Hong Jun Park

821043370365 • howay96@gmail.com

EDUCATION

University of Rochester **Rochester, NY**
Bachelor of Science, Computer Science *Anticipated Dec 2021*
- Minor, Biomedical Engineering

Georgetown Preparatory School **North Bethesda, MD**
High School Diploma *May 2015*

EXPERIENCE

Seoul Central District Prosecutor's Office **Seoul, Republic of Korea**
Service of Social Work Personnel *Aug 2018 – May 2020*
- *Managed distribution of office supplies and kept track of inventory using Microsoft Excel*
- *Distributed and disposed furnitures and printers upon replacement request*
- *Manually censored personal information of outgoing incident reports*
- *Organized documents regarding annual expenditure of the company and bound the collection into a book for future reference.*

ACADEMIC PROJECTS

Korean Hate Speech Detection **Mar 2022**
- *Using a Korean Hate Speech Dataset, fine-tuned a pre-trained KcELECTRA-base model to classify a short comment to one of three labels (Hate/Offensive/None). For multi-class classification, the model include two linear classification layer which uses the CLS token output from Electra's discriminator. Fine-Tuned model showed an accuracy of seventy four percent for its validation set. Having the model purposefully fine-tuned upon an unchanging hyper parameters, the project allowed to observe the affects of altering preprocessing data, model's parameters and different versions of the KcELECTRA model.*

MatchSum **Apr 2022**
- *Focused on preprocessing and implementing MatchSum extractive text summarization model without fine-tuning process. As BertSum trained in Korean dataset is not available, modified version of Greedy-Selection algorithm from BertSum was used to score and prune sentences for each articles. MatchSum, as described in the paper, matches CLS token vector of the original text with candidate summary created from pruned article along with gold summaries written by human. Although no functioning summarization model was produced, the project helped to educate myself with both MatchSum and BertSum along with oracle summary generating algorithms and deeper understanding on the Rouge scores.*

Korean Semantic Textual Similarity API **Apr 2022**
- *Created a model scoring semantic textual similarity between two sentences, using Klue/Roberta-large pre-trained model along with the klue-sts dataset. Roberta model was used in a weight sharing Siamese structure, producing mean pooling value for each of the two outputs, which would be used to calculate the cosine similarity between the input sentences within the range of 0 to 5. The best model found through random searching of hyper parameters showed pearsonr correlation of 0.882, along with F1 score of 0.835, given a threshold of three for the initial prediction. The model was implemented as a Rest API, predicting the STS score for 1:1 , 1:N sentence inputs.*

SKILLS

- Most comfortable programming in Python, PyTorch, Swift
- Basic programming in Java, MySQL
- Ability to conduct scientific literature searches associated with computer science, chemistry and biomedical topics