Yu-Ting Shen

405.200.2633
YuTing.Shen-1@ou.edu
linkedin.com/in/yu-ting-shen-6b730b160
github.com/ytatus94

QUALIFICATIONS

- Ph.D. degree with experience in large-scale data analysis and machine learning to build predictive models.
- Proficient in Python, C/C++, SQL and shell script. Experienced in Keras, TensorFlow, Spark, and Hadoop.
- Passionate about discovering insights hidden in data in order to tell the story of data to non-experts.
- Excellent analytic and problem-solving skills through critical thinking, and teamwork with colleagues.

EXPERIENCE

CERN Researcher — Geneva, Switzerland | 2015/03 - 2018/02

- Established data processing pipeline to optimize the analysis processes and improve key performance metrics by 70% using decision tree algorithm via Python.
- Built regression model and applied statistical methods to extract signal within 95% confidence interval.
- Improve true positive rate from 0.85 to 0.93 by developing a new classification model for the electron isolation efficiency which became the standard version for the collaboration.

Academia Sinica Research Assistant — Taipei, Taiwan | 2009/07 - 2011/07

- Discovered serious bugs in simulation programs by performing exploratory data analysis on simulated data. Programmed in C++ to redesign bug-free Monte Carlo simulation programs.
- Improved 50% of the signal significance by devising a new classification model.

Taiwan Semiconductor Manufacturing Company R&D engineer — Hsinchu, Taiwan | 2006/12 - 2009/02

- Led and supported 7 projects of various device models and built model performance documents for clients.
- Coordinated effectively with cross-functional teams and clients to resolve issues with tight deadlines.

PROJECTS

Build recommender system using Yelp data challenge round 12 dataset

- Built a restaurant recommender system to recommend top 5 restaurants using item-item similarity based collaborative filtering and matrix factorization based on users' past visits and ratings.
- Applied natural language processing (NLP) and sentiment analysis techniques to classify the positive and negative reviews, being able to understand business performance based on users' reviews.
- Identified and understood the common user preference by clustering users into groups and inspecting the cluster centroid of each group.

Users churn prediction of online music streaming platform

- Scraped 35 GB dirty data from the Internet using Python BeautifulSoup package and created databases with comprehensive information to meet the needs for analysis.
- Used Spark to build data processing pipeline including data wrangling and feature engineering.
- Analyzed user behavior patterns and trends to get business insight and build users churn predictive model by leveraging both Sci-kit learn and Spark MLlib packages and achieved the AUC 0.90 and accuracy 0.82.

Sequence-to-sequence natural language generation with deep neural networks

- Collected large amount of dirty textual data from the Internet and processed data for the text generation.
- Used TensorFlow to construct a sequence-to-sequence model using RNN architecture with LSTM cells.
- Designed an end-to-end training procedure to achieve data-driven language modeling for the generation of poems by training more than 40,000 poems.
- Conducted parameters tuning and used GPU acceleration to obtain optimal speed for training the deep neural networks.

EDUCATION

University of Oklahoma Ph.D. in Physics | Cum. GPA: 3.47 / 4.00 — Norman, OK | 2011/08 - 2018/05

- Published 4 top journal papers and 5 conference notes, and presented in 2 conferences.
- 6 semesters teaching assistant for college and graduate level courses.