Jaeyoung Choi

+82(10)8220-9061 | cjengy@gmail.com | https://jaeyoung-jane-choi.github.io

Research Interest

Digital Humanities, Data Curation, Machine Learning, Deep Learning, Natural Language Processing

Education

Sungkyunkwan University, Seoul, South Korea   Mar 2016–Feb 2021

Bachelor of Arts in Library and Information Science

- GPA: 3.93/4.50

Bachelor of Science in Data Analytics

- GPA: 4.29/4.50

Indiana University, Bloomington, IN   Aug 2019–Dec 2020

Exchange Student, Luddy School of Informatics, Computing and Engineering

- GPA: 4.0/4.0

Publications

Jaeyoung Choi, Chaeeun Han, Heeyoon Yang, Yeonkyoung Hong, Seoyoung Jeon and Yongjun Zhu. 2021. Embedding-based Neural Network Models for Book Recommendation in University Libraries in *Workshop on AI + Informetrics* (AII2021, 2021)

Jaeyoung Choi, Heeyoon Yang, Hayoung Oh. 2020. Store Sales Prediction Using Gradient Boosting Models in *Journal of the Korea Institute of Information and Communication Engineering* (JKIICE, 2020)

Projects

Recommendation System for Sungkyunkwan University Library | *Python*   Oct 2020–Present
- Utilized book and user embeddings through RoBERTa and Efficient-net
- Implemented a neural network to predict user's book preference
- Probed the recommendation system by giving student interviews

Analysis of Seoul Public Bike Usage | *Python, R*   Mar 2019–July 2019
- Administrated digital strategy to help understand predictions for public bike stations in Seoul on a daily basis
- Researched variables and enriched it by centrality and index creation
- Suggested prediction models of statistical analysis, machine learning and neural networks with entity embedding

Factor Analysis of Juvenile delinquent | *R*   Sep 2018–Dec 2018
- Retrieved relevant data from surveys provided by Korean Children and Youth Panel Survey(KCYPS)
- Conducted survival analysis to determine correlated factors for adolescents' first runaway

Class Projects

Classification on Korean petitions | *Python*   May 2020–July 2020
- Created word embedding through word2vec and pre-trained word embeddings
- Proposed classification models utilizing deep learning methods such as recurrent neural network (RNN), convolutional neural network(CNN) and long short-term memory(LSTM)

Analysis on the social trend of COVID-19 | *Python, R*   May 2020–July 2020
- Accumulated headlines regarding COVID-19 from the New York Times
- Preprocessed words to see the monthly trend of COVID-19
- Analyzed and visualized the trend using word-cloud
- Wrote short paper about the social trend through co-occurring words with COVID-19

Effects of obesity in outpatient mental health treatment | *R*   Oct 2019–Dec 2019

- Compiled survey data from the 2018 National Survey on Drug Use and Health(NSDUH)
- Interpreted statistical inference from the exposure variable: Body mass index(BMI) and outcome variable: outpatient mental health service utilization
- Formulated additive and interaction models of four covariates: alcohol consumption, suicidal thinking, education level and income
- Wrote paper regarding the models and their influence on the association between obesity and outpatient mental health treatment

Obama to Trump: The American shift | *R* Oct 2019–Nov 2019
- Analyzed the interaction between immigration attitude and demographic variables to explain the probability of switching votes from Obama to Trump
- Composed a short paper analyzing whether the interactions of demographic factors respect to immigration attitudes have effect on the probability of voting for Trump in 2016, given a vote for Obama in 2012

Exploratory Data Analysis of Life expectancy and its relationship to income | *R* Sep 2019–Oct 2019
- Scrutinized the relationship between GDP per capita and life expectancy
- Wrote short paper on the statistical inference of the overall patterns regarding GDP and life expectancy for each continent

INTERNSHIPS

Data Marketing Korea *Artificial Intelligence Development Intern* Jan 2021–March 2021
- Applied Bidirectional Encoder Representations from Transformers(BERT) on social buzz data for multi class classification and sentimental analysis on the detection of customers true voice

TEACHING EXPERIENCE

Teaching Assistant of Introduction to Artificial Intelligence Sep 2020–Dec 2020
- GCO-2002 Introduction to Artificial Intelligence in Institute for Convergence, Sungkyunkwan University
- Contained basic Python programming, Algorithms, Neural Networks, and Deep Learning Techniques
- Answered questions online, marked assignments and proctored exams

AWARDS AND EXTRA CURRICULAR ACTIVITIES

2020 Co-deep Learning Project | *3rd Place* Feb 2021
- Sungkyunkwan University, Seoul, South Korea

2020 Data Creator Camp Hackathon | *3rd Place* Oct 2020
- National Information Society Agency, Seoul, South Korea

TNT(Train aNd Test) Sep 2020–Sep 2021
- Gave lectures and made study material on weekly basis
- Participated in study groups for natural language processing and data mining

P-SAT (Power Statistical Analysis Techniques) Sep 2018–Sep 2019
- Engaged in study groups for regression and categorical data analysis
- Prepared learning materials of statistical inference and gave lectures on a weekly basis

SKILLS

Programming: Python, R, SQL, HTML, QGIS
Language: English(Fluent), Korean(Navtive)