ADITYA KANTH MANNE

(813) 357-4409   adityamanneqr@gmail.com   Florida, USA

SUMMARY

Experienced Data Science professional with experience in statistical modeling, machine learning, and large-scale empirical research using real-world, high-noise datasets. Experienced in building end-to-end research pipelines for feature engineering, model estimation, validation, and stress testing under regulatory and operational constraints. Proven ability to apply optimization, probabilistic modeling, and scenario analysis to decision systems involving risk, uncertainty, and policy-driven regimes. Technically fluent in Python, SQL, and distributed computing, with hands-on experience translating complex data into interpretable, decision-relevant insights. Seeking research-oriented roles and doctoral study focused on quantitative finance, econometrics, market dynamics, and risk modeling.

SKILLS & TECHNOLOGIES

- Python, C++, OOP, Data Structures & Algorithms, API Development, Microservices, System Design, Scalable Backend Engineering, Machine Learning (supervised/unsupervised, deep learning), Time-Series Modeling, Feature Engineering, Model Evaluation, Statistical Analysis, Distributed & Parallel Computing, Scikit-learn, TensorFlow, NumPy, Pandas; anomaly detection; regression/classification; forecasting, ETL/ELT Pipelines, Data Pipelines, Cloud Architecture (AWS, Azure), High-Performance Computing, CI/CD, Testing, Version Control, Agile Development, Data Visualization & BI.
- NumPy, Pandas, SciPy, Scikit-learn, TensorFlow, PyTorch, Statsmodels, Matplotlib, Seaborn, SQL Server, PostgreSQL, MySQL, MongoDB, Redis, Kafka, Airflow, Docker, Kubernetes, AWS (EC2, Lambda, S3, CloudWatch), Azure Functions, Azure Data Factory, Git, GitHub, GitLab, Jenkins, GitHub Actions, Tableau, Power BI, QuickSight, Jupyter Notebook, VS Code, PyCharm, Postman, JIRA, Confluence, Linux/Unix Tools.

EXPERIENCE

ELCFV   **Jul 2024 - Present**
Junior Data Scientist   *Remote, FL*

- Faced with inconsistent historical enrollment snapshots caused by full-batch overwrites, rebuilt the nightly ingestion workflow into a Python-based incremental loader with temporal versioning, reducing load time by 40% and improving time-series fidelity for modeling and backtesting.
- To quantify drivers of program attrition, engineered time-windowed behavioral and demographic features and trained logistic regression and random forest models to estimate enrollment drop-off risk, achieving a 19% lift in ROC-AUC over rule-based baselines.
- When nonlinear enrollment patterns were missed by traditional thresholds, evaluated multiple supervised ML models (logistic regression, gradient boosting) and selected stable estimators using cross-validation and cohort-based holdouts, reducing false-negative risk identification by 23%.
- When nonlinear enrollment patterns were not captured by static thresholds, evaluated multiple supervised ML models (logistic regression, gradient boosting) using 5-fold cross-validation, reducing missed high-risk cases by ~400 enrollments per cycle.
- Built an anomaly detection framework combining rolling statistical bounds with Isolation Forests, flagging 20–30 high-impact provider anomalies per month that were previously undetected until after reporting cycles.
- Built an anomaly detection framework combining rolling z-score statistics with Isolation Forests to detect abnormal provider-level behavior, reducing false positives by 28% and surfacing material operational issues weeks earlier.
- To prevent noisy upstream data from biasing ML outputs, implemented pre-model validation gates using statistical distribution checks and feature completeness thresholds, cutting data-quality-driven model failures by 60%.
- Developed a config-driven feature engineering pipeline (Python + SQL + YAML) that standardized temporal aggregates, ratios, and volatility metrics across CRM, ERP, and enrollment systems, reducing feature-rebuild effort by 45% and ensuring reproducible retraining.
- Integrated trained ML models into scheduled batch-scoring pipelines, generating probabilistic risk scores for 100K+ records per cycle, enabling consistent downstream consumption by dashboards and decision-support tools.
- Implemented model performance monitoring tracking ROC-AUC stability, prediction variance, and feature drift across monthly windows, enabling early retraining and reducing post-deployment degradation incidents by 30%.
- Faced with multi-hour preprocessing latency limiting research iteration, refactored loop-based feature calculations into vectorized Pandas and SQL aggregations, reducing end-to-end runtime from 4 hours to under 50 minutes.
- Designed rule-based constraint layers encoding regulatory eligibility thresholds as deterministic regimes, ensuring ML-driven recommendations remained compliant while improving coverage efficiency by 17%.
- Built internal CLI utilities to automate dataset extraction, feature refreshes, and validation checks, cutting new-researcher onboarding time by 50% and accelerating model experimentation cycles.
- During audit reviews, documented feature definitions, modeling assumptions, validation logic, and retraining criteria, supporting explainable and defensible quantitative decision-making.

Smart Grow Infotech   **Feb 2024 - Jul 2024**
Business Intelligence Engineer   *Tampa, FL*

- Built a churn-risk prediction pipeline using logistic regression and gradient boosting by merging behavioral logs from the Customer Engagement Platform, billing data from the Financial Ledger, and event histories from the Data Lake, increasing early-risk detection accuracy by 22%.
- Engineered daily rolling features by designing an Airflow-powered feature store that refreshed demographic, transactional, and time-windowed attributes, improving model stability across releases.
- Developed a real-time inference API deployed through the organization's API Gateway, enabling downstream systems to consume predictions without manual batch uploads or CSV dependencies.
- Converted disparate Jupyter notebooks into structured ML projects with MLflow tracking, reproducible configs, and automated artifact storage in the Model Registry, standardizing experimentation.
- Built automated model drift monitors that logged distribution shifts to the company's Monitoring & Alerting Stack, allowing for early retraining before performance dropped.
- Replaced spreadsheet-based forecasting processes with ARIMA, XGBoost, and Prophet models, pulling data directly from the Enterprise Data Warehouse, improving forecast accuracy for strategic planning.
- Constructed a batch-scoring framework that scored millions of records nightly by leveraging Spark jobs in the Cloud Data Lake, reducing runtime by 70%.
- Automated hyperparameter tuning with Bayesian Optimization pipelines tied to the Experimentation Cluster, eliminating inconsistent manual experimentation.
- Designed a cross-validation framework for multi-region datasets stored in the Regional Data Hubs, ensuring fair evaluation across demographic distributions.
- Created SHAP and LIME explainability layers that integrated into the Business Intelligence Portal, enabling non-technical teams to understand model outputs.
- Implemented automated retraining triggers based on live-data performance metrics sourced from the Operational Metrics Dashboard, ensuring continuous learning.
- Developed NLP-based classification models to analyze feedback submitted through the Customer Support System, grouping themes and reducing manual tagging workload.
- Designed transformation pipelines that normalized messy third-party vendor datasets before they entered the Analytics Warehouse, improving model input consistency.

Vesper Distributions   **Nov 2020 –Nov2021**
Junior Analyst   *Hyderabad, India*

- Automated manual spreadsheet workflows by converting multi-tab Excel models into Python-based scripts integrated with the Internal Reporting Portal, eliminating calculation inconsistencies between teams.
- Built a file-ingestion validator that checked formats, headers, numeric fields, and date ranges before loading data into the Operational Data Store, reducing data load errors by 60%.
- Wrote Python wrappers to simplify connections to SQL Server, MySQL, and Postgres, enabling junior analysts to run parameterized queries without deep SQL knowledge.
- Automated daily SFTP pulls from state partner systems and transformed them into normalized tables within the Enterprise Database, removing reliance on manual downloads.
- Replaced slow Python scripts with vectorized Pandas logic, improving performance for large datasets stored on the Shared Analytics Drive.
- Built Jupyter-based data profiling notebooks that automatically generated summary statistics, quality checks, and correlation matrices for datasets pulled from the Data Warehouse.
- Developed lightweight FastAPI microservices to expose calculation logic previously trapped inside Excel macros, enabling consistent use across multiple business teams.
- Introduced configuration-driven scripts using YAML to eliminate hardcoded paths and credentials, ensuring smoother deployment across dev, test, and production environments.
- Created batch-processing utilities able to merge thousands of CSVs from the Document Exchange System, reducing manual file preparation time.
- Developed automated email-notification scripts via Python to report pipeline results, replacing tedious manual communications.
- Implemented structured logging and traceback capture for ETL scripts feeding the Data Lake, improving issue diagnosis speed.
- Documented workflow diagrams and script catalogs detailing how data flowed across CRM, ERP, and third-party vendor systems.
- Conducted peer training sessions to help new hires understand Python syntax, debugging patterns, and company workflows.
- Participated in sprint grooming meetings by breaking down vague user stories into actionable development tasks with clear technical acceptance criteria.
- Presented weekly progress updates outlining blockers, completed deliverables, and dependency risks to maintain alignment across the engineering team.

PROJECTS

Intelligent Log Analytics & Root-Cause Classification Platform
- Developed a log analytics engine that parsed and standardized 2+ billion log lines from microservices, load balancers, and cloud infrastructure using Python, regex pipelines, and spaCy-based NLP tagging.
- Applied clustering and topic modeling (LDA, BERTopic) to group recurring log signatures, reducing manual root-cause analysis time by 50%.
- Built a supervised model (XGBoost + TF-IDF embeddings) to classify log events into 20+ incident categories, achieving 87% accuracy in automated incident tagging.
- Integrated the system with AWS Lambda + CloudWatch to generate real-time alerts, cutting incident triage time by 30–40%.

Customer Behavior Segmentation & Personalization Engine
- GMM, and hierarchical clustering applied to 500K+ user behavior records sourced from CRM, billing, and activity logs.
- Engineered 80+ behavioral features (recency-frequency patterns, clickstream paths, session metrics, purchase velocity), improving clustering separation scores (Silhouette +0.21).
- Designed a recommendation framework using cosine similarity and association rules (Apriori/FPGrowth) to generate personalized suggestions with a 27% improvement in engagement.
- Implemented a real-time scoring pipeline using FastAPI + Redis, enabling web/mobile apps to retrieve live segment IDs and recommendations under 30ms latency.

Financial Forecasting & Optimization Engine
- Created a forecasting engine for revenue, demand, and cost projections across 12 business units using SARIMA, Prophet, and LSTM sequence models, improving forecast accuracy by 18–22%.
- Developed a Monte Carlo simulation layer running 100,000+ economic scenarios (CPI shifts, interest-rate shocks, seasonality disruptions) to quantify financial uncertainty and risk exposure.
- Built a linear programming optimization model using OR-Tools to recommend budget allocations and resource mix optimizations, reducing overspending variance by 25%.
- Orchestrated ETL pipelines pulling data from SQL Server, Snowflake, and API-based financial systems into AWS S3 + Glue, reducing data prep time by 60%.

Ad Performance Anomaly Detection & Factor Sensitivity Engine
- Real-time anomalies in ad P&L went undetected, so I built ensemble z-score and IQR-based models, flagging 93% of issues within 5 seconds.
- Without campaign-level Greeks, I computed 1st/2nd-order sensitivities across 500 ads, enabling 12% faster attribution of performance changes.
- Lacked factor-based dashboards, so I developed P&L visualization tools with breach alerts, increasing advertiser confidence and retention by 11%.

EDUCATION

University of South Florida-Tampa   **Jan 2022 - Dec 2023**
Master of Science-Engineering Management
- **GPA:** 3.83/4

Jawaharlal Nehru Technological University   **Aug 2016 - Nov 2020**
Bachelor of Technology-Mechanical Engineering
- **GPA:** 3.45/4