

Hoang Le

Needed to write code everyday

Email: hoangle2806@gmail.com • <https://github.com/hoangle2806> • Street Address: San Jose, California (Open to Relocation) • Phone: 408-666-0462

Goal

A versatile and analytical data scientist looking for an entry level in software with a growing company where I can leverage programming skill and machine learning.

Highlights

- Learned fast, adapted quickly, and willing to work on a range of software-related jobs.
- Very ambitious to apply my data skill into diverse industries.
- Showed my interest in data field by completing specialization from University of Washington.
- Few data-science applications available on Github

Technical Profile

Languages/Scripting SQL and Python

Operating system Linux, Mac OS, and Windows.

Tools/Framework AWS, Django, pandas, numpy, nltk, keras, and scikit-learn

Work Experience

Siri development, Apple

Software QA

(2016)

Quality control for about 30 versions of Siri's AI to recognize English and solely responsible for Siri's localization.

- Ensured new version of Siri's AI able to recognize variations in voice commands by creating many test cases
- Created a training dataset for Siri with 3 million data points. Contributed in development of localized Siri

R&D lab, Intel Corp

Lab Assistant

(2016)

Analyzed samples from 5 suppliers using Excel with linear regression to assure quality compliance.

Applications

Goal: data product used to help social media managers to predict virality of social media posts

- Collected text and number of responses from Facebook and Twitter API using json and requests
- Transformed / processed large data set into intermediate .csv file with pandas, re, and nltk
- Final data set was loaded on local SQL database using SQLAlchemy
- Applied Word2Vec model to vectorize text with gensim and numpy
- Researched machine learning models: linear regression, LSTM using sklearn and keras
- Defined two metrics to evaluate model: MAE and F1-score
- Put together a friendly UI for the ease of use, built with Django.
- Supporting technology for this application: version control (Github), visualization (Bokeh), automation (crontab), Linux, Jupyter notebook, and data pipeline (luigi)

Mini Applications

- Analyzed restaurant's reviews of Yelp dataset to predict sentiment based on text using machine learning model logistic regression and n-grams with sklearn and gensim.
- Predicted housing price from Zillow housing dataset using decision tree and regularized linear regression.
- Classified food images with Keras framework using CNN approach.

Education

M.S. Chemistry (3.4/4.0)

University of California, Riverside

(2015)

Machine learning certificate

University of Washington

(2016)

U.S. Citizen. No visa sponsorship required.