Aditya Jayanth Vadali

Severn, MD | adityajayanthvadali@gmail.com | LinkedIn | +1 (585) 957-2989

Internships and Work Experience

Technology Intern(ML) at Fidelity Investments (FI-AICOE)
Boston, MA **June 2024 – August 2024**
Technologies Used: Python · AWS Sagemaker/EC2 · Streamlit · scikit-learn · seaborn · Snowflake · MySQL

- Collaborated on the development of a **Sales Prospecting Tool** for Fidelity Institutional sales teams, leveraging **Python**, **Streamlit**, **AWS Sagemaker** to enable data-driven targeting of financial advisors.
- Wrote complex **SQL queries** to extract and isolate advisor activity patterns and then built **interactive advisor journey visualizations** for 10,000+ firms using **Streamlit** and **Seaborn**, translating raw engagement data into actionable sales insights.
- Replaced a legacy heuristic-weighted revenue scoring system with a **logistic regression model** (scikit-learn), improving forecast accuracy by **25%** through hyperparameter tuning and validation.
- Conducted **statistical analysis** to validate model predictions, including accuracy and confidence intervals, and deployed the final solution using **AWS SageMaker** for training and **EC2** for hosting.
- Authored SQL queries and Python scripts to standardize and automate data flow processes from external sources.

Customer Experience Technology Intern at Fidelity Investments, India (PI-CKSI)
Bengaluru, KA, India **January 2023 – July 2023**
Technologies: MySQL · Snowflake · Tableau · Qliksense · Python

- Led data migration of 5 key customer experience indicators from **Oracle** to **Snowflake** using **Python scripts** with Oracle and Snowflake connectors, ensuring full data integrity and accessibility.
- Consolidated **40M+ rows** into a centralized Snowflake table, streamlining CX reporting and enabling scalable, performant analytics.
- Built a **centralized backend repository** for CX programs with optimized **SQL**, reducing query time by 40%.
- Designed and launched the CX Data Lab Sandbox in **Tableau**, allowing advanced filtering, hypothesis testing, and real-time access to key metrics.
- Cut quarterly reporting effort by **200+ man-hours** through **self-serve Tableau dashboards** with granular, role-specific insights.

Education

- **Master's in Science (M.S in Data Science)** **September 2023 - May 2025**
  Rochester Institute of Technology, Rochester, New York
  Key Courses: Database Design Implementation, Data Driven Knowledge Discovery, Software Construction, Software Engineering for Data Science, Non Relational Data Management, Applied Data Science I & II,Foundations of Data Science and Analytics, Visual Analytics, Software Construction

- **Master's in Science (MS in Business Analytics)** **October 2021 - June 2023**
  Manipal Academy of Higher Education (MAHE), Manipal, India **GPA: 8.09/10**
  Key Courses: Applied Multivariate Data Analysis, Prescriptive Analytics, Analytics using Python, Time Series Analytics, AI and ML, Data Warehousing and Mining, Financial Analytics, Minor Project I & II

- **Bachelor's in Science Economics, Mathematics and Statistics** **June 2018 - June 2021**
  Christ (deemed to be University), Bangalore, India **GPA: 3.15 / 4**
  Key Courses: Descriptive Statistics and probability Theory, Statistical Methods, R Programming, Statistical Inference, Mathematical Modelling using Python, Sampling Techniques, Linear Regression Models

Technical Skills

- **Programming:** Python, Java, Javascript, HTML R, SQL, MongoDB, bash
- **Data Engineering and Analytics:** Snowflake, Neo4J, Oracle DBMS, PySpark, Airflow, ETL, MS Excel
- **AI**: Open AI API, RAG Models, Vector Databases, FAISS, pickle
- **Machine Learning:** scikit-learn, XGBoost, TensorFlow, Keras, PyTorch, Transformers, pandas, NumPy, DEAP
- **Visualization:** Tableau, PowerBI, QlikSense, plotly, matplotlib, seaborn, ggplot, Streamlit, altair
- **Project Management:** Jira, Confluence, GitHub, GitLab
- **Cloud:** AWS EC2, AWS Lambda, AWS Redshift, AWS SageMaker
- **Machine Learning:** scikit-learn, XGBoost, TensorFlow, Keras, PyTorch, Transformers, pandas, NumPy, DEAP
- **Statistics:** JMP, statsmodels, prophet, arch, sktime, scipy.stats, linearmodels

Products

- A.L.I.N.A (All In Nutritional Assistant) **(2024 August - Present)**
  Co-*Creator and Lead Developer of an AI-powered nutrition tracker tailored for RIT Dining system*
  Technologies: Python · Flask · MySQL · FAISS · DEAP · LangChain · OpenAI API · JavaScript · Bootstrap · Selenium · Pandas
  - **Conceptualized, designed, and led development** of a personalized nutrition platform improving student health by bridging the gap between complex campus dining systems and nutritional awareness.
  - **Architected and implemented a modular, Flask-based web application** with a tri-database schema (MySQL) for user profiles, dynamic food tracking, and nutritional menu data, supporting scalable, per-user data storage and secure session management.
  - **Developed robust ETL pipelines** using Selenium, Pandas, and Regex for structured ingestion and transformation of RIT dining menus, USDA branded food databases, and franchise nutrition data into normalized relational tables with allergen and macronutrient breakdowns.
  - **Engineered a custom food recommendation engine** using the DEAP evolutionary algorithm library; designed a multi-objective fitness function to optimize food choices based on user-specific calorie, protein, carbohydrate, and fat targets.
  - **Integrated a Retrieval-Augmented Generation (RAG) chatbot** using FAISS, LangChain, and OpenAI APIs to support real-time, personalized natural language queries, with context-aware allergen warnings and semantic food item matching.
  - **Deployed the application on AWS EC2**, ensuring reliable hosting and enabling 24/7 access.

Projects

1. **Batch ETL & Forecasting Pipeline – Candy Store Sales**
   Technologies: PySpark · Apache Spark · MySQL · MongoDB · Apache Airflow · Prophet
   - Built a PySpark batch ETL pipeline to process 10 days of JSON orders, integrating MySQL/MongoDB into normalized fact/dimension tables with real-time inventory validation.
   - Generated daily KPIs, order summaries, and inventory snapshots; automated forecasting with Prophet and evaluated results via MAE/MSE.
   - Orchestrated the workflow using CI-compliant Airflow DAGs with production-ready standards.

2. **Healthcare Data Normalization Pipeline**
   Technologies: Databricks· Apache Spark · Delta Lake · MySQL · Tableau · Dimensional Modeling
   - Engineered a batch ETL pipeline in **Databricks** using **Apache Spark and Delta Lake** to normalize raw healthcare data into a Snowflake schema (10 dimension, 1 fact table), supporting scalable analytics on visits, labs, treatments, and billing.
   - Implemented modular notebooks to extract key entities (e.g., insurance plans, prescriptions), resolved nulls/duplicates, generated hashed surrogate keys, and enforced PK/FK integrity across Delta tables.
   - Integrated data from **MySQL and flat files into Databricks**, exported curated datasets to Snowflake and CSV, and built Tableau dashboards for audit teams to analyze patient demographics, visit trends, and payer mix.
   - Optimized Databricks job performance via cluster autoscaling and task parallelization, improving data load efficiency and dashboard refresh times.

3. **Credit Card Transaction System – Lambda Architecture**
   Technologies: Kafka · PySpark · MySQL · Python
   - Designed a Kafka-based Lambda Architecture pipeline to simulate 4 days of real-time credit card transactions, combining stream validation and end-of-day batch reconciliation.
   - Implemented rule-based transaction checks (amount thresholds, geo-distance, credit limits), tracked pending balances in real time, and recalculated credit scores with dynamic limit adjustments.
   - Orchestrated the pipeline with flag-based triggers, integrated MySQL for persistent storage, and exported validated results to CSV and relational tables for reporting and analysis.

4. **Fake Job Posting Detection (NLP Pipeline)**
   Technologies: Python · scikit-learn · TF-IDF · SGDClassifier
   - Built an end-to-end binary text classification pipeline to identify fraudulent job postings using TF-IDF vectorization and an SGDClassifier, optimized via RandomizedSearchCV with 200 iterations.
   - Engineered feature preprocessing using 1–5 n-grams over the “description” and “requirements” fields; applied class balancing and hyperparameter tuning to improve minority-class detection.
   - Achieved an F1 score of 0.79 on the fraudulent class within a 2.3-minute runtime cap; implemented a custom evaluation framework for precision, recall, and F1 metrics using Python.