Dr. Yu-Ting Shen

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

SKILLS

- **Programming**: Python, SQL, Spark, C/C++, Bash shell script, Scala, VBA
- **Machine learning**: scikit-learn, Keras, TensorFlow, PyTorch, time series, anomaly detection, NLP, recommendation system
- **Reinforcement learning**: Gym, Stable-Baselines, Ray,
- **Visualization & Dashboard**: matplotlib, seaborn, dash, bokeh, DataStudio, Tableau, Power BI
- **Big Data**: Apache Hadoop, Hive, Cloudera
- **Cloud**: GCP, Azure, AWS
- **Others**: Git, Jira, Docker, Jupyter Notebooks, Jupyter Lab, Visual Studio Code,, anaconda

EXPERIENCE

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

- Developed and maintained the essential deep-Q learning network (DQN) models for inventory control 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).
- Applied the time series analysis and forecasting methods using Python and predicted 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.
- Analyzed supply chain data from various clients by writing SQL queries on the Google Big Query and PostgreSQL.
- Built interactive dashboards for visualizations using Google DataStudio and Python Plotly Dash.
- Implemented an abstraction layer (API) on top of 3 major cloud platforms (GCP, Azure, and AWS) to access the cloud storage. This Python-based API provides the AP&I cross-platform functionalities.
- Designed a Python-based universal interface to submit batch jobs to Google AI platform, GKE, Azure VMSS, Azure ML, AKS, on-premises cluster, and local docker container.

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. The study was done using C++. This results became a new standard for all analysis at CERN.
- Analyzed 400 TB data from the LHC computing Grid using C++ and Python, 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.
- Implemented, trained, tested, and tuned hyperparameters of decision tree and boosted ensemble models, which are written in C++, for particle classifications with accuracy ranging 96% to 99% depending on the particle types.

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 ETL and analysis pipeline, which includes high level data cleaning, data engineering, visualization, statistical model building, and Monte Carlo simulation using C++, Bash shell script, Perl, and Excel VBA macros.

EDUCATION

University of Oklahoma   **Norman, OK**
Ph.D. in Physics   *2011/08/22 - 2018/05/11*