Bala Swapnika Gopi

bgopi@umd.edu | +1 (240) 350-7674 | balaswapnika-gopi | BalaSwapnikaGopi | USA

Education

Master of Science, Data Science — University of Maryland (UMD), *College Park, MD* — Aug 2024 – May 2026
GPA:4.0

Bachelor of Technology — Jawaharlal Nehru Technological University (JNTUH), *India* — Aug 2018 – May 2022
GPA:7.15

Relevant Coursework

- Probability and Statistics
- Principles of Machine Learning
- Principles of Data Science
- Big Data Systems
- Data Representation and Modeling
- Algorithms for Data Science
- Deep Learning
- Natural Language Processing
- Cloud Computing

Experience

Data and AI Intern, *Connyct – USA* — Jul 2025 – Present

- Automated data collection across 50+ university websites using Python, Selenium, and BeautifulSoup, reducing manual effort by **90%** and enabling faster, more reliable updates for CampusAI's platform, which significantly boosted data availability for analytics.
- Produced structured, analysis-ready datasets that accelerated model training efficiency by **25%**, accelerating product development timelines and enabling faster experimentation cycles for data science teams.

Associate Machine Learning Engineer, *Infor – India* — Apr 2022 – Aug 2024

- Built and deployed a self-service SQL generation toolkit that translated natural language into executable queries, enabling non-technical users to access data directly, which reduced reporting turnaround time by **80%**.
- Executed a DeepAR forecasting solution for sales contracts across multiple vendors and locations, achieving **25%** improvement in demand prediction accuracy, optimizing inventory planning, and saving clients millions annually in carrying costs.
- Advanced invoice automation by integrating OCR pipelines with Claude 3.5 Sonnet, achieving **95%+** field-level accuracy and cutting manual finance processing time by **70%** (200+ hours/month).

Projects

Real-Time Bitcoin Data Processing with Apache Ray *(Python, Apache Ray, API Integration, Distributed Computing)*

- Engineered a distributed real-time Bitcoin monitoring system using Apache Ray, enabling **60%** lower latency in live data pipelines, supporting high-frequency trading analytics, and ensuring scalability for large-scale financial data processing.
- Crafted interactive Plotly dashboards that delivered actionable market insights to traders, increasing user adoption for live crypto monitoring by **30%** and improving decision-making through real-time visualization of market trends.

Vision-Language Model (VLM) Safety Evaluation *(Python, VLLM, GPT-4o, Claude 3 Haiku, Gemini 2)*

- Conducted a comprehensive evaluation of multimodal models using VLLM Safety Benchmarks, achieving **96.5%** accuracy across out-of-distribution and counterfactual datasets and demonstrating the robustness of models in handling complex real-world tasks.
- Formulated strategies to mitigate hallucinations and unsafe responses, which strengthened AI system reliability, safety for enterprise applications, and ensured compliance with ethical AI guidelines.

Credit Card Fraud Detection *(Python, Scikit-learn, SMOTE, Logistic Regression, Decision Tree, Pandas, NumPy)*

- Created a machine learning–based fraud detection system, achieving over **96%** accuracy on highly imbalanced financial data, reducing false positives that previously impacted approvals, and enabling banks to improve fraud monitoring efficiency by **40%.**
- Evaluated model effectiveness using precision, recall, F1-score, and ROC-AUC metrics, delivering a comprehensive performance report that enhanced interpretability for stakeholders and supported business adoption of the solution.

Facial Expression Recognition *(Python, TensorFlow, MobileNet, Keras, OpenCV, Google Colab)*

- Developed and trained a CNN model to classify facial expressions into seven categories on the FER2013 dataset, achieving **68%** validation accuracy and demonstrating applicability in human–computer interaction systems.
- Applied preprocessing techniques and regularization methods, improving generalization by **20%** and reducing overfitting, ensuring stronger model performance on unseen data.

Skills

- **Programming & Data Science:** Python, SQL, PySpark, Java, NumPy, Pandas, Scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, Statistical Modeling, Feature Engineering, Time Series Forecasting, Large Language Models (LLMs), RAG Systems
- **MLOps & Deployment:** ETL Pipelines, Model Deployment, Model Evaluation, CI/CD for ML, FastAPI, Git, JIRA, Version Control
- **Software & Tools:** Jupyter Notebook, VS Code, Google Colab, Microsoft Excel, Figma, Data Modeling, Overleaf
- **Data Visualization & Business Analytics:** Tableau, Power BI, Matplotlib, Seaborn, Plotly, Data Storytelling, Dashboard Development
- **Big Data, Cloud & Databases:** MySQL, PostgreSQL, BigQuery, Databricks, Apache Spark, Hadoop, Hive, Apache Airflow, GCP AI Platform, AWS, Docker, Kubernetes

Achievements and Volunteer Experience

- **Silver Medal**– Programming Data Structures & Algorithms in Python (NPTEL)
- **Top Performer**- Programming with Python & Machine Learning (Internshala)
- **Core Member**-NSS JNTUHCEJ; Led 10+ social impact projects, impacting 1,000+ people
- **Event Organizer** – EMBLAZON FEST, JNTUH Department of ECE