YIFEI WANG

Work Authorization: 36-months-OPT

@ yifei.wang828@duke.edu 984-245-5566 linkedin.com/in/yifei-wang-0249b5170/ ywang512.github.io/

EDUCATION

Duke University — Durham, NC

Master of Science in Interdisciplinary Data Science — 3.9/4.0 — Aug 2018 – May 2020

Relevant Coursework: Machine Learning, Bayesian Statistics, Deep Learning, Database Management, Data Visualization, Data Ethics, Computer Vision, Text Analysis, Cloud Computing

Sun Yat-sen University — Guangzhou, China

Bachelor of Science in Applied Mathematics — 3.7/4.0 — Aug 2014 – June 2018

Exchange Experience: University of California, Berkeley

Relevant Coursework: Multivariate Statistics: Data Mining, Artificial Intelligence and Neural Net, Complex Data Modeling, Data Structure and Algorithms

EXPERIENCE

Hearful Technologies, Inc. — Chapel Hill, N.C.

Data Scientist Intern — May 2019 – Aug 2019

- Analyzed the startup's business strategy and accordingly researched on Aspect-based sentiment analysis for online reviews.
- Independently implemented an Natural Language Processing algorithm in python from scratch, which is an augmented Latent Dirichlet Allocation for the simultaneously discovery of latent topics and associated sentiment polarity from scraped website.
- Collaboratively deployed tools to thoroughly analyze model outcomes and instantly provide deep insights for a new domain.
- Achieved automatic topics extraction and sentiment polarity assignment with 0.82 F1-score on hand-labeling dataset.
- Accelerated part of the business process 5 times with 0.82 F1-score comparable to the original model.

National Supercomputer Center in Guangzhou — Guangzhou, China

Research Assistant — Oct 2017 – Jan 2018

- Contributed to "Translingual Literature Recommendation System", motivated by the difficulties in cross-language searching.
- Implemented various algorithm, including Recurrent Neural Networks (RNN), Autoencoders using python and TensorFlow.
- Built an ETL pipeline for automatic literature text mining and key features extraction on MongoDB
- Operated data and ran scripts on Tianhe-2, a supercomputer located in NationalSupercomputer Center in Guangzhou.

HANDS-ON SKILLS

- **Programming:** Python, R, Linux bash, C/C++, Swift, MATLAB
- **Database:** SQL, MongoDB, Hadoop, Apache Spark, Hive
- **Machine Learning:** scikit-learn, PyTorch, TensorFlow, Keras
- **Cloud:** AWS, GCP, CI/CD, Git, Docker, Kubernetes
- **Visuals:** Tableau, matplotlib, Plotly
- **Others:** Flask, Hugo, R Shiny, REST API, HTML, Excel

PROJECTS

Capstone Project: Image and Text Features Extraction for Auction Price Prediction — Aug 2019 - Now

- Led and coordinated regular technical meetings with external stakeholders following Agile project management.
- Extracted interpretable features from auction images using unsupervised decomposition methods and supervised CNN attention map.
- Supported business strategy making by achieving $2600 mean absolute error on price prediction tasks (typical price is $20000).
- Working on continuous deployment of our model with Flask Application.

Quantitative Analysis on Super-resolved Satellite Images for Object Detection

- Implemented and opened source of state-of-the-art Super Resolution GAN (SRGAN) to increase satellite images resolution.
- Improved the mean average precision (mAP) score on YOLO object detection model by 43.60% with the use of SRGAN.

HONORS & AWARDS

- Winner of HL7 FHIR DevDays Student Track at Seattle for an Asthma Management App on IOS — *June 2019*
- "Best Use of Outside Data" prize at Duke Data Fest on fatigue prediction (one of best 4 teams among 82 teams) — *Apr 2019*
- 7th prize of Duke Datathon on customer segmentation and analysis among 200+ students — *Nov 2018*