

Shaun (Siyeon) Kim

(434) 227-3351 | sk4973@columbia.edu | [Personal Website](#)

Education

Columbia University, New York City, NY

Expected December 2022

M.A. in Statistics - Data Science Concentration

University of Virginia, Charlottesville, VA

May 2021

B.A. in Statistics and Computer Science, graduated with Distinction

Relevant Courseworks: Cloud Computing, Artificial Intelligence, Machine Learning, Database, Time Series, Data Structure, Inference, Algorithms

Skills

Technical Skills: Machine Learning, Data Analysis, Data Wrangling, Data Visualization, A/B testing,

Tools & Languages: Python (Proficient) - TensorFlow, PyTorch, R (Proficient), Java (Proficient), SQL (Proficient), Spark (Intermediate), Tableau (Intermediate), C++ (Intermediate), Microsoft Suite

Work Experience

Capital One

June 2022 – Aug 2022

Incoming Data Science Intern

- Selected as one of the data science interns for the summer of 2022

Gentine Lab | Columbia University

June 2020 - Present

Graduate Student Researcher

- Implemented carbon dioxide prediction models of the globe by approaching the problem as deep-learning vision tasks such as image segmentation and next-frame-prediction to incorporate spatial and temporal information, advised by Dr. Gentine
- Designed and examined benchmark models with traditional machine learning algorithms such as XGboost, random forest, and feed-forward neural network
- Built data pipelines for training on Columbia's cluster and presented data visualizations for 5-dimensional geographical data

Clova AI | Naver Corporation

June 2020 – Aug 2020

Machine Learning Engineer Intern

- Developed and fine-tuned a toxic reply filtering model for NAVER news forum by utilizing BERT transformer, which resulted in an F1 score of 94.87 and had an experience of creating a baseline model using LSTM and RNN
- Created an algorithm that will select topics from a news article detecting search trend spike, built a demo using JavaScript to demonstrate the algorithm, and familiarized other Natural Language Processing tasks in doing so
- Utilized Spark to fetch and build a data pipeline from Hadoop and preprocessed the data set

UVA CS Department

Jan 2020 – May 2021

Machine Learning Teaching Assistant

- Worked closely with the professor to develop the course materials for Introduction to Machine Learning (CS 4774)
- Provided intuitive explanations for students during office hour with various machine learning algorithms and applications

Boichuk Lab | McIntire School of Commerce

Aug 2019 – May 2020

Undergraduate Student Researcher in Marketing

- Conducted times series & survival analysis using R on whether obtaining the benefit corporation certification increases company performance (revenue and public interest)
- Designed research methods through literature reviews and was sponsible for designing surveys and collecting raw data from web-scraping Google trend data

Projects

Crowd-source Data Platform

Sep 2020 – Nov 2020

- Led a team of four through web application development, which won the Social Innovation scholarship at Yonsei University
- Built a web application that allows people to crowd-source data collection process; admin can request specific data collection tasks, and the regular users can submit data and evaluate data quality

Other Projects

- pCO2 Modeling, Benefit Corporation Crawling + Analysis, Toxic Reply Filtering Modeling, Covid-19 Flight Visualization, UVA Wheelchair Access, Facebook Data Challenge, Relations Between Air Pollution and Happiness