Mohammad Salem

Data Science · AI/ML Engineering · Mathematics

⊙ Clifton, NJ   ✉ hello@mohammadsalem.com   in linkedin.com/in/mohammad-salem0   ⑂ github.com/shumude   ⊕ mohammadsalem.com

EDUCATION

Montclair State University   Sep 2022 — Apr 2026
Bachelor of Science in Data Science · Minor in Mathematics
GPA: 3.61 · Dean's List (4x recipient)

EXPERIENCE

Extern — Pfizer Supply Chain Document Processing   Feb 2026 — Present
AI Engineer Extern · Remote
- Participating in the Pfizer Supply Chain Document Processing Externship hosted by Extern, focused on developing production-ready AI solutions for pharmaceutical data extraction.
- Building AI pipelines to process pharmaceutical vendor files using OCR (PyMuPDF, Tesseract, PaddleOCR) and RAG for data extraction and classification.
- Developing a RAG-based document retrieval system with LlamaIndex, optimized with chunk tuning and open-source LLMs (Mistral, Phi-2).
- Conducting end-to-end evaluation of the document intelligence system—benchmarking OCR accuracy, RAG retrieval quality, and routing performance.

Montclair State University   Jan 2026 — Present
Machine Learning Research Assistant · Montclair, NJ · Hybrid
- Collaborating on deep learning research under the guidance of Dr. Debapriya Hazra to advance non-invasive seizure prediction and detection technologies.
- Developing a machine learning classifier to identify pre-seizure physiological patterns by analyzing large-scale EKG/ECG datasets from wearable monitors.
- Engineering and training neural networks on multi-dimensional time-series data to detect subtle autonomic shifts that precede clinical seizure onset.
- Optimizing model performance for real-time inference, focusing on deployment on smartwatch hardware and edge devices.

RWJBarnabas Health — Rutgers Health Hack   Dec 2025 — Jan 2026
Health Innovation Developer · New Brunswick, NJ
- Worked on a 6-person team to design and pitch a mobile solution to reduce treatment delays for severe maternal hypertension.
- Collaborated with the Chief of Obstetrics at RWJ to understand clinical workflow challenges and refine the product concept.
- Co-developed core app features for BP triggers, observation timers, and treatment-window tracking.
- Assisted in building role-based notifications for real-time care-team coordination.
- Helped prepare and deliver a winning pitch and live demo to judges and hospital stakeholders.

Alma — AI Study Companion   Aug 2025 — Nov 2025
Product Developer & Founder · Remote
- Developed and launched an iOS app using AI-assisted development, directing system architecture, feature specifications, and UX flows.
- Designed a content processing pipeline using Whisper transcription and AI models to generate notes, flashcards, quizzes, and podcasts from lectures.
- Built a RAG-based Q&A system enabling users to query their saved content with cited answers, improving retention and study efficiency.
- Managed end-to-end product lifecycle: App Store compliance, monetization strategy, user analytics, and privacy infrastructure.

roundtable&advisory   Jul 2025 — Sep 2025
Data Science Intern · Wayne, NJ · Hybrid
- Researched and evaluated methods for identifying high-fit companies and extracting relevant contacts using AI tools and structured search criteria.
- Designed and tested a multi-provider waterfall enrichment process to obtain accurate contact information, achieving >96% verified emails.
- Reduced projected outreach and list-building costs by >45% by replacing manual processes with a data-driven, tool-based system.
- Documented the workflow, created filtering templates, and produced reports that standardized the targeting and enrichment process.

Signs Of 2000   Jun 2020 — Aug 2025
Data Analyst Intern · Clifton, NJ · Dec 2024 — May 2025
- Cleaned and organized a database of sales and clients. Cleaning clientele data resulted in formatted addresses and contact info for marketing, as well as preprocessing for analysis and prediction.
- Generated a geographical heat map of sales to determine good and poor performing areas in the market, leading to 7+ marketing opportunities.
- Gathered insights on sales trends via data visualization. Generated 7 insights on purchase patterns with 4 visuals.
- Created an ML model to predict the next purchase category using XGBoost and deployed it with Streamlit.

Production Specialist · Jun 2020 — Aug 2025
- Collaborated & consulted with customers to create and finalize signage projects from concept to installation.
- Prioritized and managed multiple tasks operating printers, plotters & other machinery to produce final products.

PROJECTS

NYC Motor Vehicle Collision Analysis — Analyzed 2M+ NYC collision records using K-Means clustering, geospatial mapping, and feature engineering on a 500K-record sample to identify crash patterns and causal factors.

California Violent Crime Analysis — Analyzed 14 years of California violent crime data (50,000+ records) using choropleth maps, heatmaps, and time-series analysis. Applied correlation and clustering techniques to identify crime hotspots and causal drivers across counties.

Tableau Sales Analytics Dashboard — Developed an interactive dashboard with geographic heatmaps, time-series revenue trends, and sales distribution charts for regional performance analysis.

USAID Anti-Corruption Analysis — Extracted key metrics to reveal trends across international anti-corruption projects. Developed visualizations and presented findings to non-technical audiences.

TECHNICAL SKILLS

Languages: Python, SQL, R

Data: Pandas, NumPy, Statsmodels, Seaborn, Matplotlib, Tableau

Domains: NLP, Computer Vision, Deep Learning, Time Series, OCR

ML/AI: TensorFlow, XGBoost, Scikit-learn, LlamaIndex, OpenAI, RAG

Databases: PostgreSQL, Supabase, SQL Server

Tools: Whisper, PyMuPDF, PaddleOCR, Streamlit, Git, AI-Assisted Development