Tik Kei Dicky Woo

Folsom, CA 95630 United States | (916) 914-3804 | dickywoo@berkeley.edu

EDUCATION

University of California, Berkeley

Master of Information and Data Science, Data Science — Aug 2023

Computer Vision, Natural Language Processing with Deep Learning, Machine Learning at Scale, Statistical Methods for Discrete Response, Time Series, and Panel Data

B.S., Electrical Engineering and Computer Science/Materials Science and Engineering — Aug 2014

SKILLS

- Programming: Python, R, Perl, C++, Java
- Supervised Machine Learning: Regressions, SVMs, Decision Trees & Ensemble Methods, Naïve Bayes, kNNs, Recommendation Systems
- Unsupervised Machine Learning: PCA, Clustering, GMMs
- Deep Learning: CNN, RNN, LSTM, Transformers
- Statistical Modeling: Logistic Regressions, Time-Series Analysis, Panel Data Analysis, Model Selection & Diagnostics
- Data Engineering: SQL/NoSQL Databases (i.e. PostgreSQL, Redis, Neo4j), Hadoop, Spark, AWS, GCP
- Electrical Engineering: NAND Flash, Device Physics, Analog/Digital Circuit Design
- Material Characterization: SEM, TEM, Powder Diffraction
- Language: Native in Cantonese; Fluent in Mandarin Chinese and English

DATA SCIENCE PROJECTS

- **Text Summarization of Academic Paper for Title Generation:** Group project for Natural Language Processing with Deep Learning. Fine-tuned sequence-to-sequence transformer text summarization models (PEGASUS, T5, BART) to summarize arXiv academic paper abstracts to generate corresponding titles. The best model achieved ROUGE-L score of 0.49 between predicted titles vs actual titles
- **U.S. Domestic Flight Delay Prediction:** Group project for Machine Learning at Scale course. Leveraged Apache Spark on the Azure Databricks platform to develop large-scale ML models (logistic regression, gradient-boosting decision tree, XGBoost decision tree) to predict flight delays from all U.S. domestic flight in 2021 based on U.S. Department of Transportation flight dataset and NOAA weather dataset from 2015-2020 (40M+ rows of data). Developed the ELT process, feature engineering (i.e. PageRank on U.S. airports) and ML algorithm pipelines (i.e. times-series block cross-validation, hyperparameters grid search) in 5 weeks
- **Optimal Delivery Solution with BART System:** Group project for Fundamentals of Data Engineering course. Performed ELT process to process multiple data sources (customer, BART and Google Map) into PostgreSQL database and Neo4j graph database. Utilized various graph analysis and Google APIs to find the optimal delivery solution utilizing the San Francisco BART system.
- **Incident Management Predicting SLA conformance:** Group project for Applied Machine Learning course. Performed EDA and built a ML model on the IT service management data to predict a given service level agreement (SLA) can be achieved on time given the attribute of the submitted ticket.
- **Statistical Analysis on Global CO2 Concentration:** Group project for Statistical Methods for Discrete Response, Time Series, and Panel Data course. Performed time-series analysis (i.e. ARIMA model) on the Kneeling Curve and NOAA CO2 Concentration data to model and predict future CO2 concentration.

WORK EXPERIENCE

Product Engineer, Micron Technology, Folsom, CA — Mar 2015 to Present

- Design experiments to characterize device performance, power, read window budget, reliability on memory products to deliver optimal trims that meet performance/reliability specifications
- Develop state-of-art and advanced software to automate data collection and analysis on product validation and characterization experiments
- Present/share experiment and data analysis results and collaborate with other Engineering groups to problem-solve, configure, and optimize the product to improve performance, yield, cost, and reliability