

Santhosh Dayakar

#401,1260 Dr Penfield Ave, Montreal, QC H3G 1B6 | santhoshdayakar2002@gmail.com | 263-881-4020
www.linkedin.com/in/santhosh-dayakar | github.com/S-15-77

Education

Concordia University , Masters in Applied Computer Science Co-Op	Sep 2024 – Current
• Eligible for full-time Co-Op /internship (Fall,Winter 2025)	
Sri Ramachandra Institute of Higher Education and Research , B.Tech Computer Science and Engineering	Sep 2020 – May 2024

Skills and Certifications

Languages: Python, SQL (Advanced: Joins, Aggregations, Window Functions), Java, C, C++ , C#, R, Dart, JavaScript, TypeScript, Bash ,Erlang.

Data Analytics & Project Management: Data Analysis, Reporting Automation, Jira, Agile Scrum, Project Coordination, Stakeholder Communication, Capacity Planning, Workload Monitoring

Big Data & Cloud: Apache Spark, ETL Pipelines, Data Wrangling, AWS, Azure, Google Cloud, Docker, Kubernetes, Distributed Systems, Data Warehousing, Data Modeling, Streaming Data, Networking Fundamentals

Machine Learning: TensorFlow, Scikit-Learn, PyTorch, Pandas, NumPy, Matplotlib, OpenCV, MediaPipe, GenAI, Hugging Face, Hyperparameter Tuning, Model Deployment, ML Debugging

Frameworks/Technologies: FastAPI, Node.js, ReactJS, Flutter, Firebase, Selenium, REST APIs, Pytest, Linux, JIRA, VS Code, GitHub Actions, CI/CD, Argo Workflows, Helm

Certifications: Google IT Automation with Python (Coursera, 2023), AWS Academic Cloud Foundations (AWS,2022), AWS Academic Machine Learning Foundation (AWS, 2023)

Experience

Data Science Intern , Datakulture – Chennai, IND	Feb 2024 – Jul 2024
• Designed and deployed scalable ML pipelines integrating GPT-4 powered natural language to SQL translation, improving data accessibility by 50% and enabling real-time analytics for enterprise reporting.	
• Built fully automated cloud-native microservices using Docker, Kubernetes, AWS (EC2, S3, Lambda), and GitHub Actions, increasing operational efficiency by 70% while ensuring production-grade model reliability.	
• Collaborated with cross-functional teams to troubleshoot ML pipeline failures, optimize feature generation, and implement robust GenAI-based autonomous pipelines supporting digital twin readiness and distributed systems	

Projects

Image Captioning using Transformer Architecture	[Source Code]
• Built image captioning pipeline using MobileNetV3 encoder and Transformer decoder achieving BLEU-4 improvement from 0.17 to 0.32.	
• Tuned model hyperparameters and optimized data pipelines reducing GPU training time by 40% for scalable experiments.	
• Delivered model evaluation reports using BLEU, ROUGE, METEOR scores with visualized analytics for peer and stakeholder review.	
SpeakLens – Multimodal Meeting Assistant	[Source Code]
• Built real-time meeting assistant combining transcription, speaker diarization, summarization, sentiment analysis, and Q&A using GenAI models.	
• Led cross-functional coordination between ML, backend, and cloud teams to integrate distributed microservices using FastAPI, Docker, Redis, and PostgreSQL.	
• Automated weekly metric dashboards for tracking processing load, model accuracy trends, and operational issue logs to support leadership reporting.	
Gym Workout Tracker	[Source Code]
• Developed posture evaluation system using MediaPipe, OpenCV, and Random Forest, achieving 78% accuracy across dynamic exercises.	
• Automated real-time rep-counting pipelines and adaptive feedback loops for edge-device deployment. Created training data processing pipelines with visual error tracking and reporting modules for team review.	