RAJESH SHARMA

linkedin.com/rajeshsharma98 | sharmarajesh98@outlook.com | github.com/rajeshsharma98

WORK EXPERIENCE

Founding Data Scientist & Software Engineer | Upper Hand AI   **May 2023 - Present**

- Implemented LLM-powered NLP pipelines for keyword extraction, topic clustering, and semantic labeling that streamlined the analysis of 50K customer emails, reducing manual workload by 20+ hours weekly.
- Engineered a CLV scoring framework by building interpretable customer matrices integrating revenue, retention, and engagement features, uncovering behavioral patterns, and increasing user engagement by 20%.
- Architected 5 Python libraries and ETL pipelines for automated ingestion from 12+ APIs (incl. Meta), building resilient, token-authenticated systems with retry logic, cutting manual effort by 60% and saving $10K+ monthly.
- Developed KPI dashboards with drill-down analytics and sandboxed views, managing 45+ ETL scripts across 6+ clients to automate reporting, saving 10+ hours/week, and delivering real-time business insights.
- Established data engineering standards for 200+ tables, deployed geo-analytics on 10M+ records, unlocked $20K/quarter via compliant monetization through Snowflake Marketplace, and scaled reporting systems to drive growth and profitability.

Founding Full Stack & ML Engineer | Smart Charity & Project 990 (NPO)   **Jan 2023 - Present**

- Spearheaded a 9-member team to architect a multi-tenant AWS platform with OAuth and CI/CD, cutting deployment time by 50% and boosting nonprofit fundraising efficiency by 25% via analytics dashboards for 5K+ users.
- Built semi-supervised NLP pipelines with anomaly detection, sub-categorization, and LLM-assisted labeling for 700K+ records, reducing false positives by 15% and improving feature representation via vector embeddings.
- Architected and migrated from Qualtrics to OceLab.org, a scalable AWS-hosted PWA with rule-based forms, MongoDB, and cron-based reporting, supporting 3K+ users and saving $20K annually.

Backend Software Engineer | Indiana University Bloomington   **Jan 2023 - May 2023**

- Designed fault-tolerant, resumable data extraction workflows with checkpointing, retries, monitoring, and JSON-to-Excel automation, ensuring 100% completion on multi-GB datasets and reducing text processing time by 50%.
- Automated OCR data retrieval by building HathiMine, a Python library with clean architecture, reusable modules, API integration, parallel processing, and error monitoring, cutting fetch time by 70% across 10K+ pages.

Data Scientist | Indiana University Bloomington   **Aug 2022 - Dec 2022**

- Engineered an NLP pipeline with web scraping, transformer-based NER, and rule-based extraction to parse executive names and roles from 1K+ M&A press releases at 96% accuracy.
- Built robust Python ETL pipelines to ingest 84K+ crypto transactions, extract 6M+ Quora interactions (API/Selenium), and retrieve 500K+ NFT records using resilient OpenSea scripts, delivering fundraising and monetization insights.

Data Scientist | SCEE IIT Mandi   **Aug 2019 - Aug 2021**

- Developed ensemble models to predict magnetic moment and formation energy (R=0.81, RMSE=0.232; R=0.96, RMSE=0.261), accelerating discovery workflows and improving prediction efficiency.
- Analyzed magnetic properties of 12,000 manganese compounds, identifying 60% ferromagnetic prevalence, revealing inconsistencies in existing magnetic theories, and presenting findings at EASTMAG 2019 in Russia.
- Collaborated with physicists to implement ML models that reduced pre-screening time by 6+ hours for 11K+ iron compounds, enabling rapid identification of candidates with strong magnetic properties. DOI: 10.1063/9.0000498.

TECHNICAL SKILLS

Programming Languages	Python, C++, R, JavaScript, SQL, Bash
Frameworks & Libraries	Django, Pandas, NumPy, PyTorch, TensorFlow, Scikit-Learn, XGBoost, Transformers, NetworkX
Cloud & DevOps	AWS (EC2, S3, Lambda, IAM), CI/CD, Docker, Snowflake, DBT, Secure Data Sharing, Monitoring
Machine Learning	LLMs, Anomaly Detection, Classification, Regression, Clustering, Model Deployment, A/B Testing
Data Engineering	PostgreSQL, MySQL, MongoDB, Snowflake, ETL & Data Pipelines, API/Data Integration, Token Management, API Development, Backend System Design

EDUCATION

Master of Science, Computer Science| Indiana University, Bloomington | GPA: 3.68/4   **Aug 2021 - May 2023**
Associate Instructor (Machine Learning): *Project evaluation and mentoring sessions for a batch of 70+ students.*
Competitions: *Winners at IU AWS GameDay 2023 | Finalists at Colts Crossroads Classic Analytic Challenge 2023.*
Coursework: *Advanced Operating System, Data Mining, Applied Algorithm, EDA, Computational Linguistics, Graph Analytics*

P.G.Diploma, Data Science and Artificial Intelligence | IIITD, India | GPA: 9.30/10   **Dec 2020 - Oct 2021**
Coursework: *Machine Learning, Software Engineering, NLP, Linear Algebra, Deep Learning, Statistical Modeling*

B.Tech, Computer Science & Engineering | *Guru Nanak Dev University, India* | GPA: 9.01/10   **Aug 2016 - May 2020**
Coursework: *OOPs, Big Data Analysis, Operating Systems, Distributed Computing, Cloud Computing, Web Development*