Prabhjas Singh

www.prabhjass.com | (510) 565-9017 | prabhjasgill@gmail.com | Union City, CA

Summary

Pursuing a Data Science with a concentration in machine learning at UC San Diego, with a strong background in computer science and hands-on experience in programming, computer networks, algorithms, and software development.

Education

University of California, San Diego | San Diego, CA
Data Science | 06/2027
Relevant Courses: The Practice and Application of Data Science, Theoretical Foundations of Data Science, Data Science in Practice, Data Structures and Algorithms for Data Science, Computer Systems and Assembly Language and Lab, Computer Systems and C Programming, Computer Architecture, Introduction to Computer Networks, Introduction to Probability Theory

Experience

University of California - Santa Cruz | Santa Cruz, CA
Network Simulation and Security Engineer - Academic Project | 03/2024 - 06/2025
Developed a simulated network enviorment for Large Language Model startup using Mininet and POX, focusing on network topology design, traffic routing, and security enforcement.

Key Contributions:

- **Custom network topology design:** Created a detailed Mininet topology respresenting a towo-floor office building. Each floor had its own switches and subnets, interconnnected by a core switch including a dedicated subnet for the LLM server located in a data center.
- **Advanced Routing and Firewall Implementation:** Developed and deployed a POX controller to manage routing and enforce strict firewall rule. Blocked all IP and ICMP traffic from untrusted external hosts to internal hosts and LLM server. Restricted trusted external hosts from accessing specific subnets and the LLM server, while allowing free communication with designated subnets. Prevented ICMP traffic between hosts on different floors within the office building.

Projects

Economic Impact and Power Outage Severity Analysis (Python)

- Developed an end-to-end data science and machine learning pipeline to analyze and predict high-severity power outages across the United States using outage, economic, and climate data.
- Conducted extensive data cleaning, exploratory data analysis, missingness assessment, and hypothesis testing to understand structural patterns and data limitations.
- Built baseline and Random Forest classification models using scikit-learn pipelines, incorporating engineered features such as log-transformed economic indicators and climate regions. Tuned hyperparameters with cross-validation and evaluated models using F1-score. Performed a fairness analysis using permutation tests to assess performance disparities across income-based state groups.

Web crawler system (C/C++)
- Developed a sophisticated web crawler system, orchestrating efficient web page retrieval, content analysis, and data storage.
- Utilized libcurl for robust web page fetching, encapsulated within a custom wrapper for streamlined content retrieval. Implemented essential data structures from scratch, including a dynamic hashtable for efficient URL deduplication and a flexible bag data structure for managing crawl frontier.
- Developed a 'pagedir' module for orderly storage of crawled web pages, facilitating easy retrieval and indexing, while also emphasizing modular design across components, ensuring easy maintenance, extensibility, and reusability.

Skills

scikit-learning, pipeline, LLM, Python, C/C++, Linux, Java, MATLAB, Assembly, Mininet, Wireshark, Web Crawling and Data Structure, Gesture Recognition System, Pandas, Ubuntu, GitHub, GitLab, SQL