Hongling Lei

honglinglei98@gmail.com | +1 (412) 805-8510 | Website | GitHub | LinkedIn

EDUCATION

Carnegie Mellon University Pittsburgh, PA
Master of Information Systems Management, Business Intelligence and Data Analytics. GPA 3.93 12/2022

- Coursework: Deep Learning, Computer Vision, Machine Learning, Big Data and Large-Scale Computing, Unstructured Data Analytics (NLP), Distributed Systems, Object-Oriented Programming, Data Structures

Xiamen University Xiamen, China
Bachelor of Economics, Finance. GPA 3.81 06/2020

- **University of California, Berkeley** | *Semester Exchange*. GPA 3.95
- **Nanyang Technological University, Singapore** | *Summer Exchange*. GPA 5.0

PROFESSIONAL EXPERIENCE

PPG Industries Pittsburgh, PA
AI/ML Intern, AI/ML Center of Excellence 05/2022 – 08/2022

- Constructed 15+ time-series forecasting models including ARIMA, Exponential Smoothing and Prophet to predict finished goods demand for inventory optimization, improving the status-quo forecast accuracy by 17%
- Developed a Python-based Auto-Forecaster that takes any time series, experiments with 10+ forecasting algorithms, self-tunes hyper-parameters and recommends the best model, accelerating the speed of iteration by 1000 times
- Derived key drivers of demand fluctuations with explainable AI (XAI) techniques like SHAP and LIME

Tencent Shanghai, China
Data Scientist Intern, Public Data Science Department 04/2021 – 07/2021

- Conducted causal inference with algorithms like Causal Bayesian Networks, X-Learner and Causal Forest to analyze reasons behind customer behavior on Tencent's streaming platforms, increasing user satisfaction rating by 11%
- Deployed an inference pipeline that automates feature engineering, machine-learning modeling and future interventions, putting everything into production for practical decision-making with live data

DiDi Beijing, China
Data Analyst Intern, Decision Support Department 09/2020 – 11/2020

- Extracted and analyzed billions of data points on Apache Hive to support operational and strategic decisions
- Designed a market sizing model with SQL and Excel, correctly predicting driver and order growth rates during holidays and thus alleviating traffic burdens in 14 Chinese metropolitan areas
- Created and monitored business dashboards that can update data and visualize product metrics weekly

Bairong Technology Shenzhen, China
Data Scientist Intern, Financial Technology Department 05/2020 – 08/2020

- Built a semi-supervised learning model using the MixMatch algorithm with PyTorch to classify customers into different groups based on their probabilities of default, achieving 90% prediction accuracy
- Developed an automated report generation program to calculate performance indicators, graph statistical distributions, produce analytical summaries and create formatted slideshows for clients, boosting team efficiency

PROJECTS

Unsupervised Speech Recognition [Deep Learning | GANs] 04/2022

- Implemented unsupervised audio-to-text transformation with the wave2vec_U algorithm, enabling speech recognition for low-resource languages without sufficient training labels

Object Tracking System [Video Processing] 03/2022

- Constructed a target tracking system for both template matching and motion detection with the Lucas-Kanade method

Facial Recognition [Deep Learning | CNNs] 02/2022

- Built a face recognition application that achieved 82% accuracy on classification and 0.96 AUC on verification

Augmented Reality with Planar Homographies [Computer Vision] 02/2022

- Conducted real-time image and video AR projections through interest point matching and homography estimation

Grocery Master [Web Scraping] 10/2021

- Developed software that allows users to search for a product, shows available options at nearby grocery stores and compares their nutritional information by live-scraping Target, Walmart, and Trader Joe's websites

SKILLS

Programming Languages: Python, Java, SQL, R, Stata, HTML
Machine Learning: PyTorch, scikit-learn, sktime, OpenCV, pandas, NumPy
Big Data: Database (MySQL, Hive), Cloud (Google Colab, Azure, AWS), Distributed Computing (Spark, Hadoop, MLlib)