Junchen (Mark) Jin

734-882-7091 | kinkunchan@outlook.com | linkedin.com/in/junchen-jin/

EDUCATION

Northwestern University 12/2022

M.S. in Analytics GPA: 4.00/4.00

- Courses: Database & Information Retrieval (SQL), Big Data, Analytics Value Chain (A/B Testing, Docker, AWS, Flask), Natural Language Processing, Deep Learning, Reinforcement Learning

Shanghai Jiao Tong University 08/2020

B.S.E. in Electrical and Computer Engineering (Dual Degree) GPA: 3.85/4.00

University of Michigan, Ann Arbor 12/2019

B.S.E. in Computer Science (Dual Degree), Summa Cum Laude, Minor in Statistics GPA: 4.00/4.00

- Related Courses: Machine Learning, Deep Learning, Data & Graph Mining, Web Systems, Probability and Statistics

SKILLS

ML Toolkits: PyTorch, TensorFlow, scikit-learn, pandas, NetworkX **Visualization:** Tableau, Matplotlib, Plotly, ggplot, Bokeh

Big Data: Hive, Spark, Pig, Hadoop, MapReduce **Programming:** Python, R, SQL, C/C++, Java, HTML/CSS, JavaScript

PROFESSIONAL EXPERIENCE

PayPal - San Jose, CA 02/2023-Present

Machine Learning Engineer II

- Building ML/AI solutions for PayPal consumer personalization and recommendation

PayPal - San Jose, CA 06/2022-09/2022

Data Science Intern – Machine Learning Scientist

- Piloted in refreshing the debit card counterfeit detection model, with new business background and workflow from business partner
- Extracted, transformed, and loaded (ETL) the training set with 70 million records using BigQuery, Teradata, Hive and Spark
- Delicately designed sampling strategies and training weights to prevent potential model bias and mitigate class imbalance issue
- Selected 300 candidate features from 15,000 using WoE, IV, PSI, and iterative backward selection with sensitivity analysis
- Distributedly trained 20+ benchmark models with modeling tricks and various architectures, resulting in 6% enhancement in recall

TransUnion - Chicago, IL 10/2021 – 06/2022

Part-time Contractor - Data Science Consultant

- Led the practicum team to initiate, develop, prototype, and document automatic drift detection and model refitting solutions
- Researched and coded drift detection algorithms, including evaluation of uni- & multi-variate data stability (PSI, autoencoder, etc.), concept stability (KS, GINI, etc.), and stability measurement (DDM); applied to fraud detection and online streaming use cases
- System designed, implemented, tested, and released an `R` package for drift detection, internally used in TransUnion analytics teams
- Prototyped automatic performance monitoring, drift detection, and model retraining pipeline using `R targets` and `Jenkins`

CyberInsight Technology - Beijing, China 05/2018 – 08/2018

Data Scientist & Analyst Intern

- Architected and landed end-to-end SaaS solutions for windfarms to monitor wind turbine status and perform failure prediction
- Researched and implemented time-series failure prediction models for wind turbines: innovatively adopted TF-IDF for feature extraction on time-series data; built ensemble model using XGBoost and LSTM (PyTorch); model foresees risk 90 days beforehand
- Improved wind turbine yaw alignment model with Iso-Forest; reduced misalignment by 30%; published at PHMAP conference
- Designed and visualized wind turbine status and ML model outputs (Plotly.js) on a web-based interactable dashboard (JavaScript)
- Automated model running and re-training with Java; deployed to production on 100+ wind turbines with Docker on Azure cloud

RESEARCH EXPERIENCE

Graph Exploration and Mining at Scale (GEMS) Lab - University of Michigan, Ann Arbor 01/2019 – 10/2021

Research Assistant supervised by Prof. Danai Koutra

- Led an empirically study on 11 structural embedding methods about their functionality and limitation on unstructured graph data
  - Proposed and verified the feasibility of novel supervised and unsupervised embedding evaluation methods referring to NLP
  - Created Python package (`GEMS-SEMB`) for effortless embedding methods code integration and efficient experiment pipelining
- Contributed in a robustness study on Graph Neural Networks (GNNs) with heterophily-inspired designs under adversarial attacks
  - Proved theoretically heterophily-inspired designs lead to stronger robustness; integrated designs into source code of 4 prevailing GNNs; empirically showed an up to 18% accuracy increase under attacks, with a self-designed parallelized experiment pipeline