

HUA, Yuchen (April)

LinkedIn: <https://www.linkedin.com/in/yuchen-hua-70505917b>

Mobile: (+1) 413-461-8023 | Email yuchenhua96@gmail.com

EDUCATION

Brown University (GPA: 3.83/4)	M.S. in Data Science	08/2020 – 12/2021
Brown University	M.S. in Physics	08/2018 – 05/2020
East China University of Science and Technology	B.S. in Applied Physics	08/2014 – 06/2018

SUMMARY

Experienced scientist with 3 years of experience in developing end-to-end predictive models, data mining and big data analytics. Extensive experience in leading data science teams and cross-functional projects. Seeking a leadership role to guide and inspire teams in delivering impactful data-driven insights. Committed to staying abreast of the latest advancements in data science and technology to foster innovation and continuous improvement within the organization.

PROFESSIONAL EXPERIENCES

- **Citizens Financial Group Inc.** 02/2022 - Current
- **Senior Data Scientist Model lead – Fraud Detection and Risk Segmentation** 09/2023 - Current
 - Led data scientist team on deposit check fraud detection and owned project development process end-to-end. Successfully reduced deposit check fraud losses by ~7MM with final chosen model **lightgbm**. Model achieved average precision 55%, and capture rate (recall) 55% with less than 10% false positive rate. Led **cross-functional** collaboration with other teams and presented modeling insights to stakeholders, ensuring modeling objectives align with business objectives.
 - Used parallel computing tools **Dask** in Python processed ~500GB raw data in **parquet** format. Designed data structure for modeling, performed data cleaning, and **engineered features** such as demographic, statistical and trends information for modeling on different granularity. Converted preprocessing and feature engineering into modularized **data pipeline**.
 - Perform correlation analysis and permutation importance to select modeling features. Worked with fair banking team, developed frameworks to ensure compliance with data privacy regulations and fairness standards.
 - Designed and implemented **sampling scheme** to sample data from a fixed in-time and out-of-time range, a fixed target ratio in training set. **Up-sampled** training dataset based on target, efficiently improved model confidence due to highly imbalanced fraud data nature.
 - Used **Optuna** to perform model selection, and automate **hyperparameter tuning** process with Bayesian optimization, largely improved working efficiencies and reduced the time of manually tuning effort.
 - Created interactive dashboard and various segmented visualizations to evaluate model performance. Designed and proposed model monitoring plan and ensured its implementation for continuing effort to detect data and model drift.
- **Data Scientist - Cashflow Time Series Forecasting for Commercial Banking** 02/2022 – 08/2023
 - Developed multi-entity, multi-variate, and multi-step **time series forecasting** models using **TensorFlow**, employing an encoder-decoder LSTM structure to predict cash flows. Model achieved 35% sMAPE for single-day point forecast on highly active client accounts, 30% sMAPE in weekly aggregated forecasts and 25% sMAPE in monthly aggregated forecasts for overall accounts performance. Enhanced commercial clients' balance management experience and improved customer satisfaction.
 - Preprocessed a large-scale dataset (~90 TB of sequential history data) using the distributed computing tool **Dask**. Developed modular data pipeline to clean, scale, and process data into multi-dimensional sequences that adapted to **Encoder-Decoder** model. Designed and implemented custom data generators in **TensorFlow** with **stratified sampling** based on activity level, to optimize large-scale data loading and significantly improving training speed. Documented end-to-end data pipeline and generator implementation process for reproducibility and team knowledge sharing.
 - Automated model training process by setting up **containers** training on **AWS Experiments** with TensorFlow and Trail objects. Distributed different hyperparameter combination cross **EC2 instances**, wrote script enabling parallel training, hyperparameter tuning and result uploading without continuous human supervision, significantly improved training processing and result monitoring efficiency.
 - Developed model performance evaluation pipeline, created illustrative plots to evaluate weekly and monthly aggregated performance based on business needs. Interpreted and explained model strengths and weaknesses to stakeholders, provided actionable insights for adjusting business strategy accordingly.

TECHNICAL SKILLS

- **Programming language:** Python, R, SQL, Scala, HTML, Tableau
- **Tools:** Dask, PySpark, Optuna, AWS SageMaker Studio, AWS Experiment, AWS Docker, Bitbucket (Github), Tableau
- **Packages:** Tensorflow, PyTorch, Optuna, Keras, MLflow, Pandas, Numpy, Sklearn, Plotly, Dash, Matplotlib