Ruthvik Penumatcha . Jarvis Consulting

I hold a Bachelor's degree in Computer Engineering with a specialization in Software Engineering, and a Master's degree in Electrical and Computer Engineering with a specialization in Artificial Intelligence from Toronto Metropolitan University (formerly Ryerson University). My academic background has equipped me with strong technical expertise and problem-solving skills. Throughout my academic journey, I developed a keen interest in Data Engineering, DevOps, and Software Development. As a research assistant at Toronto Metropolitan University, I helped develop a novel bias measurement metric for Large Language Models using Python, PyTorch, NumPy, Pandas, and Matplotlib. I've also built personal projects using FastAPI, Pandas, and Spark SQL to analyze web-sourced data, strengthening my skills in data manipulation and scalable system design. I am currently seeking opportunities in Data Engineering, DevOps, or back-end development, where I can apply my expertise and passion for learning to solve complex technical challenges.

Skills

Proficient: Java, Python, JavaScript, Node.js, Express.js, Linux/Bash, RDBMS/SQL, Pandas/NumPy, NoSQL/MongoDB, HTML, CSS, Agile/Scrum, REST APIs, TensorFlow/PyTorch, Maven, Git

Competent: Spark, Hive, Hadoop, Docker, React, Google Cloud Platform (GCP), Next.js, Scikit-learn/Matplotlib, Postman

Familiar: MATLAB, C, Jenkins, RabbitMQ, gRPC, VHDL, LaTeX

Jarvis Projects

Project source code: https://github.com/jarviscanada/jarvis_data_eng_RuthvikPenumatcha

Cluster Monitor [GitHub]: The Linux Cluster Monitoring Resource app has been developed for the Linux Cluster Administration (LCA) team at Jarvis to monitor the system specifications and resource usage information across 10 Linux clusters. This project has been implemented by creating a PostgreSQL database instance on a Docker container to store hardware specifications and resource usage information. Shell scripts were created to handle various tasks, such as managing the container setup and data insertion. Crontab was used to automate data collection every minute, and Git was used for version control. Through automating data collection and centralizing monitoring, the system reduces manual overhead and enhances decision-making for administrators.

Java Grep App [GitHub]: Developed a Java-based tool utilizing object-oriented programming principles, which functions similarly to the grep command on Linux. It recursively traverses a user-provided directory, identifies matching regex patterns the user provides, and stores the results in a file. Utilized Java Streams and Lambda expressions to optimize the performance of the application. Deployed the application by creating a Docker image using a custom Dockerfile. Used Maven for dependency management, and Git for version control.

Python Data Analytics [GitHub]: This project was developed for London Gift Shop (LGS), a London-based online retail business. The objective of the project is to perform data analytics on historical sales data to uncover any trends and patterns that exist, which could support the marketing team in creating targeted advertisements to improve sales. The retail data sent by the client was loaded onto the PostgreSQL database to set up the Data warehouse. Jupyter Notebook was used to perform data cleaning and exploration, utilizing popular Python libraries such as NumPy, Pandas, and Matplotlib. Various key factors were analyzed as part of the data analysis, including monthly sales data, the number of new customers added per month, and segmenting clients into various sections using RFM (Recency, Frequency, Monetary) analysis. To ensure platform independence and reproducibility, Docker was used to containerize both the PostgreSQL database and the Jupyter Notebook environments. Git was used for version control.

Highlighted Projects

Big Data Analysis of Used Car Sales in Europe: Performed analysis of used car sales in Europe over the past 20 years using GCP and Spark SQL. Utilized a comprehensive dataset containing over 3 million records, including attributes such as car make, model, price, mileage, fuel type, transmission, and registration year. Conducted data cleaning to remove irrelevant columns and rows containing null or empty values. Wrote Spark SQL queries to identify trends such as the most popular car makes and models across various price ranges, and determined the least preferred vehicles by users over the past 5, 10, and 20 years.

Schema-Tune: Noise-Driven Bias Mitigation in Transformer-based Language Models [GitHub]: Co-authored the research paper and helped successfully implement the Schema-Tune framework to reduce biases in large language

models. The framework introduces curated and adapted noises to the input embeddings of language models to challenge the model's embedded stereotypes. I utilized various tools such as Python, PyTorch, Pandas, NumPy, and Matplotlib to help train the models and evaluate their performance. The paper has been published at the ACML 2024 conference.

Binary Semantic Segmentation of Drone Images: Developed a binary semantic segmentation model using the U-Net architecture to accurately identify foreground and background components. The model was successfully trained on a dataset of drone images and their respective ground truth. The model obtained high training, validation, and testing accuracy of over 98%, showing its effectiveness.

Full-Stack E-commerce Platform [GitHub]: Developed a full-stack E-commerce website using Node.js, Express.js, MySQL, and vanilla JavaScript with HTML/CSS. Implemented User Authentication (Login/Signup), product browsing and search, shopping cart functionality, and a checkout process. Designed a MySQL database to store user accounts, product inventory, and transaction history, including the purchase quantities, dates, and total price. Built RESTful APIs for front-end and back-end integration, ensuring seamless data flow and dynamic content rendering.

Professional Experiences

Software Developer, Jarvis (April 2025-present): As a software developer, I contributed to multiple projects using technologies such as Linux, Bash scripting, PostgreSQL, Docker, and Git. I played an active role in designing, implementing, and deploying scalable solutions that met client needs. I collaborated closely with a cross-functional team in an Agile/Scrum environment, participating in sprint planning, daily stand-ups, sprint reviews, and retrospectives.

Research Assistant, Toronto Metropolitan University (September 2024-December 2024): Collaborated with a research team to develop an innovative bias measurement metric that addresses the limitations of existing metrics to improve fairness in AI models. Implemented the project utilizing tools and technologies such as Python, PyTorch, Pandas, and NumPy. Incorporated quantum computing principles, including superposition and interference, to evaluate bias. Conducted extensive evaluations on popular LLMs, including GPT-2 and BERT, demonstrating the metric's effectiveness in detecting nuanced biases.

Education

Toronto Metropolitan University (2018-2023), Bachelor of Engineering, Computer Engineering - Graduated with Distinction - Dean's List (2018-2019, 2020-2021, 2021-2022)

Toronto Metropolitan University (2023-2024), Master of Engineering, Electrical and Computer Engineering - CGPA: 4.08/4.33

Miscellaneous

- Badminton Enthusiast: Played badminton regularly with friends from middle school through university during my free time. It's a sport that helps me relax and stay mentally focused.
- Cycling: Passionate about cycling during the summer months, using it as an opportunity to stay active, explore my neighborhood, and discover new local spots.