Drew Yang

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

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
- **Deployment/Pipeline:** setuptools, cythonize, Docker, Terraform, SaltStack, Kubernetes, Airflow, MLflow, Jenkins
- **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/SciOps *Houston, TX*

*** SciOps MATLAB Worker Deployment:** AWS MATLAB GPU docker

*** SciOps System DevOps:** AWS Terraform CloudWatch SaltStack docker

*** Online Workshop Event on JupyterHub:** AWS Kubernetes-KOps JupyterHub SeleniemTester docker

Data Scientist **May 2019 - July 2021**

dataVediK- Oil & Gas *Houston, TX*

*** Drilling Analytic Dashboard:** This is the first phase of a client project that I worked with a data engineer and another developer. Developed a web application that visualizing the processed data from the Airflow ETL pipeline by using Plotly Dash framework, CSS media query, Redis and sqlalchemy. Also, implemented a socket service pushing a notification from the Airflow server to trigger a process that can synchronize(refresh) the dashboard with the updated data at the end of the ETL pipeline.

*** CI/CD Pipeline:** Set up a Jenkins server working with Azure DevOps for continuous development, testing and continuous deployment. Additionally, made a Jenkins pipeline in Linux shell scripts to work with on-premise infrastructures and several Azure Pipelines in powershell to work with Azure resources.

*** 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 using Docker, miniconda and AWS EC2.

*** Prediction Task Web Manager:** Working with a front-end developer, designed and developed a production level 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 for optimizing training 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*