Akash Iyer

San Francisco Bay Area, California | 408-607-3557 | akash.v.iyer@gmail.com | Linkedin | Github

EDUCATION

University of California, Berkeley *Aug 2021 – Dec 2024*

B.A. Data Science | Concentration in Business and Industrial Analytics (GPA 3.5)

Relevant Courses: Data Mining and Analytics, Data Science Foundations, Essential Tools For Data Science, Data Structures, Structure and Interpretation of Computer Programs, Probability & Statistics, Integral and Differential Calculus, Linear Algebra, Urban Data Analytics, Linux System Administration

Certificates & Awards: Stanford University: Relational Databases and SQL, UC Berkeley DS Discovery: Best Insights

EXPERIENCE

Data Engineer Intern — *MKS Instruments* *Aug 2024 – Dec 2024*

- Built metadata-based ETL pipelines using Informatica and Informatica Cloud, processing 1M+ rows of data from OracleDB and SQL Server, write validation queries to ensure data accuracy and integrity
- Automated and scheduled 30+ data integration task flows, writing validation tests via Postman for API integrations, ensuring reliable feature engineering and data pipelines for seamless data flow across platforms

Data Science Intern — *Viasat* *May 2024 – Aug 2024*

- Improved labor-time estimates for $10M+ antenna installation contracts by querying 2M+ rows of expenditure data using SQL from Starburst data lake platform to build and integrate predictive model into Flask web application,
- Enhanced data consistency and governance by utilizing chromaDB vector embeddings and GPT-4 from Azure OpenAI API to cluster and correct inconsistencies in historical data, and presented findings to stakeholders weekly

Machine Learning Engineer — *California State Water Resources Control Board* *Aug 2023 – May 2024*

- Developed XGBoost model using socio-economic and parcel data to map lead pipe water service lines, enhancing utility removal efforts and winning UC Berkeley's DS Discovery Best Insights and Ribbon of Excellence awards
- Combined AWS based APIs with web scraping to collect data, designing final deliverable machine learning and generative AI models, visualizations, and interactive web application for residents across five U.S. states

Data Scientist — *Roots of Success* *Aug 2023 – Dec 2023*

- Analyzed education course delivery and effectiveness with sentiment analysis using NLTK, Pandas, and NumPy, and delivered impactful visualizations for the organization's website, decreasing user dropoff by 56%
- Conducted hypothesis tests, causal analysis, and topic extraction on survey data to communicate positive impacts of programs on incarcerate behavior and employment to partners and donors, increasing average donations by 30%

PROJECTS

Bookmark Dec 2024 – Present

- Developed full-stack React Native mobile app for rating and reviewing books using TypeScript & Expo
- Designed Supabase (PostgreSQL) DB for user & book data storage, sourcing book metadata via Open Library API
- Implemented user authentication, book & profile search, and Google AI's Gemini API to impute missing book data

Retina-Scan Image Sharpening — *5th Annual Datathon for Social Good* *Nov 2023*

- Enhanced quality of 400+ blurry retina images by developing CNN-based model with Blind Spot Networks (BSNs)
- Boosted peak-signal-to-noise ratio by 18% with self-supervised NOISE2VOID model on unpaired training samples

Asteroid Spectral Classification *Jun 2023 – Jul 2023*

- Classified asteroids based on 8 spectral types using XGBoost, Decision Trees, Random Forest, and neural networks
- Deployed models on AWS cloud computing infrastructure integrated with AWS Elastic Beanstalk backed web app
- Implemented modular & organized code, CI/CD with Git, logging, error handling, and interactive Tableau dashboard on joined data queried from NASA's JPL Database and Planetary Data System in Python, Javascript, and with Flask

Question Sincerity NLP *Jun 2023*

- Classified Quora forum questions as sincere or insincere using natural language processing (NLP) and transfer
- Leveraged BERT-infrastructured large language model for high precision and recall prediction and 96.35% accuracy
- Implemented word vector-tokenization and input masks, fine-tuning the pre-trained model in TensorFlow and Keras

TECHNICAL SKILLS

Programming Languages: Python, SQL, Java, Typescript, Javascript, R

Tools & Frameworks: Pandas, Tensorflow, Keras, Scikit-Learn, NumPy, React JS, Expo, Tableau, Matplotlib, Seaborn, Excel, NLTK, chromaDB, Supabase, Oracle, OpenAI API, REST API, Git, Amazon Web Services, Linux, Bash, Flask

Machine Learning: Deep Learning, Artificial Intelligence, NLP, LLMs, Vector Embeddings, Generative Models