Tian Zhang

0tianzhang0@gmail.com | 213-373-0955 | 1603 W 25th ST, LA, CA | https://dukesky.github.io/app/index.html

WORK EXPERIENCE

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

Auto-Trained Quotation Prediction Platform

- Understood business requirements from stakeholder, collected pricing, plant, date related data, use ***Pandas, Matplotlib*** to visual and analysis data and preprocess with several ***feature selection*** methods
- 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 ***RESTful APIs*** (collect data, train, predict) with ***Flask***, used ***Docker*** to package code
- 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 based on new uploaded data in ***S3***
- Set up auto test (***Pytest***), deploy work flow (***Circle CI***), error monitor (***Sentry***) to maintain the prediction platform
- Increased the quotation prediction accepted rate to 85%, greatly reduce human resource cost on making quotes
- Tuned model by applying new features and shrinking training period to face the dramatic price change caused by pandemic, the quotation prediction accepted rate return to 82% from a 61% drop to 21%

Inventory Forecast System

- Collected 10 years sales data to build a ***time series*** demand forecast model by using several demand forecast methods, accuracy(***MAPE***) has improved 10% compared with previous manually forecast MAPE(395%)
- 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 sent to subscribers at beginning of each month

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

Deploy Online Trainable Machine Learning Platform for Image Classification

- Built an 8-layer ***MobileNet*** model by ***TensorFlow***, with a test accuracy of 85% in *Stanford cat and dog* dataset
- Designed ***REST API*** of saving, loading model, building, training model and report model by using ***Flask*** while all data and models are stored and can be exported from AWS S3
- Packaged trainable model into ***Docker Container***, allow users to train and test by designed API

SKILLS

Skill: Data Analysis, Visualization, Deploy Machine Learning Model, Image Process, Android Develop, Web design
Software | Framework: Docker, Git | Serverless, TensorFlow, Flask, AWS(S3, Lambda, API Gateway, SageMaker)
Coding Language | Tool: C/C++, Python, SQL, Java | Jupyter, MySQL, Android Studio, VS Code

PROJECTS

(Deep Learning, Image Processing) Fast Super-Resolution CNN for Human Image March 2019 – May 2019

- 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 March 2019 – April 2019

- 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) Spoken Language Identification January 2019 – March 2019

- Analyzed audios records from 3 languages, generated features in every 10ms based on MFCC from ***Librosa***.
- Built a 5-layer ***LSTM RNN*** model, came with a prediction accuracy of 73.5% in 10 seconds audio fragment.

(Data Analysis, Regression) News Popularity Prediction Oct 2018 – Nov 2018

- Implemented ***ANOVA*** test and ***forward feature selection*** to test reliability and found the most influential features.
- Developed ***Logistic Regression*** model on SPSS, increased final accuracy by 34% compared to by-chance model

EDUCATION

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