Milind Jadhao
Email ID: mjadhao997@gmail.com
Contact No: +91-9370544344
Data Scientist | GenAI Engineer (5 Years)

Professional Summary

Result-oriented Data Scientist with over 5 years of experience in designing, developing, and deploying AI/ML and Generative AI solutions. Adept at building scalable data pipelines, processing structured and unstructured data, and integrating Large Language Models (LLMs) to solve real-world problems. Proven ability in document intelligence, fraud detection, OCR, and AI-based automation. Strong expertise in Python, AWS, Frappe, OpenCV, and model fine-tuning for custom business workflows.

Core Skills:

Languages & Frameworks	Python, Flask, FastAPI, SQL, Frappe, LangChain, LangGraph, Crew AI
AI/ML/GenAI	LLMs (OpenAI, Anthropic,BERT, Qwen2), LoRA,QLoRA,
OCR & CV:	PaddleOCR, OpenCV, AWS Textract, LayoutLM, DONUT
Databases:	MySQL, MongoDB,Mariadb
Cloud & DevOps	AWS (S3, EC2, SQS, Bedrock, SageMaker), Docker, Git, CI/CD
Vector Databases	FAISS, Pinecone, ChromaDB,

Professional Experience

Company Name	**Duration**
Bizmap Technologies Pvt Ltd	Dec 2023 – Present
Trakiot Solutions Private Limited	Jan 2020 – Dec 2023

Milind Jadhao
Email ID: mjadhao997@gmail.com
Contact No: +91-9370544344
Data Scientist | GenAI Engineer (5 Years)

Professional Summary

Result-oriented Data Scientist with over 5 years of experience in designing, developing, and deploying AI/ML and Generative AI solutions. Adept at building scalable data pipelines, processing structured and unstructured data, and integrating Large Language Models (LLMs) to solve real-world problems. Proven ability in document intelligence, fraud detection, OCR, and AI-based automation. Strong expertise in Python, AWS, Frappe, OpenCV, and model fine-tuning for custom business workflows.

Core Skills:

Languages & Frameworks	Python, Flask, FastAPI, SQL, Frappe, LangChain, LangGraph, Crew AI
AI/ML/GenAI	LLMs (OpenAI, Anthropic,BERT, Qwen2), LoRA,QLoRA,
OCR & CV:	PaddleOCR, OpenCV, AWS Textract, LayoutLM, DONUT
Databases:	MySQL, MongoDB,Mariadb
Cloud & DevOps	AWS (S3, EC2, SQS, Bedrock, SageMaker), Docker, Git, CI/CD
Vector Databases	FAISS, Pinecone, ChromaDB,

Professional Experience

Company Name	**Duration**
Bizmap Technologies Pvt Ltd	Dec 2023 – Present
Trakiot Solutions Private Limited	Jan 2020 – Dec 2023

Project Title:	**Document Automation System (DASH)**
Role	**Data Scientist**
Description:	● Built a semantic search and document QnA system for retrieving A large-scale document automation platform used by multiple enterprise clients, capable of processing over 1 million documents per day. **Key Contributions:** ● Built asynchronous, multi-threaded services using **Python multiprocessing** to handle massive file queues and maintain low latency. ● Engineered scalable pipelines using **AWS services** (S3, SQS, EC2) integrated with **IBM MQ** for message-driven processing. ● Designed modular pipelines to handle **OCR (PaddleOCR, Textract)** and **LLM-driven extraction** using models like **DONUT, LayoutLM, Pix2Struct**. ● Fine-tuned transformer models using **LoRA, QLoRA, and PEFT** to achieve domain-specific extraction of names, amounts, dates, and checkboxes. ● Implemented **custom rule engines** for validating extracted fields against business rules (e.g., amount mismatch, missing signature). ● Built dashboards using **Grafana and Redash** to monitor system metrics like job failure rate, throughput, and accuracy. ● Reduced document processing turnaround time by 45% and improved extraction accuracy by 38% through iterative testing and model enhancement. ● Coordinated with DevOps to set up **CI/CD** for auto-deployment of microservices, ensuring rapid feature rollout.

Project Title:	**Document Automation System (DASH)**
Role	**Data Scientist**
Description:	● Built a semantic search and document QnA system for retrieving A large-scale document automation platform used by multiple enterprise clients, capable of processing over 1 million documents per day. **Key Contributions:** ● Built asynchronous, multi-threaded services using **Python multiprocessing** to handle massive file queues and maintain low latency. ● Engineered scalable pipelines using **AWS services** (S3, SQS, EC2) integrated with **IBM MQ** for message-driven processing. ● Designed modular pipelines to handle **OCR (PaddleOCR, Textract)** and **LLM-driven extraction** using models like **DONUT, LayoutLM, Pix2Struct**. ● Fine-tuned transformer models using **LoRA, QLoRA, and PEFT** to achieve domain-specific extraction of names, amounts, dates, and checkboxes. ● Implemented **custom rule engines** for validating extracted fields against business rules (e.g., amount mismatch, missing signature). ● Built dashboards using **Grafana and Redash** to monitor system metrics like job failure rate, throughput, and accuracy. ● Reduced document processing turnaround time by 45% and improved extraction accuracy by 38% through iterative testing and model enhancement. ● Coordinated with DevOps to set up **CI/CD** for auto-deployment of microservices, ensuring rapid feature rollout.

Project Title:	**Document Automation System (DASH)**
Role	**Data Scientist**
Description:	● Built a semantic search and document QnA system for retrieving A large-scale document automation platform used by multiple enterprise clients, capable of processing over 1 million documents per day. **Key Contributions:** ● Built asynchronous, multi-threaded services using **Python multiprocessing** to handle massive file queues and maintain low latency. ● Engineered scalable pipelines using **AWS services** (S3, SQS, EC2) integrated with **IBM MQ** for message-driven processing. ● Designed modular pipelines to handle **OCR (PaddleOCR, Textract)** and **LLM-driven extraction** using models like **DONUT, LayoutLM, Pix2Struct**. ● Fine-tuned transformer models using **LoRA, QLoRA, and PEFT** to achieve domain-specific extraction of names, amounts, dates, and checkboxes. ● Implemented **custom rule engines** for validating extracted fields against business rules (e.g., amount mismatch, missing signature). ● Built dashboards using **Grafana and Redash** to monitor system metrics like job failure rate, throughput, and accuracy. ● Reduced document processing turnaround time by 45% and improved extraction accuracy by 38% through iterative testing and model enhancement. ● Coordinated with DevOps to set up **CI/CD** for auto-deployment of microservices, ensuring rapid feature rollout.

Project Title:	**Fraud/Tempered Document Detection**
Role	**Data Scientist**
Description:	A fraud detection solution focused on identifying manipulation in scanned financial and legal documents without needing training data. **Key Contributions:** ● Developed CV-based forgery detection using **OpenCV**, contour analysis, and histogram-based feature extraction. ● Designed logic to detect anomalies like pixel shifts in digital signatures, tampered font regions, cloned stamp areas, and image artifacts. ● Created a preprocessing pipeline for image normalization, DPI correction, and noise reduction to improve visual analysis accuracy. ● Built modular components for batch document scoring, alerting, and visual markup of suspected forgery zones. ● Integrated the system into an internal workflow tool for verifying uploaded documents and generating authenticity reports. ● Added support for PDF and image formats with automatic parsing and report generation in HTML/PDF for audit trails. ● Designed field validation techniques comparing extracted text against external APIs (e.g., PAN, Aadhaar, GST) for cross-verification. ●

Project Title:	**Insurance Document Extraction System**
Role:	**Data Scientist**

	An AI-based pipeline that extracts key fields from a wide variety of insurance forms for downstream processing and claims automation. **Key Contributions:** ● Engineered OCR-first document parsing pipelines using **Tesseract, EasyOCR**, and fallback image preprocessing techniques. ● Used **PDFMiner and OpenCV** to isolate structured zones such as checkboxes, signature fields, tables, and input blocks. ● Applied **document template matching** to route input documents to appropriate parsing logic, increasing template match rate by 30%. ● Developed a hybrid approach using rule-based parsing with **prompt-based LLMs (like Mistral and LayoutLM)** for improved field recognition. ● Containerized services using **Docker**, and integrated them with Kafka-based message queues for scalable ingestion. ● Enabled logging, retry mechanisms, and failure tracking for documents with low confidence scores or extraction errors. ● Built validation dashboards to allow manual correction and learning feedback loop to further refine extraction logic. ● applications. ● Implemented logging and monitoring tools for proactive issue resolution. ● Provided technical support and mentorship to junior developers.

Project Title:	Certificial
Description:	Insurance form extraction service where different types of insurance forms details has to be extracted for further downstream systems
Responsibilities	Collaborated closely with business analysts to gather and refine specifications, ensuring alignment with evolving business needs and accurately addressing defects or enhancement requests. Participated actively in the entire software development life cycle, including coding, debugging, performance optimization, and production support. Conducted high-level design and requirement elicitation sessions to translate business processes into technical specifications. Engineered a document intelligence system capable of accurately extracting

	An AI-based pipeline that extracts key fields from a wide variety of insurance forms for downstream processing and claims automation. **Key Contributions:** ● Engineered OCR-first document parsing pipelines using **Tesseract, EasyOCR**, and fallback image preprocessing techniques. ● Used **PDFMiner and OpenCV** to isolate structured zones such as checkboxes, signature fields, tables, and input blocks. ● Applied **document template matching** to route input documents to appropriate parsing logic, increasing template match rate by 30%. ● Developed a hybrid approach using rule-based parsing with **prompt-based LLMs (like Mistral and LayoutLM)** for improved field recognition. ● Containerized services using **Docker**, and integrated them with Kafka-based message queues for scalable ingestion. ● Enabled logging, retry mechanisms, and failure tracking for documents with low confidence scores or extraction errors. ● Built validation dashboards to allow manual correction and learning feedback loop to further refine extraction logic. ● applications. ● Implemented logging and monitoring tools for proactive issue resolution. ● Provided technical support and mentorship to junior developers.

Project Title:	Certificial
Description:	Insurance form extraction service where different types of insurance forms details has to be extracted for further downstream systems
Responsibilities	Collaborated closely with business analysts to gather and refine specifications, ensuring alignment with evolving business needs and accurately addressing defects or enhancement requests. Participated actively in the entire software development life cycle, including coding, debugging, performance optimization, and production support. Conducted high-level design and requirement elicitation sessions to translate business processes into technical specifications. Engineered a document intelligence system capable of accurately extracting

	critical information such as insured party details, policy numbers, coverage dates, limits, and endorsements from a wide variety of insurance documents including Certificates of Insurance (COI), policy declaration pages, and more. Utilized OCR and NLP techniques to enhance the accuracy of data extraction from scanned documents and unstructured form layouts. Integrated pre-trained models and custom rule-based logic for handling edge cases and uncommon document formats. Developed APIs to expose the extracted data to other systems and stakeholders for further processing and analytics. Set up and maintained CI/CD pipelines using tools like GitLab CI/Jenkins to automate testing, integration, and deployment workflows, reducing manual overhead and deployment time. Containerized the application using Docker, enabling consistent deployments across development, staging, and production environments. Worked with cloud services (e.g., AWS/GCP/Azure) to deploy scalable and secure environments for running extraction services.

Educational Qualifications:

- **Bachelor of Science (B.Sc.)**
- **Graduated: 2019**
- **University: PDV-,Akola**