Drew Yang

Phone: 214-727-9608 | **Email:** drew.yang.dev@gmail.com | **Working at:** Houston, TX | **Github.io:** https://yambottle.github.io/me/

Technical Skills

- **Programming:** Python, Java, R, C/C++, HTML/CSS, JavaScript, SQL, Groovy, bash shell, powershell, MATLAB
- **Python:** pandas, scikit-learn, keras, tensorflow, pyspark, matplotlib, Plotly Dash, flask, celery, sqlalchemy, socket
- **Storage/Cache:** SQL Server, MySQL, MongoDB, Redis, RabbitMQ
- **Pipeline:** Airflow, MLflow, Jenkins, Azure Pipeline
- **Deployment:** cythonize, gunicorn, Nginx, Docker, Terraform, SaltStack, Kubernetes(kOps), helmchart
- **AWS:** VPC, EC2, RDS, S3, Route53, LoadBalancer, CloudFormation, CloudWatch, Tag Editor
- **Azure:** Azure SQL, Storage, Data Lake Gen 2, Data Factory, Data Explorer(Kusto), Azure Function

Experience

Software Engineer **July 2021 - Present**
DataJoint - Neuroscience/ScienceOperation *Houston, TX*

* ***SciOps Platform DevOps:*** SciOps enables research teams to organize and automate data operations. I was assigned to work on **AWS** infrastructure provisioning and deployment using **Terraform** modules, **SaltStack** states and **Kubernetes kOps**
* ***SciOps MATLAB Worker Deployment:*** MATLAB worker is a part of SciOps platform that enables **GPU**. I focused on building a MATLAB **docker**, making a **docker-cuda** environment and deploying it on an **EC2** instance with GPU.
* ***Online Workshop on JupyterHub:*** This is a **week-long** online workshop that provides a jupyter notebook environment for **each** audience to complete several coding sessions. I worked on setting up **JupterHub** using **Kubernetes kOps** and **helmchart** on AWS. I also developed a JupyterHub load tester using **Selenium**.

Data Scientist **May 2019 - July 2021**
dataVediK- Oil & Gas *Houston, TX*

* ***Interactive Drilling Dashboard:*** This is an **enterprise** product that I worked with two more engineers. Developed a **Plotly Dash** dashboard that visualizes processed data using Bootstrap, CSS media query, **Redis** and sqlalchemy. Also, implemented a **socket** service will notify when **Airflow** pipeline finished processing in order to **synchronize**(refresh) the dashboard's data.
* ***CI/CD Pipeline:*** Set up several **Azure Pipelines** for continuous development, testing and continuous deployment in **dev, test and prod** stages. Additionally, made a **Jenkins** pipeline to work with on-premise infrastructures.
* ***ML Pipeline:*** Set up a **MLflow** server for machine learning experiment logging, parameter tuning, continuous training, model management and model serving.
* ***ETL Pipeline:*** Working with a data engineer, set up an **Airflow** server for our data ETL pipeline.
* ***Prediction Task Manager:*** Working with a front-end developer, designed and developed a **production** web application that supports job queuing and parallel processing for drilling speed prediction using JavaScript, **flask**, sqlalchemy, **celery**, RabbitMQ, gunicorn, Nginx, supervisord, Docker and AWS EC2, AWS Cognito Authentication, HTTPS
* ***Drilling Status Detection:*** Working with a domain expert, developed two **classification** models for detecting drilling status using Logistic Regression and Random Forest with the convenience of the MLflow server
* ***Drilling Speed Prediction:*** Working with a domain expert, applied Gaussian Process **Regression** for feature synthesis based on geographical information as well as **feature engineering** based on correlation matrix and F1 score ranking, built a non-linear regression model using LSTM RNN.
* ***Image Classification:*** This is a short-term **client** project that I worked with a senior data scientist. Applied k-Means clustering to **help** manual data labeling, then made a **classification** model for oil pump failure detection using Random Forest and CNN.

Education

Southern Methodis University **Aug 2017 - May 2019**
Master in Computer Science *Dallas, TX*
Qingdao University **Aug 2013 - May 2017**
Bachelor in Software Engineering *Qingdao, China*