

Vishal Singh Yadav

LinkedIn: [linkedin.com/in/tokudai/](https://www.linkedin.com/in/tokudai/)

Github: github.com/tokudai-vs

Address: NCR / Hyderabad, India

Email: vishal.singh@alumni.iith.ac.in

Mobile: +91-9555756235

SUMMARY

Data Scientist and ML Engineer with proven expertise in building, deploying, and scaling machine learning and AI systems. Specialising in NLP, predictive modelling, and large-scale optimisation, with hands-on experience in deploying production-grade models on cloud infrastructures (AWS, Azure). Strong focus on driving business outcomes with data-driven insights, from model development to seamless integration in real-world applications.

EXPERIENCE

- **Carelon Global Solutions (*formerly Legato Health Technologies*)** Hyderabad, India
Jul 2023 – Present
Associate Data Scientist
 - Designed AI-driven forecasting modules for the Cost of Care platform using anomaly detection and GenAI, converting retrospective claims data into state-wise, 3-month predictive insights to enable faster healthcare interventions.
 - Built a medical chatbot using LLAMA2 + RAG (LangChain), fine-tuned with QLoRA (4-bit) for domain-specific QA, significantly improving context relevance and response quality.
 - Developed LLM-based health condition identification algorithms, improving coverage from 48% to 81% and enabling a 15x increase in member outreach.
 - Created interpretable risk-scoring and clinical flagging models with explainable AI techniques, aiding underwriters and delivering \$100K in business impact.
 - Automated underwriting workflows using ML and LLM-powered conversational tools, reducing manual effort and enhancing operational efficiency.
 - Deployed models across hybrid infrastructure (AWS, Azure, on-prem), integrating with business rules and dashboards to support real-time decisions.
 - Applied large-scale preprocessing and optimization techniques to boost model performance, using Docker-based cloud deployments for scalability.
 - Built predictive models for High-Cost Claimants and Special Conditions, achieving a 5% lift in accuracy over baseline.
- **Qulabs Software India Pvt Ltd** Hyderabad, India
Mar 2023 – Jul 2023
Machine Learning Engineer
 - Built a domain-specific RAG-based chatbot for real-time document search and semantic matching over large unstructured datasets, enabling faster query resolution.
 - Guided junior developers on ML workflows, including data cleaning, model training, and deployment, improving overall code quality and delivery speed.
- **Krama Lab, IIT Hyderabad (*Machine Learning Research*)** Hyderabad, India
Jan 2020 – Dec 2022
Research Assistant
 - Conducted applied research on point cloud segmentation, graph neural networks, and knowledge graph completion under an industry-sponsored collaboration (GreatFour Systems).
 - Developed few-shot learning techniques for point cloud classification and designed GNN-based architectures for structured knowledge representation and reasoning tasks.

EDUCATION

- **Indian Institute of Technology Hyderabad** Hyderabad, India
Jan 2020 – Dec 2022
M.Tech (by Research) – Computer Science and Engineering GPA: 8.38
- **GGs Indraprastha University** New Delhi, India
Jul 2014 – May 2018
B.Tech – Computer Science and Engineering Percentage: 71.2

PROJECTS

- **ARROW – AI/ML-Driven Underwriting Risk Tool (Carelton):** Co-developed ARROW (Anthem Rating and Risk Optimisation Wizard), an AI-powered underwriting risk platform for Elevance Health, projected to save \$1.5M via vendor insourcing by 2025. Applied NLP, ML, and large-scale optimisation to automate risk scoring from financial, clinical, and demographic data. Achieved a 20% reduction in underwriting time and improved rating accuracy with real-time model deployments and intuitive dashboards.
- **Medical-LLM – GenAI for Medical Data Understanding (Carelton):** Built a HIPAA-compliant medical LLM pipeline leveraging QLoRA and distributed ML frameworks to automate clinical data handling. Improved processing efficiency by 10%, boosted model accuracy by 5%, and reduced manual effort by 10%. Integrated secure cloud infrastructure and GenAI-driven data analytics to improve scalability and operational reliability.
- **Compliance Suite – AI Automation for US Healthcare (Qulabs):** Led end-to-end development of an AI-powered compliance automation suite for US healthcare. Built document QA chatbots using RAG + LLMs, semantic matching algorithms, and medical rule-check engines. Enhanced compliance accuracy and reduced turnaround time using real-time inference, few-shot learning, and large-scale data preprocessing.

SKILLS

- **Programming Languages:** Python, C++, Bash
- **ML & AI Frameworks:** TensorFlow, PyTorch, HuggingFace, SpaCy, Scikit-learn, Keras, XGBoost, LightGBM
- **Cloud & Deployment Tools:** Docker, AWS, Azure, Git, GitHub, LaTeX
- **Data Science & ML Techniques:**
 - Natural Language Processing (NLP), Large Language Models (LLMs), Graph Neural Networks (GNNs)
 - Few-Shot Learning, Hyperparameter Tuning, Object Detection, Feature Engineering
 - Model Evaluation (AUC, ROC, F1 Score), Statistics, Data Preprocessing/Postprocessing
 - Data Visualization (Matplotlib, Seaborn)
- **Soft Skills:** Data-Driven Decision Making, Predictive Modeling
- **Platforms:** Linux, AWS, Windows

ACHIEVEMENTS

- **Award for Rapid Product Deployment – Carelon Global Solutions:** Recognised by the Director of Data Science for leading the accelerated deployment of a critical product, improving operational efficiency and enabling faster business insights
- **Go Above Award – Carelon Global Solutions:** Twice honoured for exceptional execution under pressure: delivered a member dashboard within 24 hours and expedited the ARROW project under stringent deadlines.
- **Mentor – AI and Advanced Technologies:** Delivered technical mentorship under an IIT Hyderabad and NSE Talentsprint initiative, guiding professionals in machine learning and GenAI fundamentals.

CERTIFICATIONS

- **Advanced Certification in Generative AI:** Carelon Global Solutions & Prizmato — Focused on deploying production-grade GenAI systems and prompt engineering best practices.
- **NVIDIA Deep Learning Institute:** Completed six specialised courses:
 - **NLP & Deep Learning:** Fundamentals of Deep Learning, Transformer-Based NLP Applications
 - **AI for Ops:** AI for Anomaly Detection, Accelerated Data Engineering Pipelines
 - **Accelerated Computing:** Fundamentals of Accelerated Computing and Data Science
- **Coursera / University Programs:**
 - **Machine Learning Specialization** — University of Washington
 - **Applied Machine Learning** — University of Michigan
 - **Deep Learning Specialization** — deeplearning.ai

PUBLICATIONS

- **Context-Aware Question Routing in Community Question Answering Sites:** *Vishal Singh Yadav, Manish Singh May 2023.* Developed a method for improving question routing on community-driven Q&A platforms using context-based embeddings.