Mohit Appari

Tallahassee, FL | P: +1 8503001952 | mohitt.appari@gmail.com | www.linkedin.com/in/moh1tt |moh1tt.xyz |github.com/moh1tt

SUMMARY

Data Scientist specializing in demand forecasting, time series modeling, and experimentation design. Built and validated a multi-domain probabilistic forecasting system (DeepAR/Temporal Fusion Transformers) with quantile forecasts and cold-start handling for new products and grid nodes, plus a causal inference platform modeling marketplace supply-demand dynamics and interference effects in A/B testing. 3+ years of production ML/ SDE experience (40+ automated pipelines, 300K+ records, 83% latency reduction) grounding research-level modeling in deployable systems.

WORK EXPERIENCE

APD Florida Tallahassee, FL
Data Scientist Jun 2025 - Current

- Designed and implemented Python/SQL-based analytical workflows to transform ambiguous operational questions into structured statistical analyses, automating SPSS execution and validation to support evidence-based decision-making across programs.
- Built and maintained 40+ automated ETL pipelines processing 300K+ records (Python, SQL Server, Azure Data Warehouse, Docker), implementing data validation and lineage tracking to ensure reliability for downstream statistical and trend analysis.
- Conducted time-series trend and anomaly analysis on operational metrics (client intake volume, application throughput, financial/rate trends) to detect spikes and shifts linked to system changes, translating findings into actionable insights for 50+ stakeholders via Power BI and Tableau dashboards.
- Optimized SQL query performance through indexing, partitioning, and caching, reducing analytical report latency by 83% (60 min to under 10 min), enabling faster iteration on data-driven analysis.

Florida State University Tallahassee, FL
Research Assistant Jun 2024 - Sep 2024

- Developed and benchmarked deep generative modeling approaches (GANs, VAEs, RNNs) to synthesize large-scale molecular datasets for lithium-ion battery materials, evaluating architecture tradeoffs for predictive performance on structured scientific data.
- Benchmarked and fine-tuned deep learning models (NequIP) via transfer learning across an 11,000+ material dataset, improving prediction consistency by 15% through iterative experimentation and hyperparameter optimization.
- Designed and executed controlled experiments across multiple model architectures using A/B testing and statistical comparison methods, tracked via MLflow, ensuring rigorous, reproducible model evaluation and informing experimentation design practices.

SP Global Bangalore, IN
Software Developer Jan 2022 - Jun 2023

- Built a production-grade data extraction platform processing 10,000+ documents/day, combining OCR and CNN-based classification models to convert unstructured PDFs and scanned documents into structured datasets for downstream analysis and modeling.
- Engineered cloud-based data pipelines (AWS S3, Athena, BigQuery) to ingest, transform, and serve large-scale datasets, enabling analysts and downstream models to access up-to-date structured data within minutes.
- Designed a modular, reusable Python extraction framework with CDC-based updates and 80% test coverage, enabling reliable, scalable data delivery across multiple business domains.
- Developed REST API integrations with schema validation and CI/CD (GitHub Actions), improving data quality and deployment reliability for downstream analytics consumers.

LG Bangalore, IN
Software Developer Intern May 2021 - Aug 2021

- Built a computer vision and IoT prototype for logistics monitoring (YOLO, REST APIs, PostgreSQL, Raspberry Pi), implementing real-time tracking and driver drowsiness detection (~90% accuracy); secured 3rd place, LG Campus Ideathon.

ITC Bangalore, IN
Software Developer Intern Jan 2021 - Mar 2021

- Built an inventory management system (Flask, MySQL, MERN stack) with role-based authentication and automated workflows to improve data management efficiency.

PROJECTS

Multi-Domain Probabilistic Demand Forecasting Engine | Python, PyTorch, M5/Walmart dataset, PJM/ENTSO-E energy data

- Built a global DeepAR model jointly trained across 2,000+ retail time series with P10/P50/P90 quantile forecasts (pinball/WQL loss) and a leakage-tested cold-start holdout (verified via automated leakage tests) showing cold-start series performing on par with warm-start series; translated forecasts into a business simulation showing a P90 inventory policy cut stockout rate from 23.9% to 1.9% at ~6x the holding cost.

Causal Inference & Experimentation Platform for Marketplace Dynamics | Python

- Built a two-sided marketplace simulator with a known ground-truth causal effect to quantify how naive A/B testing degrades under spatial interference (bias growing to -36.6% at high spillover), and validated switchback/cluster-randomized designs that stayed within ~5% bias regardless of spillover strength; also applied diff-in-diff to a non-random rollout, cutting a 1,471% naive-comparison bias to -8.5%.

EDUCATION

Florida State University, Masters in Data Science, Tallahassee, FL Aug 2023 - May 2025
Christ University, Bachelors in Computer Science, Bangalore, IN Aug 2018 - May 2022

SKILLS

Language & Backend: Python, SQL, R, Java, JavaScript, TypeScript, C++, Bash, FastAPI, Flask, Django, REST APIs, Streamlit, React
Databases & Cloud: PostgreSQL, MySQL, MongoDB, Redis, Snowflake, BigQuery, DuckDB, AWS, GCP, Azure, Docker, Kubernetes, Terraform
Data & AI: Time Series Forecasting (DeepAR, Temporal Fusion Transformer, ARIMA, Prophet), Apache Spark, Airflow, dbt, PyTorch, PyTorch Forecasting, TensorFlow, scikit-learn, XGBoost, LightGBM, Pandas, NumPy, LangChain, OpenAI API, Github, Copilot, Claude, Neo4j