Yiqun Hu

huyiqun2017@gmail.com | 815-593-4505 | LinkedIn | GitHub | Bethesda, MD 20814

Education

Johns Hopkins University   **Baltimore, MD**
Master of Science in Information Systems   Aug 2021 - Aug 2022
  Courses: Large Scale Computing with Hadoop **(AWS)**, Deep Learning **(PyTorch, TensorFlow),** Machine Learning **(Python)**, Data Science **(R, Tableau)**, Data Visualization **(R Shiny)**, Business Analysis, Web Development **(HTML, CSS, JavaScript)**

Northern Illinois University   **DeKalb, IL**
Master of Science in Statistics, GPA: 3.85   Aug 2019 - May 2021
  Thesis: Traffic fatality rate prediction based on Deep Neural Network and Bayesian Neural Network
  Courses: Time Series Analysis **(R)**, Statistical Learning, Bayesian Statistics, Longitudinal Data Analysis **(SAS)**, Multivariate Analysis, Generalized Linear Models, Statistical Software **(Python, R)**

Bachelor of Science in Statistics, Minor in Economics   Aug 2015 - May 2019
  Achievement: Awarded Business Analytics & SAS Scholarship by College of Business OM&IS Department
                        Passed Actuarial Probability Exam

Experience

Ping An Technology   **Shanghai, China**
Machine Learning Engineer Intern   Sept 2020 - July 2021
- Built a data ETL pipeline for monthly ESG data of 4,500+ China A-share listed companies since 2015 **(Python, pgSQL)**
- Collaborated with cross-functional teams to build a distinctive CN-ESG (Environmental, Social, Governance) evaluation framework including 24 themes, 154 secondary indicators, and 43 industry indicators using a variety of **Machine Learning** techniques and integrated ESG score into FactSet dataset
- Performed sentiment analysis with **Deep Learning** using BERT on over 20 million public opinion data and improved the model performance (F1 score increased by 3%) by fine-tuning
- Developed **RESTful APIs** for trained ML models deployment into **Linux** production environment **(Python, Flask, CentOS)**
- **Patent:** Zhu, Shizhuo; Shao, Xi; **Hu, Yiqun**. Enterprise ESG Index Determination Method Based on Clustering Technology and Related Product. CN Patent CN113190683A, filed July 02, 2021, and issued July 30, 2021.

Northern Illinois University   **DeKalb, IL**
Math SI Leader   Aug 2018 - May 2019
- Facilitated out-of-class SI sessions to provide guidance and review course materials on a weekly basis for 30+ students in Fundamentals of Mathematics I&II.
- Earned positive feedback from my students regarding classroom instruction, office hours, and student learning success.

Projects

Wine Quality Prediction with Statistical Learning (Python)   Sept-Dec 2020
- Implemented EDA in Python to explore the influence of 11 predictors on wine quality by analyzing heavy datasets collected from UCI Machine Learning Depository
- Predicted wine quality data by using methods of Linear Regression, KNN, Random Forest, Ridge & Lasso Regression, and optimize the final Random Forest model to improve the prediction accuracy

Factors Affecting the Number of Prescription Drugs by Longitudinal Regression Analysis (SAS)   Oct-Dec 2019
- Led a team of 4 students to construct models to investigate the relationship between Prescription Medicines and 15 independent variables based on 4000+ data records from The LSOA II Wave 2 Survivor data (CDC)
- Utilized SAS to build multiple regression models (Binomial, Normal, Negative Normal, Poisson, Forward Logit, Backward Logit) to get the best fit model and visualize the analyses in a 20-page paper

Monthly Milk Production Forecast by Time Series Analysis (R)   Sept-Nov 2018
- Consolidated and cleaned the datasets of monthly milk production (pounds per cow) from January 1962 to December 1975.
- Completed the data visualization, model building, and residual analysis in regression and predicted the data for the next 10 years with ARIMA model

Programming Skills: Python (Pandas, NumPy, sklearn, Keras, TensorFlow, PyTorch), R, SAS, SQL, C++, HTML, JavaScript, CSS
Data Science & Miscellaneous: Data Science Pipeline (Cleansing, Visualization, Modeling, Deployment, Interpretation), DBMS, Spark, Linux, Git, Microsoft Office Suites
Languages: English, Mandarin (Native)