SULYUN LEE

Data Scientist

sulyun.lee.careers@gmail.com | (+1) 319-512-5979 | Los Angeles, CA
sulyunlee | SulyunLee | sulyunlee.github.io

SKILLS

Machine Learning
Deep Learning
Graph Mining
Predictive Modeling
Network Embedding
Link Prediction
Image processing
NLP
Database
Classification
Clustering
Recommender System
Social Network Analysis
Data Visualization

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

NATURAL LANGUAGE PROCESSING

BERT
Transformer
LDA
Word2Vec

WORK EXPERIENCE

Data Scientist | Happiest Baby, Inc.

06/2022 – Present
Los Angeles, CA, USA

- Developed state-of-the-art ML/DL models that predict customer retention with 80% accuracy
- Researched sleep modeling to provide insights into infant sleep quality
- Collaborated with the marketing team and healthcare researchers to statistically demonstrate the outstanding performance of company products for better infant sleep
- Reported data insights for business strategy from big data

Statistical Consultant & Instructor | Iowa Social Science Research Center

08/2019 – 05/2022
Iowa City, IA, USA

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

Graduate Research Assistant | University of Iowa

08/2017 – 05/2021
Iowa City, IA, USA

- Achieved 40% gains in predicting the survival rate of elderly heart attack patients via a novel deep learning architecture for treatment optimization
- Discovered the risk factors of genetic diseases through statistical inference from massive medical records
- Provided data-driven insights for diseases to combine with the knowledge of health-care professionals

Graduate Teaching Assistant | University of Iowa

08/2021 – 12/2021
Iowa City, IA, USA

- Course: Analyzing Data for Informatics

EDUCATION

Ph.D., Informatics | University of Iowa

08/2017 – 08/2022
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, Network embedding

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

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

TOOLS

Python
SQL
DBT
AWS EC2
AWS Redshift
Google BigQuery
Jupyter Notebook
R
Spark
Hadoop
Tableau
Java
C/C++
Git
SPSS

PACKAGES

PyTorch
Keras
PyG
Deep Graph Library (DGL)
Scikit-Learn
Statsmodel
Tensorflow
Numpy
Pandas
Scipy
Matplotlib
Seaborn
NLTK
Gensim
Igraph
NetworkX

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 Healthcare Embeddings for Improving Patient Care |

PyTorch, Python, Network embedding

- Proposed a GCN model for heterogeneous co-evolving dynamic networks
- Achieved 48% gains on the mortality risk prediction in hospitals

Team Success Prediction Among Research Scholars |

Python, Regression analysis, 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 the scholar's expertise

Link Prediction in an Online Health Community |

Python, Scikit-learn, Keras, NLP

- Proposed a model that predicts future interactions among online health community users
- Achieved 8% gains in prediction using Logistic Regression, Random Forest, AdaBoost, and Neural Networks

Customer Satisfaction Prediction on Crowdfunding Platform |

Python, Regression analysis, Scikit-learn, Sentiment analysis

- Predicted production delays (AUC: 0.9) of crowdfunding businesses from social media posts using Random Forest, AdaBoost, and XGBoost
- Demonstrated the appropriate entrepreneurs' responses on delays for higher customer satisfaction using regression analysis and sentiment analysis

PUBLICATIONS

H. Jang, **Sulyun Lee**, D. M. H. Hasan, P. M. Polgreen, S. V. Pemmaraju, B. Adhikari. "Dynamic Healthcare Embeddings for Improving Patient Care" *In Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2022 | Paper

J. Lee, **Sulyun Lee**, W. N. Street, L. A. Polgreen. "Machine Learning Approaches to Predict the 1-year-after-initial-AMI Survival of Elderly Patients" *BMC Medical Informatics and Decision Making*, 2022 | Paper

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

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