Keyang Yu

Columbus, Ohio • Phone:614-3978038 • Email: yu.2049@osu.edu • https://www.linkedin.com/in/keyang-yu-8b704a111/

EDUCATION

The Ohio State University, Columbus, Ohio **Expected May 2020**

- **Double Major:** Data Analytics (Computational Track) & Computer Information Science, **Minor:** Mathematics GPA: 3.92
- **Relevant Coursework** in Machine Learning, Data Mining, Artificial Intelligence, Data Management in Cloud, Data Structures and Algorithms, Operating System, Software Component, Development and Design, Linear Algebra, Ordinary and Partial Differential Equation, Statistical Modeling, Statistical Learning, Data Visualization and Database System
- **Awards:** Edward G. Mayers Scholarship, the Veeam Software Endowed Scholarship in Data Analytics and Dean's List

SKILLS

Programming Language: Proficient in Java, Python and R; Familiar with JavaScript, Html, CSS, C/C++ and C#
Databases: MySQL, SQL Server, SQLite, PostgreSQL, Impala SQL and MongoDB
Tools & Frameworks: Node.js, Bootstrap, jQuery, Docker, Git, Spark, Hadoop, MapReduce, VTK, D3, Leaflet and Tableau

WORK EXPERIENCE

IT Data Analyst Intern, CenturyLink, Monroe Louisiana **May 2019 – July 2019**

Infinera Card Failure Analysis (Project 1)

- Utilized Impala SQL to query all failure events happened 10 days before an Infinera Card equipment failure and conducted EDA to find potential factors that leads to an equipment failure.
- Visualized frequency of card equipment failure by locations and time using heatmap and animated time-heatmap.
- Applied PrefixSpan algorithm with PySpark.mllib to find frequent prefix event patterns before an Infinera card failure.
- Self-defined distance metrics based on prefix patterns found and implemented hierarchical clustering algorithm to find potential groupings of card nodes.

Abnormal Job Flow Prediction (Project 2)

- Designed and built a reusable model to monitor and predict abnormal job streams in Data Lake scheduled flows.
- Preprocessed the data using PySpark and deployed a standard pipeline for various ML classification algorithms and resampling methods to deal with imbalanced data and evaluated by metrics and visualizations.

Teaching Assistant (Data Management in Cloud), The Ohio State University, Columbus Ohio **January 2019 – April 2019**

- Evaluated submissions of quizzes and homework along with providing effective feedbacks.
- Tutored student for problems in the fields of MapReduce, Data Warehousing, query optimization and NoSQL.

Data Analyst Intern, GridSum Holding Inc., Shanghai, China **May 2018 – July 2018**

- Utilized MS SQL Server to perform data extraction and manipulation for daily analysis and department heads.
- Developed a linear regression model by R for business performance of products on Chinese e-commerce platforms
- Performed market basket analysis by R to uncover associations between items in transactions.

RESEARCH EXPERIENCE

Assistant Researcher, The Ohio State Center for Aviation Studies, Columbus Ohio **August 2019 – Present**

- Developing predictive analysis tools for a web-based interactive visualization and analysis framework based on aviation data.
- Managing the database with over 1TB data via PostgreSQL and the code base, including Python framework and Javascript (jQuery, D3, Leaflet) application, on a Linux Server.
- Applying hierarchical clustering with self-defined distance metric based on geo-spatial data to predict and categorize flight routes and visualizing the clustering result using D3 on the web interface.
- Comparing and testing clustering algorithm's performance based on various system architectures.

SELECTED PROJECTS

Netflix Movie Recommender System **May 2019 – June 2019**

- Implemented a Netflix movie recommendation system for users based on the dataset with size up to around 1GB provided in the Netflix competition to find users' favorite movies based on their watching history using Java, Hadoop and Docker.
- Designed and applied Item Collaborative Filtering algorithm by multiplying normalized co-occurrence matrix between movies and users' rating matrix to compute and generate recommendation list.
- Developed a workflow that efficiently realizes inner product between matrices in MapReduce.

Super Mario Bros, The Ohio State University **January 2019 – April 2019**

- Used C# and Monogame framework to re-create Super Mario Bros game with additional self-designed features.
- Designed, developed and refined a high-quality 2D platformer game engine utilizing object-oriented design patterns.
- Designed and implemented game physics engine and collision system and optimized their time complexity by developing reducing the number of objects' collision checks.
- Worked in agile software development and team management approach using Azure Devops and version control using Git.

Book Store Inventory System, The Ohio State University **October 2017**

- Used C to establish an inventory system and construct linked list to store and process data and business records for a bookstore.
- Utilized function pointers to execute various functions for inventory transaction and statistics calculation.
- Created a Makefile that organized code compilation with all dependencies and generated an executable program.