

ROHIT BANKAR

(+1) 765-767-2732 | rohitbankar2000@gmail.com | [LinkedIn](#) | [GitHub](#) | [Medium](#) | [Portfolio](#)

EDUCATION

Purdue University

Master of Science in Computer and Information Technology (4.0/4.0)

January '24 - Dec '25

West Lafayette, IN

Indian Institute of Technology (IIT) Indore

Bachelor's in Technology (BTech) (8.77/10.0)

July '18 - May '22

PROFESSIONAL EXPERIENCE

Amazon (AWS Bedrock)

Software Development Intern

May '25- August '25

Seattle, USA

- Developed **Hubble**, a centralized service for managing Anthropic Claude LLM models (Opus, Sonnet, Haiku) on AWS Bedrock, integrating with Trainium-based infrastructure to optimize deployments and reduce manual intervention by 90%.
- Enhanced the Forklift team's operations by optimizing **inference latency, reducing throttling, and integrating API-based capacity management with automated rebalance planning**, reducing issue-resolution time by 50%.
- Implemented and integrated **15+ Restful APIs**, enhancing visibility and management efficiency for Claude replicas.

Razorpay (South Asia's Largest Payment Gateway)

Software Development Engineer (Backend)

July '22- December '23

Bengaluru, India

- Interacted with 5+ teams to lead the re-architecture of Harvester, a real-time Spark-based data service, **saving \$10-11k /month**.
- Coded custom whitelisting logic for Maxwell and Debezium CDC collectors, securing ~90% of company data.
- Migrated SRM dashboard **from TiDb to Pinot, reducing lag from 5 to 1 minute**, enhancing experience of 5+ teams & clients.
- Integrated Schema Registry into entity onboarding flows and ETL pipelines by implementing necessary transformation logic and handling horizontal and vertical scaling of pods, resulting in better versioning and 100% faster retrieval of schemas.
- Improved on-call bot's data flow, adding features for real-time updates, achieving 40% faster resolution.

Jaguar Land Rover

Software Development Intern

June '21- August '21

Bengaluru, India

- Designed and coded a Machine Learning Pipeline to extract critical information from handwritten sentences using standard NLP libraries **like Stanford-openie and Textacy, achieving 70% improvement** over previous algorithms.
- Made exponential smoothing and **LSTM-RNN models, boosting failure prediction accuracy from 80% to over 90%**.
- Wrote a model to quantify optically scanned data using string-matching algorithms like Levenshtein-Damerau and Jaro-Wrinkler, achieving desired results with 95% accuracy.

Purdue University

Software Development Engineer (GRA)

January '24 - Present

West Lafayette, USA

- Built 8+ critical data pipelines to ingest thousands of daily data points from Qualtrics, & Salesforce to on-premise databases.
- Created and maintained 10+ data pipelines for secure, efficient data transfer. Generated data visualizations for 20+ clients.
- Develop, host, and maintain a scalable website for the NSF-funded project [Advancing Cybersecurity Education](#).

RESEARCH & PROJECTS

Compute and Memory resource estimation for ML pipelines

MS Research - Thesis Project

January '24 - Present

- Designed a metamodel for predicting machine learning models' CPU/GPU & memory requirements, reducing cost by 25%.
- Optimized resource allocation by proposing and validating a new CORE-ML framework for various stages of a non-inference pipeline, reducing **CPU overprovisioning by up to 80% and cutting costs by up to 50%**.
- Implemented a real-time resource scaling framework (Deep Scale) using Kubernetes VPA for training pods, reducing total compute costs by 60% and **CPU hours by up to 85% through high-frequency scaling**.

Detection of Suicidal Ideation through Deep Learning

Research Project - IIT Indore

March '23 - October '23

- Developed a model using BERT, Graph Sage, and Hierarchical Attention to classify suicidal texts with an **F1 score > 0.975**.
- Enhanced model accuracy by integrating lexicons and emotional analysis, achieving a total F1 score increase of 6-7%.

Open-Source Contributions

January '21 - December '23

- Programmed a Go authentication module for Trino gateway, enhancing security & access control for 2000+ Razorpay users.
- Integrated credential authentication into Query Book's UI, benefiting 10000+ users.
- Engineered a tool to convert LookML to Cube.js at Razorpay's annual hackathon, automating the process and reducing resource consumption by **at least 70%; received organization-wide recognition and bounty**.

SKILLS

- Languages:** Python, Golang, Scala, C++, SQL, Java, JavaScript, TypeScript, R, HTML, CSS, Rust, MATLAB.
- Tools:** Databricks, AWS, GCP, Docker, Cassandra, GitHub, CI/CD, Git, MySQL, PostgreSQL, Redis, MongoDB, Kafka.
- Technologies:** Spark, OOP, Hadoop, Terraform, SDLC, DBT, Hive, PyTorch, Kubernetes, Airflow, API, Docker, DynamoDB.
- Extra-curricular/Awards:** Head Chess Club IITI, Exceptional Work Award for HMS Upgrade, Inter-IIT Sports Meet.