

# Sahil Dadhwal

(707) 863-2820 | [sahildadhwal2001@gmail.com](mailto:sahildadhwal2001@gmail.com) | [linkedin.com/in/sahildadhwal](https://www.linkedin.com/in/sahildadhwal) | Fairfield, CA

## EDUCATION

### University of California, Davis

*Master of Science in Computer Science*

Sept. 2024 – June 2027 (Expected)

*Davis, CA*

### University of California, San Diego

*Bachelor of Science in Computer Science — A.S. in Applied Mathematics & Economics*

Sept. 2019 – June 2023

*San Diego, CA*

## RESEARCH & TECHNICAL PROJECTS

### Dimensionality Reduction for CNN Features | *PyTorch, Deep Learning*

Oct. 2025 – Nov. 2025

- Conducting comparative research on neural network-based dimensionality reduction (Autoencoders vs. Transformers) for compressing 2048-D ResNet50 features to 128-D while preserving image classification accuracy on CIFAR-10/100 datasets
- Architected novel transformer-based compression model with 4 layers and 8 attention heads using self-attention mechanisms to identify discriminative computer vision features, hypothesizing 2-4% higher accuracy than autoencoders
- Implemented end-to-end machine learning pipeline with comprehensive evaluation framework measuring classification accuracy, reconstruction quality, and interpretability across multiple compression ratios (32D-256D)

### 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

## EXPERIENCE

### Data Engineer Intern

*BTIS*

April 2024 – Present

*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

## 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