Nabeel Sabzwari - Computer Science, M.S.

· nabeelsabz101@gmail.com · (408) 218-8885

Skills

Languages: C/C++, Java, JavaScript, MATLAB, Python, R, SQL, TypeScript
Technologies: AWS, Cursor, Copilot, Docker, Django, Firebase, Flask, Git, GCP, Hugging Face, HPC, LangChain, Linux, LlamaIndex, Next.js, Pytest, PyTorch, RAG, React.js, Snakemake, Tensorflow

Experience

Software Engineering Researcher (University of California, Davis) - Davis, CA Nov 2024 - June 2025
- Developed data analysis and visualization pipelines in Python and R on an HPC cluster in Dr. Megan Dennis's lab, processing 10+ TB of genomic data to extract actionable biological insights.
- Designed and deployed over 20 automated, data-processing pipelines with Snakemake, Bash scripting, and Slurm, significantly reducing manual effort and runtime variability.
- Integrated open-source tools containerized with Docker to ensure reproducibility and efficiency across diverse bioinformatics tasks within the genomic pipeline.

Machine Learning Researcher (University of California, Davis) - Davis, CA Mar 2024 - Oct 2024
- Developed, trained, and evaluated deep learning models (CNN, TCN) using TensorFlow in Dr. Laura Marcu's lab to predict clean fluorescence signals from noisy tumor data, looking at R^2 and domain-specific metrics for model validation.
- Ensured model robustness by performing a random hyperparameter search using Optuna and K-Fold cross-validation with a held-out test set for final model training.
- Synthesized findings into a lab-reviewed scientific report that generated interdisciplinary interest and initiated collaborative research efforts.

System Engineering Co-Op (iRhythm Technologies) - San Francisco, CA Jan 2023 - Sep 2023
- Conducted over 20 system verification and validation tests for the Zio AT heart monitoring device, assessing BLE and cellular performance in varying coverage conditions by using DC power analyzers, oscilloscopes, and multimeters.
- Leveraged knowledge in Python, data structures, and serial communication to automate a test case, decreasing testing time by 80%.
- Designed and implemented a functional battery model for the heart monitoring device by creating a Python-based Monte Carlo simulation that more accurately represented field conditions, using Pytest for unit testing and adhering to SDLC practices.

Data Scientist Intern (Xtrava Health) - Santa Clara, CA June 2022 - Sep 2022
- Deployed a parsing script to process raw data into a neat Excel spreadsheet from 500+ digital COVID-19 sensors using Python and MATLAB.
- Implemented and presented Python visualization scripts to pinpoint improvement in Xtrava's COVID-19 detection algorithm.

Projects

Broom Service - San Jose, CA July 2025 - Present
- Leading the design and development of a mobile-first house chore management app that automates task rotation, sends real-time reminders, and tracks completion for shared households.
- Designing an interactive React Native frontend in Expo Go featuring push notifications, offline sync, and dynamic task dashboards.
- Developing AWS Lambda functions integrated with DynamoDB for scalable, serverless task management and user authentication.

Skin Cancer Lesion Classification - Davis, CA Mar 2025 – June 2025
- Developed a CNN-based computer vision pipeline in PyTorch to classify skin lesions and evaluate model bias across diverse skin tones.
- Fine-tuned a ResNet50 model on prelabeled MSKCC skin lesion data to classify six skin tone categories, achieving a 91% within-one accuracy, 75% precision, and 72% recall.
- Curated a diverse skin tone image dataset by training StyleGAN2 on MSKCC and DermaMNIST-E data, supporting robust model training and enabling fair cancer classification benchmarks.

AI Agent for Medicinal Data Insights - San Jose, CA June 2024
- Developed an AI agent using RAG to analyze medicinal data from a 50k-entry CSV file sourced from Kaggle.
- Integrated Pandas for parsing the CSV into a dataframe and developed tools for the AI agent to extract insights from both the CSV and a PDF document on amoxicillin.
- Implemented LlamaIndex for efficient data retrieval and utilized OpenAI API (ChatGPT 3.5 turbo) as the language model for generating responses and insights.

Education

Masters of Science, Computer Science (University of California, Davis) Sep 2023 - June 2025
- GPA: 3.7
- Relevant Coursework: Probability Models for Computer Science, Machine Learning and Discovery, Genomics

Bachelor's of Science, Bioengineering (University of California, Berkeley) Aug 2020 - Dec 2022
- GPA: 3.8
- Relevant Coursework: Python Fundamentals, Data Structures and Algorithms, Data Science, Robotics

Recommendations

See https://www.linkedin.com/in/nabeel-sabzwari/ for recommendations.