# SULYUN LEE

## PhD Candidate

sulyun-lee@uiowa.edu (+1) 319-512-5979 Iowa City, IA
sulyunlee SulyunLee sulyunlee.github.io

## SKILLS

Machine Learning
Deep Learning | Database
Data Mining | Graph Mining
Classification | Clustering
Recommender System
Social Network Analysis
Predictive Modeling
Language Modeling
Data Visualization
Data Preprocessing

## MACHINE LEARNING ALGORITHMS

Random Forest | XGBoost
AdaBoost | Decision Tree
Naive Bayes | PCA
Support Vector Machine
Logistic Regression
K Nearest Neighbors
K-means Clustering
Linear Regression

## DEEP LEARNING ALGORITHMS

GNN | GCN | GAT
CNN | RNN | LSTM
ANN | Encoder-Decoder
Autoencoder

## NATURAL LANGUAGE PROCESSING

BERT | Transformer
LDA | Word2Vec
Word Embedding

## WORK EXPERIENCE

### Statistical Consultant & Instructor | Iowa Social Science Research Center

08 2019 – Present — Iowa City, IA, USA

- Data Management and Analysis Workshop
- Network Analysis with NetworkX Workshop
- Introduction to Programming Workshop

### Graduate Research Assistant | University of Iowa

08 2017 – 05 2021 — Iowa City, IA, USA

- Achieved 40% gain on predicting the survival rate of heart attack patients on Medicare via a novel deep learning architecture for treatment optimization
- Predicted patient survival (AUC: 0.8) using Logistic Regression, Random Forest, SVM, and Neural Networks
- Validated the statistical significance in risks via regression analyses to communicate with health professionals with data-driven evidence

### Undergraduate Research Assistant | Handong Global University

06 2015 – 02 2017 — Pohang, South Korea

- Optimized classroom assignments to minimize students' travel distance
- Achieved 30% reduction in the total travel distance by using a genetic algorithm
- Translated a Python textbook by Prof. John V. Guttag from English to Korean

### Graduate Teaching Assistant | University of Iowa

08 2021 – Present — Iowa City, IA, USA

- Course: Analyzing Data for Informatics

### Undergraduate Teaching Assistant | Handong Global University

02 2016 – 12 2016 — Pohang, South Korea

- Course: Introduction to Big Data
- Python Camp for university students

## EDUCATION

### Ph.D., Information Science | University of Iowa

08 2017 – 05 2022 — Iowa City, IA, USA

### B.S., Computer Science and Engineering | Handong Global University

03 2013 – 02 2017 — Pohang, Korea

## DATA SCIENCE PROJECTS

### Representation Learning in Hierarchical Collaboration Networks |

*PyTorch, Python*

- Introduced a novel Graph Neural Network model that predicts team performance from hierarchical collaborations
- Achieved a 9% gain in predicting team success using the NFL coach dataset

Sentence Embedding
TF-IDF | Bag-of-words
Sentiment Analysis

## TOOLS

AWS EC2
Python | MySQL | SQLite
Jupyter Notebook
Google Colab | Rstudio
Spark | Hadoop | Java
C/C++ | Git | SPSS

## PACKAGES

PyTorch | Keras | PyG
Scikit-Learn | Statsmodel
Numpy | Pandas | Scipy
Matplotlib | Seaborn
NLTK | Gensim | Igraph
NetworkX
Deep Graph Library

## PROFESSIONAL SERVICE

Session Chair at Data Mining on Networks | INFORMS 2021

10 2021

Leader of Big Data Conference | Handong Global University

03 2016 – 12 2016

## CERTIFICATES

Coursera courses

- Structuring Machine Learning Projects Link
- Improving Deep Neural Networks Link
- Neural Networks and Deep Learning Link
- Machine Learning Link

### Dynamic Embedding Learning using Autoencoder

*PyTorch, Python*

- Proposed an auto-encoding heterogeneous co-evolving dynamic neural network
- Achieved a 48% gain on the mortality risk prediction task

### Team Success Prediction Among Research Scholars |

*Python, Regression analyses, NLP*

- Performed regression analyses and topic modeling to identify the collaborative patterns leading to scholars' team success
- Presented the increase of research team success by 50% with scholar's expertise using statistical tests

### HIV/AIDS Prediction |

*Python, Scikit-learn*

- Proposed decreasing HIV/AIDS by 7% using Linear Regression and Random Forest when a policy is implemented
- Provided data-driven evidence of cost reduction and public health benefits

### Link Prediction in an Online Health Community |

*Python, Scikit-learn, Keras, NLP*

- Achieved an 8% gain in predicting future interactions among online health community users with Logistic Regression, Random Forest, AdaBoost, and Neural Networks

### Customer Satisfaction Prediction on Crowdfunding Platform |

*Python, Regression analyses, Scikit-learn*

- Predicted production delays (AUC: 0.9) of crowdfunding business from posts using Random Forest, AdaBoost, and XGBoost
- Suggested appropriate entrepreneurs' responses on delays for customer satisfaction

## PUBLICATIONS

**Sulyun Lee** and K. Zhao. "Hierarchy2vec - Representation Learning in Hierarchical Collaboration Networks for Team Performance Prediction" *INFORMS Data Science Workshop*, 2021 (Best Paper)

**Sulyun Lee**, H. Jang, K. Zhao, M. Amato, and A. Graham. "Link Prediction in an Online Health Community for Smoking Cessation" *KDD workshop on Mining and Learning with Graphs*, 2020 | Paper

**Sulyun Lee**, H. Jang, K. Zhao, M. Amato, and A. Graham. "Multi-Relational Link Prediction for an Online Health Community" *INFORMS Data Science Workshop*, 2019 | Paper

L. A. Polgreen, N. Street, **Sulyun Lee**."Treatment Combinations for Elderly Patients and Those With Comorbidities After an Acute Myocardial Infarction" Circulation, 2019 | Link

## HONORS AND AWARDS

Best Paper Award Nominee | INFORMS Data Science Workshop

11 2021

Ballard & Seashore Dissertation Fellowship | Graduate College

02 2022 – 06 2022

Graduate Fellowship | Interdisciplinary Graduate Program in Informatics

09 2020 – 08 2021