Sahil Dadhwal

(707) 863-2820 | sahildadhwal2001@gmail.com | linkedin.com/in/sahildadhwal | Fairfield, CA

Education

University of California, Davis Sept. 2024 – June 2027 (Expected)
Master of Science in Computer Science *Davis, CA*

University of California, San Diego Sept. 2019 – June 2023
Bachelor of Science in Computer Science — A.S. in Applied Mathematics & Economics *San Diego, CA*

Experience

Data Engineer Intern April 2024 – Present
BTIS *Sacramento, CA*

- Engineered Python ETL pipelines processing 15+ GB datasets from 50+ MongoDB collections to SQL Server, reducing data inconsistencies by 90% through software engineering best practices
- Designed end-to-end CI/CD pipeline using Jenkins, Vault, and Orkes Conductor, automating ETL deployment across 3+ environments with containerization and scheduled orchestration
- Developed 50+ Power BI dashboards for CTO using Snowflake data warehouse, supporting $10M+ in strategic decisions while leading coordination among 15+ stakeholders across 8 critical migration projects
- Built comprehensive data validation framework and architected WCPOLS compliance solution processing 2M+ weekly records

Research & Technical Projects

Dimensionality Reduction for Image Classification | *PyTorch, Deep Learning* Oct. 2025 – Dec. 2025

- Conducted comparative research on dimensionality reduction methods (PCA, UMAP, Autoencoder) for compressing 3072-D CIFAR-100 images to 512-D, evaluating impact on classification accuracy across 12 experimental conditions
- Implemented three classifier architectures (ResNet-50 with transfer learning, Transformer, Autoencoder) achieving 65.82% top accuracy on raw features and discovering UMAP fails for high-dimensional targets
- Built end-to-end ML pipeline with t-SNE visualizations revealing that simple linear PCA (26.57%) matched complex Autoencoder reduction (25.69%) while reducing training time by 50-98%

Bird Species Distribution Tracker | *Python, React, FastAPI, MongoDB, ML, GIS* April 2025 – May 2025

- Built full-stack dashboard analyzing 5M+ bird observation records to visualize climate change impact on species migration patterns using GBIF and PRISM climate APIs
- Implemented machine learning forecasting models predicting future species distribution shifts with interactive geographic heatmaps and seasonal analysis visualizations
- Architected FastAPI backend with MongoDB for processing multi-gigabyte geospatial datasets and React frontend with real-time data filtering across 10+ North American species

AI Toxicity Detection in Developer Communities | *Python, BERT, NLP* Jan. 2025 – March 2025

- Analyzed correlation between toxic communication and developer productivity across GitHub repositories using GitHub Archive data and Incivility Dataset with 1M+ interactions
- Implemented BERT-based natural language understanding model with API rotation system for toxicity detection, processing repository metrics, commits, and issue resolution patterns
- Conducted statistical analysis (Spearman/Pearson correlations) revealing significant relationships between communication patterns and project outcomes in software engineering teams

Technical Skills

Programming Languages: Python, SQL, Java, JavaScript, C++, HTML/CSS
Data Engineering: ETL/ELT, Data Modeling, MongoDB, PostgreSQL, SQL Server, Snowflake, Power BI, Tableau
Machine Learning: PyTorch, scikit-learn, BERT, Transformers, NLP, Computer Vision, Deep Learning, Statistical Analysis
Tools & Frameworks: Jenkins, Docker, Vault, Orkes Conductor, Git, pandas, NumPy, FastAPI, React, REST APIs, CI/CD