Yanjie He

Portfolio of Projects: https://yanjiehe.github.io/
LinkedIn: https://www.linkedin.com/in/yanjiehe/

Mobile: +1 (202)733-7796
Email : heyanjie0@outlook.com
Github: https://github.com/YanjieHe

Education

- **The George Washington University** — Washington DC, USA
  Master of Science in Data Analytics (Computer Science track); GPA: 3.59/4.00 — *Anticipated May. 2020*
- **Shanghai University of International Business and Economics** — Shanghai, China
  Bachelor of Arts in Economics; GPA: 3.71/4.00 — *Sept. 2013 – June. 2017*

Skills

- **Coding**: C++, Python, Java, C#, R, SQL, Scheme/Racket
- **Front-end**: React.js, HTML/CSS, JavaScript, Bootstrap
- **Back-end**: Spring, Hibernate, MySQL, Linux, Docker, OpenCV, AWS (EC2, RDS)
- **Technologies**: Machine Learning, Computer Vision, Compiler Construction

Experience

- **Software Engineer Intern - Computer Vision** — Reston, VA, USA
  ScientiaMobile, Inc. — *Jun. 2019 - Aug. 2019*
  - **Image Classification**: Developed an image classification system using Python and C++. It is part of the ImageEngine project. It classified more than **4 million images** on the server.
  - **ImageEngine**: ImageEngine (https://www.scientiamobile.com/products/imageengine/) is a framework for mobile devices and website image optimization, widely used by industry leaders, including **Amazon, Google, Oracle, and Willis Towers Watson**.
  - **Docker**: Standardized the C++ and Python dependencies using Docker. Deployed and maintained the service on a Linux server.
  - **Computer Vision**: Researched how to extract image features and classify images. Completed a C++ program with OpenCV to compute image and color statistics in high performance.
  - **Machine Learning**: Built an SVM classifier to identify images that are not suitable for overly compressed. According to the result of 10-fold cross-validation, the precision is **95%**, and the overall accuracy is more than **80%**.
  - **Data Wrangling**: Manually tagged 3,000 images for training the models. Utilized Python to process the collected data, and used the Tesseract OCR engine to detect text on the images.
- **Teaching Assistant - Programming for Analytics** — Washington DC, USA
  George Washington University — *Sep. 2019 - Dec. 2019*
  - **Course**: Tutoring and grading 23 students in Professor John Helveston's Programming for Analytics course. Helping the professor to design the course and exams once a week.
  - **R Programming Tutoring**: Teach students data wrangling, text manipulation, vectorization, and data visualization using R (tidyverse and ggplot2). Tutor students to build data pipelines and create reproducible research using R Markdown. Open sessions for students twice a week.
  - **Front-end & RStudio Addin**: Completing a front-end using React.js for problems posting and student management. Developing an RStudio add-in for students to test and submit their solutions.
  - **Back-end**: Building a RESTful back-end service using Java Spring, Hibernate, and MySQL. Designed data models.
- **Computational Social Scientist** — Washington DC, USA
  George Washington University — *Oct. 2018 - Jun. 2019*
  - **Social Network Analysis**: Worked for Professor Vontrese Pamphile's social science research project. Applied mathematical and statistical techniques to novel data. Measured reputation premium gained from social connections.
  - **Data Analysis**: Reviewed academic papers. Cleaned datasets and run the models using Python (numpy, pandas, matplotlib, seaborn, and networkX) and R (ggplot2 and igraph). Wrote analysis reports using LaTeX and Jupyter Notebook.

Selected Projects

- **A Compiler and a Virtual Machine**: Developed a compiler for a statically typed language, a bytecode disassembler, and a virtual machine in C++.
- **A Movie Recommender System**: Link: https://yanjiehe.github.io/Movie-Recommender-System/ Developed a responsive web app for retrieving 26,631 movies' information and recommendations using Scala (Play, Akka, Spark). Deployed the web service on the AWS EC2. Managed MySQL database on the AWS RDS.