

Taehee Jung

Minneapolis, MA 55414

☎ 412 | ✉ jungtaeh at amazon.com | 🏠 theejung.github.io | 📄 Google Scholar

Education

University of Pittsburgh

PH.D., DEPARTMENT OF STATISTICS

- Advisor: Lucas Mentch

Pittsburgh, PA, USA

Aug 2018 - Aug 2022

University of California, Berkeley

MASTER OF ARTS, DEPARTMENT OF STATISTICS

Berkeley, CA, USA

Aug 2016 - May 2017

Korea Advanced Institute of Science and Technology (KAIST)

BACHELOR OF ARTS, DEPARTMENT OF MANAGEMENT SCIENCE

- Academic excellence scholarship (2007 - 2011)

Daejeon, Korea

2007 - 2011

Professional Experience

Amazon, Alexa AI

APPLIED SCIENTIST

- Full-time Applied Scientist at Amazon Alexa AI

Seattle, WA

Sep 2022 - Present

APPLIED SCIENTIST INTERN

- Collaborators: Sungjin Lee, Joo-kyung Kim
- Applying text generation method to improve tail label performances in Extreme Multi Label Classification Task. Published in EACL' 2023.

Summer 2021

APPLIED SCIENTIST INTERN

- Collaborators: Liyuan Zhang, Tommy Powers
- Generated user's synthetic utterances for Alexa model training, using bi-directional transformer language models (BERT) and variational autoencoder (VAE).

Summer 2020

3M HIS M*Modal, Natural Language Understanding Machine Learning team

RESEARCH NLU INTERN

- Collaborators: Thomas Schaaf
- Developed a part-of-speech tagger for medical documents, using recent deep learning techniques (FLAIR).
- Interviewed with 3M Career website with respect to the internship experience [\[link\]](#).

Pittsburgh, PA

Summer 2019

Korea Credit Bureau (KCB)

DATA SCIENTIST

(1) BIG, PUBLIC DATA ANALYSIS, FUNDED BY MINISTRY OF GOVERNMENT ADMINISTRATION & JEONJU CITY

- Analyzed and visualized boarding histories, local floating population, petition documents, and home to office O/D information
- Constructed new strategies for Jeonju city's bus rearrangement plan [\[article\]](#) [\[report\]](#) (written in Korean)

Sep - Dec 2015

(2) FRAUD DETECTION SYSTEM DEVELOPMENT, FUNDED BY WOORI BANK

- Segmented customers' transaction histories and developed predictive ensemble models on regression, neural net, and decision tree
- Compared with previous FDS, overall detection rate increases more than 20% [\[article\]](#) (written in Korean)

Dec 2014 - Aug 2015

(3) MIGRANT ANALYSIS FOR MAGOK NEW TOWN, FUNDED BY LG ECONOMIC RESEARCH INSTITUTE

- Developed demographic database of migrants and classified them using decision tree for potential cultural demand to the new town

Jun - Sep 2014

(4) LOCATION BASED DECISION SUPPORT SYSTEM, FUNDED BY KOREA CREDIT BUREAU & R114

- Developed an integrated database (e.g., location, real estate, residential's demographic and financial data) called K-atlas

May - Aug 2012

Publication

- 2023 **Cluster-Guided Label Generation in Extreme Multi-Label Classification**, [Taehee Jung](#), Joo-Kyung Kim, Sungjin Lee, Dongyeop Kang, EACL
- 2022 **Quantifying Uncertainty in context of Natural Language Processing**, [Taehee Jung](#), PhD Dissertation
- 2020 **Posterior Calibrated Training on Sentence Classification Tasks**, [Taehee Jung](#), Dongyeop Kang, Hua Cheng, Lucas Mentch, Thomas Schaaf, ACL
- 2019 **Earlier Isn't Always Better: Sub-aspect Analysis on Corpus and System Biases in Summarization**, [Taehee Jung*](#), Dongyeop Kang*, Lucas Mentch, Eduard Hovy (*equal contribution), EMNLP

Skills

- Programming** Python, Java, R, SAS, SQL, Unix/Linux
- Language** English (Fluent), Korean (Native), Spanish (Beginner)