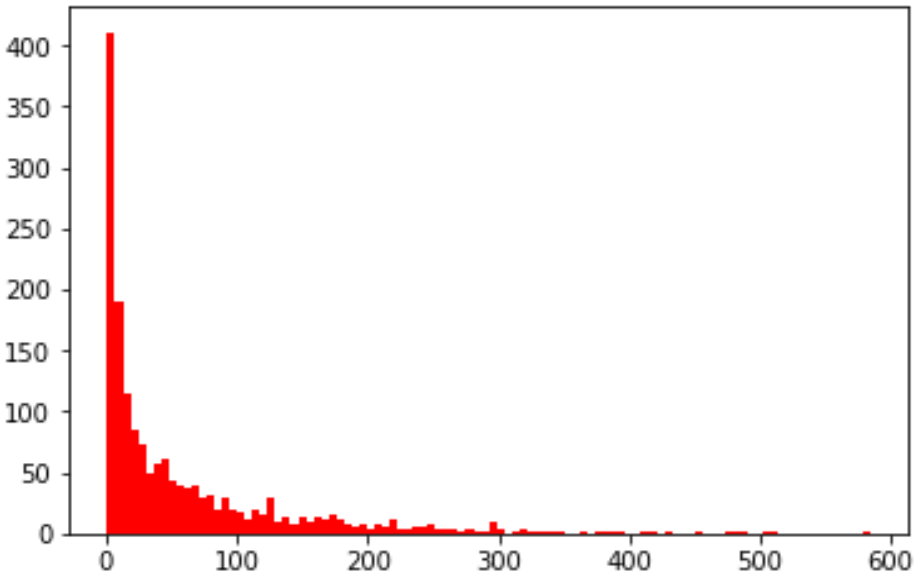
- 2020~2022년, 코로나의 여파로 사람들이 집에서 대부분의 시간을 보냄에 따라 '밀키트'라는 것의 수요가 증가하는 추세이다. 한 마케터는 단순 추측이 아닌 확실한 정보에 근거하여 그 원인을 파악해내고 신제품 제안서를 작성하고자 한다.
- 1. 네이버 데이터랩, '밀키트' 연관어 데이터 csv 추출
- 2. 파워쿼리 활용 테이블 구조화 & 데이터 전처리 (R, Python)
- 3. 변수들 간 상관관계수 & 회귀계수 도출
- 4. 상관/회귀계수 분석 & 유의미한 연관어 정렬
- 5. 최상위 3개 연관어 활용 광고 컨셉 제안서 작성

밀키트 연관어 분석



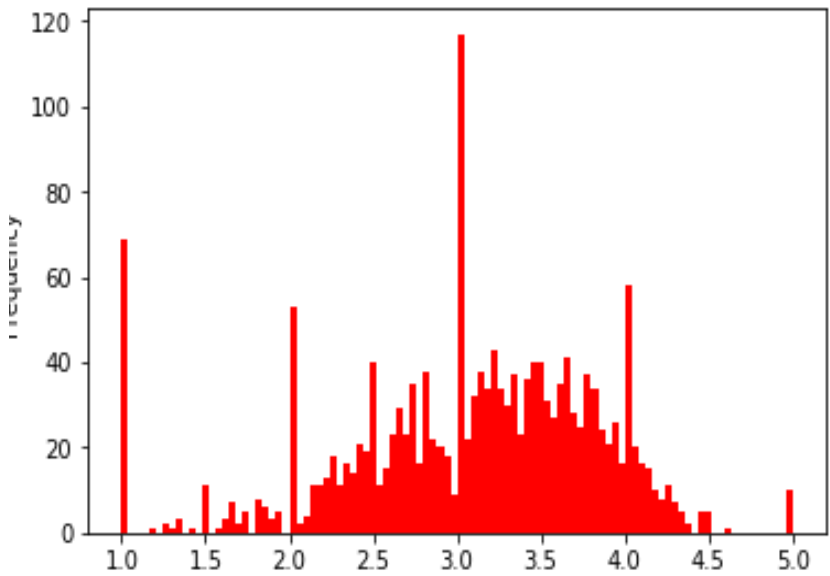
| title                     | count | mean     |
|---------------------------|-------|----------|
| 'Til There Was You (1997) | 9.0   | 2.333333 |
| 1-900 (1994)              | 5.0   | 2.600000 |
| 101 Dalmatians (1996)     | 109.0 | 2.908257 |
| 12 Angry Men (1957)       | 125.0 | 4.344000 |
| 187 (1997)                | 41.0  | 3.024390 |

[영화 데이터셋] count는 영화가 평가받은 횟수, mean은 영화 평점입니다.



[영화 시청 빈도수] x축: 시청 횟수, y축: 시청 횟수의 빈도수

|   | user_id | item_id | rating | title            |
|---|---------|---------|--------|------------------|
| 0 | 0       | 50      | 5      | Star Wars (1977) |
| 1 | 290     | 50      | 5      | Star Wars (1977) |
| 2 | 79      | 50      | 4      | Star Wars (1977) |
| 3 | 2       | 50      | 5      | Star Wars (1977) |
| 4 | 8       | 50      | 5      | Star Wars (1977) |
| 5 | 274     | 50      | 5      | Star Wars (1977) |
| 6 | 227     | 50      | 4      | Star Wars (1977) |
| 7 | 99      | 50      | 5      | Star Wars (1977) |



[영화 평점] 각 영화 별 고객 평가 점수 분포

- 영화 시청 횟수의 빈도수와 영화 평점 분포를 시각화하여 EDA 직관성을 높였다. 대다수의 영화가 대략 한두 번 이내로 고객에게 평가받았고, 고객들은 대체로 영화 평점 3점이 근사하는 점수를 주는 특징을 발견할 수 있다.

|                  | count | mean     |
|------------------|-------|----------|
| title            |       |          |
| Star Wars (1977) | 584.0 | 4.359589 |
| Contact (1997)   | 509.0 | 3.803536 |
| Fargo (1996)     | 508.0 | 4.155512 |

|                              | Correlation | count |
|------------------------------|-------------|-------|
| title                        |             |       |
| Titanic (1997)               | 1.000000    | 350.0 |
| River Wild, The (1994)       | 0.497600    | 146.0 |
| Abyss, The (1989)            | 0.472103    | 151.0 |
| Bram Stoker's Dracula (1992) | 0.443560    | 120.0 |
| True Lies (1994)             | 0.435104    | 208.0 |

- 각 영화별 총평가 횟수와 그 평점을 열로 갖는 하나의 테이블 형태로 통합한다. 영화 'Titanic'을 시청한 고객에게 알맞은 영화를 추천하기 위하여 영화별 'Titanic'과의 상관관계수 값을 추가하고 Correlation 값으로 테이블을 정렬한다.
- 따라서, 우리가 추천해줄 영화는 자신을 제외하고 'Titanic'과 가장 높은 상관관계를 보인 'River Wild, The (1994)'이다.

영화 추천 시스템

# 기타

[HCI Report Papers](#)

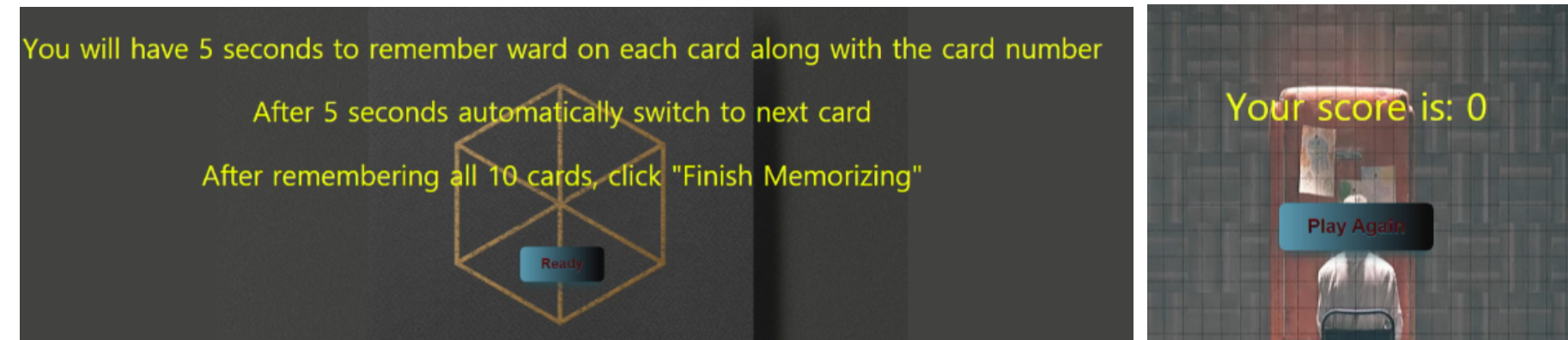
[운영체제](#)

[자료구조](#)

## C++ COVID 자가진단 소프트웨어

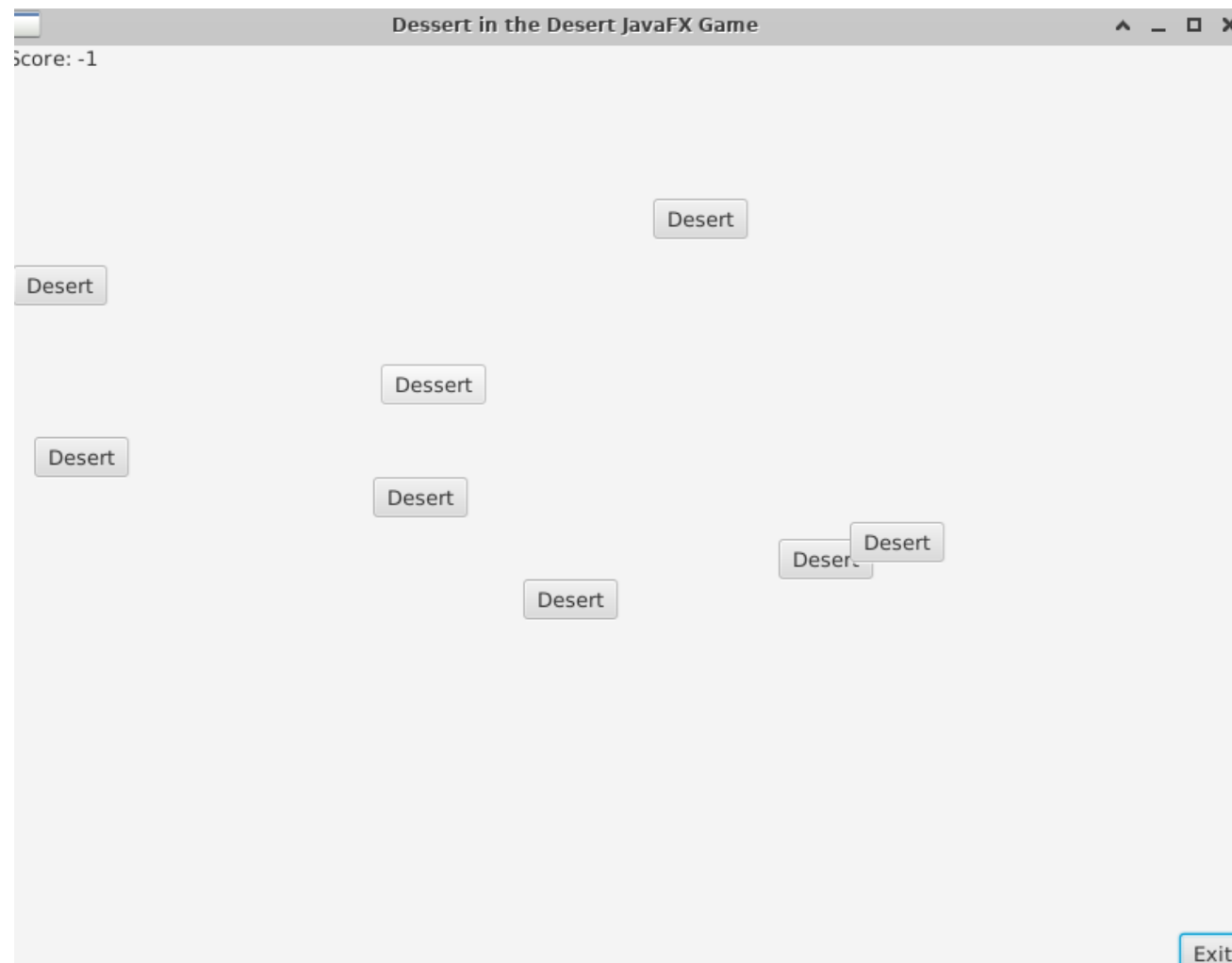
```
*****
Welcome (back) to Safe-ish Badgers!
[1] View testing locations
[2] Make test reservation
[3] Use symptom checker
[4] Check for Covid - 19 Exposures
[5] View my reservations
[6] View my personal data
[7] Adjust my personal data
[8] Change settings
[9] Exit program
Please enter a menu item:
*****
```

## 순간 기억력 온라인 테스트

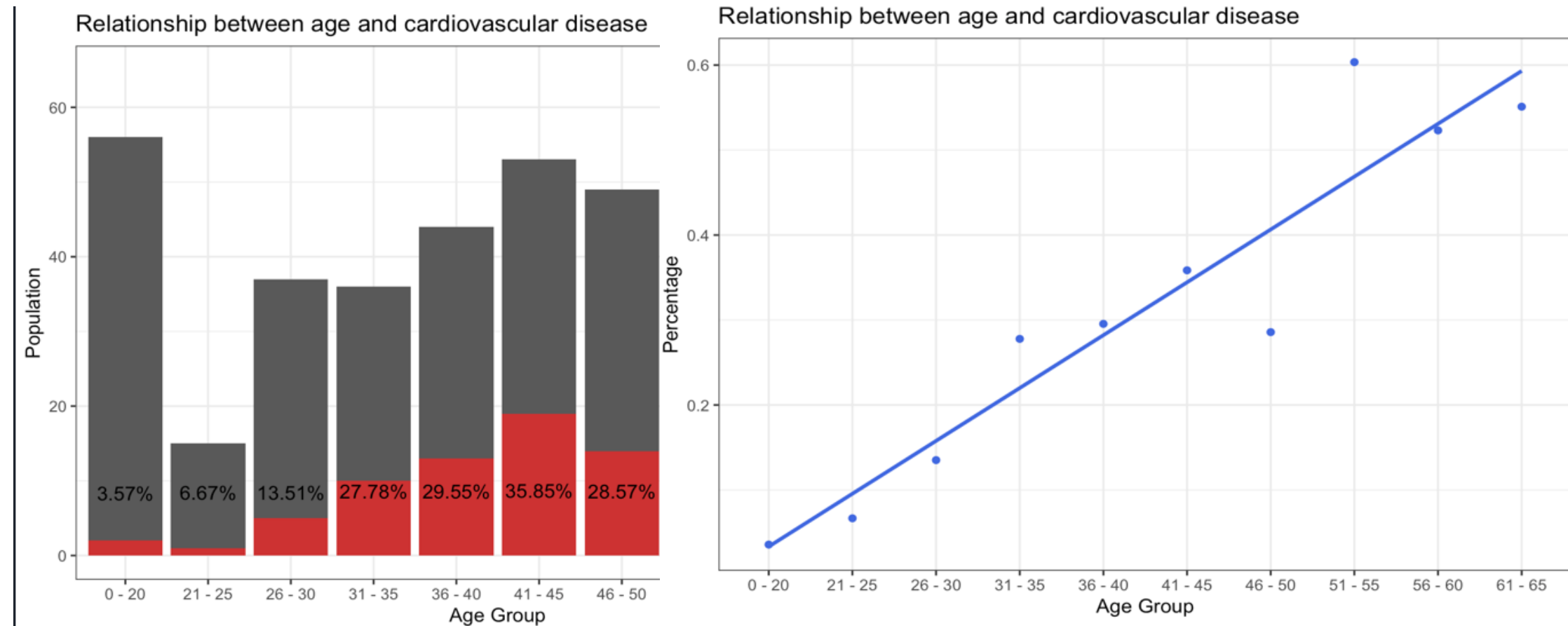


사용자는 테스트를 시작하기 전 별명을 입력한다. 테스트를 시작하면 5초 간격으로 연이어 (단어, 숫자) 10쌍을 보게 된다. 이후, 랜덤하게 나오는 단어 혹은 숫자를 보고, 몇 초안에 알맞은 counterpart를 입력해야 한다. Java, CGI로 구축한 웹 서버 DB에 저장된 유저들의 점수로 마지막에 LeaderBoard를 Popup으로 보여준다.

## JavaFX 미니게임



## 심혈관 질환 통계학적 원인 분석



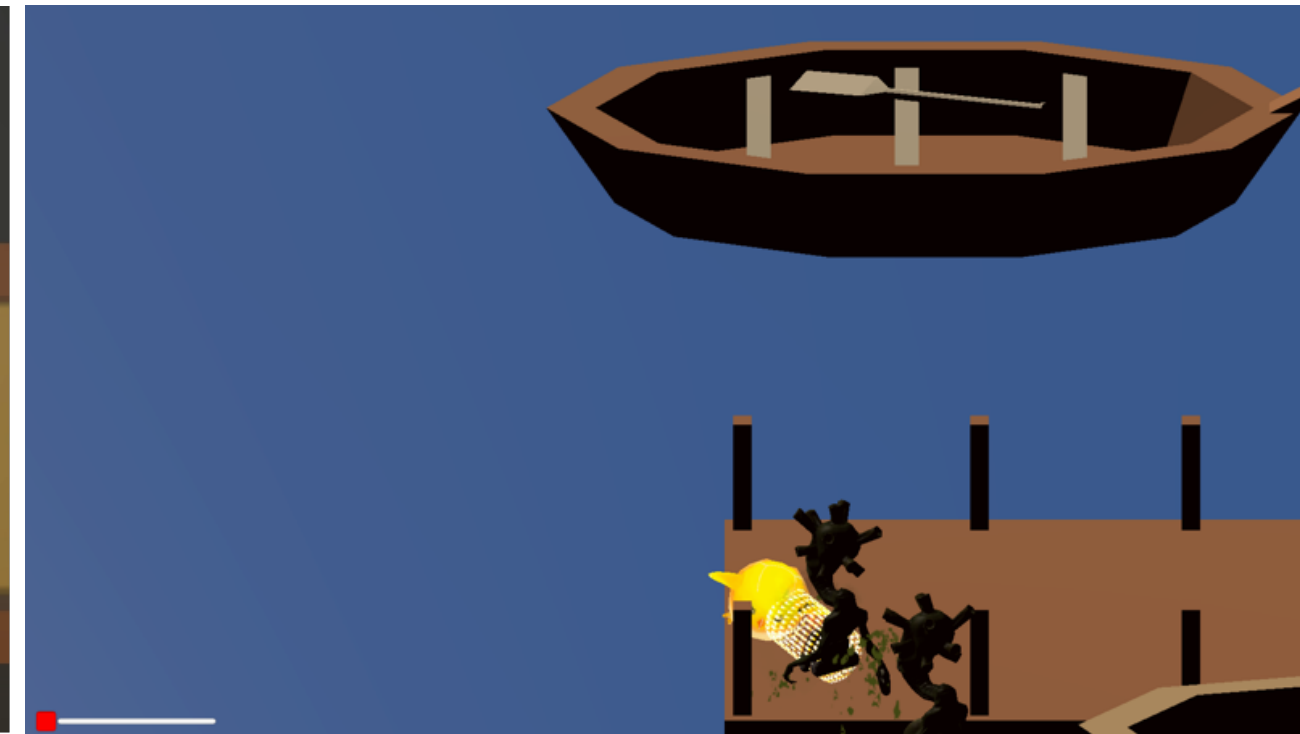
해당 보고서는 Kaggle에서 손쉽게 찾을 수 있는 심혈관 데이터셋을 참조한다. R 언어로 데이터 전처리 과정을 거치고, 머신러닝 회귀와 통계 분석을 통하여 심혈관 질환의 가능성 큰 위험 요인을 받침 하는 시각화 자료를 도출한다.

# 기타:

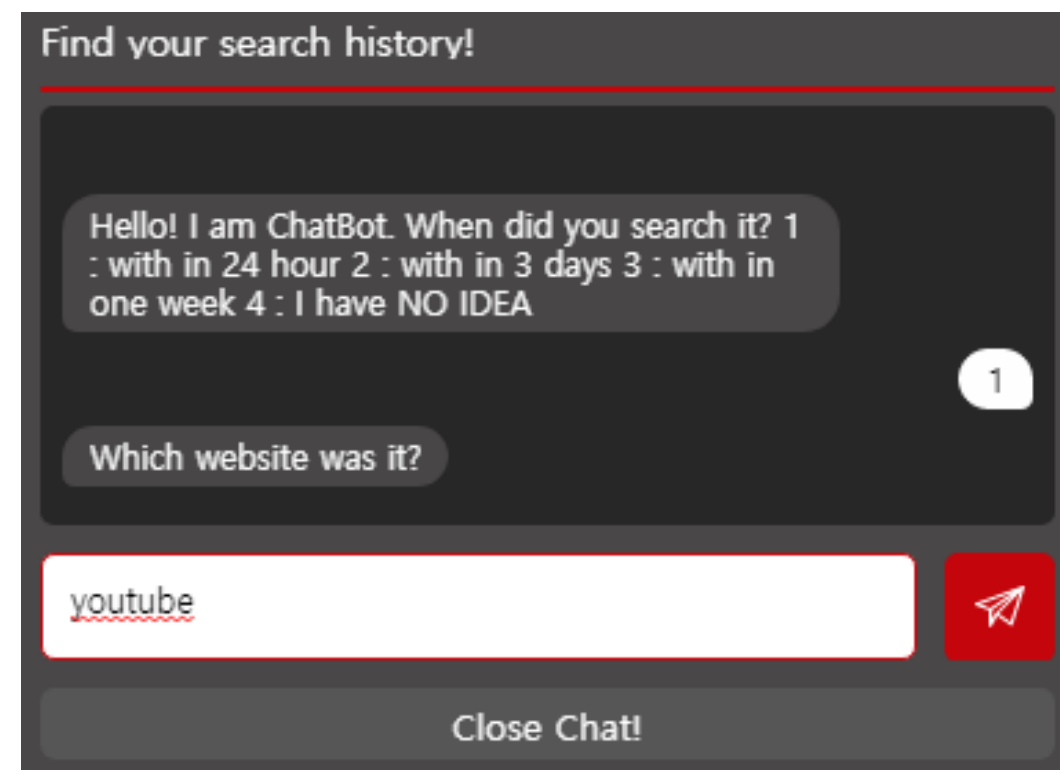
C# Unity 슈팅게임



C# Unity 로그라이크 게임

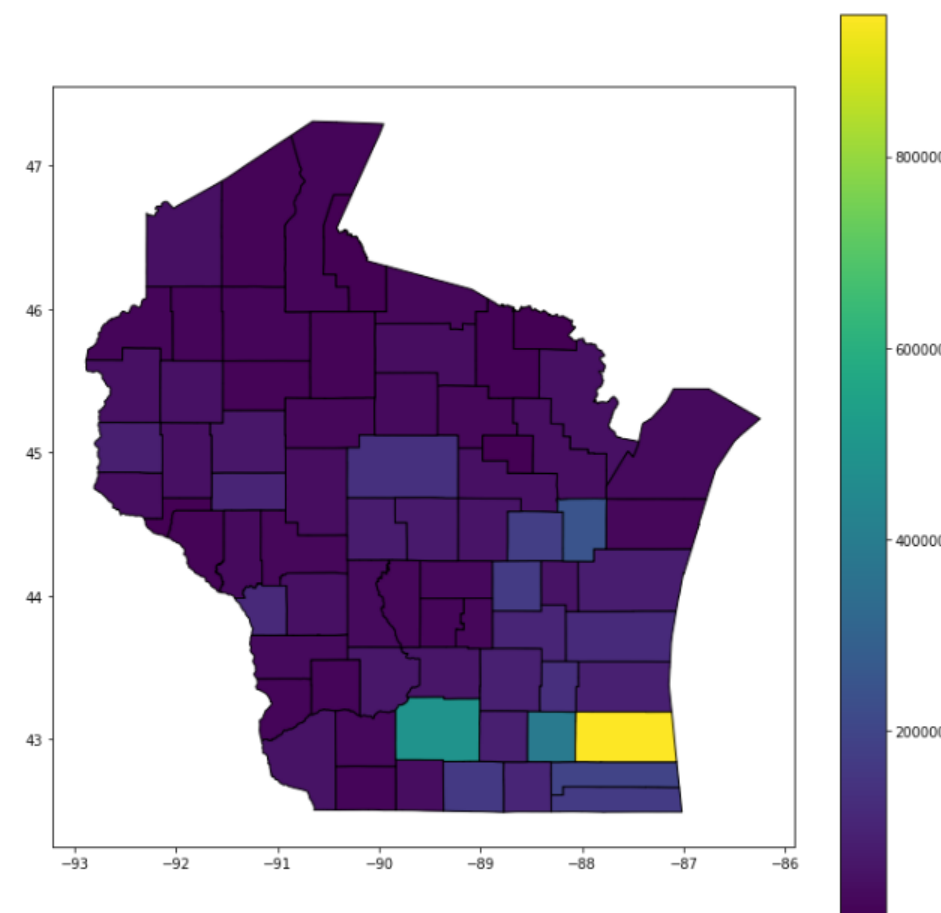


Crome Extension App

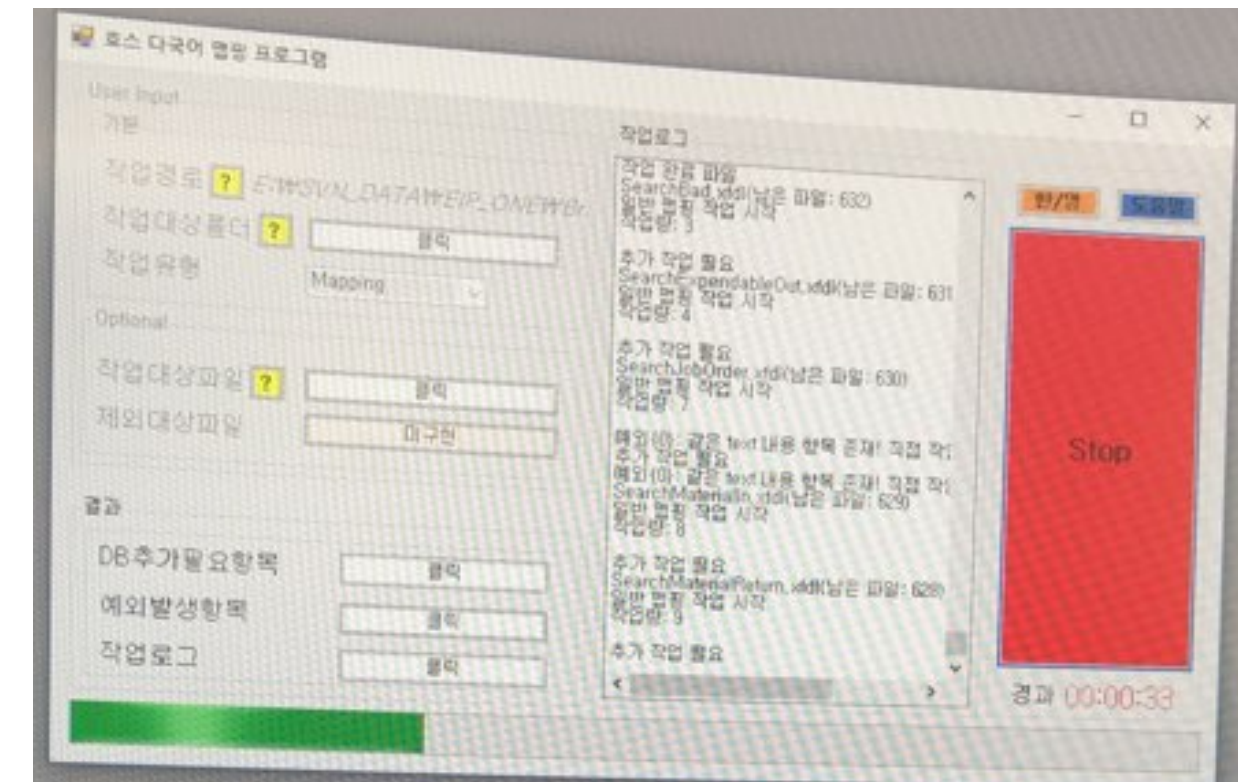


A chrome extension to search forgotten webpages

데이터 모델링



Window C# Forms: 자동화 프로그램



[산업체 전산실 ERP 솔루션 앱] 당사 프로그램에서 언어 전환 시 다수 컴포넌트의 text가 변하지 않았다. 이것은 수동 작업을 필요로 하는데, 당시 부서에서 낮은 우선순위로 처리가 밀리고 있었던 업무였다. 자동화 프로그램 개발로써 해결했다.



# 자연어 처리 Tasks

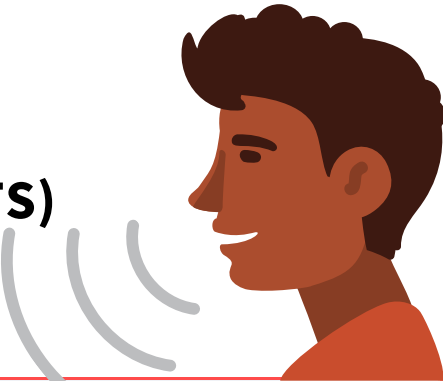
## 음성인식 번역기 (STT/TTS)

번역할 문장을 넣어주세요

반가워요

Schön, dich kennenzulernen

[음성인식 번역기] 음성언어 Encoding 후 텍스트 변환(STT) & 타언어로 Decoding하고 자동 발음 (TTS)



## 개체명 인식 (NER)

홍길동은 2022년에 대학교를 졸업한다.

홍길동 - 사람  
2022년 - 시간  
대학교 - 조직

[개체명 인식] 사람, 조직, 시간에 대한 NER 모델

James is working at Disney in London.

(S (PERSON James/NNP), is/VBZ, working/VBG, at/IN, (ORGANIZATION Disney/NNP), in/IN, (GPE London/NNP))

[개체명 인식] 품사 태깅

## GPT-2 문장 생성

"There are times when we are really tired of people but we feel lonely too"

[인풋 문장]

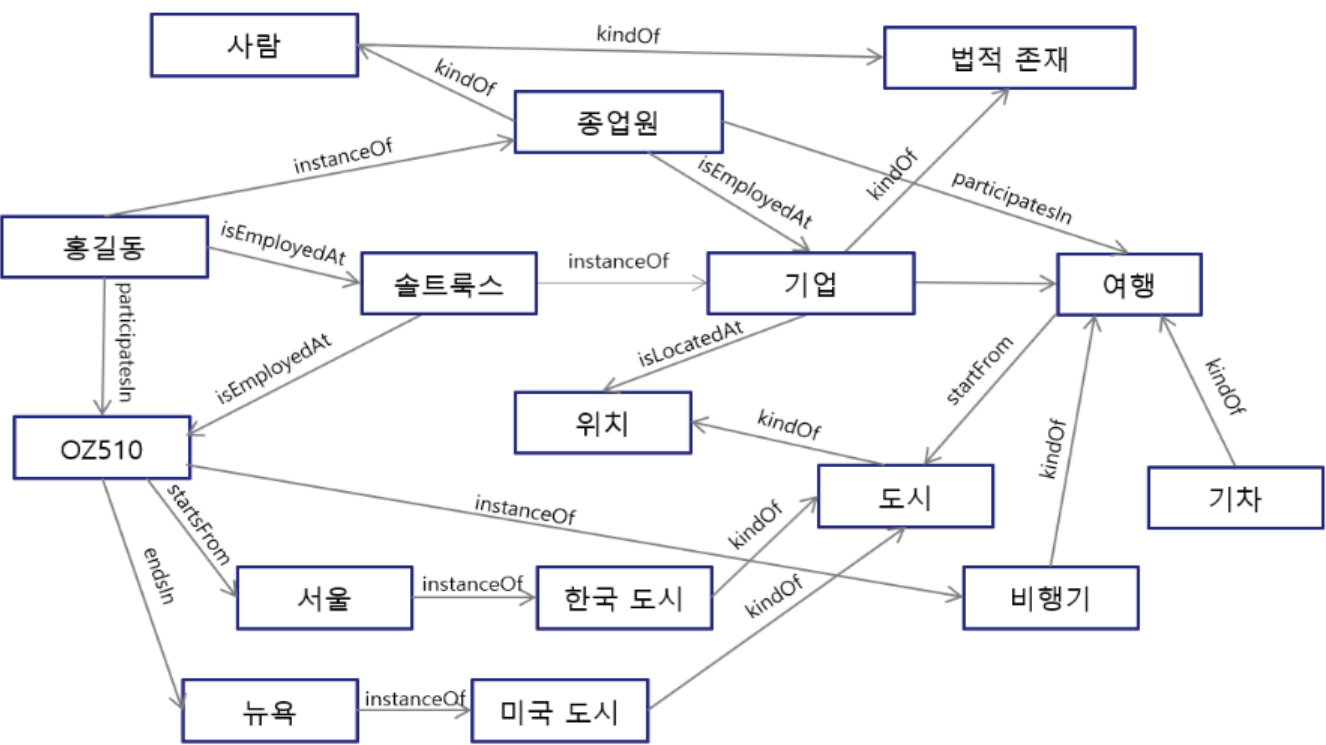
Output:

There are times when we are really tired of people but we feel lonely too. There are times when we're not sure what to do and want to leave but there is a lot of pressure. There are times when we're frustrated and don't know what to do. There are times when we don't know if we're right or wrong.

There are times when we are really tired of people but we feel lonely too," he said. He has no doubt that the Chinese government will try to silence the opposition in any way possible. "The Chinese government is going to try to silence us with the use of the law, that's my feeling," he said.

[인풋 문장] 위키 문장 데이터셋을 학습한 GPT-2 모델을 활용해서 주어진 문장과 자연스럽게 이어지도록 다음 문장을 생성합니다.

## KoBERT 연관어 추출



[연관어 추출] KAIST 제공 데이터셋(gold-standard-v1) - 정확도 90% ↑

# 자연어 처리 Tasks

## KoBERT 네이버 영화 리뷰 감성분석

6270596 굳 ㅋ 1  
 9274899 GDNTOPCLASSINTHECLUB 0  
 8544678 뭐야 이 평점들은 . . . . 나쁘진 않지만 10점 짜리는 더더욱  
 아니잖아 0  
 6825595 지루하지는 않은데 완전 막장임 . . . 돈주고 보기에는 . . . . 0  
 6723715 3D만 아니었어도 별 다섯 개 줬을텐데 . . 왜 3D로 나와서 제  
 심기를 불편하게 하죠?? 0  
 7898805 음악이 추가 된, 최고의 음악영화 1  
 6315043 진정한 쓰레기 0  
 6097171 마치 미국애니에서 튀어나온듯한 창의력없는  
 로봇디자인부터가, 고개를 찢게한다 0  
 8932678 갈수록 개판되가는 중국영화 유치하고 내용없음 품잡다 끝남  
 말도안되는 무기에 유치한cg남무 아 그림다 동사서독같은 영화가 이건  
 3류아류작이다 0

[네이버 영화 리뷰 감성분석] 0: 부정, 1: 긍정 - 정확도 90% ↑

## 기사 한/영 요약번역 프로그램

### Article Summarizer & Online Translating Tool (ENG/KOR)

Press your URL here.

https://en.wikipedia.org/wiki/Korea|



35/500

Summary

The modern South Korean school system consists of six years in elementary school, three years in middle school, and three years in high school. Goryeo (also spelled as Koryŏ), whose name developed into the modern exonym "Korea", was a highly cultured state that created the world's first metal movable type in 1234. Thus, one born immediately before New Year's Day may only be a few days old in western reckoning, but two years old in Korea. Accordingly, a Korean person's stated age (at least among fellow Koreans) will be one or two years more than their age according to western reckoning. Toward the end of the 1st millennium, Goguryeo was resurrected as Goryeo, which defeated the two other states and unified the Korean Peninsula as a single sovereign state.

Translated

한국의 현대 학교 시스템은 초등학교 6년, 중학교 3년, 고등학교 3년으로 구성되어 있습니다. 현대식 이름인 "Korea"로 이름이 발전한 고려는 1234년에 세계 최초로 금속활자를 만든 고도로 교양된 국가였습니다. 따라서 설날 직전에 태어난 사람은 고작 며칠 밖에 되지 않을 수도 있습니다. 서양식으로는 늙었지만 한국에서는 2살. 따라서 한국인의 명시된 나이(적어도 동포 사이에서는)는 서구식으로 그들의 나이보다 1~2년 더 많을 것입니다. 1천년 만에 고구려가 고려로 부활하여 두 나라를 무찌르고 한반도를 하나의 주권국가로 통일하였다. 아이들은 유치원 1년, 초등 교육 4년, 중등 교육 6년을 거쳐 대학에 진학합니다. 두 주는 여전히 공통의 전통 문화와 냉전 이전 역사를 공유하고 있지만 분할 이후 문화적으로나 경제적으로 크게 갈라졌습니다.

Made by Eric Choi

[텍스트 요약번역] 프로그램 세부 옵션 지정 후, 기사 URL 입력 시 요약본과 번역본을 보여준다.

# 다양한 ML/DL Tasks

04

## Teeko 보드게임

```

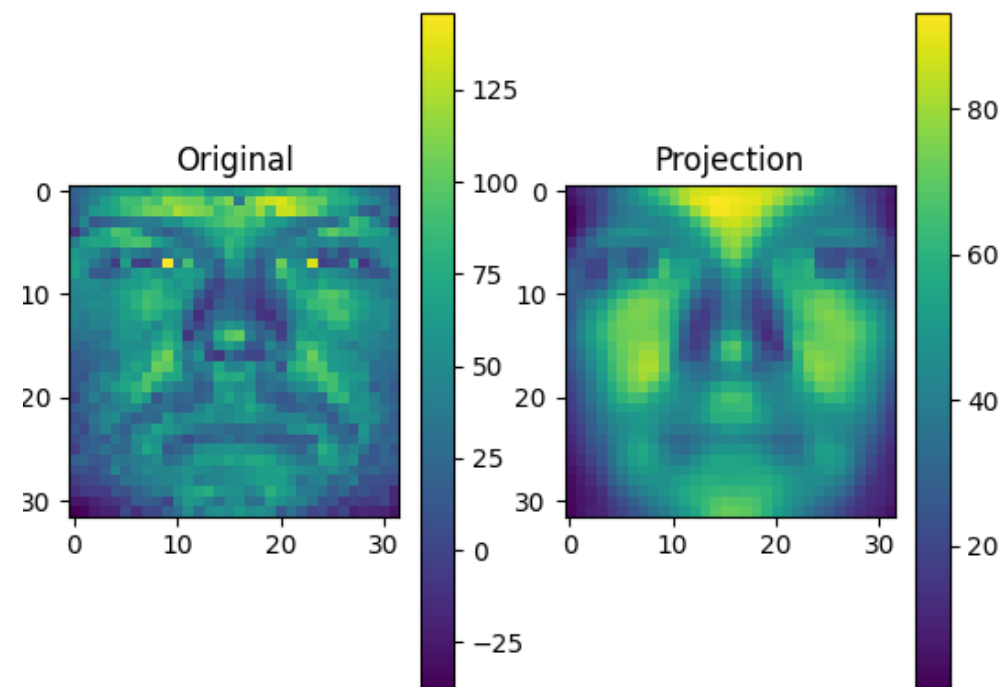
Hello, this is Samaritan Move (e.g. B3): B3
0:      0:
1:      1:
2:      2:
3:      3:  b
4:      4:
   A B C D E      A B C D E
[]
r moved at C2
0: r b b
1:  r
2:  r
3:  b
4:
   A B C D E
b's turn
Move (e.g. B3):
  
```

[게임 이론] Minimax 알고리즘 활용  
Teeko 보드게임

Teeko 보드게임은 플레이어와 AI 로봇이 번갈아 가며 마커를 놓는다. 한쪽이 먼저 4개의 마커를 수직, 수평, 대각, 혹은 2x2박스 형태로 놓으면 승리한다.

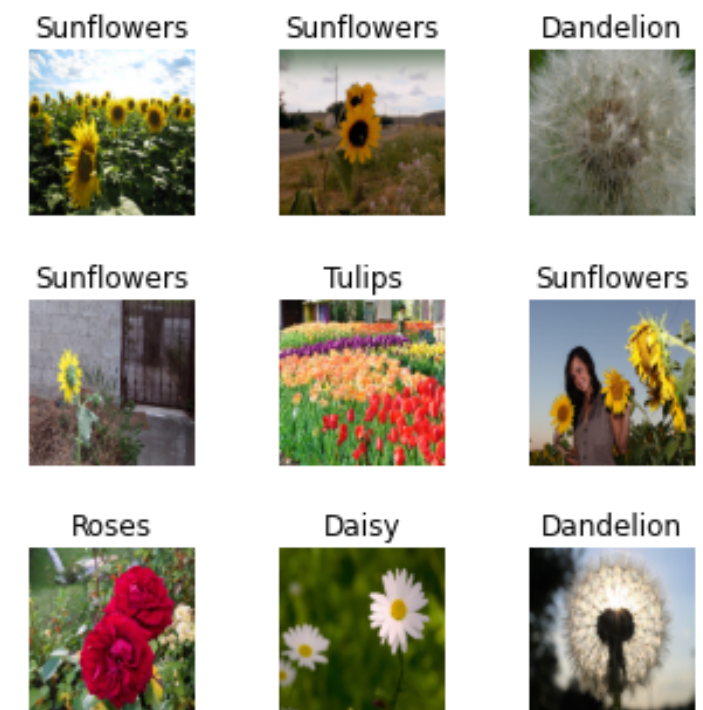
AI 로봇은 플레이어의 마커를 피해서 가장 빠른 성공 공식인 대각선을 침투하는 모습이다.

## 이미지 압축



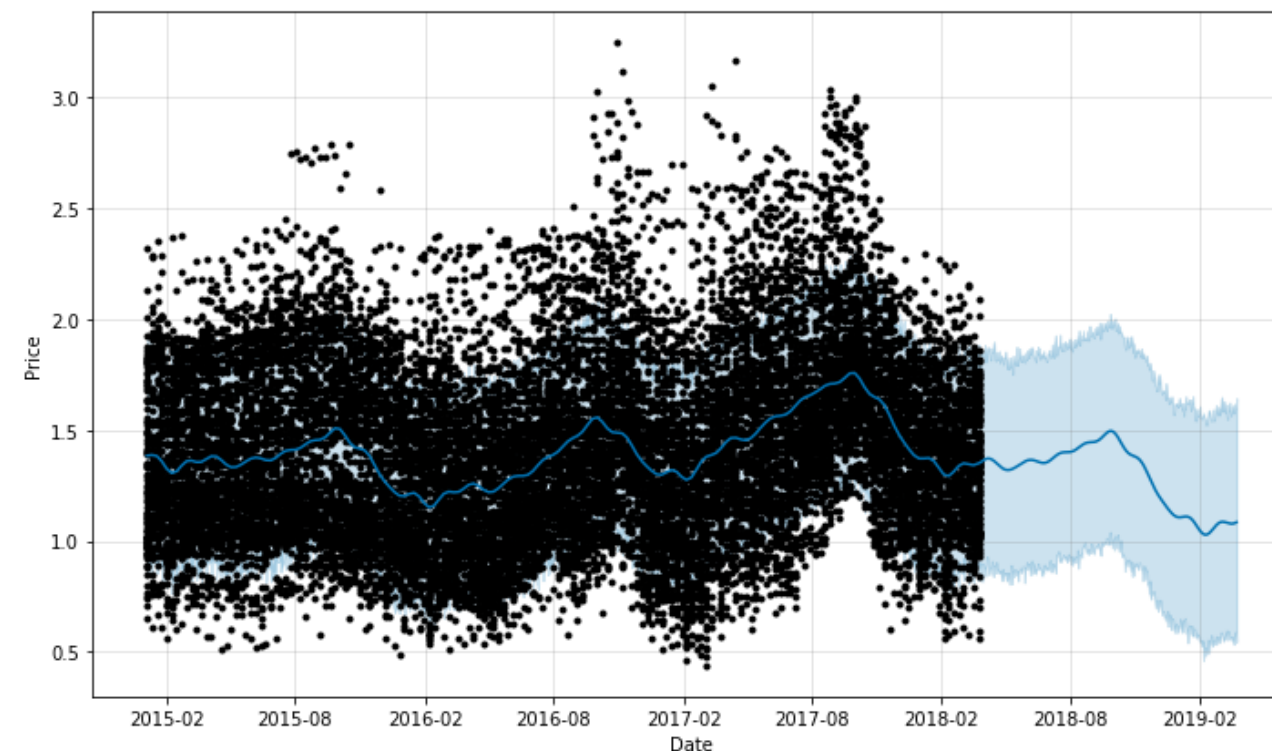
[이미지 압축] PCA 차원축소 활용 이미지 features 조절

## 전이학습



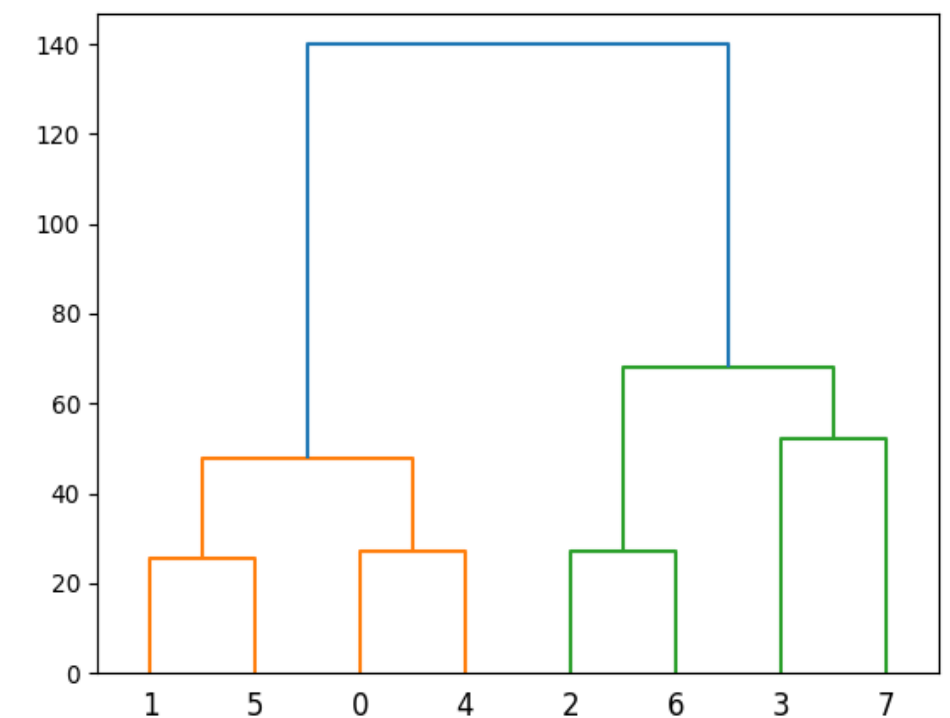
[전이학습] MobileNet 사전학습모델 fine-tuning 적용해서 원하는 분류 결과 도출

## 시계열 분석



[시계열 분석] LSTM/페이스북 Prophet 활용 미래 데이터 예측하기

## 포켓몬 군집화



[HAC] x축: 포켓몬 ID, input: 군집 개수  
HAC Complete-linkage 활용 군집 개수만큼 유사 포켓몬 묶어주기