Nagesh Gupta (AI Engineer | Data Scientist)

+ 44 7874329884| nageshgupta0806@gmail.com | Linkedin | Github | Portfolio

Professional Summary

As a Data Scientist with 4 years of experience in designing and implementing AI- and ML-driven solutions. Holds a Master's degree with a specialization in cutting-edge Deep Learning and Machine Learning techniques. Skilled in developing data-driven solutions to solve complex business problems. 'crane'

Skills

Programming/Data Tools:	**Python (Expert)**, PyData Stack (Pandas, NumPy, Scikit-learn), SQL, Vector DB, MySQL
ML & Frameworks:	**Generative AI (LLM, RAG, Agentic AI,** Prompt Engineering, **LangChain)**, Deep Learning (PyTorch, TensorFlow), Classification, Regression, Statistical Modeling
MLOps/Deployment:	Docker/Kubernetes, FastAPI (API Deployment), Azure (Functions, Cosmos DB), CI/CD, Git
Front-End:	React JS, Angular, JavaScript (for UI integration), HTML, CSS

Projects

Doc Query - Generative AI (RAG System) *(Python, FastAPI, ReactJS, FAISS, GPT-2, DistilGPT-2)* (Github Link)

- Designed and deployed **Doc Query**, a Retrieval-Augmented Generation (RAG) system that answers questions from unstructured documents (PDF, DOC) using LangChain, FastAPI, and GPT-2/DistilGPT-2.
- Integrated FAISS for vector search and implemented **retriever-reader chaining** with early agent-like control flow, improving contextual relevance and enabling scalable response generation from multi-doc queries
- Added a rule-based output validation layer to score hallucination risk by comparing generated answers against source chunk overlap, increasing trust in generated outputs during testing

Crick-Vision - Video Analysis on Cricket Videos *(YOLOv5, OpenCV, Python, Roboflow)* (Github Link)

- **Developed a state-of-the-art Computer Vision system** (YOLO/OpenCV) to analyse batsman performance, accurately detecting the ball and pitch in real-time.
- **Successfully delivered key functionalities** including accurate **Ball Detection** and **Ball Speed Calculation**, providing detailed insights into delivery types and dynamics.

Customer Churn Prediction:(End-to-End Solution) *(Python, Classification, pandas, scikit-learn, joblib, FastAPI) (Github Link)*

- Developed and deployed an end-to-end machine learning pipeline for customer churn prediction, demonstrating proficiency in data preprocessing, model training, and API integration.
- Engineered a predictive model using **Logistic Regression** and implemented a **StandardScaler** for feature normalization to improve model performance and generalization.

Experience

AI Automation Engineer (Intern) - SpeedNetworkNow, London, UK — Nov 2025 - Present

- Designing and automating event outreach and lead engagement workflows using tools like **n8n** and **Smartlead**, reducing manual workload for the sales and operations teams by over 40%.
- Integrated email sequencing, lead assignment, and CRM updates into end-to-end pipelines using n8n's workflow builder and custom webhooks.
- Developed automation logic to handle lead enrichment, segmentation, and follow-up tracking, increasing outreach speed and contact accuracy across campaigns.

Data Scientist (Associate Professional Application Delivery) - DXC Technology, India, — Mar 2022 – Sep 2023

- Predicted cost savings from international employee relocations by training an XGBoost regression model on HR and compensation data, increasing scenario planning accuracy by **27%.**
- Reduced model error by **18%** by applying one-hot encoding, currency normalization, and feature filtering to clean and transform migration records for consistent cross-market input.
- Enabled end-to-end real-time cost estimation by deploying the trained model via a **FastAPI** microservice and **Docker**, allowing integration with internal HR dashboards and tools.
- Fine-tuned model hyperparameters using **Bayesian optimization with Gaussian process search** across tree depth, learning rate, and column sampling, reducing prediction variance by 12%.

ML Engineer/ Software Engineer – Lentra AI Pvt. Ltd. India, July 2020 – Mar 2022

- Designed and implemented an OCR-based document verification system to extract structured KYC data (PAN, Aadhaar, DOB) from uploaded images and PDFs using Tesseract and OpenCV.
- Achieved over **90% field-level extraction accuracy** by applying adaptive thresholding, noise removal, and bounding-box filtering to improve OCR reliability.
- Developed a Python microservice using FastAPI to compare extracted values with user-entered form data and return real-time match confidence and anomaly flags.
- Improved KYC validation time by over 60% in tests, reducing manual review effort while maintaining audit-ready output with JSON field-by-field comparisons.

Education & Certificate

MSc in Advanced Computer Science, **(Distinction, Grade 1:1)**, Oxford Brookes University, UK Sep 2023 - Dec 2024

BTech, Computer Science, (**Honours**), RGPV, India July 2016 – June 2020

Certificates: Prompt Pro (Google), Certified Scrum Developer (CSD), Neural Networks and Deep Learning.