Yatin Kande

yatink@umich.edu | Dearborn, MI | +1 313-413-8327 | linkedin.com/in/yatin-kande | github.com/YatinKande

Skills

- **Programming Languages:** Python, SQL, R
- **Databases:** PostgreSQL, MySQL, SQLite, MongoDB
- **Machine Learning & AI:** Supervised & unsupervised learning; deep learning (CNNs, RNNs, LSTMs); computer vision (YOLO, object detection); model evaluation (precision, recall, F1, ROC-AUC); hyperparameter tuning
- **Data Science:** EDA; feature engineering; statistical analysis; hypothesis testing; A/B testing; dashboarding reporting (Power BI, Tableau, Qlik Sense)
- **Data Engineering:** ETL; data preprocessing & transformation; API integration; data quality validation
- **Frameworks & Libraries:** Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow, OpenCV, Transformers, FAISS, FastAPI, Plotly/Dash
- **Cloud & DevOps:** AWS (S3, Lambda, API Gateway, DynamoDB, SQS/SNS, Lex), Docker, Git/GitHub
- **Certifications:** OCI AI Foundations (2025); Google Data Analytics (Coursera)

Professional Experience

Data Analyst Intern (AI/ML) – DataZymes, Bengaluru *Feb 2024 – Jul 2024*

- Improved autonomous field perception by training and tuning YOLO-based object detection in PyTorch/OpenCV on 1,000+ multispectral images, applying preprocessing and augmentation to increase plant vs. weed detection reliability.
- Ensured consistent model outputs by building an end-to-end CV workflow with data-quality validation, GPU training, and metric-driven evaluation using accuracy, precision, recall, and F1 to quantify performance and error patterns.
- Accelerated operational decision-making by prototyping reinforcement-learning path optimization and delivering Plotly/Dash KPI dashboards that translated model performance into ROI-oriented insights for cross-functional stakeholders.

Machine Learning Intern – SmartKnower, Bengaluru *Mar 2022 – Jun 2022*

- Built a supervised learning pipeline on the UCI Census Income dataset (32,000+ records) by cleaning data, engineering features, and training baseline models in Python, Pandas, and Scikit-learn for structured prediction tasks.
- Selected the best classifier by benchmarking Logistic Regression, Decision Trees, and Random Forests using accuracy, precision, recall, F1, ROC-AUC, and confusion matrices to identify trade-offs and reduce misclassification risk.
- Enabled deployment decisions by summarizing experiment results into stakeholder-ready findings, documenting assumptions, metrics, and outcomes to support informed selection of a production-ready approach.

Projects

AutoRAG: Multi-Modal Generative AI Diagnostic Assistant *Nov 2025 – Present*

- Architected production-oriented RAG service to answer technical diagnostic queries by implementing FastAPI + FAISS retrieval and routing responses to Google Gemini with an Ollama (local Llama 3) fallback for offline, cost-aware inference.
- Improved answer grounding for multi-format knowledge sources by creating ingestion for PDFs, DOCX files, and technical images (e.g., OBD scans, wiring diagrams) and generating normalized descriptions to strengthen indexing and retrieval relevance.
- Increased system reliability by adding circuit-breaker, retry, and failover controls to handle quota limits and connectivity errors, achieving 94% uptime and reducing failed-response scenarios under external dependency failures.

Dataset Recommender Bot (AWS Cloud Application) *Aug 2025 – Oct 2025*

- Delivered dataset discovery from natural-language queries by architecting a serverless chatbot with Amazon Lex, API Gateway, Lambda, and DynamoDB, integrating Kaggle and Hugging Face APIs for dataset retrieval.
- Improved personalization by storing user profiles and query history in DynamoDB, deploying a lightweight web UI on S3 and triggering automated email delivery using SQS/SNS to share recommended dataset results.
- Measured feature impact by instrumenting logging and analytics to track engagement and support A/B testing on recommendation logic, enabling iterative optimization based on observed user behavior signals.

LipNet-Based Visual Speech Recognition (3D CNN + Bi-LSTM) *Apr 2025 – Jul 2025*

- Built a visual speech recognition model using 3D CNNs and Bi-LSTMs with CTC loss on the GRID dataset in PyTorch/TensorFlow, achieving 15% word error rate to validate sequence-learning performance.
- Standardized training data by extracting and normalizing 30,000+ video frames and generating stratified train/validation splits, enabling controlled experimentation and repeatable comparisons across runs.
- Enabled experiment traceability by designing an SQLite schema for metrics and training logs, running SQL-based tracking across 50+ iterations to compare results and automate performance reporting.

Education

University of Michigan–Dearborn *Dearborn, MI*
Master of Science in Data Science

Vellore Institute of Technology *Bangalore, India*
Post Graduate Program in Data Science

Jain University *Bangalore, India*
Bachelor of Science (Honors) in Data Science