Namratha Tiptur Manjunath

namrathatm.ms@gmail.com | LinkedIn | GitHub | Website

EDUCATION

Master of Science, Information Systems | Syracuse University, Syracuse, NY — *Jan 2025 – Dec 2026*
GPA: 4.0 / 4.0

Bachelor of Technology, Computer Science and Engineering | Reva University, Bangalore, India — *Aug 2019 – Aug 2023*
GPA: 3.5 / 4.0

PROFESSIONAL EXPERIENCE

Associate Software Engineer – Data Engineering & Analytics | Carelon Global Solutions — *Aug 2023 – Aug 2024*
- Architected **Informatica PowerCenter ETL pipelines** integrating 20+ source systems with **Python (Pandas, NumPy, SQLAlchemy)** and **MySQL** staging tables, processing **1TB+ daily healthcare data** at 99.5% accuracy.
- Reduced ETL runtime by **40%** through Informatica parallel processing and MySQL query optimization with strategic indexing on high-volume tables.
- Developed Python **data quality framework** using Pandas validations and NumPy-based anomaly detection, improving data quality metrics by **30%**.
- Built 15+ MySQL analytical queries using **window functions and CTEs** for complex aggregations, automating reports that saved **25 hours/week** of manual analysis.

Data Analyst Intern | Belgian Waffle Factory — *Sep 2022 – Nov 2022*
- Developed 8 interactive **Tableau dashboards** with calculated fields, dynamic parameters, and action filters for regional-to-store drill-down analysis, supporting data-driven management decisions.
- Extracted and transformed **50K+ records** using MySQL window functions and **Tableau Prep Builder** to create a unified sales and inventory data model.
- Automated weekly reporting using MySQL stored procedures and **Python (Pandas, openpyxl)**, reducing report generation time by **60%**.

DATA SCIENCE & AI PROJECTS

RAG Observatory – LLM Evaluation & Monitoring System
Python, LlamaIndex, Streamlit, Groq, OpenAI, Ollama, ChromaDB
- Built 8 custom **evaluation metrics** (relevance, grounding, coverage, diversity) using **LlamaIndex**, enabling detection of **hallucination and retrieval failures** across 1,000+ test queries with configurable severity thresholds.
- Developed **Streamlit** monitoring dashboard with live querying and real-time metric visualization, reducing RAG pipeline debugging time by surfacing root-cause failure patterns across multiple LLM providers.

Cogni Research – Autonomous AI Research Assistant
Python, LangGraph, Claude (Anthropic), MCP, ChromaDB, FastAPI, Streamlit, Tavily
- Architected **MCP server ecosystem** with 4 specialized servers (web search, memory, database, filesystem), enabling modular tool integration that reduced manual research time from **hours to minutes** per topic.
- Integrated **ChromaDB vector database** with cosine-similarity retrieval and **LangGraph** conditional routing, enabling persistent research memory that eliminated redundant queries across sessions.
- Built production-ready dual-interface application (1,800+ lines) with **FastAPI REST backend** and Streamlit UI, using **Pydantic** config management and structured logging for maintainability.

PCOS Subtype Discovery – Clustering Robustness Research
Python, Scikit-learn, SciPy, Pandas, NumPy, Matplotlib
- Benchmarked 5 clustering algorithms with 100 bootstrap iterations, identifying 4 clinically distinct PCOS subtypes at **82% stability and 0.73 ARI** — extending published **Nature Medicine (2025)** findings.
- Implemented uncertainty-aware classification flagging 27% ambiguous cases via **GMM probability thresholding**; validated cross-dataset generalization achieving **78% subtype consistency**.

TECHNICAL SKILLS

AI / LLM & GenAI: RAG Systems, LLM Evaluation, Hallucination Detection, LangGraph, crewAI, LlamaIndex, MCP, ChromaDB, Vector Databases, Agent Orchestration, OpenAI API, Anthropic Claude API

ML & Data Science: Scikit-learn, XGBoost, Clustering Algorithms, TensorFlow, PyTorch, Keras, NLP, Feature Engineering, Statistical Analysis, Bootstrap Validation, Model Evaluation

Programming & Data Engineering: Python (Pandas, NumPy, SQLAlchemy, FastAPI, Pydantic), SQL, R, Java, Informatica PowerCenter, ETL Pipelines, MySQL, MongoDB, AWS S3, AWS EC2, GCP BigQuery

Visualization & BI: Tableau, Streamlit, Power BI, QuickSight, Matplotlib, Seaborn, Excel

Methods & Tools: A/B Testing, Hypothesis Testing, Experimental Design, Data Quality Monitoring, Healthcare Analytics, HIPAA Compliance, Git, Agile

CERTIFICATIONS

Google Data Analytics Professional Certificate – Google | **Data Analytics & Business Intelligence with Advanced Excel** – Coursera