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 forecast, 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, A/B test, Monte Carlo simulation

EXPERIENCE

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

- Developed and deployed a deep-Q learning neural network (DQN) in reinforcement learning, which optimizes the supply chain management and provides recommendations to customers. This DQN model helps customers saving $2.05M in the inventory costs and increasing 44% annual turnover rate.
- Analyzed customers' supply chain data using SQL/Python/Spark and built ensemble models for time series forecasting using ETS/ARIMA/RandomForest/XGboost/LSTM. The model accuracy ranges from 93% to 100% for various products for different customers.
- Built interactive dashboards for visualizations using Google DataStudio and Dash, which is a Python package from Plotly. The dashboards not only visualize data but also summarize key metrics, present prediction results and analytical outcomes.
- 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.
- Mentored less experienced data scientists in the team and helped them to fix issues in projects.

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.
- Developed, tuned, and deployed a classification model for leptons in different energy scales using statistical and machine learning methods. The model retained the recall 98% and reduced 20% of the false positive rate.

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*