Drew Yang

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

Technical Skills

- **Programming Languages:** Python, bash, SQL, Markdown, Groovy, powershell, Java, R, HTML/CSS, JavaScript, C/C++
- **Continuous Integration/Deployment:** Github Actions, Azure DevOps, Jenkins, ArgoCD
- **Infrastructure as Code:** AWS, Azure, Terraform, Packer, Ansible
- **Configuration Management:** cloud-init, SaltStack
- **Containerization:** Docker, Podman, Kubernetes(kOps, kubectl, helm), Dev Container
- **Monitoring:** OpenTelemetry, CloudWatch, Datadog, Grafana
- **Workflow Orchestration:** Flyte, MLflow, Airflow

Experience

Software Engineer(DevOps) **July 2021 - Present**

DataJoint - Science Operation for neuroscientist. [Opensource] and [Commercial] *Houston, TX*

* ***CI/CD:*** Developed generic **Github Action** reusable workflows for **30+** repositories to automate Python packages and Docker images build, test, release/publish
* ***SciOps Kubernetes Cluster:*** Setting up and maintaining **Kubernetes** clusters for dev/test/production using **kOps and k3d**; configuring Nginx ingress controller, let's encrypt cert manager, Cillium Container Network Interface(CNI); deploying JupyterHub, flask, ReactJS, etc. using **helm** or **kubectl**
* ***SciOps Worker Clusters:*** Configuring **Terraform**, **cloud-init** and **SaltStack** to provision CPU and GPU(**Nvidia/CUDA**) workers in different scales as needed
* ***SciOps Operation Automation:*** Automating SciOps operations such as onboarding new customers, managing customers' permission, generating usage and billing report, etc. using **Python**, **SQL** and **bash**

Software Engineer(MLOps) **May 2019 - July 2021**

dataVediK - Optimize Oil & Gas operations by Machine Learning. [DrillVedik] *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.

Education

Southern Methodist University **Aug 2017 - May 2019**

Master's of Computer Science *Dallas, TX*

Qingdao University **Aug 2013 - May 2017**

Bachelor's of Software Engineering *Qingdao, China*