Dr. Yu-Ting Shen

yu-ting-shen-6b730b160 ytshen2020@gmail.com
ytatus94 (405).200.2633

EXPERIENCE

Senior Data Scientist **Seeloz Inc**, **San Jose, CA** **2019/04 - present**

- Analyzed large supply chain datasets using **SQL** and **BigQuery** to identify root causes of business inefficiencies and performed data cleaning and ETL using **Python** and **Pandas**
- Generated data visualizations and prepared analytical reports using **Jupyter**, **Matplotlib**, **Seaborn**, and **Plotly** packages which improved data visibility and supply chain understanding
- Built predictive models for demand forecasting using **Scikit-Learn**, **XGBoost**, **LightGBM**, and **TensorFlow** and achieved $r^2 = 0.92$ which facilitated operational planning
- Implemented **Reinforcement Learning** models to automate supply chain procurement and reduce cost by ∼30% through inventory level and stock outs optimization
- Automated machine learning training and inference workloads with Linux Shell and **Docker** using cloud platforms such as Azure ML Studio, Google AI Platform
- Developed internal Python libraries to facilitate data ETL across cloud and on-prem platforms such as **GCP**, **Azure**, **AWS** which improved developer efficiency by 3x and minimized security risks
- Created dashboards using **Google Data Studio**, **Tableau** and **Power BI** in collaboration with engineering, product, and executive teams which show supply chain metrics for stakeholders to help decision making
- Implemented industry standard procurement models including **EOQ**, **TPOP**, and **ROP** to compare and contrast behavior with trained AI models and reduced model evaluation time by 2x

Data Scientist **CERN**, **Geneva, Switzerland** **2015/03 - 2018/03**

- Improved the electron isolation efficiency from 83% to 99% (a 19% increase) which set a new benchmark for all analysis at CERN
- Designed, optimized, and implemented a high-performing **classification model** for real leptons across multiple energy scales, leveraging both statistical and machine learning methodologies. This resulted in a significant improvement in the model's recall, which increased from 62% to 98%.
- Conducted a comprehensive analysis of an extensive 400 TB dataset and employed **decision tree**, **multi-dimensional regression**, and **statistical models**, to deliver sophisticated solutions that effectively addressed complex project requirements.

Research Scientist **Academia Sinica**, **Taipei, Taiwan** **2009/07 - 2011/07**

- Created a **Monte Carlo simulation** model using **C++** and increased 20% precision

R&D Engineer **TSMC**, **Hsinchu, Taiwan** **2006/12 - 2009/02**

- Performed rigorous **statistical analysis** to develop models for advanced IC devices with cutting-edge technology

SKILLS

- **Programming**: Python, SQL, C/C++, Spark, Bash Shell script,
- **Machine learning, Deep learning**: Scikit-learn, Keras, TensorFlow, PyTorch,
- **Reinforcement learning**: Gym, Stable-Baselines, Ray,
- **Visualization**: Matplotlib, Seaborn, Dash, Bokeh, Google DataStudio, Tableau, Power BI,
- **Cloud**: Google Cloud Platform, Microsoft Azure,
- **Others**: Git, Docker, Jupyter, Databrick, Visual Studio Code, Jira,
- **Soft skills**: Collaboration, Communication, Problem-solving, Leadership,

EDUCATION

Ph.D. in Physics **University of Oklahoma**, **Norman, OK** **2011 - 2018**