Tian Zhang

0tianzhang0@gmail.com | 213-373-0955 | Los Angeles, CA, 90007 | https://dukesky.github.io/app/index.html

WORK EXPERIENCE

Data Scientist @ JM Eagle | Los Angeles, CA Dec 2019 – now

Sale Quotation Prediction

- Understood business requirements from stakeholder, collected prediction related data, use ***Matplotlib, Pandas*** to visual and analysis data and preprocessed with ***feature selection*** methods to reduce irrelevant variables
- Built 4 separated ***XGBoost*** and ***Random Forest*** models to predict quotation price and suitable shipping plant, order by date, estimated shipping date based on 210,000 quotation data, price has an accuracy of 96%(***mape***), plant has an accuracy of 98.5%, date has a ***RMSE*** of 3.74 days
- Rolled up models into ***APIs*** (get data, train, predict) by ***Flask***, ***Dockerized*** files that could serve local or cloud
- Deployed ***Docker Container*** into ***AWS*** by ***Serverless*** framework with ***AWS API Gateway, Lambda*** to receive and manage web request, set ***SageMaker*** to train and update model daily
- Set up automatic test pipeline by ***Pytest***, deploy pipeline by ***CI/CD tools***(***Circle CI)***, error monitoring and notification by ***Sentry*** to simplify future update and maintaining
- Increased the prediction accepted rate to 85%, reduce hundreds hours of human resource cost weekly
- Tuned and updated model by applying new features and shrinking training period to face the dramatic price change caused by COVID, the prediction accepted rate return to 82% from a huge drop to 21%

Monthly Sales Forecast

- Collected 10 years products sales data to build a monthly demand forecast model by using ***DeepAR*** model, accuracy(***MAPE***) has improved 10% compared with previous manually forecast(65%)
- Deployed the forecast pipeline to AWS with usage of ***Lambda***, ***AWS Forecast***, ***Step Functions***
- Design ***REST APIs*** for training forecast model and sending recent forecast result to subscribers by email (***AWS SNS***) and set monthly forecast report training and sending to subscribers at beginning of each month

Machine Learning Engineer Intern@ AiTmed | Anaheim, CA May 2019 – July 2019

Online Trainable Image Classification Platform

- Built an 8-layer ***MobileNet*** model by ***TensorFlow***, with a test accuracy of 85% in *Stanford cat and dog* dataset
- Designed ***APIs*** of saving, loading model, building, training model, reporting model performance and predicting by using ***Flask*** while all data and models are stored and can be exported from AWS S3
- ***Dockerized*** the platform and served at local, allow users to train and test new models with new uploaded images

SKILLS

Skill: MLOps, Data Analysis, Visualization, Image Process, NLP, Forecasting, Recommendation System, AWS
Software | Framework: Docker, Git | Serverless, TensorFlow,Pytorch, Flask, CI/CD
Coding Language | Tool: C/C++, Python, SQL, Java | Jupyter Notebook, MySQL, Android Studio, VS Code

PROJECTS

(Deep Learning, Image Processing) Fast Super-Resolution CNN for Human Image

- Imported ***MobileNet*** into ***Fast Super Resolution CNN(FSRCNN),*** reduced model parameters by 65%
- Maintained the image resolution **(PSNR:31.9 SSIM:0.858),** reduced 30% of image generation time

(Unsupervised Learning, NLP) Copycat App Detection

- Used ***NLTK*** to extract nouns and verbs from 40,000 App descriptions, vectorized each app by ***bag-of-words*** model
- Applied ***TF-IDF*** and ***PCA*** to extract top 10% features, applied ***Hierarchical Clustering*** to find Copycat Apps.
- Detection of copycat Apps in a designated threshold (top 50 similar Apps) has an average accuracy of 83%

(Deep Learning, NLP) Sentiment Analysis Web App

- Downloaded 50000 user reviews from ***IMDB datasets***, used ***BeautifulSoup*** and ***NLTK*** to tokenized each review into a fixed size vector by ***bag-of-words*** model, apply ***TF-IDF*** to preprocess data
- Built a 5-layer ***LSTM*** model by ***Pytorch***, used ***SageMaker*** to train/test/deploy model with 73.5% final accuracy
- Set up ***Lambda*** and ***API Gateway*** to update model endpoint to web page

EDUCATION

University of Southern California **Los Angeles, US** **Master :Electrical Engineering :** 3.9/4.0 2017 – 2019
Beihang University **Beijing, China** **Bachelor: Electrical Engineering:** 3.5/4.0 2013 – 2017