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 supply chain data from ERP system using **SQL** and **Python**, and create data ETL pipeline using **PySpark**.
- Visualized supply chain trends and inefficiencies using **Matplotlib**, **Seaborn**, **Bokeh**, and **Plotly**.
- Created interactive dashboards using Python **Dash** and deployed on **Azure** using **Docker**. Experienced and proficient in **Tableau**, **Google DataStudio**, and **Power BI**.
- Implemented **Reinforcement Learning** models to optimize supply chain management and inventory control. Lowered annual inventory by 26% and increased annual turnover rate by 44%.
- Built **anomaly detection** models for customers' orders and found 3% $\sim$ 5% anomaly across products.
- Built **predictive models** to forecast customer's demand using ETS, ARIMA, Exponential-Smoothing (**statsmodels**), and **Prophet**. Achieved model accuracy ranging from 83% to 97% across products.
- Improved 3% $\sim$ 17% accuracy in **time series forecasting** model results for different products by using RandomForest (**scikit-learn**), **XGBoost**, **LightGBM**, and LSTM (**TensorFlow**).
- Developed a Python API offering cross-platform capabilities for accessing storage blobs on **GCP**, **Azure**, and **AWS**.
- Developed a **Bash shell script** for submitting model training jobs to Google AI platform, Azure VMSS, and on-premises cluster.

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

- Improved the electron isolation efficiency from 83% to 99% (a 19% increase) by restricting the transverse energy and momentum distributions within a topological cone of 0.2 in spherical coordinate. The outcome 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**