Abdoulfatah Abdillahi

Github | *Linkedin* |*Portfolio* | *abdoulsfsu@gmail.com* |*+1 317-828-1996*|

EDUCATION

Master of Science (MS), Data Science and Artificial Intelligence ***2024-2025***
San Francisco State University — San Francisco, CA

Bachelor of Science (BS), Computer Science ***2018–2023***
San Francisco State University — San Francisco, CA

INDUSTRY EXPERIENCE

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

- Implemented CI/CD pipelines (GitHub Actions) to automate testing, validation, and deployment of data processing scripts.
- Applied parallel processing across HPC (SLURM) nodes, reducing preprocessing time by 60%.
- Deployed containerized workloads with Docker, optimizing reproducibility and scalability.
- Preprocessed and analyzed 150 synthetic patient PK/PD trajectories across multiple dose levels, enabling faster and more accurate neural SDE model training.

RESEARCH EXPERIENCE

Graduate Student Researcher, San Francisco State University, CA **Sep 2024 – Present**
Advisors: Dr. Pleuni Pennings & Dr. Sara El Alaoui
Research Project: Machine Learning-Driven Pan-genome Pipeline for Drug Resistance Prediction in *E. coli*

- Developed automated pipeline to extract gene annotations from NCBI BLAST results and applied a 5% minor allele frequency (MAF) threshold to filter SNP features across 15,695 genes, improving neural SDE model training throughput by 40%.
- Implemented and compared traditional ML algorithms (Random Forest, SVM) with deep learning architectures (Neural Networks) for drug resistance prediction, achieved over 85% accuracy in cross-validation.
- Designed advanced feature engineering strategies and predictive models specifically optimized for accurate amoxicillin resistance prediction in E. coli, contributing to antimicrobial resistance surveillance.

ACADEMIC PROJECTS

Course Projects:

Appliance Energy Consumption Prediction (Statistics Learning and Data Mining, 2025)

- Implemented STL decomposition and SARIMAX modeling for time-series forecasting.
- Developed regression models (Random Forest, XGBoost, LASSO) for energy prediction.
- Integrated environmental and weather data to improve prediction accuracy.
- Technologies: Python, Jupyter Notebook, Machine Learning, Time-series Analysis.

Alzheimer's Disease Classification Using PET Images (Deep Learning Course, 2024)

- Implemented VGG16, InceptionV3, and ResNet50 models achieving 95%+ accuracy.
- Applied transfer learning and data augmentation on ADNI dataset.
- Classified cognitive stages: CN, EMCI, AD.

TECHNICAL SKILLS

Programming Languages: Python, Java, C/C++, JavaScript, C#, SQL, MATLAB
Machine Learning & AI: TensorFlow, PyTorch, Keras, Scikit-learn, NLTK, Deep Learning, Neural Networks, Computer Vision, NLP, LLMs
Web Development: React.js, Node.js, Express.js, Django, Flask, HTML/CSS
Databases & Cloud: MongoDB, MySQL, Firebase, NoSQL, Netlify
Data Science & Analytics: NumPy, Pandas, Jupyter, Data Mining, Data Analysis, Statistical Modeling
DevOps & Infrastructure: Git, Docker, Airflow, Linux/Unix, HPC, CUDA, DevOps, AWS, CI/CD (GitHub Actions)
Design & Tools: Figma, Agile, Data Structures & Algorithms, Discrete Math, Computer Networking, Cryptography
Languages: French and Somali