Prateek Prasad Deshpande

prateekprasaddeshpande@gmail.com | +91 8296837773
https://github.com/ppd05 | Portfolio | https://www.linkedin.com/in/prateek-p-a36148223/

Summary

Results-driven data professional with strong expertise in clinical data programming, management using SAS, SQL, Oracle APEX, and Python. Proven success in leading end-to-end clinical study operations, database locks, and data validation for regulatory compliance. Delivered automation and cloud-based solutions, including SAS-to-Python migration. Currently upskilling in AI/ML engineering with hands-on project experience in TensorFlow, PyTorch, and cloud platforms (AWS, GCP). Actively transitioning career towards AI/ML, focusing on developing and integrating scalable AI applications and LLM-driven solutions such as RAG, MCP and Agentic AI.

Skills

Languages: Python, SQL, R programming (basics), Streamlit (basics), SAS

Technologies & Tools: PostgreSQL, My SQL Tableau, PowerBI, My SQL, Medidata RAVE, Veeva EDC, Veeva Vault (clinical), LangChain, LangGraph, Docker(basics), Kubernetes(basics), HuggingFace, ML Flow, Prompt Engineering, Vector Databases, HuggingFace, LLM's, Fast API, Retrieval-Augmented Generation (RAG), Model Context Protocol (MCP).

Cloud: AWS, GCP, Vertex AI

Electronic Data Capture (EDC) Systems: Medidata Rave, Veeva EDC, Veeva Vault (Clinical).

Clinical trial technology: eCOA (electronic Clinical Outcome Assessments), ePRO (electronic Patient-Reported Outcomes).

Work Experience

Independent AI/ML Projects and Coursework [Aug 2025 – Present]

Self-Directed Learning (Remote)

- Completed end-to-end Machine Learning, DeepLearning projects using Python, Scikit-Learn, and XGBoost.
- Built RAG-based clinical decision support and surgical assistant projects.
- Practicing and learning AWS, GCP, Vertex AI for ML ops along with Docker and Kubernetes and developing proficiency in TensorFlow, PyTorch, and Generative AI applications along with MCP [model context protocol] and Agentic AI.

IQVIA RDS (India) Private Limited, Bengaluru [Jun 2024 - Aug 2025]

Database Management Analyst

- Utilized SAS to process and deliver clinical datasets, aligning with protocol objectives and completing critical study deliverables with **98%** on time submissions, which resulted in completion of study production activities with **95%** accuracy**.**
- Identified and resolved **500+** clinical data ingestion and data processing issues in the new Python-based CDI system, reducing error rates by **85%** and significantly improving data quality.
- Collaborated with cross-functional teams to define and validate data acquisition requirements, resolve technical issues, and implement effective solutions, achieving **97% data** accuracy.
- Led the test planning and execution for the Clinical Data Ingestion tool and Clinical Data Extraction tools, in identifying and resolving all bugs prior to UAT, resulting in a **100%** test case pass rate and reducing data processing time from 6 hrs to **2 hrs**.
- Supported server migration to AWS and efficient application testing, achieving a **100% success** rate with a highly controlled cutover that limited total downtime to **60 minutes** during the 4 AM window.
- Served as SME/POC for Clinical Data Ingestion and Clinical Data Extraction tools, identifying **200+ critical bugs** and supporting resolution with **98% diagnostic** accuracy to enhance system reliability.
- Developed reports and dashboards using Oracle APEX (SQL) to support data visualization and reporting needs for internal teams.

IQVIA RDS (India) Private Limited, Bengaluru [Sep 2022 - Jun 2024]

Clinical Data Programming assoc1

- Drove a Corrective and Preventive Action (CAPA) for regulatory data, conducting Root Cause Analysis (RCA), executing data reprocessing, and updating SOPs/WIs to prevent recurrence.
- Reduced document approval cycle time by **25%** by successfully optimizing workflows within Veeva Vault.
- Managed end-to-end clinical trial data and partnered with clients and biostatistics to draft requirement specifications for data acquisition using SAS datasets.Ensured data quality and compliance across all trials by reviewing, analyzing, and processing data transfers to meet project specifications.

Awards and Certificates

- Received recognition and a Bronze Award for exceptional contributions in UAT testing of the Clinical Data Ingestion tool, ensuring error-free implementation and timely completion of test cases.

Education

Dr Ambedkar Institute of Technology [Aug 2018 - Aug 2022]

B.E in Electronics and Communications Engineering, CGPA**:** 8.9 / 10

Relevant Coursework: Object Oriented Programming, Databases, Discrete Maths, Data Structures and Algorithms, Operating Systems, Computer Networks, Machine Learning, Deep Learning, Tableau, R programming.

Course Certifications

- Udemy, Python for Absolute Beginners, Python Beginner to Pro.
  Credential URL: https://www.udemy.com/certificate/UC-71ad5042-f4dd-41b9-a940-bf5025b78be7/
- Geeks for Geeks, DS (Data Structures) basics using Python Programming.
  Credential URL: https://media.geeksforgeeks.org/courses/certificates/e415317ba59e1aaf2395139e588ccc92
- HackerRank, Problem Solving (Basic)
  Credential URL: https://www.hackerrank.com/certificates/bd9cd87a0dc0
- HackerRank, Python (Basic)
  Credential URL: https://www.hackerrank.com/certificates/8457bcd809b6
- Udemy, R programming For Data Science
  Credential URL: https://www.udemy.com/certificate/UC-998581f1-59b7-4a4b-a3b8-a238ef8e8031/
- Udemy, Python-for-data-science-and-machine-learning-bootcamp
  Credential URL: https://www.udemy.com/certificate/UC-8aa6475a-9c0d-49eb-9005-ce761e67af01/
- Udemy, Machine Learning A-Z: AI, Python & R
  Credential URL: https://www.udemy.com/certificate/UC-c6efec38-c4f4-4430-bed0-3a61a6ca726e/
- Udemy, LangChain - Develop AI Agents with LangChain and LangGraph (ongoing)
- Udemy, Docker and Kubernetes (ongoing)

Project Work

- **Clinical Notes AI assistant:** Developed a RAG based AI application that converts unstructured clinical notes to structured (JSON) data and uses this structured data as knowledge base to summarize and act as a chatbot.
  Link: https://clinical-notes-rag-ai.streamlit.app/
- **Surgical Complication AI assistant:** Engineered a RAG pipeline for a surgical knowledge base, integrating LangChain, cost-effective HuggingFace embeddings, and FAISS vector store with a Streamlit UI for high-speed, fact-based clinical query responses. **Link**: https://surgical-complication-ai-assistant.streamlit.app/
- **Infrastructure Crack Detection System** - Developed ResNet101-based deep learning application achieving 95% accuracy in detecting structural cracks; built Streamlit web interface for real-time image analysis with <2s inference time.
  Link: https://infrastructure-crack-detection.streamlit.app/
- **Hotel Restaurant Review sentiment predictor:** A Streamlit web app that trains and evaluates multiple ML models to predict sentiment from restaurant reviews. **Link**: https://restaurant-review-sentiment-analyzer.streamlit.app/
- **Brain Tumor Classification:** Developed a custom CNN model in TensorFlow/Keras to classify brain tumor MRI scans into 4 categories (Glioma, Meningioma, Pituitary, No Tumor), achieving 78% test accuracy.
- **Spaceship_Titanic_project competition:** Built and tuned an XGBoost model with Stratified K-Fold CV, achieving ~80% validation accuracy, with full reproducibility from pre-processing to modeling achieving rank of 572.
- **Data Visualization and dashboard using Tableau on:** you tube channel statistics, King County Washington House sales, Bank customers in UK, Movie Ratings and Cars datasets.
  Tableau repo: https://public.tableau.com/app/profile/prateek.pd/vizzes