Yanjie He

Mobile: +1 (202)733-7796
LinkedIn: https://www.linkedin.com/in/yanjiehe/
Email : heyanjie0@outlook.com
Github: https://github.com/YanjieHe
Portfolio of Projects: https://yanjiehe.github.io/

Education

- **The George Washington University** — Washington D.C., USA
  Master of Science in Data Analytics (Computer Science track) — *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*

Skils

- **Programming Languages**: : C++, C#, Java, Scala, Python, R, SQL, Scheme/Racket, HTML/CSS/JavaScript
- **Modeling Skills**: : Machine Learning, Statistics, Social Network Analysis, Natural Language Processing
- **Database**: : MySQL, SQL Server, Sqlite3
- **Frameworks & Libraries**: : Qt5, React.js, D3.js, Spark, Eigen, OpenCV
- **Tools & Technologies**: : Linux, Gephi, Recommender System, Compiler Design

Experience

- **Machine Learning Researcher** — Washington D.C. Metro Area, USA
  George Washington University — *Oct. 2018 - Anticipated Feb. 2019*
  - **Project Title**: Artificial Intelligence and User Behavior for Robust Near Real-Time Recommendations
  - **Position**: Group research collaborator for Dr. Benjamin Harvey, faculty of the George Washington University
  - **Recommender System**: Designing a graph-based recommender system, utilizing collected user behaviors data.
  - **Software Development**: Developing the recommender system using C++, Java, and Python. Implementing machine learning algorithms to build robust models. Using Github for collaborative development and version control. Utilizing Restful Api to provide interoperability between front-end and back-end.
  - **Information Retrieval**: Applying methods in Natural Language Processing to extract information from user activities in the browser.
- **Computational Social Scientist** — Washington D.C. Metro Area, USA
  George Washington University — *Sept. 2018 - Anticipated Apr. 2019*
  - **Social Network Analysis and Data Analysis**: Working with Professor Vontrese Pamphile at the George Washington University, applying mathematical and statistical techniques to novel data.
  - **Analytical Programming**: Reviewing academic papers. Using Python and R to clean datasets and run the models. Utilizing packages including NetworkX and igraph.
- **Software Engineer Intern** — Shanghai, China
  Kantar Media CIC — *July. 2016 - Feb. 2017*
  - **Data Collection**: Collected online comments with team members utilizing Python, and provided data cleaning solution using C# and SQL Server for Chanel APAC project.
  - **Data Visualization**: Developed Data Visualization Solution in C# for GroupM television show, to plot Venn Graph according to given data automatically.
  - **Text Mining System**: Developed a text mining system in C# with team members. The system was used by more than 50 data analysts in the company to make data analysis solution for L'Oréal, Chanel, Volkswagen, and Dell. Saved more than 10,000 dollars of outsourcing cost. The system was processing millions of text data records every month.
  - **Sentiment Analysis**: Improved the accuracy of sentiment analysis by 17% compared to the previous tool used in the company.

Selected Projects

For more details about my projects, please click the link https://yanjiehe.github.io/.

- **A Compiler and a Virtual Machine**: Developed a compiler for a statically typed language, a bytecode disassembler, and a virtual machine in C++.
- **A Content-based Movie Recommender System**: Developed a content-based movie recommnder system in C++ and Python. Extracted information of colors and audio from movie trailers, and utilized the network relationship between movies.
- **Sorting Algorithms Visualization using D3.js**: Visualized the procedure of sorting algorithms with D3.js.
- **Text Co-occurrence Network Analysis for The Hunger Games**: Completed a text analysis project where implemented text co-occurrence network to visualize the relationship between the main characters in the novel *The Hunger Games*. Written the program in Python and used packages including nltk, pandas, matplotlib and wordcloud. Plotted the network by using Gephi.