

# Nagesh Gupta (Data Scientist / AI-ML Engineer)

+ 44 7874329884 | [nageshgupta0806@gmail.com](mailto:nageshgupta0806@gmail.com) | [Linkedin](#) | [Github](#)

## Professional Summary

**Data Scientist** with 3+ 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.

## Skills

<b>Programming/Data:</b>	<b>Python (Expert)</b> , PyData Stack ( <b>Pandas, NumPy, Scikit-learn</b> ), SQL, Vector DB, MySQL
<b>AI/ML:</b>	<b>Generative AI (LLM, RAG, Prompt Engineering, LangChain)</b> , Deep Learning ( <b>PyTorch, TensorFlow</b> ), NLP, Random Forest, XGBOOST, Classification, Regression, Statistical Modeling
<b>MLOps/Deployment:</b>	<b>Docker/Kubernetes, FastAPI</b> (API Deployment), <b>Azure (Functions, Cosmos DB)</b> , CI/CD, Git
<b>Front-End:</b>	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](#))

- Developed a Generative AI tool (**Doc Query**) using the **LangChain framework** and the **SmolLM2-1.7B LLM** to automate question answering from structure and unstructured document types (PDF, DOC).
- Successfully integrated a **FAISS vector database** with the RAG pipeline to enhance document retrieval accuracy, providing users with contextually grounded, real-time responses.

### 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.

### Loan Approval Prediction - (Python, Decision Trees, Ensemble Methods) ([Github Link](#))

- Developed a core predictive model to automate loan approval decision-making, which achieved **98.13% accuracy** and is scalable for implementation within a high-volume financial environment.
- Managed the end-to-end data pipeline, including data cleaning, feature engineering (creating **CIBIL\_rating**), scaling, and **Label Encoding** of categorical variables for model readiness.

### 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

### Predictive Client Retention | Freelance Consulting Project for an Individual [Life Coach](#), UK

Dec 2024 – July 2025

- Developed a **Client Retention Prediction Model** (Random Forest Classifier) using a proprietary dataset of 500+ client histories, achieving **90% Recall** to identify at-risk clients, leading to a **12% reduction in client churn**.
- Engineered key features by applying **Natural Language Processing (NLP)** techniques (sentiment analysis and topic modeling using Python) to unstructured session notes to quantify client motivation and risk factors.
- Designed and implemented the **SQL database schema** via the website back-end to collect and structure critical client data, ensuring clean, optimized input for downstream predictive modeling and reporting.

### Data Scientist (Associate Professional Application Delivery) - DXC Technology, India,

Mar 2022 – Sep 2023

- Developed **Diamond**, an AI-powered web platform to automate employee migration cost savings, **replacing manual Excel workflows** with scalable rule-based processing.
- Built an **NLP pipeline** using Tabula and Python to **extract and parse quarterly updated migration rules** from 250+ page PDFs, achieving **95% rule extraction accuracy** and **reducing manual updates by 90%**.

- **Designed and implemented a data ingestion pipeline** with **Azure Functions** to refresh rules every quarter, integrating **Cosmos DB** for scalable storage and **cutting rule update time from days to hours**.
- **Optimized production systems** by replacing manual processes with a custom **Python-FastAPI** application, **cutting data access and reporting time by 60%**.

**ML Engineer/ Software Engineer – Lentra AI Pvt. Ltd. India,**

July 2020 – Mar 2022

- Developed LaviOsa, an **ML-driven loan approval system**, automating PAN card data extraction and credit scoring, which **cut approval time by 35%** for fintech clients.
- Built an **OCR pipeline** using **Tesseract and OpenCV**, preprocessing images to extract PAN details (name, PAN, DOB) with **88% precision**.
- Trained a **Random Forest model (scikit-learn)** with engineered features (debt-to-income ratio, credit score), **optimising hyperparameters for 83% approval prediction accuracy**.
- Enhanced operational integrity by conducting exploratory data analysis and developing a data-driven deduplication function, **eliminating 95% of duplicate loan applications**.

## **Education & Certificate**

---

**Master's in advanced computer science, (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.