Nikhil Karthikeyan

nikhilskarthik@gmail.com | nkarthikeyan@ucdavis.edu | linkedin.com/in/nikhil-karthikeyan/

Education

University of California, Davis **Davis, CA**

B.S. in Data Science | Minor: Computer Science *June 2027*

- Courses: Data Structures & Algorithms, Statistical Data Science, Data Pipelines, Regression Analysis, Machine Learning

Technical Skills

Languages: Python, R, SQL, Java, JavaScript, TypeScript, HTML/CSS
Tools & Platforms: Git, Docker, PowerBI, Tableau, MS Excel, MS Word, Vercel, Jupyter, PostgreSQL, MongoDB, GitLab CI/CD, Databricks, Informatica, Spark, Microsoft Azure, Supabase
Libraries & Frameworks: Pandas, NumPy, PySpark, Seaborn, Scikit-learn, Flask, React, Angular, Node.js, Spring Boot, Streamlit

Professional Experience

Data Engineering Intern **July 2025 – September 2025**

Kaiser Permanente *Remote*

- Extracted healthcare data from multiple sources into **Databricks** ensuring **HIPAA** compliance for centralized analytics.
- Orchestrated **ETL pipelines** in **Databricks/Informatica** using **PySpark/SQL**, loading **Parquet** datasets via **Azure Data Factory** into **Synapse Analytics**.
- Built automated **Tableau/Power BI** dashboards delivering real-time clinical insights, improving decision-making by 40%.

Software Engineering Intern **June 2025 – August 2025**

San Diego Supercomputer Center *San Diego, CA*

- Built **ETL pipelines** from **SQL/AWS S3** using **PySpark/Dask**, producing **Parquet** datasets for high-throughput training.
- Fine-tuned **YOLOv8/Vision Transformers** on imaging datasets with **PyTorch Distributed** for large-scale training.
- Containerized preprocessing/inference with **Docker**, deploying on **Kubernetes** at SDSC for reproducible vision pipelines.

Machine Learning Researcher **January 2024 – Present**

UC Davis Department of Public Health Sciences *Davis, CA*

- Analyzed **UK Biobank** data using **Python** to identify genetic risk factors for AMD under Professor Prabhu Shankar.
- Applied **Neural Networks/Random Forest** algorithms to predict AMD progression using **OCT imaging** data.
- Developed visualization tools with **Matplotlib/Seaborn** to represent retinal layer thickness across patient demographics.

AI/ML Developer **November 2023 – November 2024**

Aggie Sports Analytics at UC Davis *Davis, CA*

- Structured 10,000+ volleyball records (2016-2022) using **Pandas/NumPy/PySpark**, leveraging **TruMedia** for stats collection.
- Built Win Prediction model with **Scikit-learn (Neural Networks/XGBoost/AdaBoost)**, achieving 83% match outcome accuracy.
- Created interactive front-end using **Next.js/TailwindCSS**, hosting on **AWS** with real-time match simulations.

Projects

CryptoWatch | *Python, PySpark, Azure, Power BI, Databricks* **June 2025**

- Built a cryptocurrency price tracker using the **CoinGecko API**, storing real-time and historical data in **Azure Blob Storage** and **Delta Lake** on **Databricks**.
- Processed data with **PySpark** and visualized trends and anomalies using **Streamlit**, enabling insights into rolling averages and price volatility.

Volare | *Python, Next.js, Scikit-learn, Groq* **May 2025**

- Developed a full-stack AI-powered interview preparation platform using **Next.js**, **FastAPI**, and **PostgreSQL**, featuring real-time mock interview simulations and personalized feedback using **LLMs**.
- Integrated **HumeAI**'s emotional intelligence API with **Groq**'s high-throughput **LLM** inference to analyze user sentiment and tailor question difficulty, improving recommendation precision.

PitchPerfect |*Python, Pandas, NumPy, Scikit-learn* **February 2025**

- Processed 1TB+ baseball game files using **Pandas** and **NumPy**, achieving 85% prediction accuracy with **Scikit-learn** models.
- Optimized models (**AdaBoost**, **XGBoost**, **RandomForest**, **Logistic Regression**), reducing error by 15% and improving reliability.