Kyu Cho

St. Louis, MO, USA • qcho1010@gmail.com • 636.579.3120
qcho1010.com • linkedin.com/in/qcho1010 • github.com/qcho1010

Accomplished Full Stack Data Scientist and Machine Learning Engineer as a TechLead with deep expertise in quantitative research and the development of predictive models using advanced ML and optimization algorithms. Project work includes engagements for Fortune 500 clients, healthcare systems, and 120 personal projects.

EXPERIENCE

Lead Data Scientist, Object Computing, Inc. (OCI) - St. Louis Jan 2019 - present

OCI is a software consulting firm and official Google partner focused on Machine Learning, Blockchain, IoT, and Cloud.

Serve in a client-facing consultant role as a lead developer with additional contributions to technical and business decision making. Client and internal projects have included:

- ***Internal*** - Built the custom AI accelerator that rapidly converts subtle data patterns into actionable and scalable insights, resulting in saving 40% of the time of building, testing and deploying the AI products.
- ***For Mastercard*** - Built scalable and preventive anomaly detection model used for billions of transactions with insights visualized on a dashboard, resulted in a 65% reduction in total customer impact. Leveraging Splunk technology with sequential Deep Learning (LSTM) and NLP (BERT) techniques.
- ***For Purina*** - Contributed to multiple video image analysis and segmentation projects as lead developer. Successfully delivered models in production with significant penetration into the business resulted in landing three more follow up contracts with the client. **1)** Wrote a pet matching recommendation system that provides similar-looking dogs available for adoption based on a user's image upload for a desired breed, size, and color. **2)** Developed a model that accurately detects dog and cat breeds including mixed breeds with over 85% accuracy. **3)** Built a model that calculates dog health based on the image of the body shape. These projects led the opportunity to accurately recommend Purina products to customers, based on the breeds, size, and health, which resulted in a 12% estimated annual sales increase.
- ***For Bayer*** - Wrote a recommendation system for agriculture products that enables users to identify optimal yields based on historical weather, soil, and geospatial data. Developed a pricing suggestion system that helps clients to maximize the overall profit by 18% and reducing the risk of having low yields.

Senior Data Scientist, Ascension - St. Louis Oct 2017 - Jan 2019

Ascension is the second-largest healthcare organization and hospital network in America.

Hired as first Senior Data Scientist and interviewed candidates as the team grew to nine. Led team through several projects and studies designed to improve healthcare outcomes for Ascension patients. Highlights include:

- Created a recommendation system for a web-based dashboard that identifies optimal Sepsis treatments while reducing the total direct cost, which resulted in 36% (~$300MM) in estimated annual savings to Ascension and 11.7% reduction in overall mortality rate.
- Wrote an AI system that provides predictive, interpretable, and explainable results beyond the typical black box solutions by developing custom Machine Learning algorithms that help physicians by providing the transparent recommended treatments.
- Built a model to identify Opioid abusers to combat the Opioid crisis by predicting human behavior.

Principal Data Scientist, Missouri Institute of Mental Health (MIMH) - St. Louis Jan 2017 - present

Serve as Lead Investigator for all ML-related grants. Project work has included:

- Performed advanced ML analysis to identify insights for HIV disease prevention and treatment on projects, resulted in a grant awarded from the government including the National Institutes of Health.

- Collaborated with the State of Missouri and Medicaid on the study to prevent rehospitalization by predicting patient outcomes using ML.
- Led advanced analysis of brain neural network images using ML to identify network clusters associated with frailty. Also, identified the causal relationship between substance abuse related to brain damage.

Assistant Data Scientist, University of Missouri - St. Louis Jan 2016 - Dec 2016

- Identified 28 key gene expressions of 8500 that related to Alzheimer's Disease under Dr. Sharlee Climer, leading to the publication of
- Developed automated theorem proving web platform using A.I. under Dr. Wenjie He.

PUBLICATION

- Cho KS; Paul RH, *"Machine-learning classification of neurocognitive performance in children with perinatal HIV initiating de novo antiretroviral therapy"* AIDS (2019): doi:10.1097/QAD.0000000000002471.
- Cho KS, *"Ensemble learning with feature selection for Alzheimer's disease prediction"* International Journal of Scientific & Technology Research (2017): 0117-16025.
- Cho KS; Paul RH, *"Ensemble machine learning classification of frailty in persons living with HIV"* Journal of Acquired Immune Deficiency Syndromes (2019): Submitted.

EDUCATION & CERTIFICATION

MS Computer Science, University of Missouri - St. Louis

- Researched in Investment returns prediction with risk management using ML and optimization algorithm.

BS Computer Science, University of Missouri - St. Louis

- Coursework: Artificial Intelligence, Evolutionary Computation, Machine Learning, Data Mining, Cloud Computing, Advanced Algorithm & Data Structure.

Google Certified Professional Data Engineer (digital certificate here)

Coursera: Data Science (Johns Hopkins), Big Data (UC San Diego), Data Science & Engineering with Spark (UC Berkeley), Machine Learning (University of Washington), Data Science at Scale (University of Washington), Statistics with R (Duke University), Data Analysis & Interpretation (Wesleyan University)

EdX: Tackling the Challenges of Big Data (MIT), The Analytics Edge (MIT), Advanced Statistics with R for the Life Sciences (Harvard), Android App Development (University of Illinois Champaign-Urbana)

PROJECTS & HACKATHONS

- **120 personal projects** related to finance (alpha, trade, investment) and business, health, and social issues, automation, and optimization available on GitHub.
- **First Place**, Hackedu at UMSL - Intelligent Location Tracker/Analyzer to Prevent the Future Loses
- **Top 10**, Global Hackathon 6 at SLU - Analytic Platform for Future Homeless Prevention, Efficient Resource Distribution and Management System
- **Third Place**, Arch Hacks at WSU - Self-Motivated A.I. Health Care System for Heart Disease and Cancer

SKILLS

Machine Learning: Keras/TensorFlow, BERT, Attention, CNN(InceptionResnet, YOLO), RNN(LSTM, GRU), LDA, CatBoost, XGBoost, Light-GBM, Scikit-learn, K-NN, SVM, Naive Bayes, SciPy, Stats, A/B testing, H20, MLLib, NLTK

Big Data: Spark, Hadoop, MapReduce, Hive, Google Cloud (GCP), AWS, Azure, Splunk, Distributed-computing, Dask, Multi-processing, CUDA, numba, ETL data cycle (extract, validate, transform, clean, aggregate, audit, archive), EMR

Languages: Python, R, C++, C, Java, SQL, NoSQL, Tableau, Shiny, Linux shell,

Concepts: Full-stack Development, Recommendation Systems, Deep Learning, Transfer Learning, Ensemble Learning, Natural Language Processing (NLP, tokenization, tagging, sentiment analysis, entity recognition, summarization), Reinforcement Learning (Q-learning), Supervised / Unsupervised Learning (Clustering, Outlier Detection, PCA, ICA, NNMF, SVD), Evolutionary Computation, Bayesian Optimization, Linear Programming, Dynamic Programming, Text Mining, Data Visualization, Data Interpretation, Statistical Modeling, Multivariate Regression (ANOVA, MANOVA), Performance Monitoring, Recalibrating models

Portfolio Optimization using ML / Artificial Intelligence, Risk Modeling (Sharpe Ratio, Alpha, Beta, etc), Credit & Fraud Modelling, Financial Modeling, Quantitative Research, Quantitative Development, Time Series Forecasting,

Business Opportunity / Value Finding, Leadership, Management(Agile, Scrum)