YONGHYEON KWEON

852. w main St, Charlottesville, VA 22903, USA
Phone: (804) 497-9166
Email: yk4we@virginia.edu
GitHub: https://github.com/BeDSmonster

AREAS OF INTEREST

Machine Learning, Deep Learning, Security, Federated Learning, Adversarial Learning, Decentralized Machine Learning, Data Science, Data Analytics

EDUCATION

UNIVERSITY OF VIRGINIA, Charlottesville, VA, USA **07/2019-05/2020**

Master of Science in Data Science (GPA:3.8/4.0)

***Relevant Courses:** Machine learning, Bayesian Machin Learning, Data Mining, Text Analytics, Programming and Systems for Data Science

UNIVERSITY OF VIRGINIA, Charlottesville, VA, USA **08/2016-05/2019**

Bachelor of Arts in Statistics, (GPA:3.34/4.0, major GPA: 3.8/4.0)

***Relevant Courses:** Linear Algebra, Mathematical Statistics, Regression Analysis, Mathematical probabilities, Data Mining

RESEARCH EXPERIENCE

Link Lab, School of Engineering, University of Virginia, Charlottesville, VA, USA **05/2020-Present**

Research Assistant

Project Title: "Route Guidance Recommendation System" by Prof. Brian B. Park

- The purpose of this project is to predict driver route choice behavior using a federated learning approach
- Designed a federated learning model to mimic driver characteristics such as route choice behavior with TensorFlow Federated (TFF)
- Implemented clustering methods based on individual SVM models to apply to a federated learning framework
- Demonstrated robustness of the federated learning approach when predicting driver route choice behavior compared to a centralized model

McIntire School of Commerce, University of Virginia, Charlottesville, VA, USA **05/2020-Present**

Research Assistant

Project Title: "Dynamics in Movie Reviews" by Prof. Natasha Foutz

- The purpose of this project is to capture the dynamics of movie reviews and trends
- Scraped movie reviews and box-office data from multiple websites using Beautiful Soup and Selenium
- Conducted the cleaning, merging, and matching of data from different sources to build a database of box-office and review information by genre

School of Data Science, University of Virginia, Charlottesville, VA, USA **08/2019-05/2020**

Researcher

Project Title: "Deep Learning for Protein Structural Class" by Professor Cameron Mura and Professor Philp Bourne

- The purpose of this project was to restructure protein classification using 3D representation and examine classification properties of the existing CATH scheme
- Trained an autoencoder model for protein classification using 3D representation of protein structure
- Leveraged 3D convolutions to deal with data sparsity and make the problem tractable and resource-efficient
- Produced loss function as a similarity metric between different protein classes (superfamily) to augment novel protein classification
- Findings implied the existence of important factors, such as other levels of structural organization beyond purely geometric and structural similarities, supporting the recent notion of 'Urfold'

RELEVANT COURSE PROJECTS

Project Title: "Real-Time Face Detection of Gender, Age and Emotion" **01/2020-05/2020**

- Machine Learning, Supervisor: Dr. Jonathan Hughes
- The purpose of this project was to classify age, gender, and emotion of human faces in real time using different model architectures and pipelines
- Implemented different architectures and pipelines: stacked/combined pipelines with 5 different architectures including miniXception, bigXception, xceptionFinetune, InceptionV3 and resNet50
- Measured the performance of the results in terms of accuracy and time efficiency and concluded that stacked pipeline obtained 10% higher accuracy and leads to a 13 millisecond decrease in detection time
- *Technical skills:* **Python, UVA cloud computing (Rivana), Keras, openCV**

Project Title: "News Content Analysis with Natural Language Processing" **01/2020-05/2020**

- Text Analytics, Supervisor: Professor Rafael Alvarado
- The purpose of this project was to compare news content from two sources on different ends of the political spectrum
- Analyzed news content from two different data sources using NLP to compare sentiment and word usage
- Conducted sentiment analysis, LDA topic model and word embedding for comparison, and visualized results with t-SNE
- Found that proper nouns used in news article are insignificant in contrast to the important role they play in literature
- *Technical skills:* **Python, NLTK, Gensim, seaborn, plotly_express, SciPy, word2vec**

Project Title: "Prediction Model of Graduate School Admission with Bayesian Inference" **08/2019-12/2019**

- Bayesian Machine Learning, Supervisor: Professor Donald E. Brown
- The purpose of this project was to identify the importance of undergraduate school ranking in predicting the probability of being admitted to graduate school
- Modeled Bayesian linear regression and utilized hierarchical/multilevel modeling for national ranking of candidates' undergraduate schools
- Evaluated the Bayesian linear regression and hierarchical models with WAIC score and concluded that undergraduate school ranking does not have a significant effect on graduate school admission
- *Technical skills:* **Python, UVA cloud computing (Rivana), PyMC3, scikit-learn, seaborn, Matplotlib**

Project Title: "Data Mining for Banknote Authentication" **01/2019-05/2019**

- Data Mining, Supervisor: Professor Xiwei Tang
- The purpose of this project was to evaluate different classifiers for counterfeit detection
- Used various supervised learning models such as SVM, LDA, and random forest for classification using the banknote authentication data from UCI ML Repository
- Evaluated models using AUC score metrics and examined the accuracy and sensitivity for each model
- Six classifiers - LDA, logistic regression, random forest, SVM, and Adaboost - were used in the projects and SVM was judged the best model based on its predictive power and algorithmic characteristics
- *Technical skills:* **R**

PUBLICATION

[1] Jaiswal, M., Saleem, S., **Kweon, Y.**, Draizen, E. J., Veretnik, S., Mura, C., and Bourne, P. (2020). Deep Learning of Protein Structural Class: Any Evidence for 'Urfold'?. *2020 Systems and Information Engineering Design Symposium (SIEDS).*

- Proposal accepted for 2020 TomTom Applied Machine Learning Conference
- Research abstract accepted for ISMB2020 (Intelligent Systems for Molecular Biology) Abstract

[2] Kweon, Y., Sun, B., Park, B. Modeling Route Choice Behavior: A Federated Learning Approach, *2021 Journal of Transportation Research Record.* [Under Review].

- Accepted paper for UCK2020 poster presentation
- Manuscript accepted for 2021 TRB Annual Meeting

WORK EXPERIENCE

Propel Career Accelerator, University of Virginia Career Center **06/2020-07/2020**

Team Member

- Earned 1st place for highest quality recommendation in consulting capstone competition, University of Virginia
- Collaborated with a 5-student virtual team to analyze the challenges that the client, Cardboard Live, faces: it is in a concentrated market with dedicated fans. Presented research-supported recommendations and solutions
- The goal was to suggest a marketing strategy using SNS or influencers on SNS for Cardboard Live, live streaming with the company in an AI-enabled interactive experience to enhance brand image and reach a more diverse segment of viewers
- Designed and delivered a succinct 10-minute PowerPoint presentation for client

Extreme Network, Seoul, Korea **05/2014-07/2014**

Intern

- Acted as a liaison for the engineering team's ongoing technical project and delivered a presentation to the business team
- Assisted in the preparation of conference/meetings with clients or other Ethernet companies

CONFERENCE PRESENTATION

2021 TRB Annual Meeting (expected, Jan/2021), Modeling Route Choice Behavior: Federated Learning approach

VOLUNTARY/ EXTRACURRICULAR ACTIVITIES

Missionary/Voluntary Work, Guatemala **07/2015**

- Worked as a team to provide medical service and recorded/organized prescriptions
- Built stoves for local residential houses

HONORS & AWARDS

Dean's List, University of Virginia **Spring 2019**

SKILLS

COMPUTER

Deep learning libraries: PyTorch, TensorFlow

Programming Language: Python, R, C/C++, SAS

LANGUAGES

Korean (Native), English (Fluently)