SULYUN LEE

Data Scientist

@ sulyun.lee.careers@gmail.com　　📞 (+1) 319-512-5979　　📍 Salt Lake City, UT

in sulyunlee　　○ SulyunLee　　🌐 sulyunlee.github.io

SKILLS

Machine Learning

Deep Learning

Graph Neural Networks

Predictive Modeling

Statistical Modeling

Image Processing　NLP

Database Management

Recommender System

Social Network Analysis

Data Visualization

MACHINE LEARNING ALGORITHMS

Random Forest　XGBoost

AdaBoost　Decision Tree

Linear Regression

Logistic Regression

Naive Bayes　SVMs

KNN　PCA

K-means Clustering

SHAP Analysis

DEEP LEARNING ALGORITHMS

GAN　CNN　RNN

LSTM　Encoder-Decoder

TOOLS

Python　R　SQL　dbt

AWS EC2　AWS Redshift

Google BigQuery　Git

Jupyter Notebook

Spark　Hadoop　Tableau

WORK EXPERIENCE

Senior Data Scientist | Castell

📅 04/03/2023 – Present　　　　📍 Salt Lake City, UT, USA

- Developing a pipeline that utilizes a Large Language Model (LLM) that extracts patient health screening results from image-based PDF reports.
- Developing a prediction model that predicts patients' medication adherence to introduce interventions and help address the non-adherence problem.
- Developed a machine learning model that predicts the mortality of patients based on claims and EHR data to support decision-making in hospice eligibility.
- Developed a risk-adjusted model that predicts staffing needs in each clinic using a tree-based machine learning model.

Data Scientist | Happiest Baby, Inc.

📅 06/13/2022 – 11/15/2022　　　📍 Los Angeles, CA, USA

- Developed and deployed a CNN model that predicts customer retention based on longitudinal customers' product usage patterns.
- Validated the performance of product features in improving infant sleep quality using regression analysis, A/B testing, and visualization.
- Constructed and deployed database pipelines for production and modeling using dbt and SQL.
- Reported trends of product sales and product usage insights and recommended better business strategies to company executives.

Statistical Consultant & Instructor | Iowa Social Science Research Center

📅 08/19/2019 – 05/13/2022　　　📍 Iowa City, IA, USA

- Provided consultation on analyzing ego-centric network data to find how different types of social interactions influence individuals' decision-making.
- Provided consultation on collecting data from NGO websites using web scraping tools in Python.
- Designed and opened workshops for training students and faculties at the university with data science skills – Data management, analysis, predictive models, and network analysis.

Graduate Research Assistant | University of Iowa

📅 08/21/2017 – 05/14/2021　　　📍 Iowa City, IA, USA

- Developed a novel architecture using MLP that recommends personalized medications for heart attack patients with a 40% increase in survival probability.
- Provided statistical inference on risk factors of genetic diseases from massive medical claims data.
- Collaborated with doctors and pharmaceutical experts to write medical research papers to provide data-driven insights into diseases.

DATA SCIENCE PROJECTS

Graph Neural Networks for Team Performance Prediction | ○

- Developed a novel Graph Neural Network model that predicts team performance based on hierarchical collaborations among team members using PyTorch.

PACKAGES

CERTIFICATES

- Neural Networks and Deep Learning Link
- Improving Deep Neural Networks Link
- Machine Learning Link
- Advanced Learning Algorithms Link
- Build Basic Generative Adversarial Networks (GANs) Link
- SQL for Data Science Link

- Achieved 9% improvements in predicting NFL team wins from coaches' collaboration patterns using the NFL coach lineup dataset collected by web scraping techniques.

Improving Healthcare Using Deep Learning on Patient Events Graph

- Developed a Deep Learning model that learns representations for dynamic and heterogeneous graphs using PyTorch.
- Achieved a 48% increase in mortality risk prediction by applying the proposed model to Electronic Health Records (EHR) data.

Team Success Prediction for COVID-19 Research |

- Proposed a statistical model that predicts the success of COVID-19 research papers using academic collaboration graphs constructed from the CORD-19 database. Python statsmodel and scikit-learn libraries were used.
- Achieved an increase in prediction performance by 50% using the NLP topic modeling technique for analyzing the contents and values of research papers.

Predicting links in an Online Health Community |

- Proposed machine learning models that predict future links among the online health community users using Logistic Regression, Random Forest, AdaBoost, and Neural Networks implemented with Scikit-learn and Keras.
- Achieved an 8% increase in link prediction with multi-modal information from graphs using network analysis and the DeepWalk algorithm.

Customer Satisfaction Prediction on Crowdfunding Platform |

- Predicted customers' satisfaction on a crowdfunding platform with 90% test AUC based on different entrepreneurs' business strategies using Random Forest, AdaBoost, and XGBoost.
- Collected data from a crowdfunding website using a web scraping technique and stored it as structured data for analysis.
- Performed sentiment analysis on texts from comments and updates forums to extract customers' satisfaction scores.

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, South Korea

PUBLICATIONS

Sulyun Lee, W. N. Street, B. L. Carter, L. A. Polgreen. "Evaluation of Medication Combinations for Older Patients and Those with Comorbidities After an Acute Myocardial Infarction Using Machine Learning" *Journal of the American Medical Informatics Association* **(In review)**

Sulyun Lee, L. M. Harris, A. C. Miller, J. E. Cavanaugh, J. M. Nizar, J. E. Simmering, M. H. Abou Alaiwa, L. A. Polgreen, P. M. Polgreen. "Risk for Dehydration and Fluid and Electrolyte Disorders Among Cystic Fibrosis Carriers" *American Journal of Kidney Diseases*, 2023 | Paper

H. Jang, **Sulyun Lee**, D. M. H. Hasan, P. M. Polgreen, S. V. Pemmaraju, B. Adhikari. "Dynamic Healthcare Embeddings for Improving Patient Care" *IEEE/ACM Advances in Social Networks Analysis and Mining (ASONAM)*, 2022 | **Best Paper Awards (Runner-up)** | 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 **| Best Student Paper Nominee** | 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