Tian Zhang

0tianzhang0@gmail.com | +1(213)373-0955 | 1247 W 30th ST, LA, CA | https://dukesky.github.io/app/index.html

WORK EXPERIENCE

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

  Deploy Auto-Trained Quotation Price Prediction Platform on AWS
- Built a *XGBoost* model to predict quotation price and suitable shipping plant based on 110,000 quotation data
- Rolled up models in *RESTful APIs* (collect data, train, predict command) with *Flask*, used Docker to package code
- Deployed *Docker Container* into *AWS*, setup server in serverless framework by using *AWS Lambda* to manage web request and *AWS SageMaker* to train and update model daily based on new uploaded data in *AWS S3 bucket*
- Set up auto deploy workflow by *Circle CI* and error monitor by *Sentry* to maintain the prediction platform

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

  Deploy Online Trainable Machine Learning Platform for Image Classification
- Built a 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
- Packed trainable model as a *Docker Container*, allow users to train and test by designed API

  Android App Development of Vital Sign Transmission from Bluetooth Medical Device
- Implemented data transmission of several medical device by Bluetooth, coded in Android Studio by *Kotlin*

Software Develop Engineer - Intern@ Digital China | Beijing, China   May 2016- August 2016
- Implemented and tested activities included registration and login account, adding friends in web page by *Java*

SKILLS

Skill: Data Analysis, Visualization, Deploy Machine Learning Model, Image Process, Android Develop, Web design
Software | Framework: Docker, Git | TensorFlow, Pytorch, Flask, AWS(S3, Lambda, EC2, ECR, SageMaker)
Coding Language | Tool: C/C++, Python, SQL, Java | Jupyter, MySQL, Android Studio

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%
- Extracted features from human face images by pretrained model, imported these features as inputs into dense layer.
- Maintained the image resolution **(PSNR:31.9 SSIM:0.858),** meanwhile 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 *TD-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 within 3 type of languages, generated features in every 10ms based on MFCC from *Librosa*.
- Built a 5-layer *LSTM* based *RNN* model, came with a prediction accuracy of 73.5% in 10 seconds audio fragment.

(Deep Learning, Image Processing) Facial Emotion Detection   Jan 2019 – Mar 2019
- Built a 16-layer *VGGNet CNN* model by *Keras* to classify 11 imbalanced classes of 13,000 emotion facial images
- Trained on *AWS EC2* after oversampling minority classes with an accuracy of 84.5%(1 top) and 95%(2 top)

(Supervised Learning, Regression) Movie Revenue Prediction   Nov 2018 – Dec 2018
- Analyzed more than 40,000 movies from IMDB, vectorized information into numerical value, applied *semi-supervised KNN* to fill missing data, Applied PCA to extract important features.
- Trained models by several regression methods, tuned parameters by *cross validation*, chose the best prediction model (*Random Forest*) with an average R^2 score of 0.67(increased 0.2 from logistic regression).

(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

M.Sc University of Southern California, US	**Electrical Engineering :** 3.9/4.0		2017 – 2019
B.S Beihang University, China	**Electrical Engineering:** 3.5/4.0		2013 – 2017