





Dr. Yu-Ting Shen

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

SKILLS

- **Programming:** Python, SQL, C/C++, Spark, Scala, Bash shell script, VBA,
- **Machine learning, Deep learning, Reinforcement learning:** Scikit-learn, Keras, TensorFlow, PyTorch, Gym, Stable-Baselines, Ray,
- **Analysis & Modeling:** ETL, EDA, Predictive modeling, Time series forecast, Anomaly detection, Monte Carlo simulation, RandomForest, XGBoost, LightGBM, LSTM, clustering,
- **Database:** SQL, Big Query, PostgreSQL, Incorta,
- **visualization:** Matplotlib, Seaborn, Dash, Bokeh, Google DataStudio, Tableau, Power BI,
- **Cloud:** Google Cloud Platform, Microsoft Azure, Amazon AWS, Cloudera,
- **Others:** Git, Docker, Anaconda, Jupyter Notebooks, Jupyter Lab, Databrick, Visual Studio Code, IntelliJ, Jira,
- **Soft skills:** Collaboration, Communication, Problem-solving, Leadership,

EXPERIENCE

Seeloz Inc

Senior Data Scientist

San Jose, CA

2019/04 - present

- Developed and maintained a reinforcement learning model using deep-Q learning neural network (DQN) to optimize inventory control and supply chain efficiency for a leading auto parts manufacturer in Malaysia. This model lowered the client's annual total inventory by 26% from \$7.82M to \$5.77M and increased annual turnover rate by 44% from 15.42 to 22.27.
- Utilized SQL, Python, and Spark to analyze supply chain data and developed ensemble models using ETS, ARIMA, RandomForest, XGboost, and LSTM for time series forecasting. Achieved model accuracy ranging from 83% to 97% across products and boosted coverage of products by 10 times
- Designed a novel approach for tackling non-convergence issues in sparse time series data by leveraging a double random forest model with postponed action and accomplished a 68% reduction in MAE.
- Developed an anomaly detection model to uncover 3-5% anomalies in customers' ordering. Utilized this model in data engineering to improve accuracy by 9.8% in a time series forecasting model.
- Developed multi echelon single product reinforcement learning model for the world's largest oil company to optimize its oil transportation scheduling plan, reduce inventory costs, and fulfill customer demands more efficiently.
- Developed a Python-based API that offers cross-platform capabilities for accessing cloud storage on GCP, Azure, and AWS. The API was adopted in all Seeloz products and significantly enhanced workflows' efficiency and functionality.
- Created dynamic dashboards with Google DataStudio and Python Dash package, enhancing accessibility and usability of critical data insights for stakeholders.
- 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.
- Collaborated cross-functionally to design and implement sophisticated analytical solutions at scale, delivered high-quality outcomes that exceeded project requirements and met stakeholder needs.
- Led junior data scientists on the team and facilitated resolution of project-related challenges.

CERN (Organisation Européenne pour la Recherche Nucléaire)

Data Scientist

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. Employed various modeling techniques, such as decision tree, regression, and statistical models, to deliver sophisticated solutions that effectively addressed complex project requirements.

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

- Created a Monte Carlo simulation model using C++ and GEANT4 that resulted in a 20% increase in precision.
- Developed decision tree and boosted ensemble models using C++ and ROOT, and applied 3-fold cross-validation to enhance particle classification accuracy by approximately 10%.
- Led cross-functional collaboration between different departments to establish remote control capabilities for experiments.

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

- Performed rigorous statistical analysis to develop models for advanced IC devices with cutting-edge technology.
- Achieved a 40% improvement in performance by designing and deploying an ETL and analysis pipeline on-premises.
- Managed 7 projects by defining project scopes, allocating resources, scheduling and monitoring project progress, and coordinating cross-functional teams to ensure timely and high-quality project delivery.

EDUCATION

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