Varmin Singh

US Citizen ❖ singh.varmindip@gmail.com ❖ (559) 903-2599 ❖ Berkeley, CA ❖ linkedin ❖ github ❖ portfolio

EDUCATION

University of California, Berkeley **Aug. 2022 – May. 2026**
BA: Data Science, Emphasis: Economics, Machine Learning Berkeley, CA
- **Relevant Coursework:** Applied Machine Learning, Data Engineering, Intro to Machine Learning, Data Science for Economists, Probability for Data Science, Structure and Interpretation of Computer Programs, Principles and Techniques of Data Science, Data Structures, Discrete Math & Probability, Computer Architecture, Artificial Intelligence, Designing and Understanding Deep Neural Networks
- **Extracurriculars:** BIDS (Berkeley Institute for Data Science), Data Science Discovery Program, Cal Bhangra

Skills and Interests

- **Languages:** Python, SQL, R, Java, C, C++, JavaScript,
- **Tools:** Pandas, AWS, Docker, OpenCV, NumPy, Selenium, PostgreSQL, Scikit-Learn, PyTorch, TensorFlow, Git
- **Interests:** Deep Learning, Data Engineering/Analysis, Back-End, Basketball, Hiking, Weight Lifting

Professional Experience

Penserra **Jun. 2024 – Pres.**
Data Engineer Intern Orinda, CA
- Automated CP data management using Python and Bloomberg API, reducing manual entry time by 80%, and integrated BL API pages with Excel, enabling real-time updates and customized client reports for 50+ issuers.
- Designed and implemented a document parsing pipeline using Azure Computer Vision, OpenCV, and Tesseract OCR, reducing data upload time by 70% and processing over 500 flex trading records daily.
- Migrated financial data from Excel workflows to a PostgreSQL database and scaled company-wide, improving data handling processes and query efficiency, and enhancing the usability of critical economic data for multiple teams.

University of California, Berkeley **Aug. 2024 –Dec. 2024**
Academic Intern for Computer Architecture Berkeley, CA
- Assisting during office hours to help students debug labs/projects and answer technical/conceptual questions.
- Collaborating with other academic interns and student instructors to develop better information delivery methods.

University of California, Berkeley **Jan. 2024 – May. 2024**
Undergraduate Research Assistant Berkeley, CA
- Built predictive models for AWS EC2 outages using a one-year dataset, analyzing CloudWatch metrics, ELB logs, Auto Scaling events, and Lambda functions to improve system reliability.
- Utilized Bayesian Networks and Gradient Boosting Trees to identify causal links between outages and alerts, pinpointing root causes and uncovering component interactions.
- Identified root causes of EC2 outages, reducing average downtime by 30% and enhancing cloud system resilience.

Projects

Automated Fault Detection in Github Commits
- Built an end-to-end pipeline to classify GitHub commits as faulty or non-faulty using real-world repository data.
- Preprocessed data using feature standardization, one-hot encoding, and imputation for missing values, ensuring consistent input for model training and reducing data inconsistencies by 40%
- Trained logistic regression and feed-forward neural network models in PyTorch, achieving an AP of 0.03, and evaluated performance using precision-recall curves, average precision, and F1 score to optimize thresholds.

Recurrent Neural Networks for Sequence Prediction (Research)
- Built an RNN model to predict running averages from a sequence of numbers, with a regression layer to predict values at each timestep and the final timestep, testing different prediction approaches.
- Gained hands-on experience building and optimizing RNNs for time-series prediction tasks.

AI Skin Disease Detection (Berkeley AI Hackathon)
- Developed a convolutional neural network (CNN) using Pytorch to apply computer vision techniques on over 25,000 images for skin disease detection, achieving over a 90% accuracy in classifying 8 types of skin diseases.
- Implemented a user-friendly web interface for easy image uploads and real-time AI-driven disease predictions.