Yueyuan He

yhe187@syr.edu | 315-450-4485 | linkedin.com/in/yueyuanhe/ | github.com/heyueyuan

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY May 2021

Master of Science, Applied Data Science

Relevant Courses: Big Data, Data Analytics, Business Analytics, Data Warehouse, Natural Language Processing

Beijing University of Posts and Telecommunications, Department of Communication & Information Engineering June 2018

Bachelor of Telecommunication Engineering

Relevant Courses: Linear Algebra, Probability Theory and Stochastic Process, Data Structures and Applications

SKILLS

Languages: Python, SQL, R, Java, Scala

Data Science: Data Analysis, Data Warehouse, Machine Learning, Deep Learning

Big Data: Hadoop, Spark, Hive, Kafka, Flink, Azkaban, BigQuery

Business Intelligent: Tableau, PowerBI

Others: Git, Docker, Shell, Scrapy, Network, Web, FastAPI, Cloud (AWS, GCP)

EXPERIENCE

Research Assistant May 2020

Cyber-Physical Systems Lab, Electrical Engineering & Computer Science, Syracuse University

The main work of the lab is anomaly detection and recovery. My job is to write car operating programs based on Donkey Car which is an open source and OOP car software. Besides, I Applied machine learning (Linear Regression, KNN, LSTM) and analyzed the relationship between sensors and speed. This project is based on Python and I mainly used PyCharm and Jupyter Notebook. After a lot of experiments, we can predict a new speed by simulating the attack speed sensor on our experimental car, so that the car can be stable and safe. Github

Network Engineer July 2018 – Aug 2019

Housing Information Service Center, Capinfo Company Ltd.

- Maintained network infrastructure which contains hundreds of devices; Improved the Infrastructure through redesign and configuration; Troubleshooted dozens of network glitches weekly. Analyzed network packets using Wireshark, etc.
- Implemented automated network configuration backup through SecureCRT using VBScript, which reduced 1/2-day configuration backup work to 90 seconds.

PROJECTS

Fudge Corporation Data Warehouse and BI Solution Nov 2020

- Designed detailed dimensional model based on high level dimensional modeling
- Implemented ETL process with Microsoft SQL Server Integration Services
- Created dashboards with PowerBI for datasets like, “What is the contribution of each product towards yearly sales?”

E-commerce Real-Time Analysis by Flink Sept 2020

- Project was implemented with Scala and Flink framework
- Utilized DataStream API connected with Kafka and Flink SQL to achieve streaming analysis
- Calculated hot items every 5 minute during last 1 hour by time window functions
- Analyzed website page visits and calculated the number of unique visitors in streaming data
- Additional tools used are IntelliJ IDEA and Apache Maven

Pricing analysis and prediction of Airbnb in Seattle April 2020

- Applied PySpark to finish batch processing and design data pipelines that Spark can use directly for data processing
- Provided useful business solutions by exploring data analysis, such as the most popular house and its location, etc.
- Cleaned a 10k public dataset about house information of Airbnb, such as missing value, categorical features to number
- Built recommend systems based on users using ALS algorithm and another based on houses using K-Means, so that the system can recommend more suitable houses through sorting the distance between two different house descriptions.

CERTIFICATION

Data Engineering with Google Cloud