ROHIT BANKAR

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

EDUCATION

Purdue University *Expected Dec'25*
Master of Science (MS by Research) in Computer and Information Technology ***(4.0/4.0)*** *West Lafayette, IN*

Indian Institute of Technology (IIT) Indore *July'18 - May'22*
Bachelor's in Technology (BTech) ***(*8.77/10.0)***

PROFESSIONAL EXPERIENCE

Razorpay (South Asia's Largest Payment Gateway) *July'22- December'23*
Software Development Engineer (Data Platform) *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 retrieving of schemas.
- Developed software for the on-call bot's data flow, adding features for real-time updates, achieving 40% faster resolution.

Jaguar Land Rover *June'21- August'21*
Software Development Intern (Data) *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 *January'24 - Present*
Software Development Engineer (Data) *West Lafayette, IN*

- 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 the website for the NSF-funded project Advancing Cybersecurity Education.

RESEARCH & PROJECTS

Compute and Memory resource estimation for ML pipelines *January'24 - Present*
MS Research - Thesis Project

- Designed a metamodel for predicting machine learning models' CPU/GPU & memory requirements reducing cost by 25%.
- Optimized resource allocation & deployed 5+ LLM and CV pipelines on Kubeflow in a multi-node Kubernetes cluster.
- Devised a custom method for resource allocation of different pods in ML Pipelines saving 40-50% of resources on average.

Detection of Suicidal Ideation through Deep Learning *March'23 - October'23*
Research Project - IIT Indore

- 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.

Gender Prediction and Image Classification via Flask Web App *April' 21- May'21*
Open Source

- Programmed and hosted a web application with Flask on the PythonAnywhere server.
- Implemented a Support Vector Classifier for gender prediction (99 % accuracy) and an Image Classifier using Stochastic Gradient Descent (98 % accuracy), so anyone can try the deployed models.

Face Mask Detection *January'21*
Open Source

- Determined if a person is wearing a face mask in a live video feed with minimal lag and a high accuracy of 95%.
- Built with CNNs, OpenCV, and Haar cascades using a large dataset of 50,00 images.

SKILLS

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