

SUYASH

Data Scientist | AI Architect

[LinkedIn](#) | [GitHub](#) | [Portfolio](#)

suyashkumarthakur@gmail.com ❖ +1(469) 428-6430 ❖ Dallas, Texas, USA

Data Scientist & AI Architect with 8+ years delivering \$250M+ in validated business impact. I specialize in exploring uncharted territory—generating novel AI architectures, rapidly prototyping them into production systems, and proving their business value before anyone else has a blueprint. I am equally equipped to engineer the production pipelines and aggressively scale proven systems from 1 to 100.

WORK EXPERIENCE

Albertsons

Feb 2024 – Present

Data Scientist IV(Principal Platform Architect)

Pleasanton, CA

- **Enterprise Platform Adoption:** Democratized causal data science by pioneering an enterprise-wide experimentation infrastructure across **4 distinct business units** (In-Store, Pharmacy, Digital, and Private Brands); scaled cross-functional testing capacity from 3 to 8+ concurrent experiments/week, driving **\$150M+ in decentralized annual business decisions**.
- **Statistical Standardization:** Operationalized an automated distributed Spark and CUPED variance reduction pipeline enterprise-wide to achieve a **35% error reduction**; compressed experimental cycle times by 40% (from 28 to 12 days) for all cross-functional product and analytics teams.
- **Data Democratization:** Designed and deployed a custom GenAI Analyst Agent (SQL+RAG via LangChain and Gemini) over a **50TB+ enterprise dataset**; empowered non-technical stakeholders to translate natural language queries into verified SQL, reducing post-experiment analysis latency from days to hours.
- **Operational Velocity & Scaling:** Built an Auto-Summary Agent in partnership with business leadership to autonomously capture experiment telemetry and benchmark results; unlocked **240–400 hours/week of collective organizational bandwidth** by automating executive stakeholder reporting loops.
- **Algorithmic Optimization:** Redesigned the core assignment algorithm from the ground up to deliver 100% reliable, **30x faster randomization** by minimizing Mahalanobis distance as the loss function.

Applied AI Research Lab

Jan 2023 – Present

Principal Engineer & Product Architect

Remote, US

- **Architecture Incubation:** Founded an applied AI research and development lab to proactively build and stress-test emerging technologies and extract proven architectures to **de-risk enterprise deployments**.
- **ClinicOS (SaaS Deployment):** Engineered an AI-powered healthcare management SaaS platform currently live in **8 clinics across rural India**; conducted structured workflow mapping with physicians to implement real-time consultation transcription, automated prescription generation, and automated inventory pipelines.
- **QuizBeef (RAG Production):** Architected an AI-native document comprehension platform scaling to **700+ users**; synthesized learning science literature to deploy LLM semantic parsing and FAISS-powered vector retrieval for adaptive multi-source question generation.
- **mana-health (Agentic Systems):** Developed an advanced multi-agent biometric orchestration platform implementing **Critic-Reflexion loops**; enabled autonomous agent self-correction and output verification across specialized physiotherapy and nutrition nodes.

Discover Financial Services

Aug 2022 – Oct 2023

Senior Data Science Analyst

Chicago, IL

- **Overtaken decade-old risk policy:** Used XGBoost + Propensity Score Matching to prove employment verification had zero predictive value for high-FICO borrowers ($R^2 < 0.01$), then designed surgical A/B test that unlocked \$114M annualized volume and 300+ incremental customers/month through friction reduction.
- **Enabled \$1.8M in loan approvals :** Identified and resolved critical iOS bugs through funnel analysis on the Automated Loan Approval platform, collaborating with engineering and UI/UX teams to restore conversion flow.
- **Infrastructure Modernization:** Migrated legacy SAS workflows to GCP BigQuery, reducing infrastructure costs by 12% and improving decision speed by 10+ hours/week.

Vodafone

Jun 2016 – Feb 2019

Assistant Manager, Data Science

Pune, IN

- **Global Campaign Optimization:** Orchestrated data-driven email personalization strategies for a **45M+ user base in Vodafone Germany**; designed an advanced predictive logistic regression and A/B testing framework that expanded CTR by 28% (from 1.4% to 1.8%) and **generated \$11.2M in incremental revenue**.
- **Customer Lifetime Value Expansion:** Developed complex linear regression models to predict multi-tier retention curves across varying subscription price points; strategically redirected marketing spend to optimize customer lifetime value (CLV) by **5%**.

Senior Executive, Data Science

- **Enterprise Metrics Governance:** Partnered with executive product leadership to instrument **15+ custom performance metrics**; engineered automated SQL pipelines feeding core Tableau dashboards to stabilize long-term feature evaluation and model health tracking.
- **Production System Reliability:** Engineered an automated time-series anomaly detection system (Prophet) to monitor model error rates across core infrastructure; **reduced Priority-2 production incidents by 60%+** and reclaimed ~20 hours/week of core engineering bandwidth.

Kaizen

Jun 2021 – Dec 2021

Data Science Engineering Intern

Atlanta, GA

- **Hybrid Predictive Modeling:** Outperformed the existing production baseline by **improving forecasting accuracy by 8.5%** for Toyota vehicle sales; architected a hybrid time-series model combining Prophet and Long Short-Term Memory (LSTM) deep learning networks to isolate complex residual seasonality
- **Security Risk Mitigation:** Mitigated authentication friction by constructing an automated anomaly detection pipeline on AWS SageMaker; engineered a dynamic risk-based thresholding system that **slashed 2FA false positives by 13%** without degrading user sign-up flows.

EDUCATION

The University of Texas at Dallas

Jan 2021 - Jun 2022

MS, Business Analytics

Dallas, TX

- Dean's Scholar; President, Data Science Club — led workshops and speaker series.
- Top Student Mentor for Data Science track subjects. Answered 800+ questions and coached 70+ students.

Pune University

Jun 2016

BS, Computer Science and Engineering

Pune, IN

SKILLS

GenAI & Agentic Systems: LangChain, RAG (Retrieval-Augmented Generation), Context Engineering, LLMs, Prompt Engineering, Autonomous Agents, SQL+RAG, dbt, Snowflake

Causal Inference & Experimentation: Causal Inference, CUPED, Bayesian Methods (PyMC), Propensity Score Matching, Synthetic Controls, Heterogeneous Treatment Effects (HTE), Uplift Modeling, A/B Testing, SRM Detection

Data & Infrastructure: Python, SQL, PySpark, BigQuery, Airflow, Spark, GCP (Vertex AI), AWS (S3, SageMaker), Docker, Pandas, NumPy, Scikit-learn, TensorFlow

Visualization & Tools: Tableau, Streamlit, Git, Azure