Abdoulfatah Abdillahi

GitHub | Portfolio | Linkedin | abdoulsfsu@gmail.com | **San Jose, CA**

EDUCATION

Master of Science (MS), Data Science and Artificial Intelligence ***Expected Dec 2025***
San Francisco State University — San Francisco, CA
***Key Coursework*:** Artificial Intelligence, Deep Learning, Natural Language Processing, Probability, Statistical Learning & Data Mining, Probability and Statistic I & II, Data Visualization, Human-Computer Interaction

Bachelor of Science (BS), Computer Science ***Dec 2023***
San Francisco State University — San Francisco, CA
***Key Coursework*:** Data Structures & Algorithms, Operating Systems, Software Engineering, Data Science and Machine Learning, Software Development

RESEARCH EXPERIENCE

Thesis: Predicting E. coli Ciprofloxacin Resistance with Machine Learning ***Ongoing, expected Dec 2025***

- Owned the end-to-end development of a machine learning pipeline to predict Ciprofloxacin (CIP) resistance in 588 *E. coli* strains.
- Engineered two distinct, high-dimensional feature sets from raw genomic data: a 6,093-gene presence/absence matrix and a 166,415-SNP variant profile.
- Trained and tuned a suite of ML models (Logistic Regression, Random Forest, SVM etc…), applying rigorous k-fold cross-validation and regularization to prevent overfitting.
- Led a comparative analysis that evaluated the predictive power of gene-presence vs. SNP features, delivering key insights into the genomic drivers of resistance.
- Check more in Portfolio

Publications & Presentations

Machine Learning-Driven Pangenome Pipeline for Predicting E. coli Drug Resistance ***Presented at SACNAS Conference, Phoenix Arizona | Nov 2024***

- Architected an end-to-end pangenome data pipeline to process and aggregate three distinct genomic datasets (Mills, Gladstone, Kalloneen) for downstream analysis.
- Engineered a novel feature-extraction process to unify multiple data types (SNPs, Gene Presence/Absence) from the pangenome framework, enhancing the input for predictive models.
- Benchmarked a suite of machine learning models (Logistic Regression, Random Forest, GBDT) to validate the hypothesis that a combined pangenome approach yields more generalizable resistance predictions.
- Spearheaded future-work planning to integrate advanced NLP models (e.g., Nucleotide Transformer, BioBERT) for sequence-based resistance prediction.

Predicting Amoxicillin (AMC) Resistance in E. coli with CNN and VAE Architectures ***Presented at The Allied Genetics Conference, Washington D.C. | March 2024***

- Led the preprocessing of dual genomic modalities (pangenome alignments, gene presence/absence) to convert raw data files into deep-learning-ready image tensors.
- Implemented and benchmarked a Variational Autoencoder (VAE) against a CNN, demonstrating the VAE's +15-25% higher F1-score and accuracy in predicting resistance.

- Engineered a custom, color-map-based data pipeline, which enabled 40% faster model training and informed the adoption of the VAE for larger-scale analysis.

INDUSTRY EXPERIENCE

MLE Intern, Genentech, South San Francisco, CA **June 2023 – September 2023**
Project Intern: ***Unveiling Treatment Effects from Digital Health Data through Pharmacology-Informed Neural-SDE***

- Gained hands-on experience with neural stochastic differential equation (SDE) model to analyze stochastic data from digital health technologies (DHT) during my internship.
- Demonstrated initiative by actively seeking clarification through regular discussions, improving my ability to ask insightful and targeted questions.
- The code was written in Tensorflow, I had the opportunity to switch to a PyTorch to enhance the model training efficiency. Additionally, this project also provided an opportunity to strengthen my coding skills and gain hands-on experience in bioinformatics.

Mentorship & Leadership

Grader, San Francisco State University

- CSC 665/865: Artificial Intelligence **Present - Dec 2025**
- CSC 408: Data Science for Personalized Medicine (Python) ***Sep 2024 – Dec 2024***
- CSC 219: Data Structures for Data Science Application Development (Python) ***Jan 2024 – May 2024***
- CSC 101: Introduction to Computing and Programming (Java) ***Jan 2024 – May 2024***

Braven (Remote), *Grader* **Sep 2024 – May 2025**

- Graded assignments and projects for a cohort of 24 students using AU 220 rubrics.
- Provided detailed feedback to enhance students' career and leadership development.
- Held weekly virtual office hours to support learning, address concerns, and foster academic success.
- Reviewed instructional materials including slide decks, projects, and rubrics to ensure alignment with learning outcomes.

Research Peer Mentor — PINC Summer Program, SF State ***Summer 2024***
Advisor: Dr. Pleuni Pennings

- Mentored five undergraduate researchers students in R programming and guided them through a full research pipeline on quinolone use & ST131 E. coli across ten European countries.
- Supported data collection from European CSC & Enterobase and statistical analysis workflow.
- Presented findings and reinforced that multifactorial approaches are needed to address antibiotic resistance.

HONORS & AWARDS

NIH – UCSF/SFSU Bridges to the Doctorate Fellowship, SFSU ***2023–2025***
Awarded ~$32,000 in a two-year research training fellowship to master's students from underrepresented groups in science aimed at transition into biomedical research-focused doctoral programs.

BMC Scholarship, SFSU ***September 2024***
Awarded $1,750 by the Computer Science Department for academic excellence and commitment to research.

National Diversity in STEM (NDiSTEM) Travel Scholarship, SACNAS Conference, Phoenix, AZ ***October 2024***
Received full travel support including roundtrip flight, lodging, and conference registration to attend the 2024 NDiSTEM Conference.

Award of Recognition, SFSU PINC Program ***May 2024***
Recognized for exemplary performance and service to students and faculty, contributing significantly to the success of the PINC community.

Certificate

Data Science and Machine Learning for Biotechnology Certificate ***2022–2023***
San Francisco State University (in collaboration with Genentech)

- Completed a 12-unit, two-semester training program designed in partnership with Genentech to address technical needs in the biotech and pharmaceutical industries.
- Gained hands-on experience in high-demand data science and machine learning skills tailored for solving complex problems in biotechnology.
- Trained in cutting-edge tools and diverse computational approaches relevant to real-world biotech applications.

TECHNICAL SKILLS

Languages: Python, Java, C++, HTML, CSS, Linux
Data Engineering & Infrastructure: Dask, SLURM/HPC clusters, Docker
ML/Deep Learning: PyTorch, TensorFlow, scikit-learn,
Data Processing & Visualization: NumPy, pandas, Matplotlib, Plotly, Mlflow
Bioinformatics Tool: Roary, Prokka, BLAST, FastQC

ACADEMIC PROJECTS

Energy Consumption Modeling – Regression & Clustering Project | Project
MATH 448: Statistical Learning | San Francisco State University ***Feb 2025 – May 2025***

- Investigating the Appliance Energy Prediction dataset (UCI ML Repository) to explore relationships between environmental variables and household energy consumption.
- Applying regression techniques including linear models, decision trees, and neural networks to forecast energy usage based on temperature, humidity, and weather data.
- Using clustering algorithms to uncover patterns in consumption behavior for insights into sustainable energy management.

Creator & Maintainer – database-mongodb Python Package | PyPI • GitHub ***Apr 2025* – May 2025**

- Developed a lightweight Python package to automate MongoDB connections, database creation, and bulk data insertion from CSV/Excel files.
- Packaged and deployed the module to PyPI (v0.0.11), configured with tox, setuptools, requirements.txt, and pyproject.toml.
- Designed a modular structure with unit tests and GitHub Actions for CI/CD, ensuring maintainability and ease of installation via pip.

Full-Stack Developer – Educlouds App ***Jan 2024 - May 2024***
AI Entrepreneurship Course Project Project

- **Frontend Development**: Designed and implemented a responsive user interface using modern web technologies, enabling students to easily access and request computational resources.
- **Backend Engineering**: Developed RESTful APIs and backend logic using Flask, Python, and SQLAlchemy to manage data flow and facilitate secure transactions.
- **Project Showcase**: Successfully presented the completed Educlouds platform during the end-of-semester CSC 890 showcase, demonstrating full-stack functionality and real-world applicability.

Data Visualization Project – Shopping Trends and Customer Behavior ***Jan 2024 – May 2024***
CSC 805: Data Visualization Project

- Built an interactive web dashboard using Dash, Plotly, and Python to visualize and explore customer behavior from a Kaggle dataset (~3,900 records).
- Engineered modular visualizations including a product insights bar chart, seasonal trends graph, U.S. heatmap of popular items, and demographic filters (age, gender) for purchase behavior.

- Delivered a fully functional data exploration tool that enables users to dynamically filter by age, gender, and product type, offering real-time insights into customer trends and purchase patterns.

Fitness - FitBuddyPal ***Sep 2023 - Dec 2023***

Software Engineer Course Project Fitness App

- Led the integration of the ChatGPT API to automated and personalize workout and nutrition plans within the FitBuddyPal, enhancing user engagement and satisfaction
- Led Agile Practices: Acted as Scrum Master, organizing bi-weekly meetings to review team progress, address bottlonecks, and ensure on-time delivery of project milestones.

REFERENCES

Dr. Sara El Alaoui
Assistant Professor, Department of Computer Science, San Francisco State University
Thesis Supervisor
elalaoui@sfsu.edu

Dr. Pleuni Pennings
Professor, Department of Biology, San Francisco State University
Co-Thesis Supervisor
pennings@sfsu.edu

Dr. Ilmi Yoon
Teaching Professor and Director of Computing Programs, Khoury College of Computer Sciences, Northeastern University
i.yoon@northeastern.edu