

Linh Hoang Truong

CCA175 Cloudera Spark and Hadoop Certified Developer

★ <https://mrthlinh.github.io/>

✉ mrthlinh@gmail.com

in truonghoanglinh

🌐 mrthlinh

☎ 4694078033

EDUCATION

University of Texas at Dallas, Richardson, TX

MSc Computer Science with concentration on **Data Science**

Jan 2017 - Dec 2018

Danang University of Technology, VN

BSc Electronics and Communication Engineering

Sept 2010 - June 2015

SpringBoard Data Science Program

6-month intensive program in Data Science

Apr 2018 - Oct 2018

EMPLOYMENT

DATAECONOMY, Data Engineer, Dublin, OH

Jan 2019 - Present

- Installation and maintenance of Hadoop cluster in AWS, building a Pipeline for a DataEconomy client which can handle twitter data and extracting useful information from it using Spark Streaming.

- Using HDFS, Sqoop, Python, Spark, Hive, Flume to handle data ingestion, processing and consumption.

Superior Data Science, Data Scientist Intern, Dallas, TX

Aug 2018 - Nov 2018

- Developed production-level machine learning model to predict task completion time in company's flagship project management

- Delivered 8% reduction in number of task delayed over time.

- Worked with Flask Framework and Tensorflow Serving to deploy trained model

Erik Jonsson School of Engineering and Computer Science, Research Assistant, Richardson, TX

Jan 2017 - Apr 2018

- Applied Deep Learning to help software engineers writing code more efficiently, fixing bugs automatically, parameter recommendation, automatic type resolution of variables.

- Co-author of two accepted research papers in top-tier conference (acceptance rate around 20%)

Robert Bosch Engineering and Business Solutions, Software Engineer Intern, HCM city, Vietnam

Mar 2015 - July 2015

- Integrated Fuzzy Logic and Genetic Algorithm to Car Simulator in order to improve the accuracy of Antilock Braking System and Traction Control System by 5%

SKILLS

Coding: Python, Java, R, Javascript, Scala

Machine Learning: Scikit-learn, Keras, Tensorflow, PySpark, Deeplearning4J

Data Engineer: HDFS, Hadoop, Spark, Hive, Sqoop, Flume

Database: MySQL, MongoDB, HBase

OS: Linux, Window

Cloud Services: AWS, Google Cloud

PROJECTS

All code and presentations are available at my personal website <https://mrthlinh.github.io/>

Agile User Story Point Estimation (Tensorflow, Python)

Aug 2018 - Nov 2018

- Developed an end-end solution (Natural Language Processing) for effort estimation in Agile development without manual feature engineering

- Experimented with Word2Vec, Doc2Vec, Tf-idf for feature extraction and LSTM model for deep learning model.

- Average error of getting correct story point is 1.9, outperform manual feature engineering by 8%

FIFA World Cup 2018 Winner Prediction (Sklearn, Pandas, Numpy)

Apr 2018 - June 2018

- Predicted win / lose/ draw with 56% accuracy and 33% for goal difference for matches in the tournament.

- Experimented with many machine learning algorithms such as logistic regression, SVM, Random Forest, Neural Network, Gradient Boosted Tree.

Spotify Playlist Continuation Recommender (Pyspark, Sklearn, Pandas)

July 2018 - Oct 2018

- Developed a recommender system for automatic playlist continuation on 66 million ratings dataset.

- Compared performance between various type of recommenders, K-Nearest Neighbor and Collaborative Filtering on Pandas and PySpark

- Increased the chance of users getting their favorite songs in a playlist to 78%

RELEVANT COURSES

Statistical Methods for Data Science, Statistical and Machine Learning, Machine Learning, Big Data Management and Analytics, Database Design

PAPERS

- Co-author "Complementing Global and Local Contexts in Representing API Descriptions to Improve API Retrieval Tasks", *Foundations of Software Engineering (FSE) 2018 - Acceptance Rate (21%)*

- Co-author "Statistical Learning of API Fully Qualified Names in Code Snippets of Online Forums publication date", *International Conference on Software Engineering (ICSE) 2018 - Acceptance Rate (24%)*

CERTIFICATES

End-to-End Machine Learning with TensorFlow on GCP

June 2019

CCA175 Cloudera Spark and Hadoop Developer

Apr 2019

Deep Learning Specialization (Neural Network, Convolution Neural Network, Sequence Model) - Coursera

Dec 2018