

PATHIK RUPWATE

Data Scientist & Machine Learning Engineer

Mumbai, India | +91 9321194480 | pathikrds@gmail.com | pathikrupwate.com | linkedin.com/in/pathik-rupwate | github.com/pathik10

PROFESSIONAL SUMMARY

Results-driven **Data Scientist and Machine Learning Engineer** with 4+ years of experience architecting and deploying end-to-end predictive systems, Generative AI agentic workflows (Model Context Protocol / FastMCP, LangChain, RAG), and production computer vision pipelines. Proven track record of translating complex AI architectures into measurable business KPIs—driving a **+30% campaign ROI lift**, boosting credit risk anomaly detection by **+25%**, and achieving **78.85% predictive accuracy** on institutional macroeconomic forecasting. Holds an M.S. in Applied Data Science from the **University of Chicago**.

CORE COMPETENCIES & TECHNICAL TOOLING

Generative AI & Agentic Architectures: Model Context Protocol (FastMCP 21-Tool Server), Autonomous ReAct Loops, Groq LPU (Sub-50ms Inference), Retrieval-Augmented Generation (RAG), Vector Embeddings (Cosine Similarity, ChromaDB), LangChain, Google Gemini API, ElevenLabs API, Pydantic Schema Validation.

Machine Learning & Statistical Modeling: Classification, Regression, Time-Series Forecasting, Extra Trees Ensembles, Tabular GANs, SMOTE Imbalance Handling, Anomaly Detection (Isolation Forests, One-Class SVM), Causal Inference, K-Means Clustering, PCA, Silhouette (s=0.68) & Elbow Optimization, SHAP Local Feature Attribution (XAI).

Computer Vision & Deep Learning: MediaPipe 468 3D Landmark Meshes, VGG-16 CNN, MobileNetV2, OpenCV, TensorFlow, PyTorch, Scikit-Learn, Keras, WebGL Client-Side GPU Inference.

Data Engineering & Cloud Infrastructure: Python, SQL, R (ggplot2, caret), Docker, Flask APIs, AWS (S3, EC2 Incident Recovery), GCP, Linux, Git / GitHub CI/CD, MongoDB.

Analytics & Experimentation: A/B & Multivariate Testing, Tableau, Power BI, Executive BI Dashboards.

PROFESSIONAL EXPERIENCE

ABM Knowledgeware

Data Scientist

Mumbai, India
Mar 2026 - Present

- Autonomous Municipal Revenue Agent (FastMCP Upgrade):** Architected and deployed an autonomous enterprise agent leveraging a 21-tool Model Context Protocol (FastMCP) server and sub-50ms Groq LLM inference (Qwen 2.5), executing live ReAct reasoning loops to audit district utility billing, detect high-value commercial arrears, and automate statutory recovery notices across municipal wards.
- Automated Video Localization Pipeline:** Engineered an end-to-end multi-lingual tutorial video generation and dubbing engine in Python using the ElevenLabs API and dynamic audio alignment, completely eliminating manual voiceover production and saving **80+ engineering hours per release cycle**.
- Disaster Recovery & Large-Scale Data Parsing:** Built algorithmic pattern-matching parsers and schema reconstructors to salvage **400,000+ unstructured files** recovered from an AWS storage failure, restoring directory hierarchy, file signatures, and metadata for **80% of the corrupted dataset** to avert enterprise data loss.
- Automated AppSec Vulnerability Triage (OWASP DAST):** Designed an automated security testing pipeline integrating OWASP ZAP with Gemini AI and Pydantic validation schemas (`TriageReport`, `TriagedIssue`), enabling real-time detection of injection vectors (CWE-89) and synthesizing verified code remediation patches.
- Intelligent Document Processing (IDP):** Deployed a deep learning document processing pipeline utilizing fine-tuned MobileNetV2 and OCR to classify scanned legal affidavits at **98.4% confidence** and extract cadastral land parcel acreage for municipal land registry verification.
- Odisha Administrative AI Agent:** Leading the architectural design and deployment of a scalable citizen-service AI agent for the Odisha Government to automate regional administrative workflows for millions of residents.

PROFESSIONAL & ACADEMIC EXPERIENCE (CONTINUED)

Kearny Bank
Data Scientist

Austin, TX

Oct 2023 - Jan 2026

- Customer Segmentation & Marketing Lift:** Led unsupervised clustering (K-Means, DBSCAN, PCA) validated via Silhouette analysis (s=0.68) and Elbow WSS optimization (-68.8% inertia reduction) across retail banking personas; integrated Latent Dirichlet Allocation (LDA) and TF-IDF topic modeling on customer feedback to drive a **+30% advertising campaign ROI lift**.
- Loan Default Anomaly Detection:** Developed and deployed unsupervised Isolation Forests and One-Class SVM models to identify rare credit risk anomalies under severe class imbalance, boosting early default forecast detection by **+25%**.
- Explainable AI (XAI) & Risk Attribution:** Implemented local SHAP feature attribution and counterfactual optimization algorithms, enabling credit officers to decompose anomaly scores into marginal risk drivers and generate actionable "path to approval" criteria for review-tier applicants.
- A/B Testing & Executive BI:** Designed and analyzed randomized A/B and multivariate tests across product feature rollouts; engineered live interactive executive dashboards in Tableau and Power BI visualizing conversion funnels, transaction velocity, and portfolio risk distributions.

University of Chicago
Data Science Researcher (Deep Learning & Computer Vision)

Chicago, IL

Jan 2023 - Jun 2023

- Live Biometric Face-to-BMI Prediction Pipeline:** Engineered an end-to-end computer vision pipeline combining MediaPipe 468-point 3D facial mesh extraction with 22 calibrated anthropometric ratios and a fine-tuned VGG-16 convolutional neural network regression head.
- Production Serving & Inference Optimization:** Deployed the model via a high-throughput Flask API and optimized video stream processing using OpenCV and TensorFlow GPU acceleration, reducing inference latency to sub-20ms.

Aiolux
Data Scientist (Quantitative Research & Forecasting)

Chicago, IL

Jan 2022 - Dec 2022

- S&P 500 Sector Forecasting Capstone:** Conducted quantitative research supervised by Roger Moore, MBA, predicting sector outperformance across all 11 S&P 500 sectors over 50, 100, and 200-day horizons; achieved a benchmark accuracy of **78.85%** utilizing Extremely Randomized Trees (Extra Trees).
- Synthetic Data Generation with Tabular GANs:** Synthesized 25,000+ macroeconomic datapoints using Tabular Generative Adversarial Networks (T-GANs) and SMOTE to resolve extreme regime imbalance, increasing model out-of-sample robustness by **+20%**.
- Macroeconomic Regime Stress-Testing:** Quantified causal sensitivity of sector returns against CPI inflation shocks and Federal Reserve interest rate shifts, modeling defensive rotations into Energy/Staples during high-inflation regimes.

Early Engineering Roles - Beloit College
Database Engineer (Powerhouse) & Software Engineer (Open Energy Dashboard)

Beloit, WI

Jan 2020 - May 2021

- Database Engineer (Beloit Powerhouse):** Designed and deployed a SQL/Python relational inventory tracking system managing 5,000+ assets with automated reporting pipelines.
- Software Engineer (Open Energy Dashboard):** Built interactive time-series visualizations and multilingual modules for an open-source energy analytics platform using Docker, React, and Linux.

SELECTED TECHNICAL PROJECTS

- Autonomous Multi-Tool Agent Studio (FastMCP & ReAct):** Full-stack agent sandbox demonstrating autonomous tool planning, standard FastMCP protocol JSON dispatch across 21 tools, and structured Pydantic synthesis with sub-50ms Groq LLM inference.
- In-Browser Machine Learning Labs & Interactive Portfolio:** High-performance portfolio featuring 5 client-side ML demonstrations: live WebGL MediaPipe Face-to-BMI inference, interactive SHAP waterfall loan anomaly scorer, S&P 500 GAN simulator, 2D semantic embedding RAG visualizer, and dual-mode cohort clustering.
- Automated DAST AppSec Scanner:** Python & OWASP ZAP vulnerability scanner featuring AST pattern matching and LLM-assisted remediation generating validated code diffs for CWE security flaws.
- AI-Powered Portfolio Chatbot:** In-browser semantic knowledge engine paired with Google Gemini 2.0 Flash / 1.5 Flash API integration for live contextual retrieval and conversational Q&A.

EDUCATION & LEADERSHIP

University of Chicago
Master of Science in Applied Data Science

Chicago, IL

Jun 2023

Beloit College
Bachelor of Science in Computer Science & Quantitative Economics

Beloit, WI

May 2021

Leadership: *President, Beloit Student Government* - Allocated and managed a \$250,000 annual funding budget for campus organizations, student initiatives, and capital investments.