Yanjie He

LinkedIn: https://www.linkedin.com/in/yanjiehe/
Github: https://github.com/YanjieHe

Email : heyanjie0@outlook.com
Portofolio: https://yanjiehe.github.io/
Phone: (202)733-7796

EDUCATION

- **The George Washington University** — Washington D.C., USA
  Master of Science in Data Analytics (Computer Science Track) — *Anticipated July. 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
- **Analytical Skills**: : Data Analysis, Machine Learning, Social Network Analysis, Natural Language Processing
- **Database**: : MySQL, SQL Server, Sqlite3
- **Frameworks & Libraries**: : Qt5, React.js, Spring MVC, Spark, Eigen, OpenCV
- **Tools & Technologies**: : Linux, Recommender System, Compiler Design, Gephi, Git (Github)

EXPERIENCE

- **Machine Learning Researcher** — Oct 2018 - Anticipated Feb 2019
  George Washington University — *Washington D.C. Metro Area, USA*
  - **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.
  - **Information Retrieval**: Applying methods in Natural Language Processing to explore information from user activities in the browser.

- **Computational Social Scientist** — Sep 2018 - Anticipated Apr 2019
  George Washington University — *Washington D.C. Metro Area, USA*
  - **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 the dataset and run the models.

- **Data Engineer Intern** — Jul 2016 - Feb 2017
  Kantar Media CIC — *Shanghai, China*
  - **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.
    1. Used jieba package in **Python** to cut the Chinese sentences into separated words.
    2. Selected words as features that are important for classification by χ^2 test.
    3. Applied the logistic regression to classify the sentiment of user comments.

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 Content-based Movie Recommender System**: Designed a content-based movie recommnder system applying methods in Machine Learning, Computer Vision, Audio Analysis, Natural Language Processing and Social Network Analysis. Devloped the recommender system in **C++**, and build a website using **React.js**. Extracted information from movie trailers using **OpenCV** and **pyAudioAnalysis**.
- **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.