Dr. Yu-Ting Shen

yu-ting-shen-6b730b160 ytatus94@yahoo.com.tw
ytatus94 (405).200.2633

SKILLS

- **Programming language**: Python, C/C++, Bash shell script, Scala,
- **Machine learning**: scikit-learn, Apache Spark, Keras, TensorFlow, PyTorch,
- **Reinforcement learning**: Gym, Stable-Baselines, Ray,
- **Database**: SQL, Big Query, PostgreSQL, TablePlus, Incorta,
- **Visualization & Dashboard**: matplotlib, seaborn, bokeh, DataStudio, Tableau, Power BI
- **Big Data**: Apache Hadoop, Hive, Cloudera
- **Cloud**: GCP, Azure,
- **Others**: Git, Jira, Docker,

EXPERIENCE

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

- Developed and maintained the essential deep-Q learning network (DQN) models in SCAS Autonomous Procurement & Inventory (AP&I) platform to optimize profitability and minimize supply chain inefficiencies. The annual total inventory values are reduced from $7.82M to $5.77M (26% lower) and the annual turnover rate is increased 44% from 15.42 to 22.27.
- Reduced the inventory levels 30% $\sim$ 70%, which varies by warehouses and products, and retained low stock-outs by introducing the purchase-procurement splitting and postponed action methods into the model-based reinforcement learning (RL).
- Migrated the AP&I platform from Stable-Baselines to Ray.
- Built interactive dashboards for visualizations and quick inspections. The dashboards were constructed on the DataStudio at the beginning, and then ported to Bokeh now.
- Implemented an abstraction layer on top of Azure and GCP to access the Azure storage account and Google Cloud storage. This layer provides the AP&I cross-platform functionalities.
- Designed an universal interface to submit batch jobs to Google AI platform, GKE, Azure VMSS, Azure ML, AKS, on-premises cluster, and local docker container.
- Applied the time series analysis and forecasting methods to predict customers' demands and orders. The models been used including ETS, ARIMA/SARIMA, VAR, long-short term memory (LSTM), and double random forest (double RF). The r^2 score was improved from 0.15 (using ETS model) up to 0.92 (using double RF) and the number of supported products have been increased by a factor of 20 times.

CERN (Organisation Européenne pour la Recherche Nucléaire) **Geneva, Switzerland**
Data Scientist *2015/03 - 2018/03*

- Improved the electron isolation efficiency from 93% to 98% by introducing the momentum distributions in spherical coordinate. This results became a new standard for all analysis at CERN.
- Analyzed 400 TB data from the LHC computing Grid, built decision tree and regression models, applied statistical methods to extract the signal within 95% confidence interval.

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

- Developed a new Monte Carlo simulation program in C++ and GEANT4 for germanium detector and implemented decision tree models in C++ and ROOT for particle classifications with an accuracy of 96%.

TSMC (Taiwan Semiconductor Manufacturing Company) **Hsinchu, Taiwan**
R&D Engineer *2006/12 - 2009/02*

- Improved 40% of the performance by creating and deploying on-prem analysis pipeline, which includes high level data cleaning, engineering, visualization, statistical model building, and MC simulation.

EDUCATION

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