

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



I'm Choi Hojun, and I obtained my B.S. in computer science and data science from the University of Wisconsin-Madison. I took various AI-related courses during my academic career, where I developed my subject knowledge, statistical data analysis skills, and project capabilities. I also worked as an undergraduate researcher in the Sensing group of the self-driving research team. Beyond the school, I successfully completed several AI boot camps, such as the LG Research AI Program and NLP development. Additionally, I excelled in a slew of data mining hackathons.

[My Research Interests]

I am broadly interested in big data, model compression, hyper-scale AI, graph/data mining, and machine/deep learning. Building a reliable and expansive AI/ML system for sustainable development is the main objective of my study.

PROGRAMMING LANGUAGES

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

EDUCATION/EXPERIENCE

| | | |
|--------------|-----------------|----------------|
| SNU | PhD (AI) | 2023 - Present |
| UW-Madison | BS (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 | eric970412@gmail.com |
| Website | https://hchoi256.github.io/ |

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

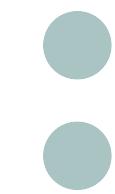
13 -14

English

2016 - current

Hojun (Eric) Choi

B.S. at UW-Madison (expected)
Highest level in ESL at SUNY Albany
Academic Mentor (1y +)
Online Translator



INTENSIVE ENGLISH LANGUAGE PROGRAM

UNIVERSITY AT ALBANY State University of New York

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

W. BESS VIVERET

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



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

지식IN 활동증명서

쪽스 (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 |
| 단별수 | 2013개 (단별 차태루 71.1%) |

- #### • 주요경력 교육기부

지식 iN eXpert

I am it may concern,
ter confirms our job offer to Hojun (Eric) Choi for the undergraduate student position of
nic Mentor at the Center for Academic Excellence (CAE). Please see the contact informa
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 goals.

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

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

Digitized by srujanika@gmail.com

 reach out with any question

10.000-15.000 € (excl. VAT) per month.

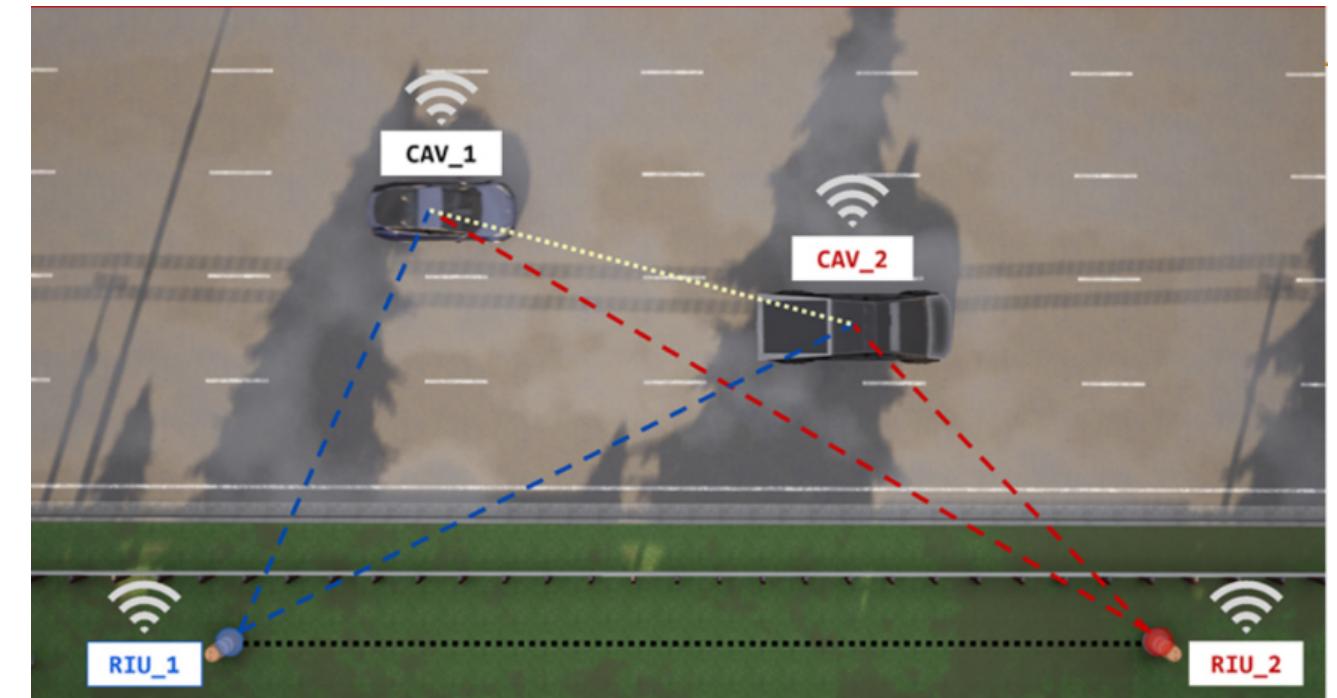
Research Experience

2022 ~ current

Hojun (Eric) Choi

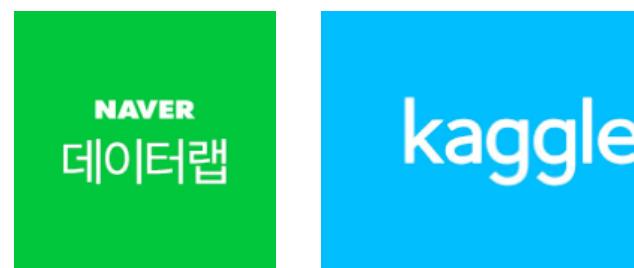
[CARLA Research Project](#)

An autonomous driving simulator

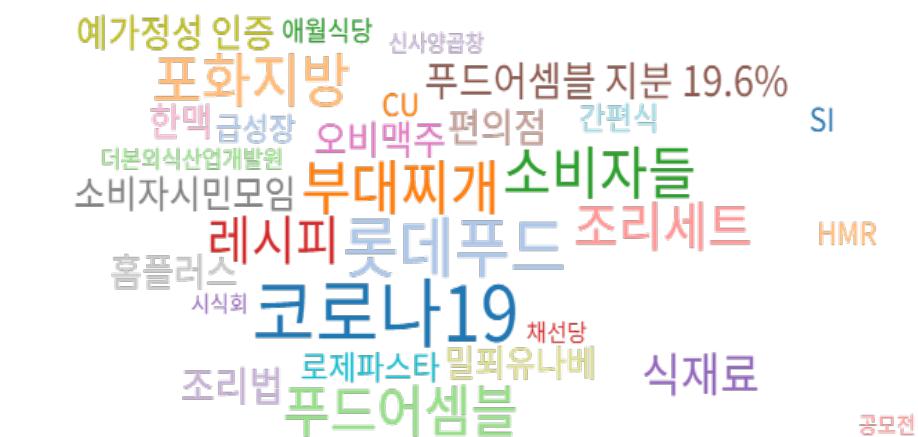


ETL & EDA :

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

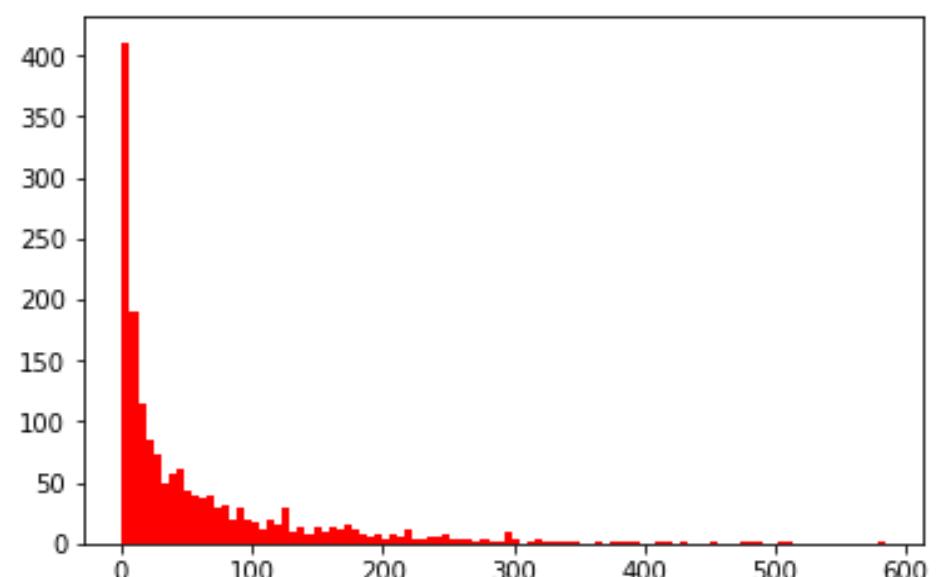
ETL & EDA

| | 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

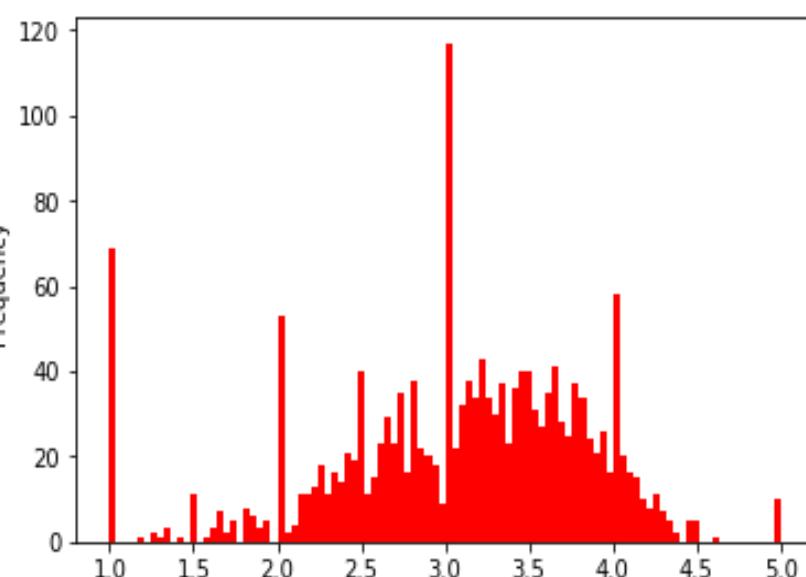
mean is the movie rating.



[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

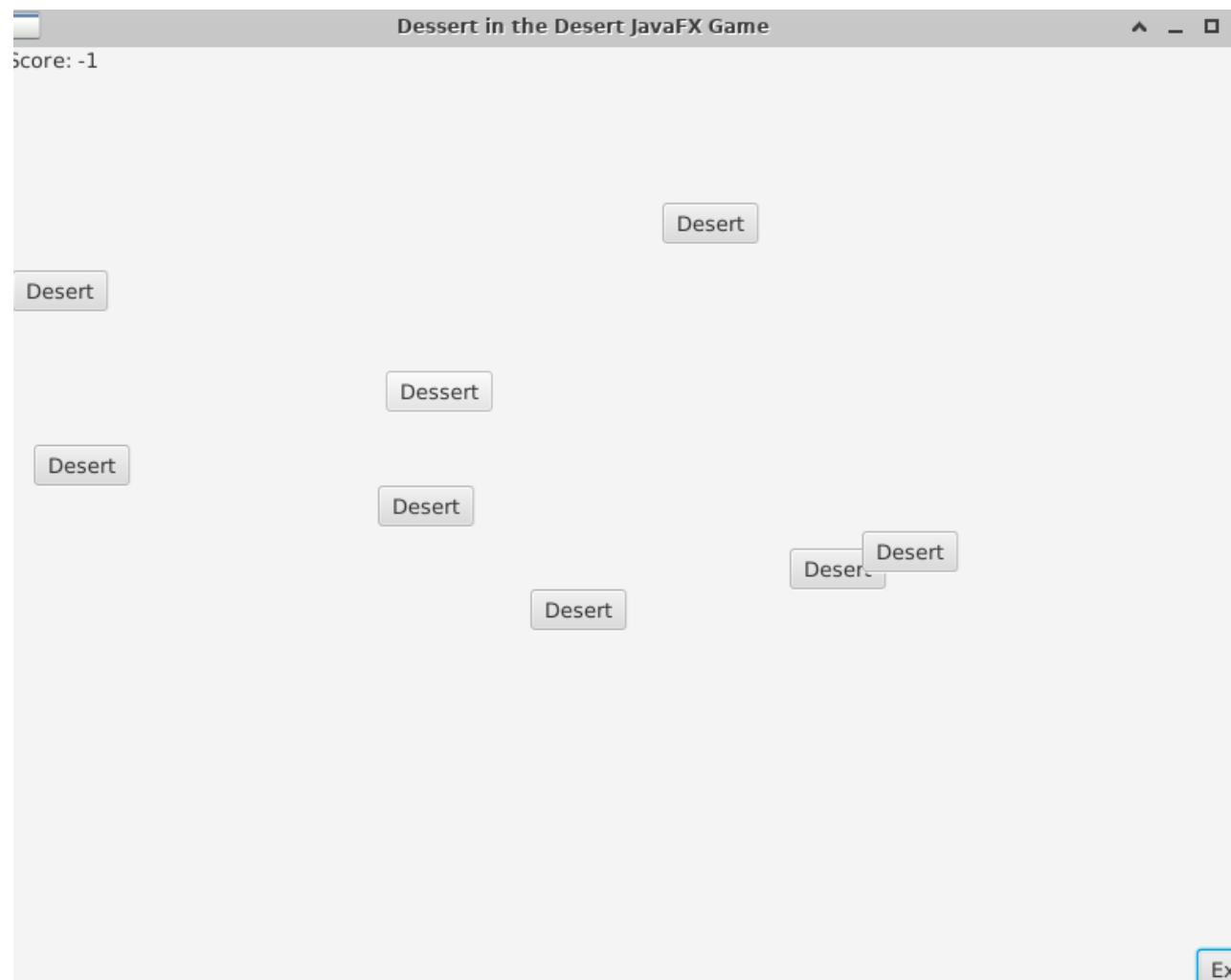
Others :

[HCI Report Papers](#)

[Operating Systems](#)

[Data Structure](#)

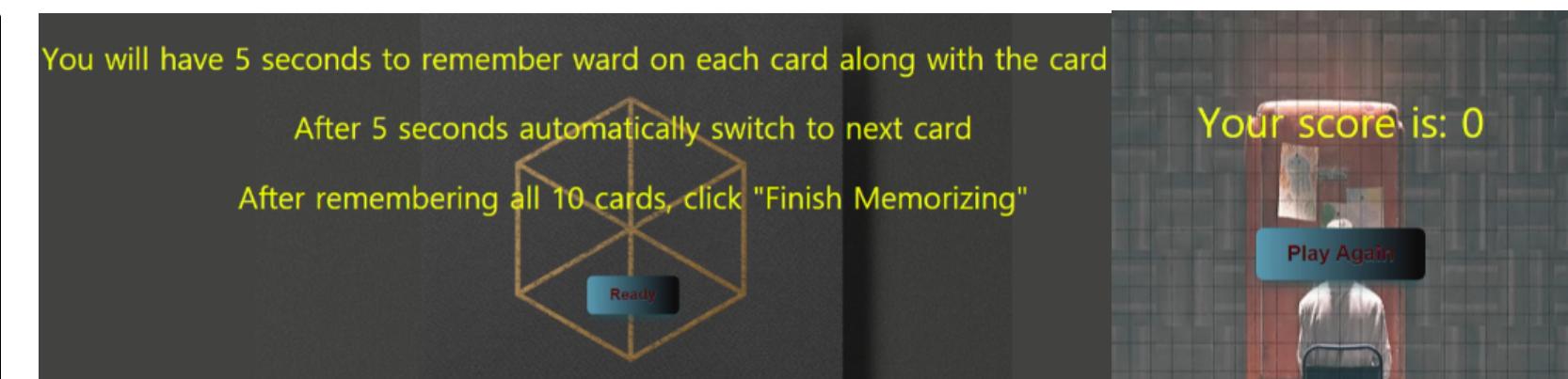
JavaFX Minigame



C++ COVID Self-Checking

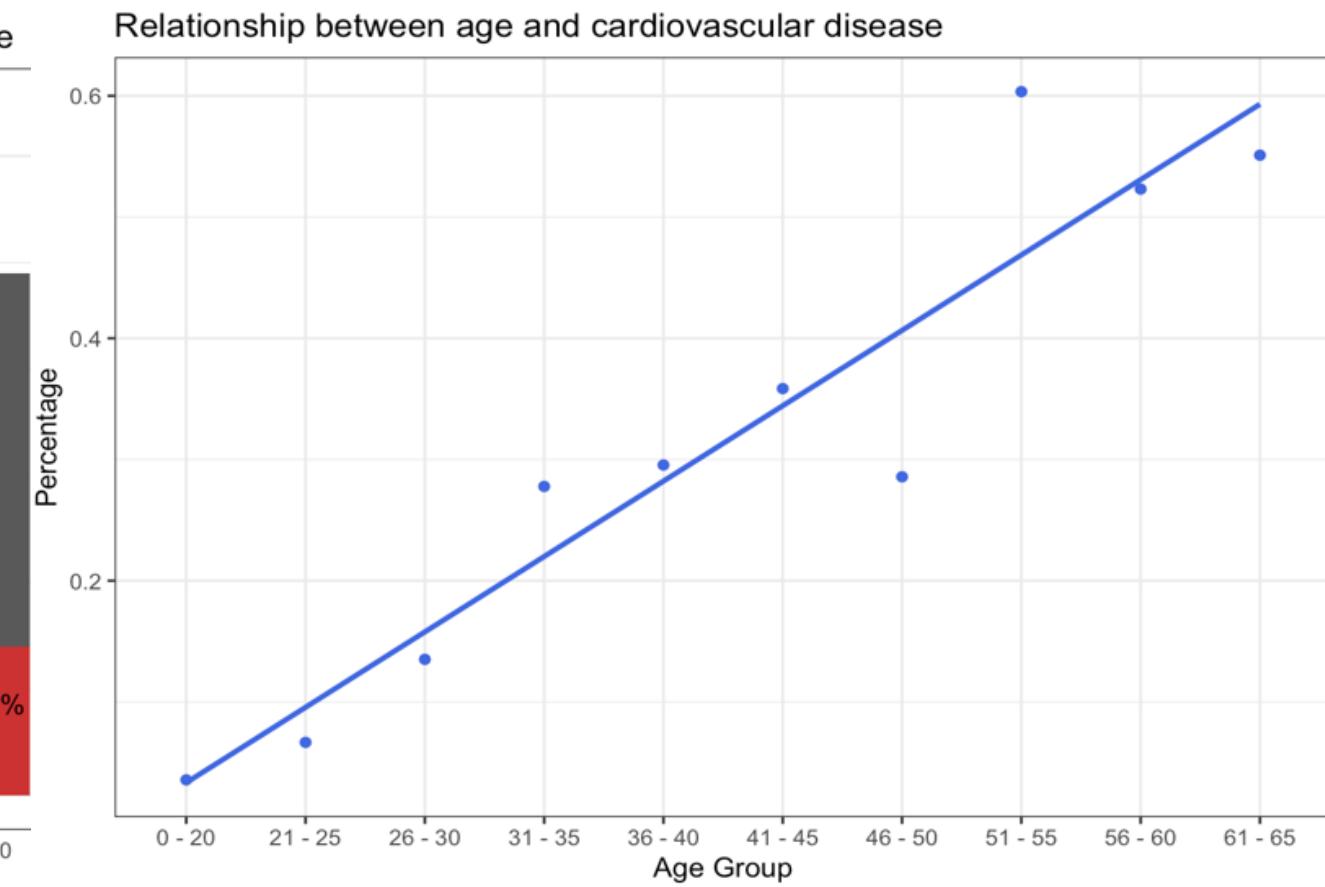
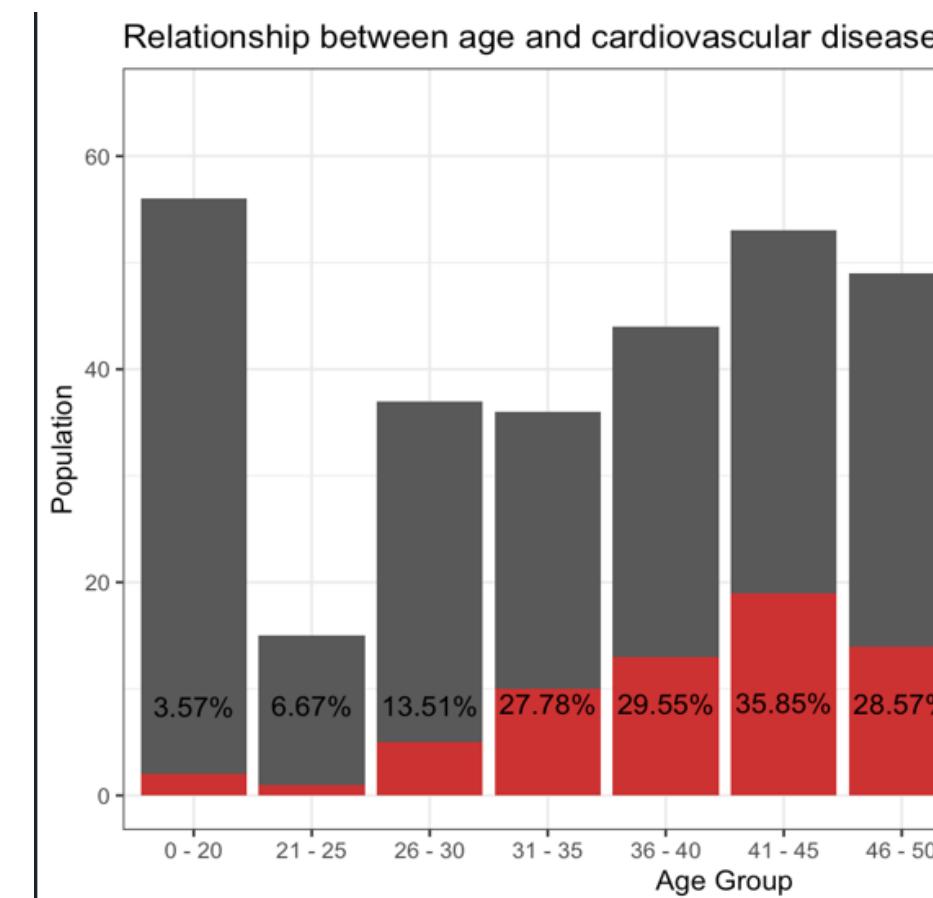
```
*****
Welcome (back) to Safe-ish Badgers!
[1] View testing locations
[2] Make test reservation
[3] Use symptom checker
[4] Check for Covid - 19 Exposures
[5] View my reservations
[6] View my personal data
[7] Adjust my personal data
[8] Change settings
[9] Exit program
Please enter a menu item:
*****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

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.

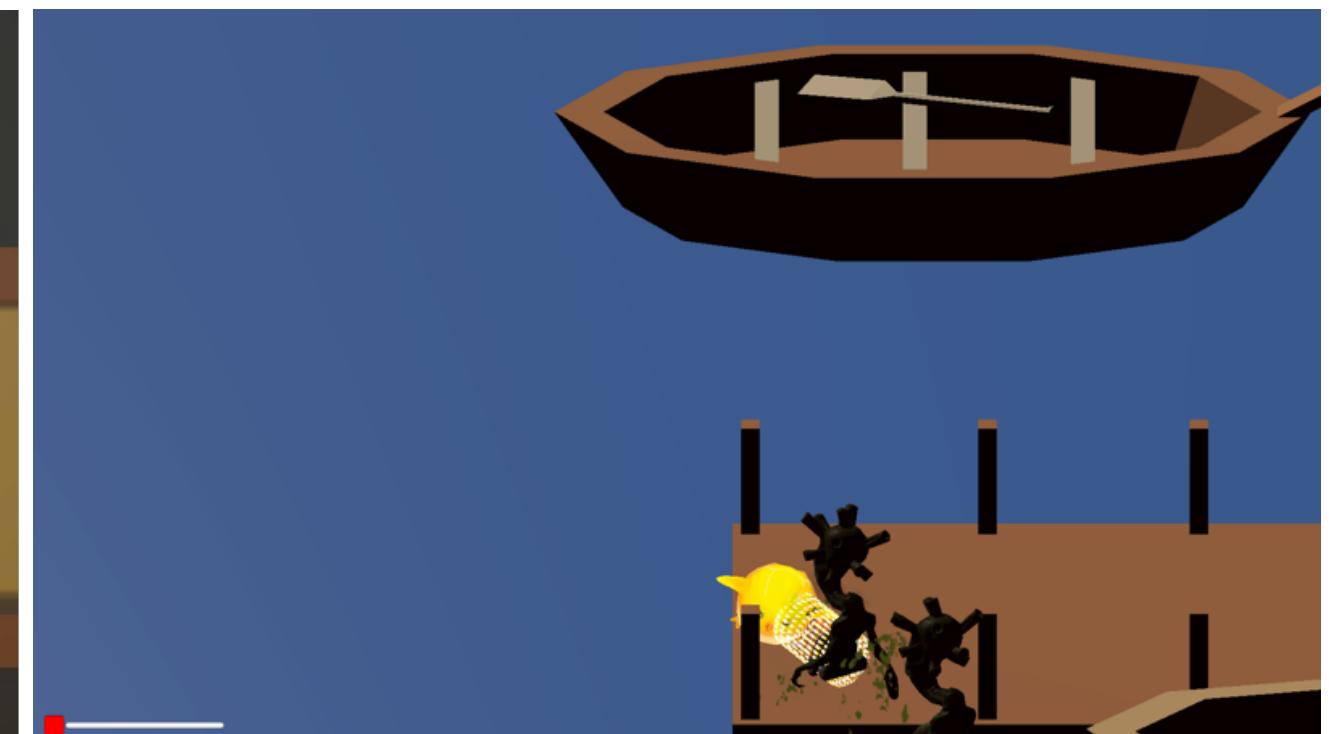
Techniques Used: data preprocessing in R language, machine learning techniques, and statistical analysis

Others :

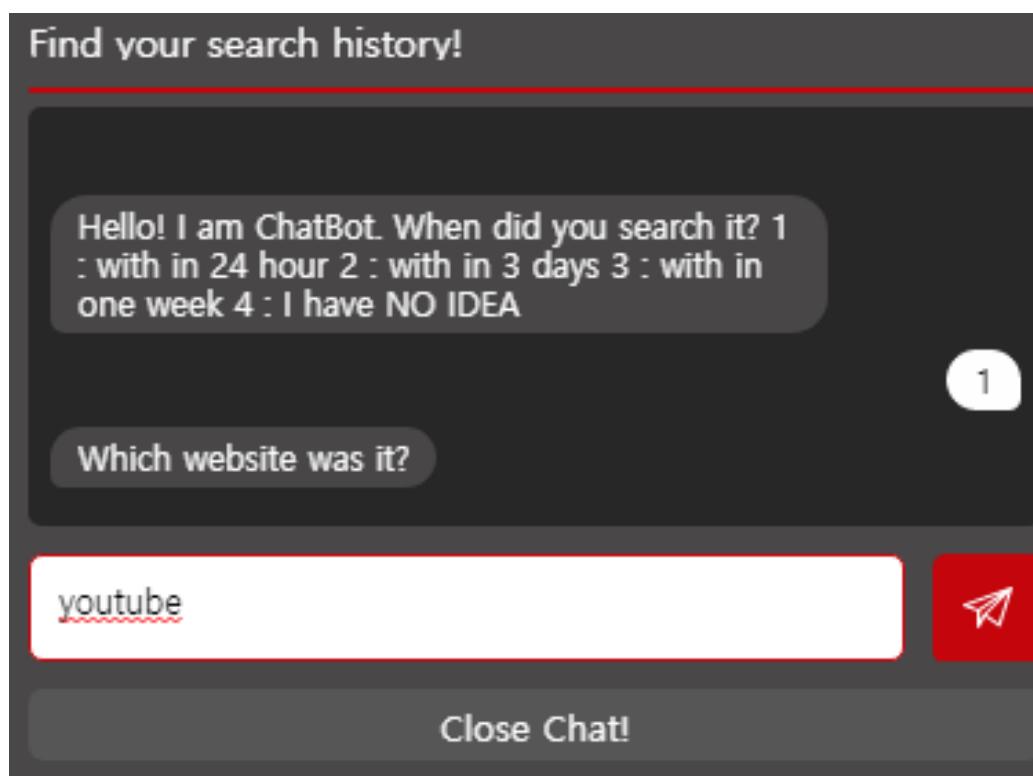
C# Unity Shooting Game



C# Unity Roguelike Game

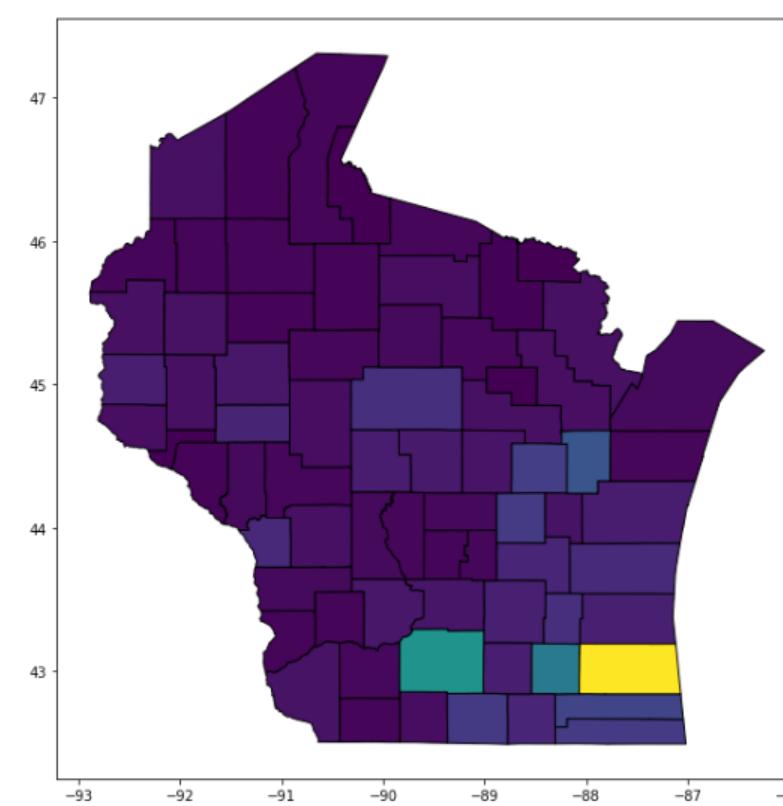


Chrome Extension App

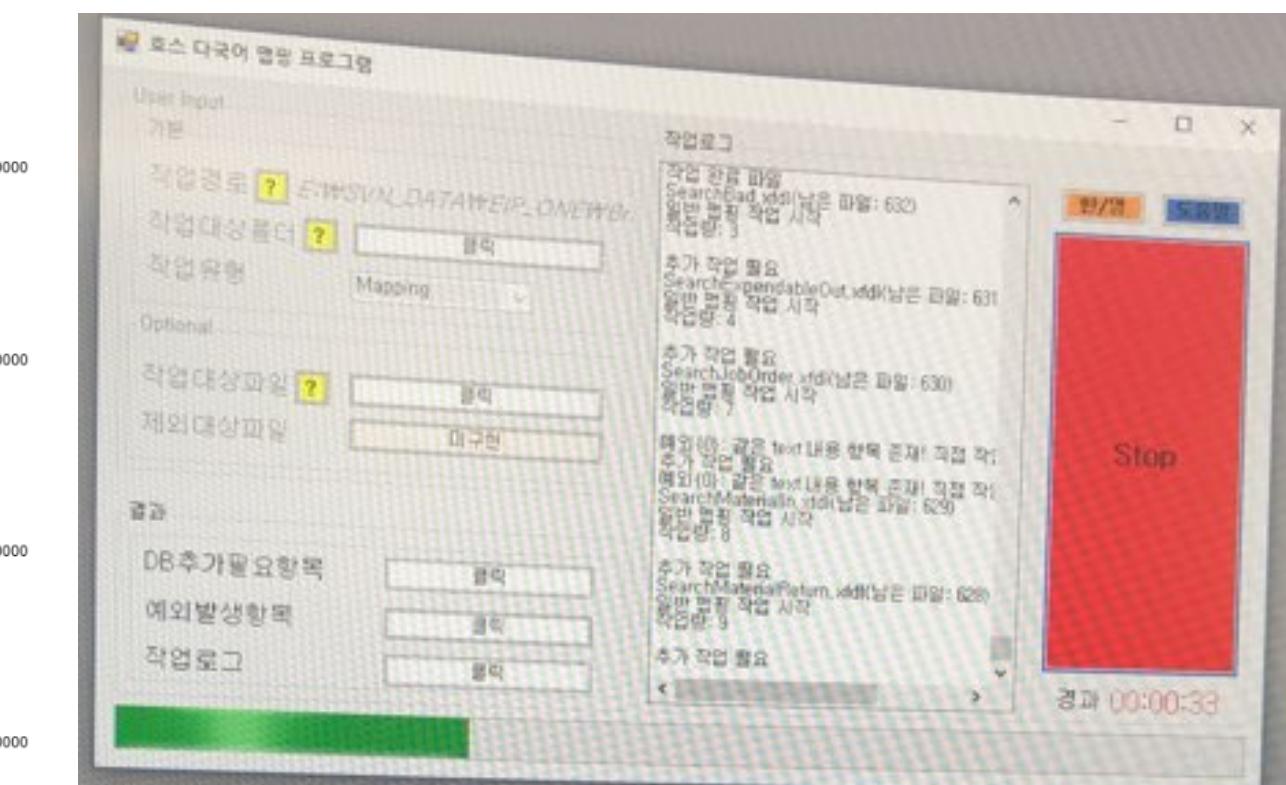


A chrome extension to search forgotten webpages

Data Modeling



Window C# Forms: Automation Program



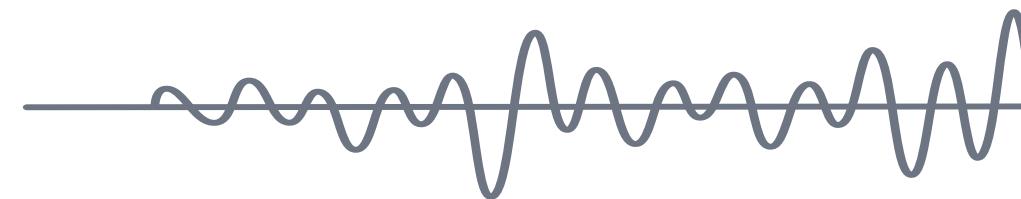
Solved manual work by developing the automation program.

NLP Tasks

Voice Translation (STT/TTS)

반가워요

Schön, dich kennenzulernen



NER

ABC will graduate from university in 2021

ABC - People

2022 - Tim

College - Organization

NE

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]

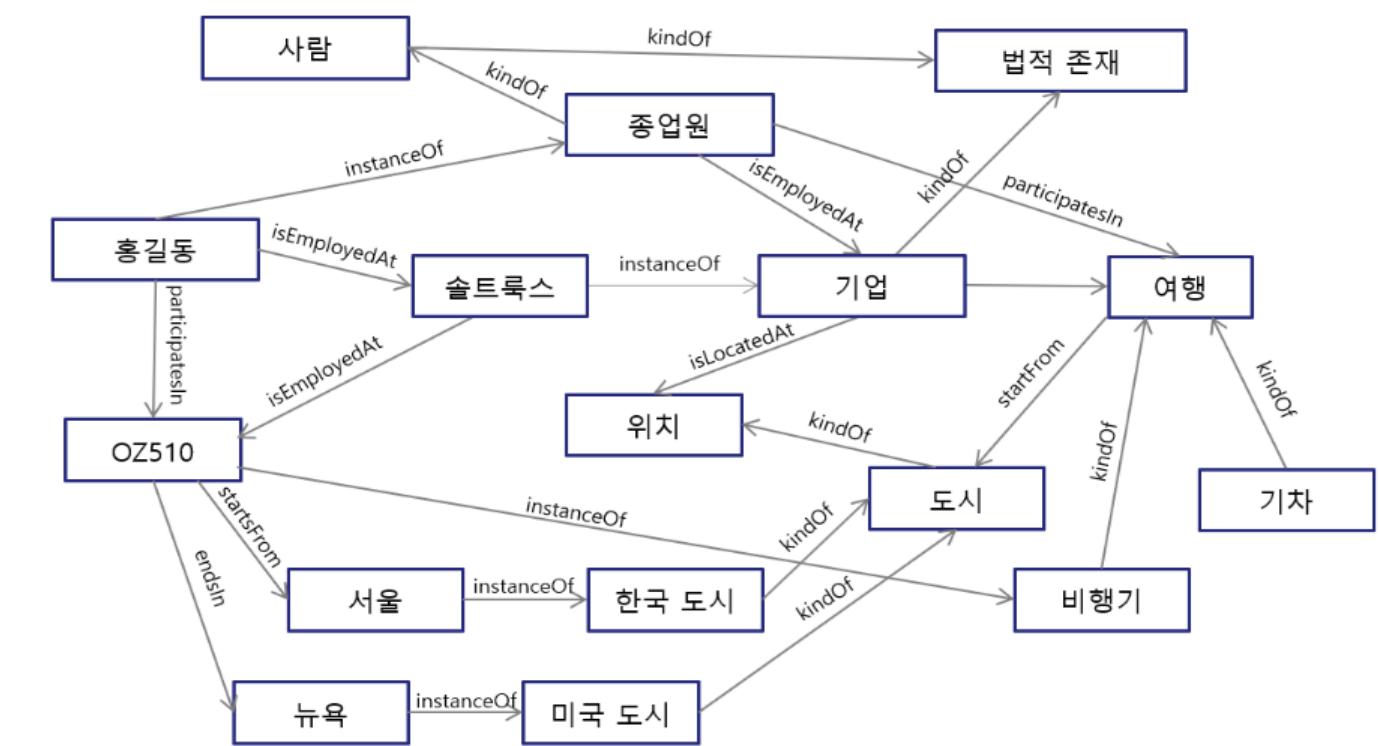
Output

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.^{1078/10000,}

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.^{417/10000,}

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
[]

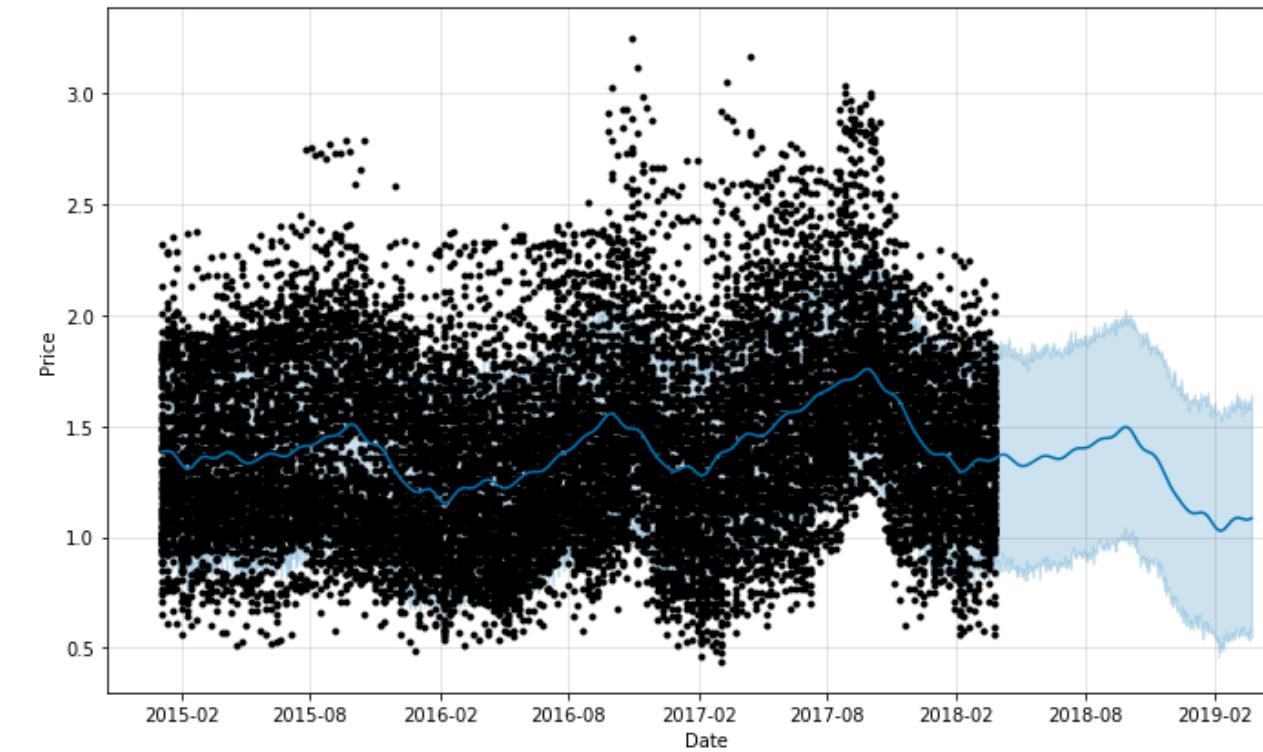
r moved at C2
0: r b b
1: r
2: r
3: b
4:
A B C D E
b's turn
Move (e.g. B3):
```

[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.

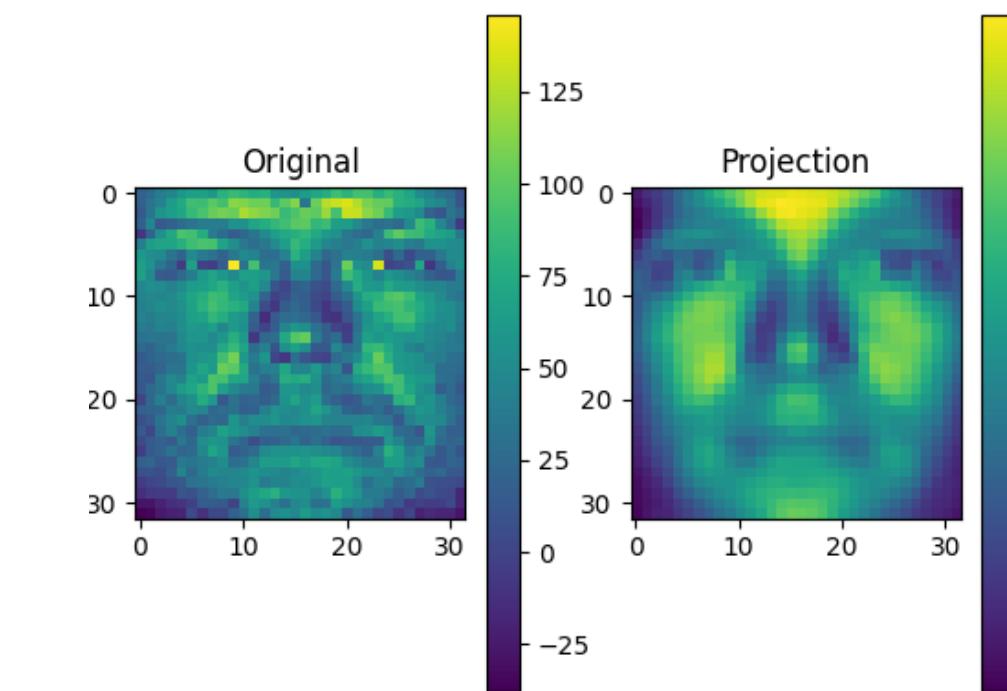
AI robots dodge the player's markers and penetrate diagonally.

Time Series



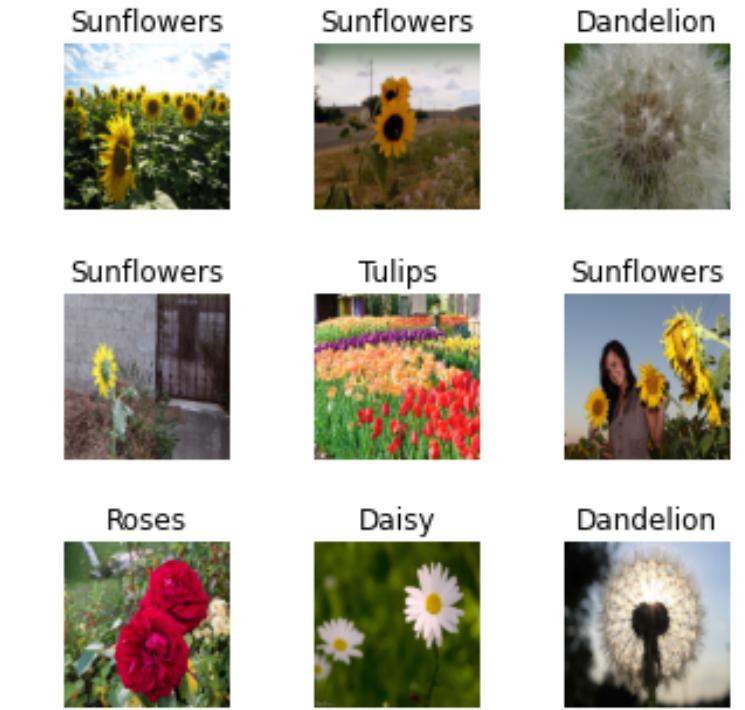
[Time Series] LSTM/Facebook Prophet

Image Compression



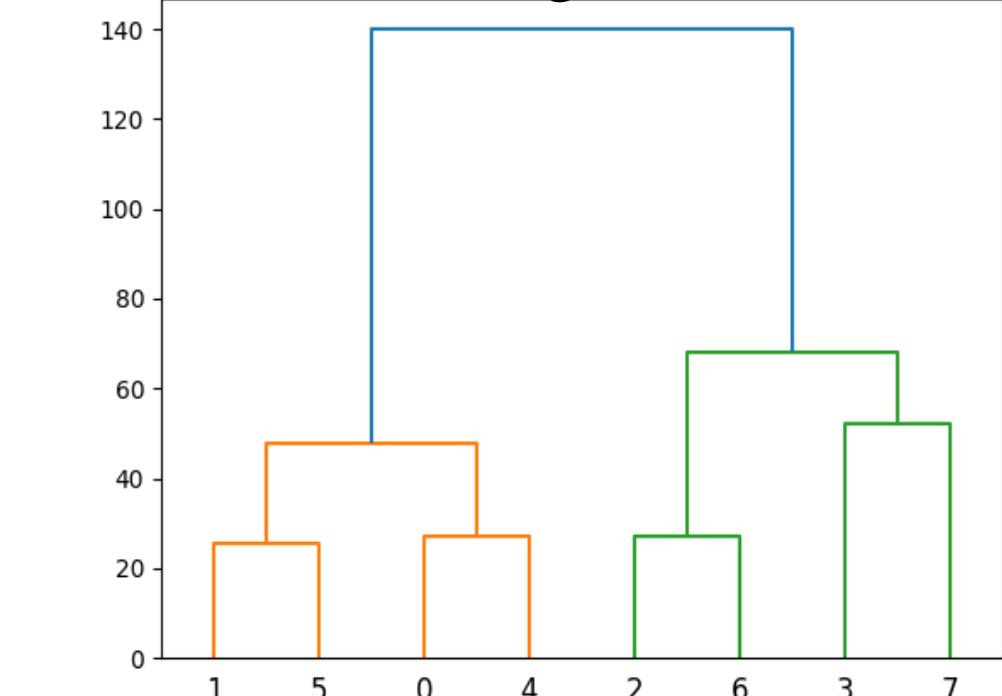
[Image Compression] Controlling # Features on PCA

Transfer Learning



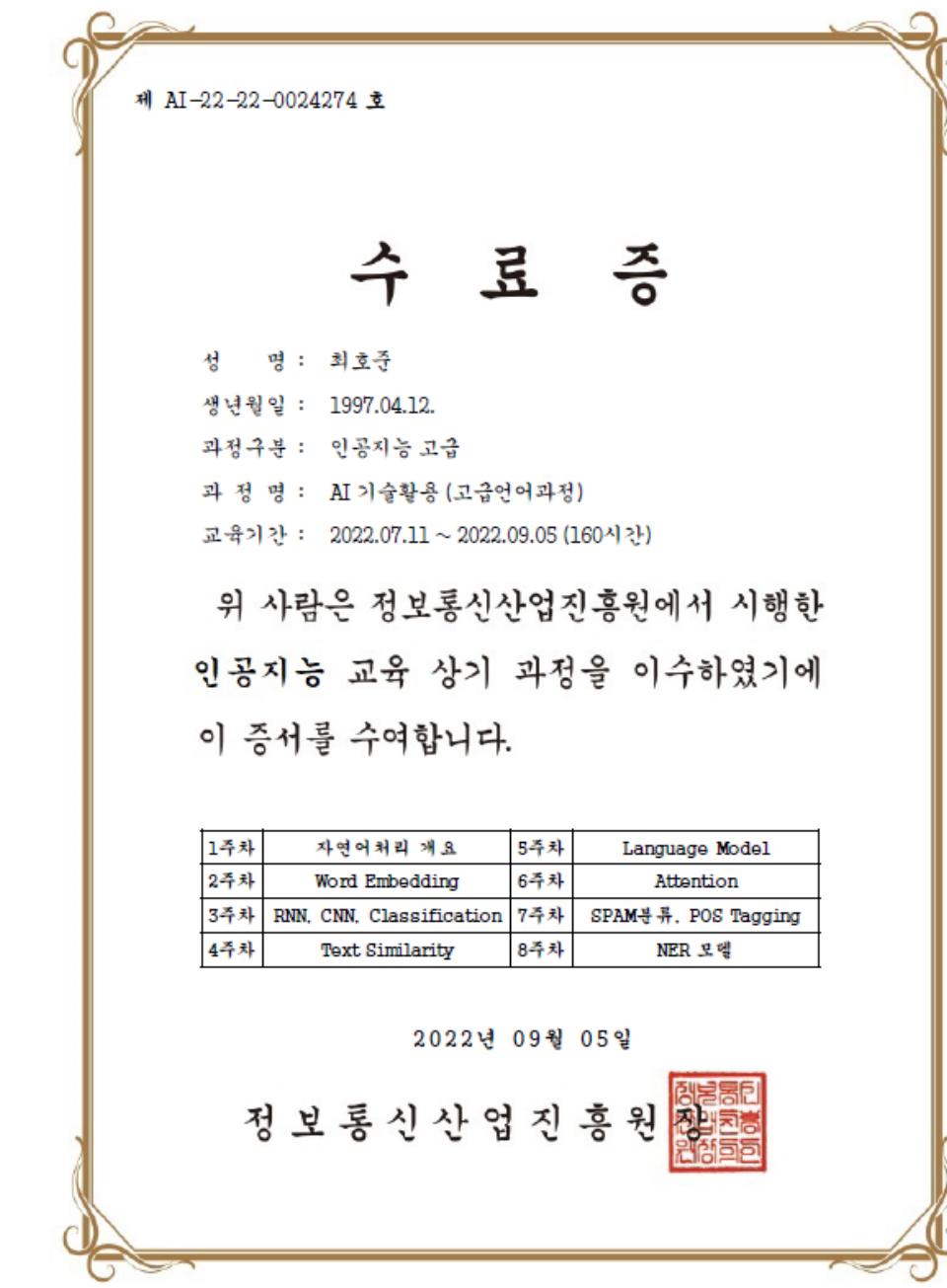
[Transfer Learning] Fine-tuning MobileNet

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