Rohit Vernekar

774-777-1969 | rohitvernekar76@gmail.com | rohit-vernekar.github.io | rohit-vernekar | rohit-vernekar | Piscataway, NJ

Machine Learning and Data Engineer with 4+ years of experience in developing and deploying ML models, building scalable data pipelines, and automating workflows in cloud environments. Strong statistical background and proficiency in SQL and NoSQL databases, along with a solid foundation in algorithms as an ACM-ICPC regionalist. Proven ability to reduce costs and enhance system performance. Skilled in AWS, Kubernetes, and Python, seeking to drive data-driven insights and innovation in top tech firms.

Education

Rutgers University — New Brunswick, NJ, USA
Master of Science in Data Science – GPA: 4.0/4.0 — *Expected Dec. 2024*

Visvesvaraya Technological University — Belagavi, KA, India
Bachelor of Engineering in Computer Science – GPA: 8.55/10.0 — *Jul. 2019*

Experience

Lutron Electronics — Philadelphia, PA, USA
Data Engineering Intern — *Jun. 2024 – Aug. 2024*

- Developed a serverless streaming pipeline with **Amazon Kinesis** and **Firehose** to optimize log processing and retention.
- Automated ETL using **AWS Glue** and **Redshift**, improving query performance and reducing processing costs by 80%.
- Used Infrastructure as Code (**IaC**) with **CloudFormation** and **Terraform** to efficiently manage pipeline infrastructure.
- Created **QuickSight** dashboards, reducing **New Relic** telemetry costs by 75% and enabling more advanced visualizations.

Aera Technology — Pune, MH, India
Machine Learning Engineer — *Jan. 2021 – Aug. 2023*

- Built an ML deployment pipeline using **Redis** as a message broker for seamless model deployment on **Kubernetes**, with a web service for API interaction, supporting both real-time and batch inference.
- Reduced ML model deployment time by 95%, from 10 minutes to 30 seconds, by integrating pre-built **Docker** images.
- Developed an **Autosklearn**-based framework that autonomously generates optimal ML models for varying datasets, enhancing model accuracy and reducing development time.
- Built a data validation and drift detection system to continuously monitor and ensure the performance of ML models, which helped maintain model accuracy and reliability over time.
- Created a high-performance asynchronous task execution framework, using **Redis** as a message queue for inter-process communication, reducing resource usage by 75%.

Tata Consultancy Services — Pune, MH, India
Systems Engineer — *Jul. 2019 – Jan. 2021*

- Built an interactive dashboard framework for automotive software components using **Matlab**, enhancing system monitoring and reducing debugging time by 30%.
- Automated Matlab model testing and report generation with **GitLab-CI**, reducing testing time by 15 minutes per model and accelerating project delivery. Received On-the-Spot award for outstanding project impact.

Intern — *Jan. 2019 – Mar. 2019*

- Implemented a Spatio-Temporal Autoencoder to detect anomalous patterns in video data, improving the accuracy of anomaly detection in client projects.
- Designed a real-time 3D visualization system using three.js, enhancing object movement analysis for client applications.

Projects

Time Series Analysis with Conformal Prediction | *Python, Yahoo Finance APIs* | Link

- Built a forecasting tool using Conformal Prediction, improving uncertainty estimates for stock market predictions.
- Used quantile regression to improve prediction accuracy by leveraging temporal dependencies in financial data.

Autonomous Traffic System | *Python, OpenCV, Flask, HTML, CSS* | Link

- Built a real-time traffic light control software with image processing, reducing traffic congestion by 15% in simulations.
- Awarded "Best Final Year Project" for technical innovation and societal impact.

Technical Skills

Languages: Python, R, Java, C/C++, Bash, PowerShell

Databases: SQL (MySQL, Amazon Athena and Redshift), NoSQL (Neo4j, MongoDB, Redis)

Machine Learning & AI: Ensemble Methods (Bagging, Boosting), AutoML (Auto-sklearn), Deep Learning (TensorFlow, PyTorch), Natural Language Processing (spaCy, NLTK, LLMs, LangChain, OpenAI API, Hugging Face Transformers).

Cloud and Infrastructure: AWS (ECS, S3, Lambda, Glue, Kinesis, Firehose, CloudFormation, Step Functions, Firelens), Microsoft Azure, Docker, Kubernetes, Jenkins, Gitlab-CI, Github Actions, Terraform, Apache Spark

Software and Libraries: Numpy, Pandas, Scikit-learn, XGBoost, Flask, Celery, OpenCV, YOLO, Three.js, Celery

Data Visualization: Matplotlib, Seaborn, Plotly, Amazon Quicksight, Tableau, New Relic