

v2022

AI Researcher

Hojun Eric Choi



<https://github.com/hchoi256>



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



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

# PORTFOLIO :

# Leading the Future

## Next AI Researcher Eric Choi



Hello, I'm Choi Hojun, majoring in Computer Science and Data Science at the University of Wisconsin-Madison. I was deeply fascinated by language models installed in autonomous robots and search engines, which led me to dream of doing research on AI with infinite value.

I have honed statistical data analysis skills in school and built domain knowledge with project capabilities through many AI courses. Beyond school, I completed an AI boot camp, an NLP development course, and an LG Research AI training program. I also had some successful hackathons that I made to the finals; autonomous sensors prediction, software development, 3d image classification, and autoML. I am currently working as an undergraduate research student in the Sensing group, the self-driving research team.

I have high expectations for the possibility of ultra-large AI research (albeit the need for a high computing environment). My research interests are large-scale AI systems, deep NLP, XAI, and object detection/segmentation.

### PROGRAMMING LANGUAGES

Python	C	JavaScript
Java	C++	SQL
R	C#	

### EDUCATION/EXPERIENCE

UW-Madison	CS & DS	2020 - 2022
MAPAL HiTEC.	S/W Dev. Intern	2018 - 2020
SUNY Albany	Transferred	2016 - 2017

### AWARDS

Ongoing	AutoML Decathlon	2022
Top 10%	LG Research Hackathon	2022
Final Round	UPL Hatathon	2022
Dean's List	Academic Excellence Award	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 :

*Move to the page on click*

---

## **Basic**

### Engligh Proficiency

Certificates

4

### ETL & EDA

ETL & EDA

6 - 7

### Research Experience

5

### Others

Statistics, OS, HCI, Unity, Web, Chrome Ext  
n., ...

8 - 9

## **AI Dev.**

### NLP Tasks

Text Generation, NER, Voice Recog.,  
Sentiment Analysis, ...

10 - 11

### ML/DL Tasks

Regression, Classification, Clustering, PCA,  
Time Series, Transfer Learning, ...

12

### Certificates

13 -14



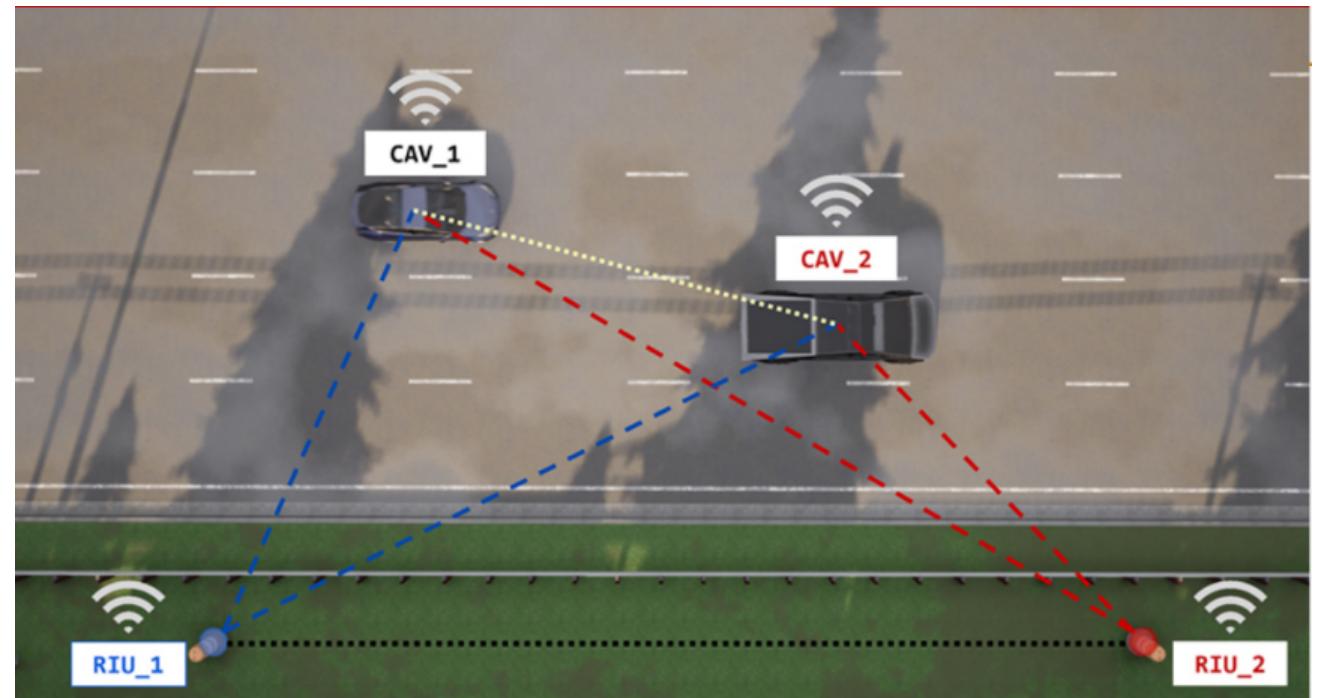
# Research Experience

2022 ~ current

Hojun (Eric) Choi

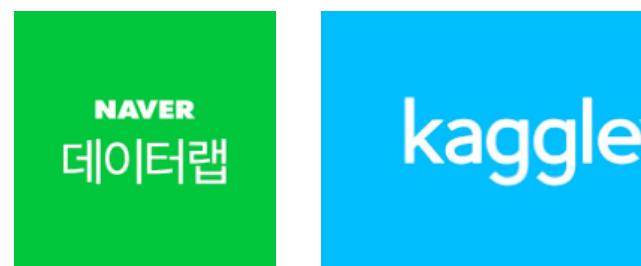
[CARLA Research Project](#)

An autonomous driving simulator

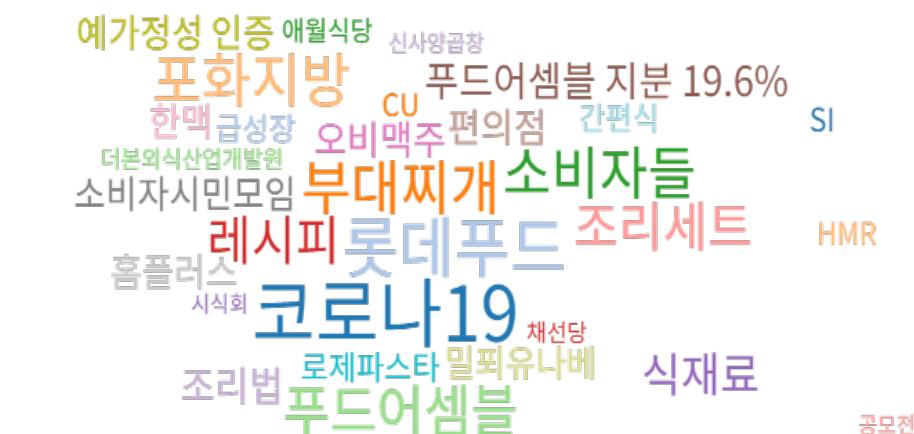


# ETL & EDA :

2022  
Hojun (Eric) Choi



['Meal kit' Word Association Analysis](#)  
[Item-based collaborative filtering](#)



: 밀푀유나베, 코로나19, 간편식 + α

• 신제품 컨셉을 한 마디로 표현하면?

: 안전제일! 혼자서도 즐기는 초간단 밀푀유나베

• 신제품 고객 타겟층

: 자취생, 자가격리중인 사람, 1인가구



In 2020-2022, as people spend most of their time at home in the aftermath of the coronavirus, the demand for 'meal kit' is on the rise. A marketer wants to find out the cause and promote a new product based on solid information.

1. Naver Data Lab, 'Meal kit' related word data csv extraction
2. Table structuring & data preprocessing
3. Correlation & regression coefficient between variables
4. Sort significantly associated words
5. Write an ad proposal with the top 3 related words

Meal Kit

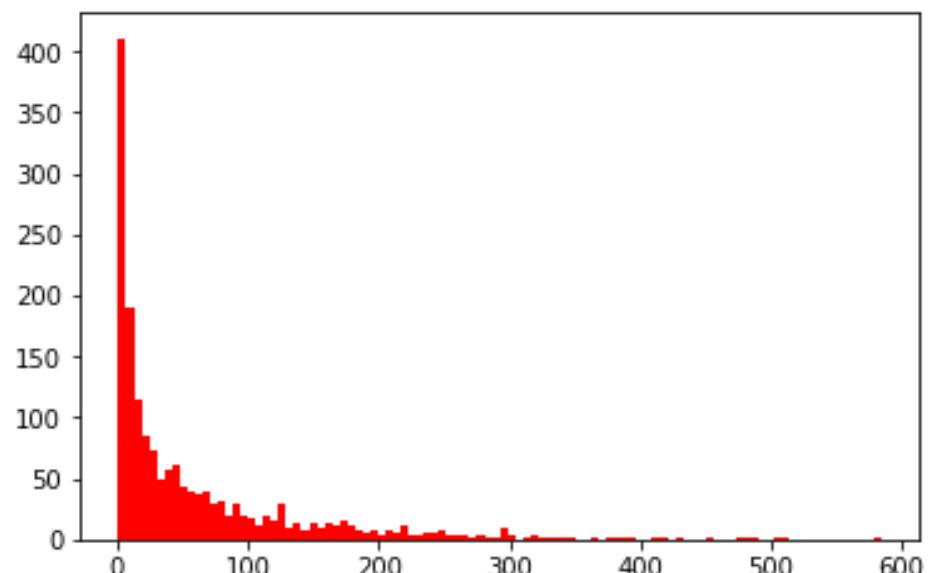
## ETL & EDA

	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	

[Movie dataset]

count is the number of times a movie has been rated

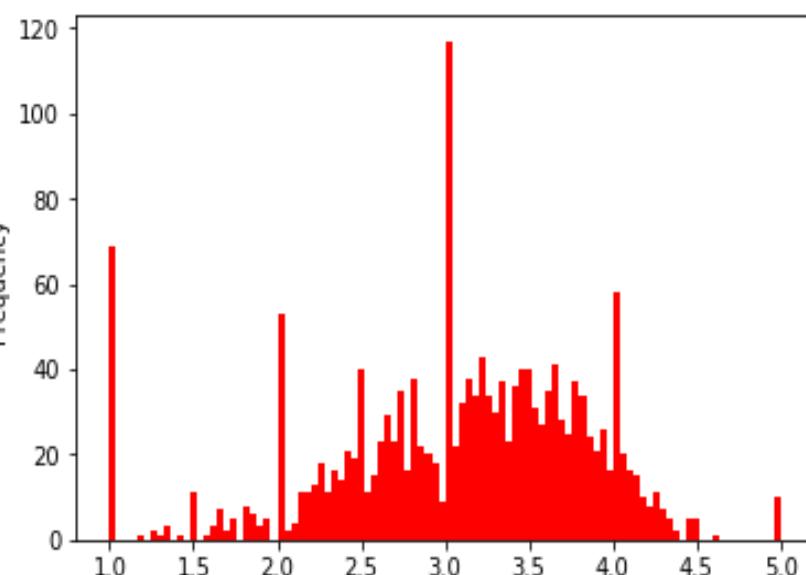
mean is the movie rating.



[Frequency of movie viewing]

x-axis: number of views, y-axis: frequency of viewing

	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)



[Movie Rating] Distribution of customer evaluation scores for each movie

Enhanced EDA intuition by visualizing the frequency of movie viewing and the distribution of movie ratings.

The vast majority of movies have been rated by customers about once or twice, and customers usually give them a score that approximates a movie rating of 3 stars.

	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

The total number of ratings for each movie and the ratings are combined into one table with columns. In order to recommend a suitable movie to customers who have watched the movie 'Titanic', a correlation coefficient value with 'Titanic' is added for each movie and the table is sorted by correlation value. Therefore, the movie we would recommend is 'River Wild, The (1994)', which has the highest correlation with 'Titanic' except for himself.

## Movie Recommendation System

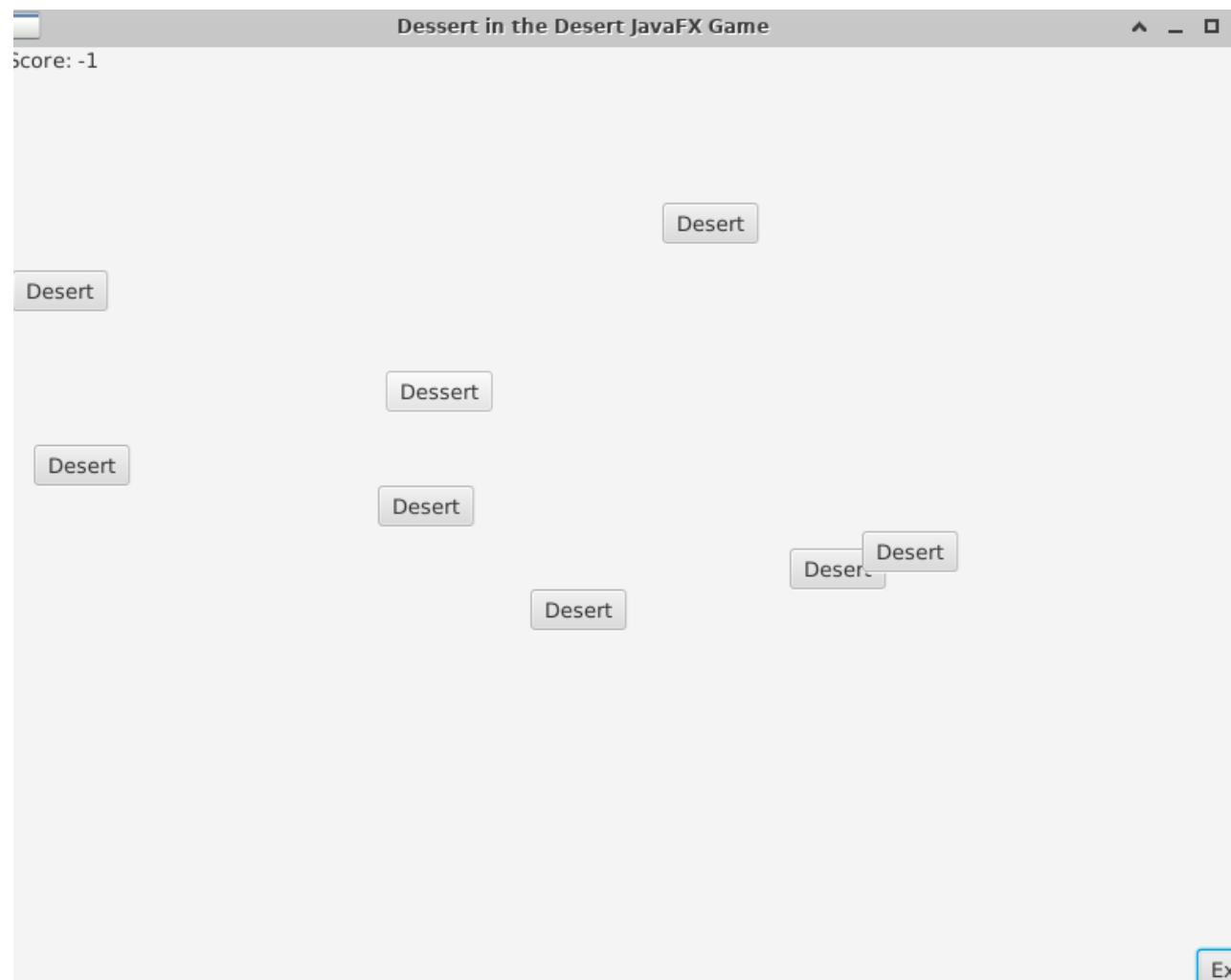
# Others :

[HCI Report Papers](#)

[Operating Systems](#)

[Data Structure](#)

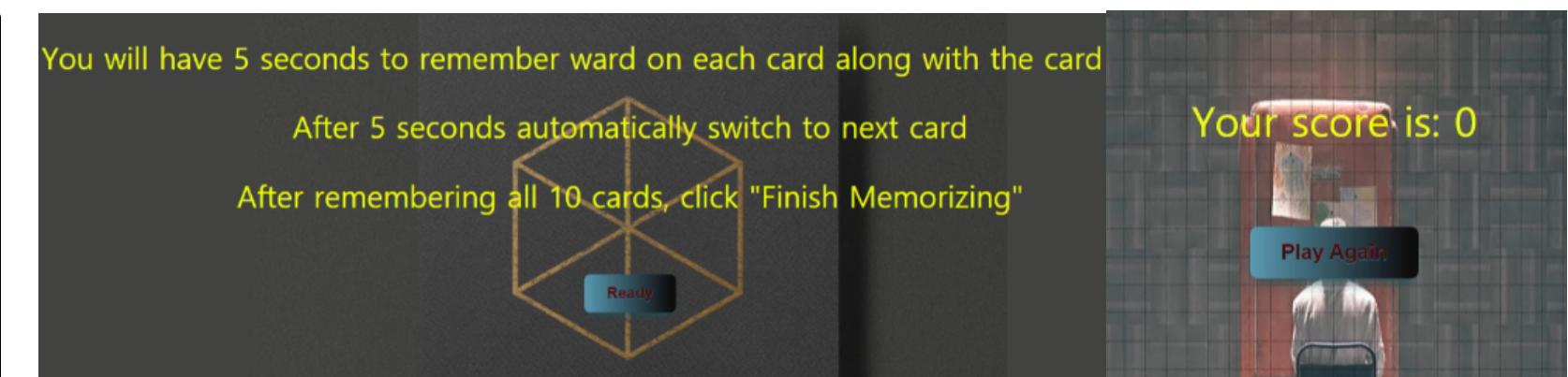
## JavaFX Minigame



## C++ COVID Self-Checking

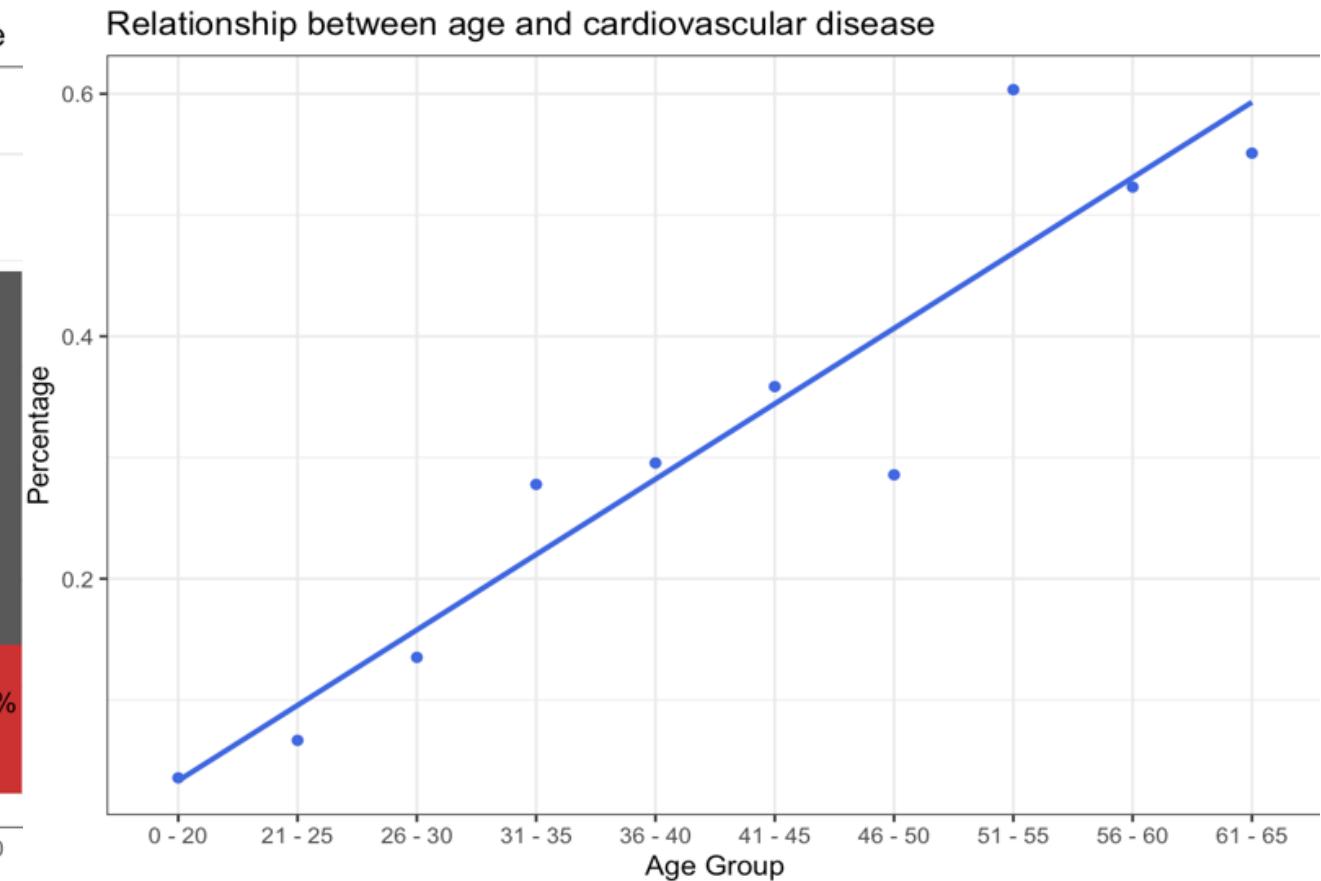
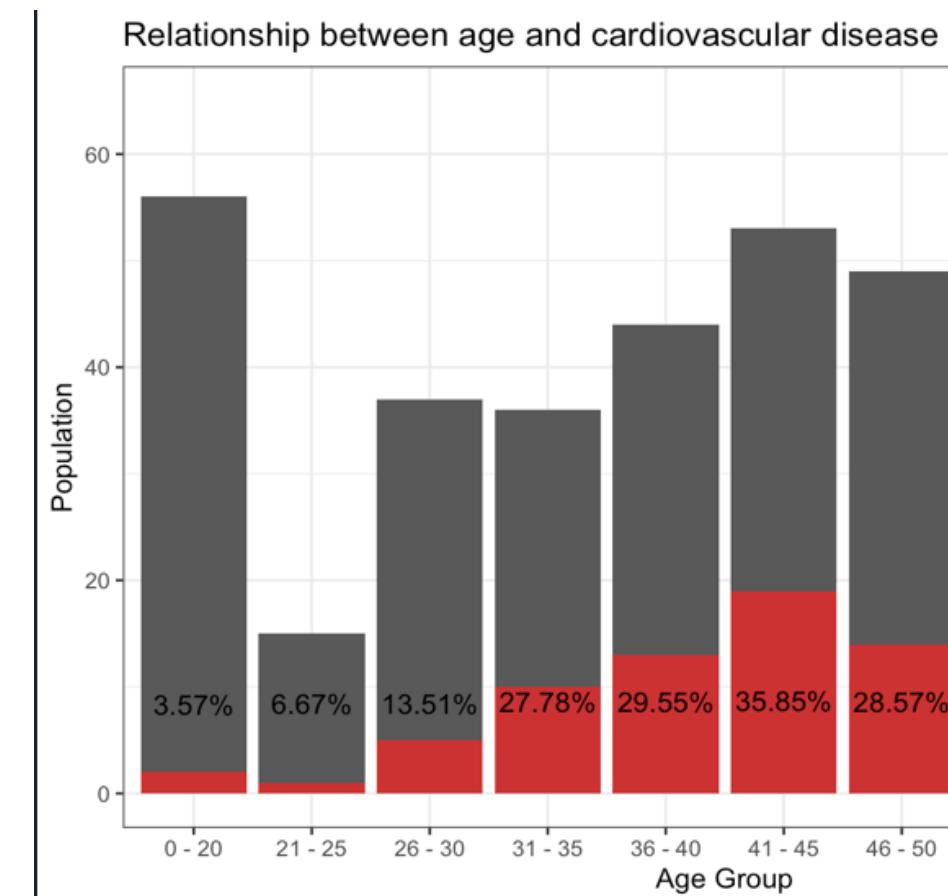
```
*****
Welcome (back) to Safe-ish Badgers!
[1] View testing locations
[2] Make test reservation
[3] Use symptom checker
[4] Check for Covid - 19 Exposures
[5] View my reservations
[6] View my personal data
[7] Adjust my personal data
[8] Change settings
[9] Exit program
Please enter a menu item:
*****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

## Short-term Memory Test



1. Enter nickname.
2. Upon Game Start, 10 pairs of (words, numbers) in a row at 5 second intervals.
3. After that, either a word or number appears randomly, and you have to enter the appropriate counterpart within a few seconds.
4. Displayed the leaderboard at the end with the scores of users stored in the web server DB.

## Cardiovascular Disease Statistical Cause Analysis



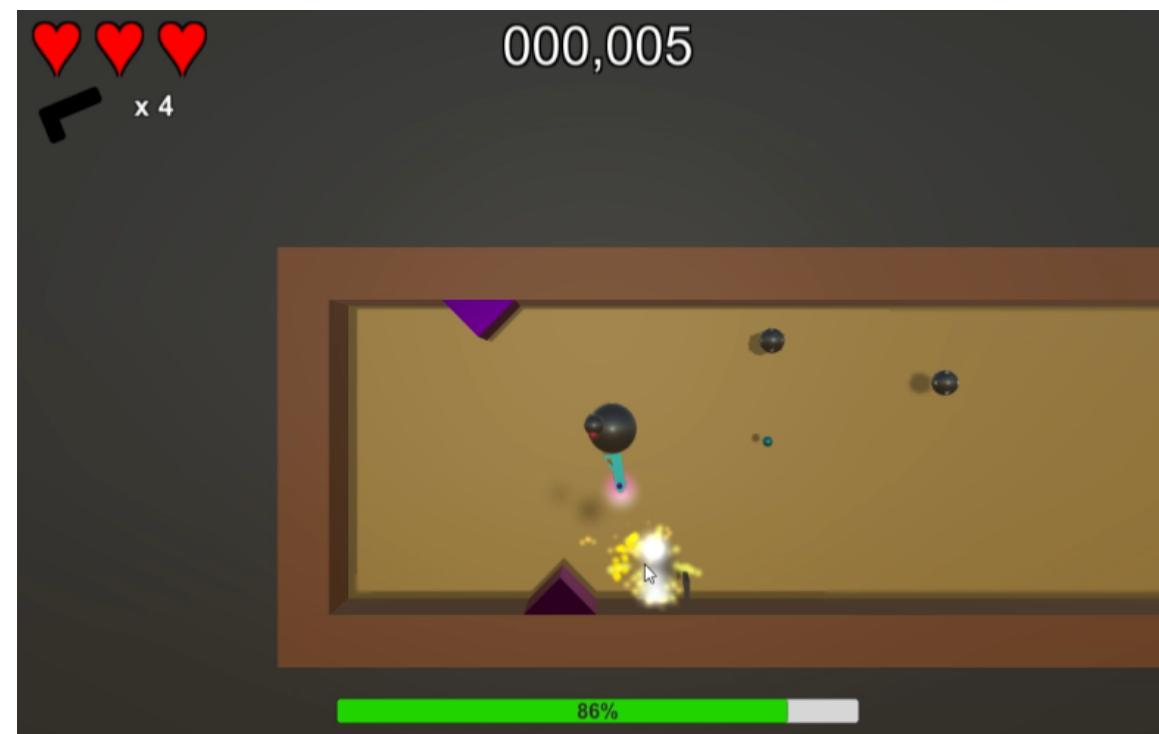
Referenced by cardiovascular dataset on Kaggle.

Derived visualization data supporting the high risk factors of cardiovascular disease.

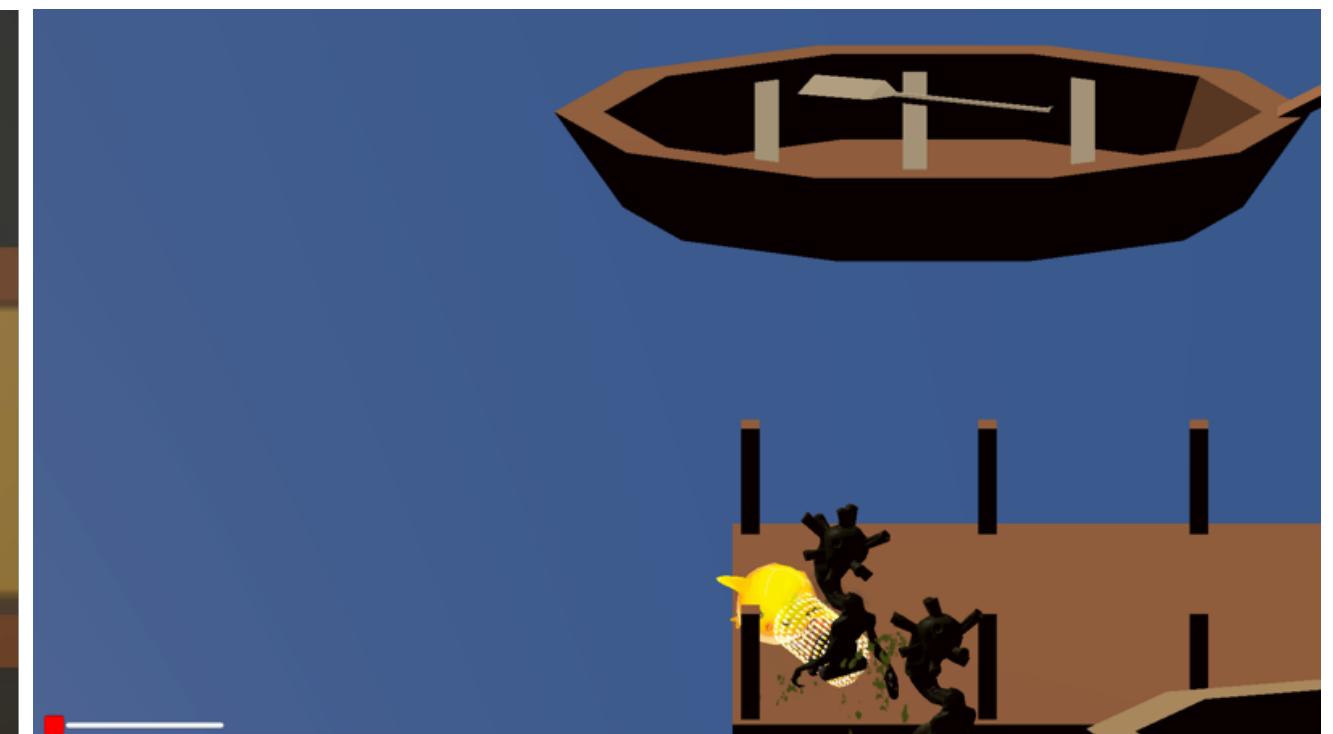
Techniques Used: data preprocessing in R language, machine learning techniques, and statistical analysis

# Others :

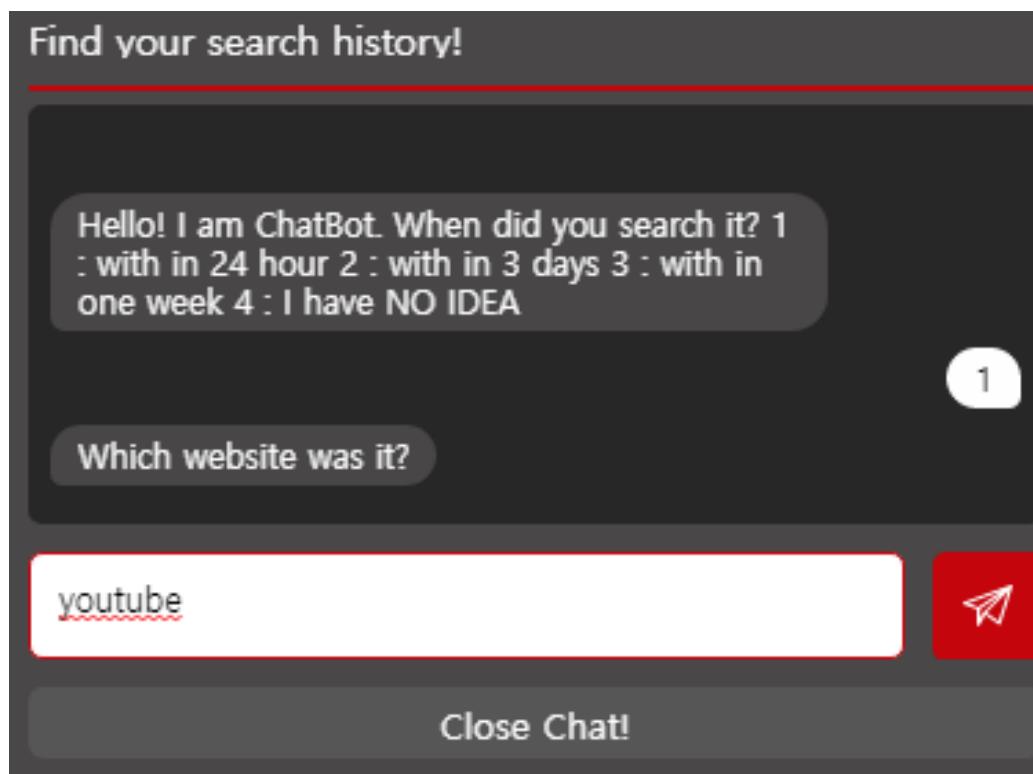
**C# Unity Shooting Game**



**C# Unity Roguelike Game**

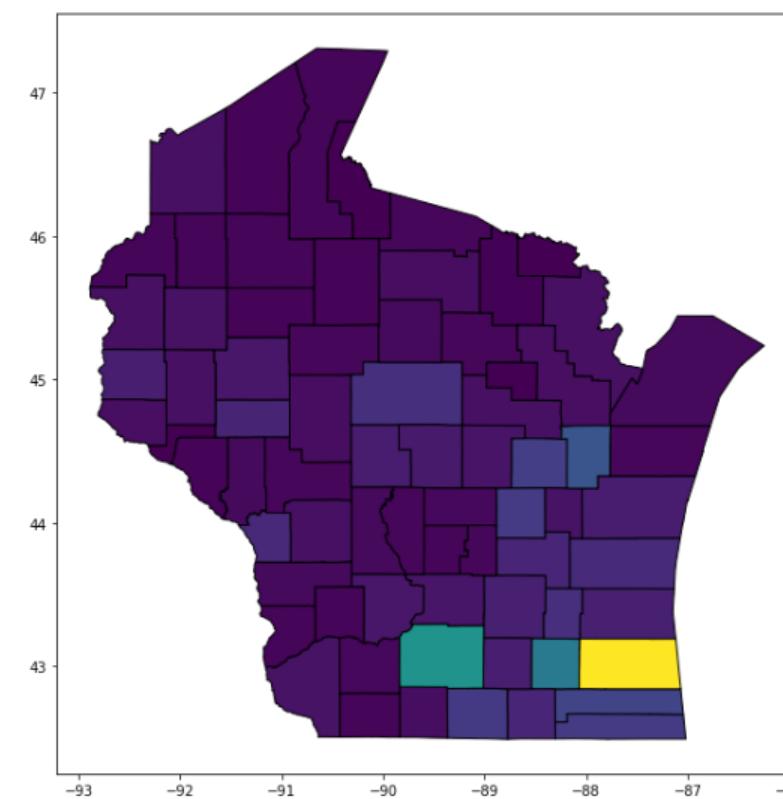


**Chrome Extension App**

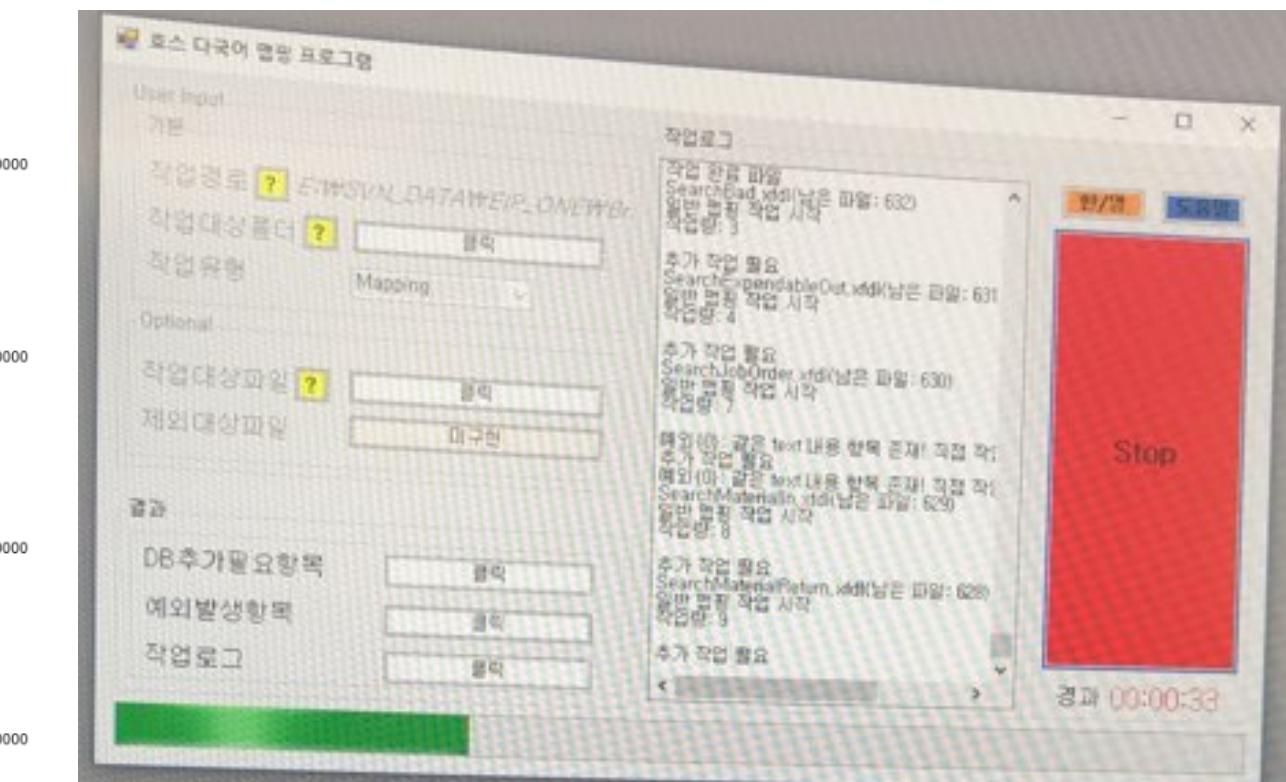


A chrome extension to search forgotten webpages

**Data Modeling**

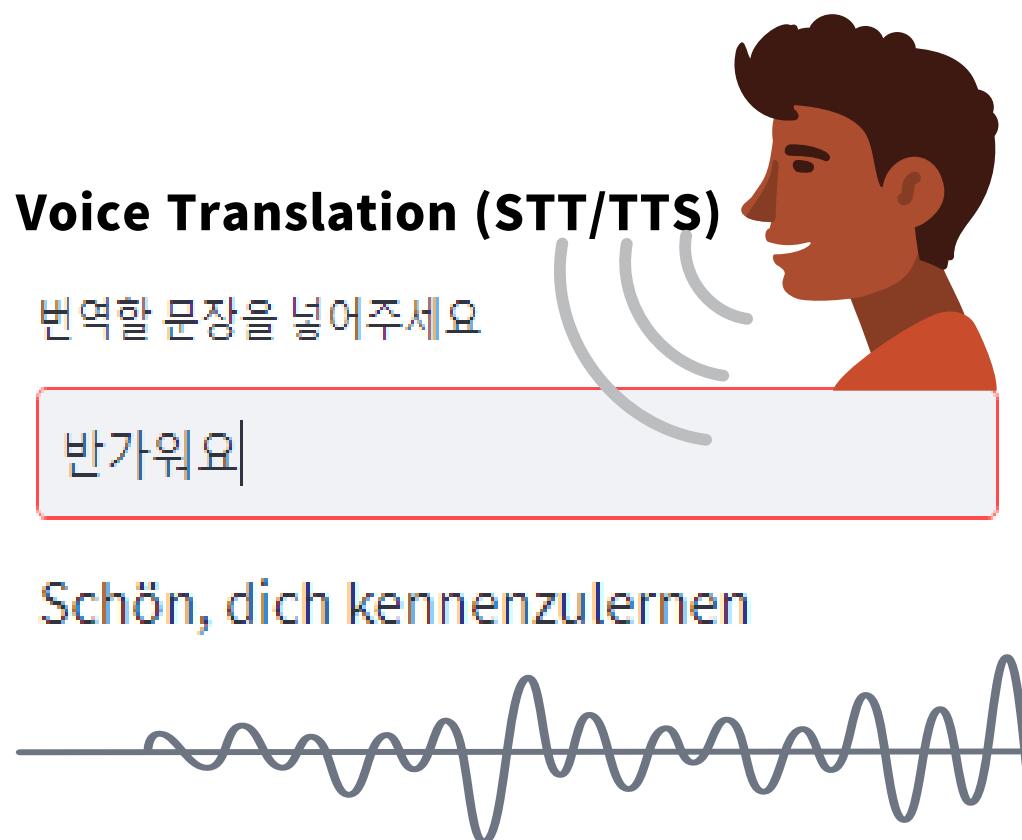


**Window C# Forms: Automation Program**



Solved manual work by developing the automation program.

# NLP Tasks :



## NER

ABC will graduate from university in 2022.

ABC - People

2022 - Time

College - Organization

## NER

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))

POS Tagging

## GPT-2 Text Generation

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

[Input]

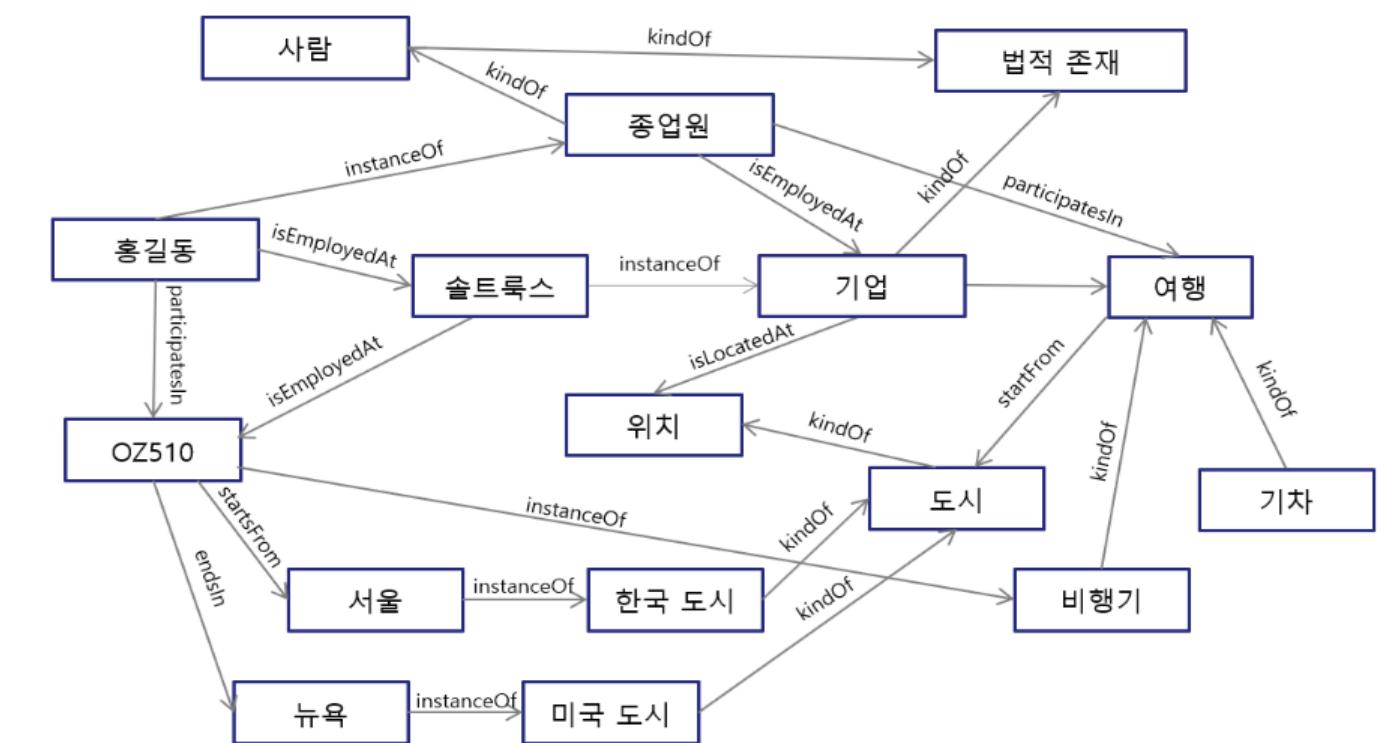
Output:

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.

Using the GPT-2 model trained on the wiki sentence dataset, the next sentence is generated so that it naturally connects with the given sentence.

## KoBERT Keyword Extraction



Used the dataset *gold-standard-v1* from KAIST, Accuracy 90% ↑

# NLP Tasks :

## KoBERT: Sentiment Analysis on Movie Reviews

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 주가 된, 최고의 음악영화 1  
 6315043 진정한 쓰레기 0  
 6097171 마치 미국애니에서 튀어나온듯한 창의력없는  
 로봇디자인부터가, 고개를 젖게한다 0  
 8932678 갈수록 개판되가는 중국영화 유치하고 내용없음 품잡다 끝남  
 말도안되는 무기에 유치한 cg남무 아 그립다 동사서독같은 영화가 이건  
 3류아류작이다 0

[Sentiment Analysis] 0: Negative, 1: Positive, Accuracy 90% ↑

## Article Summarization & Translating Tool

### 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 Koryō), 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.<sup>1078/10000,</sup>

#### 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 등포 사이에서는)는 서구식으로 그들의 나이보다 1~2년 더 많을 것입니다. 1천년 말에 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 역사 를 공유하고 있지만 분할 이후 문화적으로나 경제적으로 크게 갈라졌습니다.<sup>417/10000,</sup>

# ML/DL Tasks:

## Teeko Board Game

```
Hello, this is Samaritan Move (e.g. B3): B3
0:
1:
2:
3:
4:
A B C D E
[]

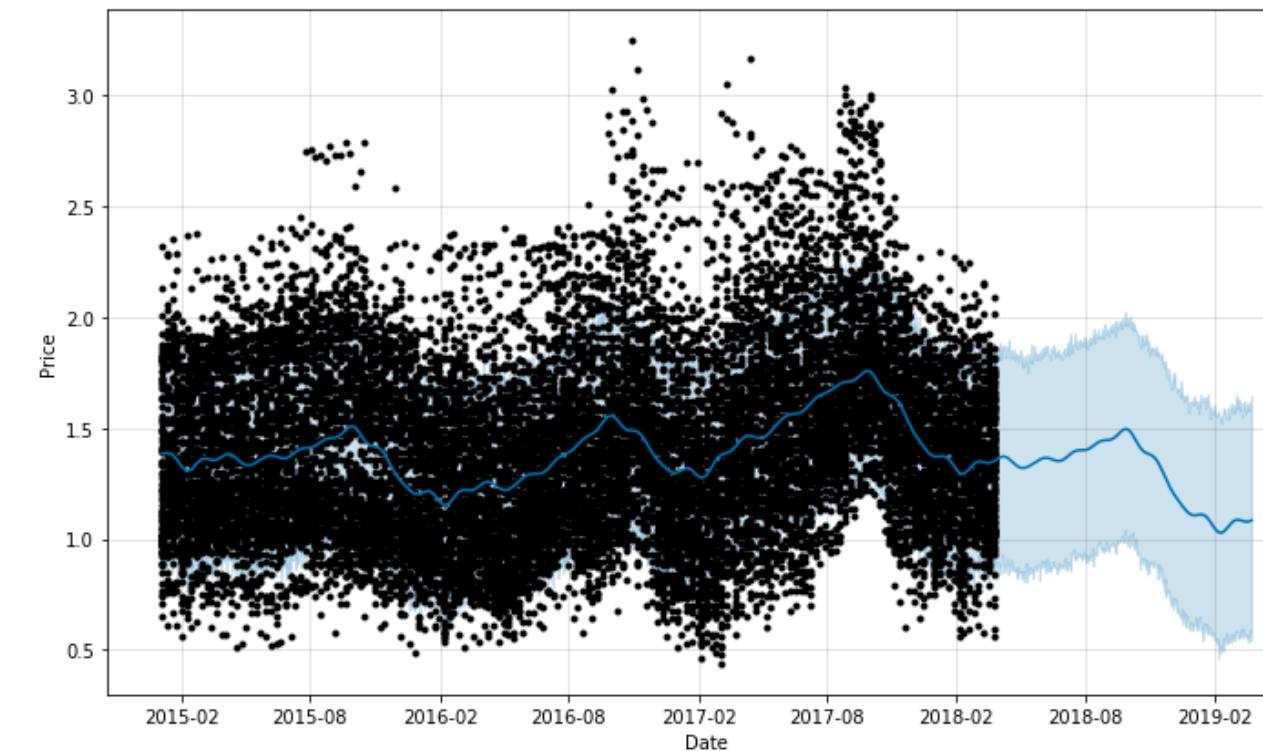
r moved at C2
0: r b b
1: r
2: r
3: b
4:
A B C D E
b's turn
Move (e.g. B3):
```

[Game Theory] Teeko Game using Minimax Algorithm

In the Teeko board game, players and AI robots take turns placing markers. If one side first places 4 markers vertically, horizontally, diagonally, or in a 2x2 box, it wins.

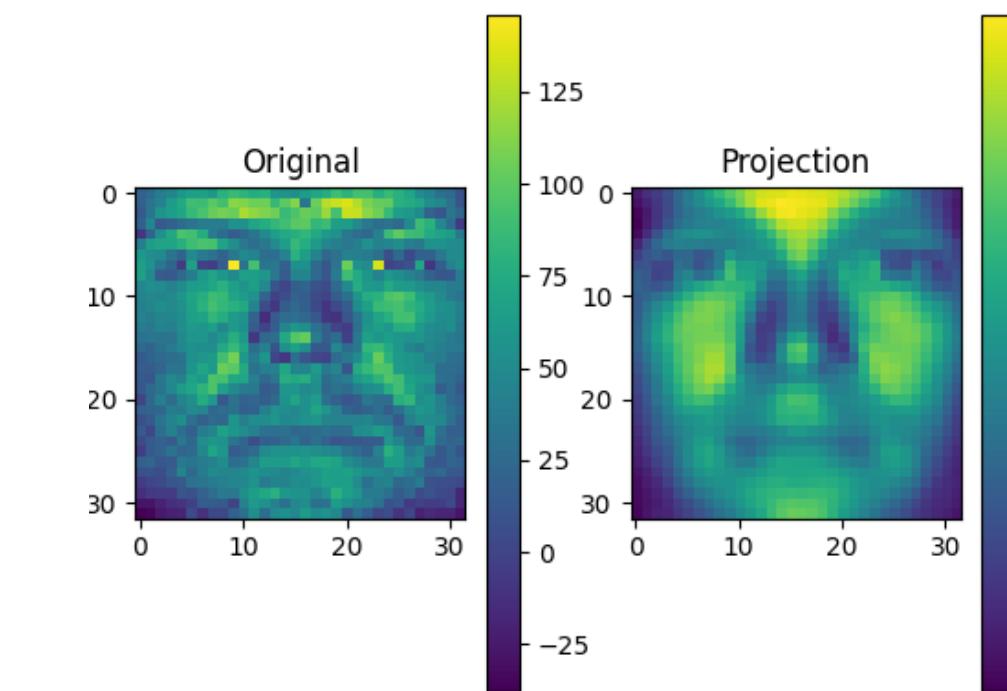
AI robots dodge the player's markers and penetrate diagonally.

## Time Series



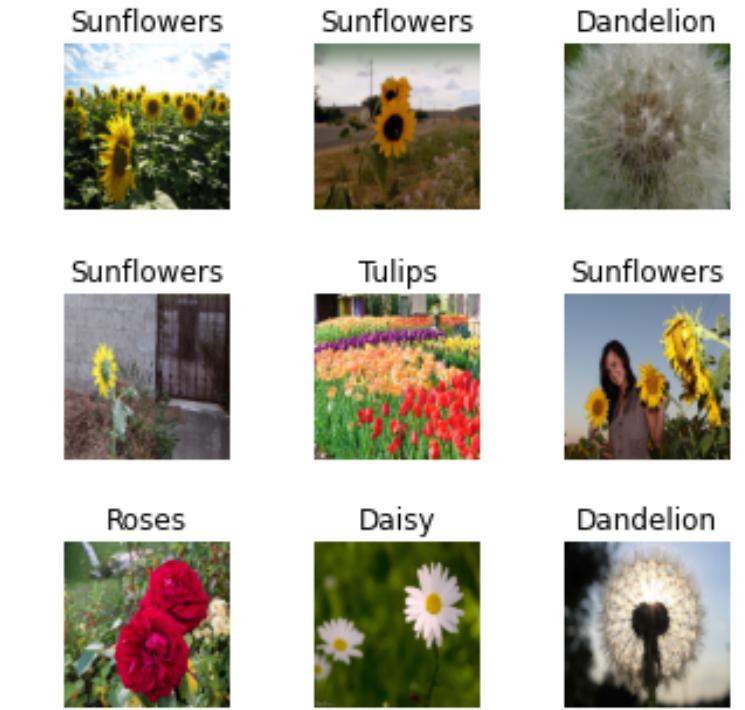
[Time Series] LSTM/Facebook Prophet

## Image Compression



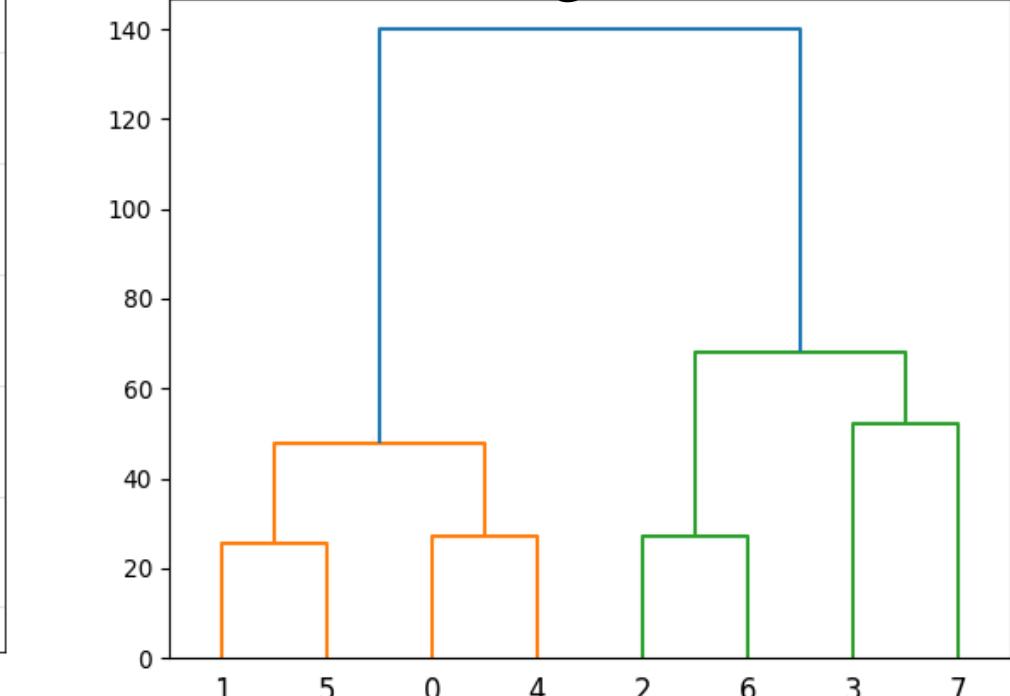
[Image Compression] Controlling # Features on PCA

## Transfer Learning



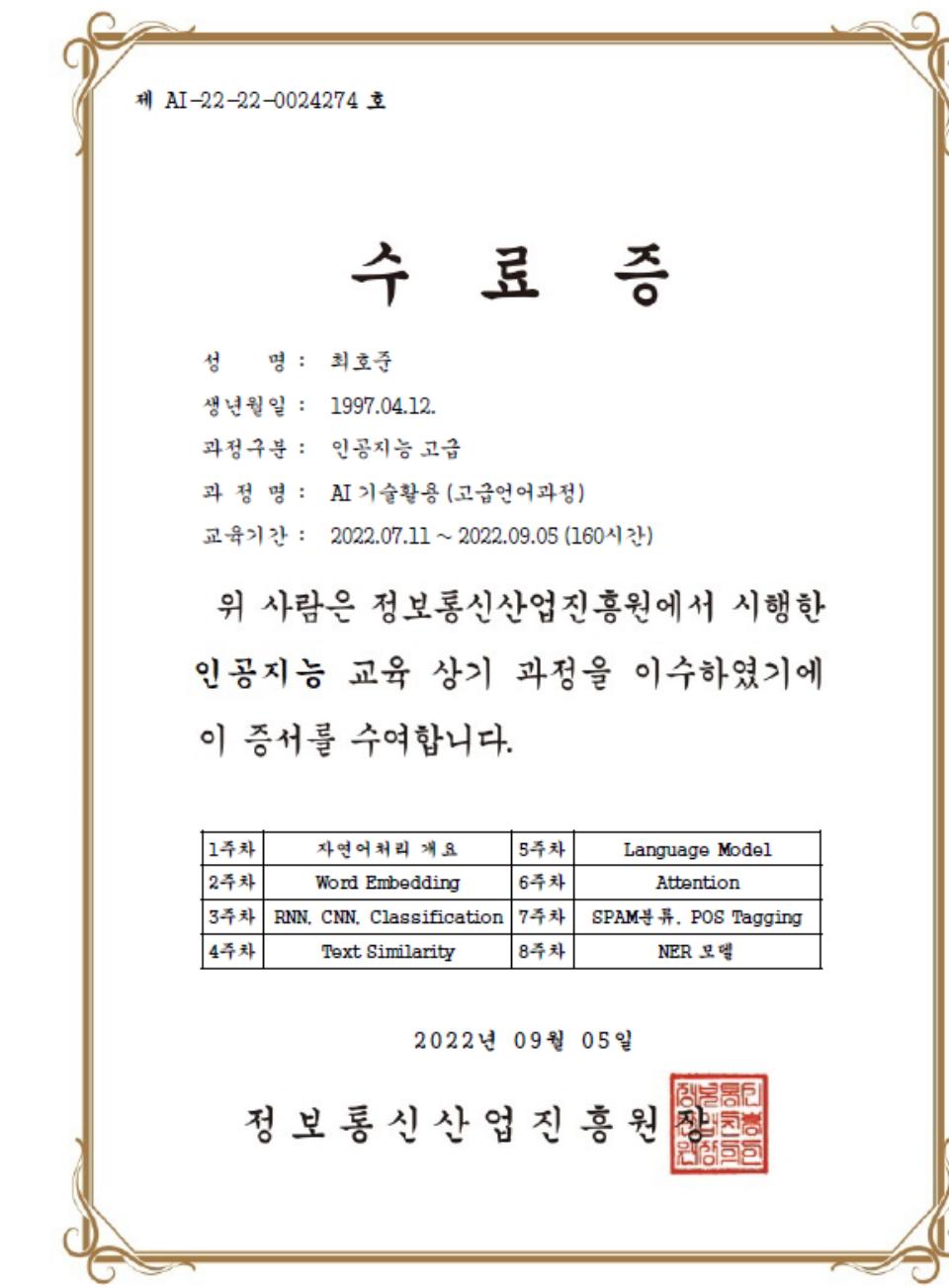
[Transfer Learning] Fine-tuning MobileNet

## Pokemon Clustering



[HAC Complete-linkage] x: Pokemon ID, input: # Clusters

# Certificates :



# Certificates :



Certificate no: UC-58160105-b6a9-4785-ae18-418eb41e1cc0  
 Certificate url: ude.my/UC-58160105-b6a9-4785-ae18-418eb41e1cc0  
 Reference Number: 0004

CERTIFICATE OF COMPLETION

**【한글자막】 머신러닝의 모든 것 with Python, R**

Instructors Kirill Eremenko, Hadelin de Ponteves, 웹진씽크빅 글로벌, Ligency I Team

**Hojun Choi**

Date July 21, 2022  
 Length 44.5 total hours



Certificate no: UC-04498d66-6d70-4340-8184-c07afcd9197a  
 Certificate url: ude.my/UC-04498d66-6d70-4340-8184-c07afcd9197a  
 Reference Number: 0004

CERTIFICATE OF COMPLETION

**【한글자막】 Machine Learning 실전 개발 | 8개의 실용 프로젝트**

Instructors Dr. Ryan Ahmed, Ph.D., MBA, Ligency I Team, Mitchell Bouchard, Ligency Team, 웹진씽크빅 글로벌

**Hojun Choi**

Date July 26, 2022  
 Length 14.5 total hours