PRIYA GANESAN

+1-857-339-8452 | shanmugapriyamurugesanganesan@gmail.com | LinkedIn: Shanmuga Priya M G | Boston, MA - 02119

Education

- **Khoury College Of Computer Sciences, Northeastern University** — *Sept 2024 - Dec 2026*
  Masters in Artificial Intelligence - CGPA: 3.67/4.0 — Boston, United States
  - Current Coursework - Machine Learning, Algorithms, Program Design Paradigm
- **Sri Sivasubramaniya Nadar College Of Engineering, Anna University** — *Aug 2018 - May 2022*
  Bachelors in Computer Science And Engineering - CGPA: 8.42/10 — Chennai, India

Technical Knowledge

- **Languages:** Python, Scala, Java, SQL
- **Data Science:** Big Data, Hive, Hadoop, Spark, Kafka
- **Python Frameworks:** Scikit-learn, Numpy, Pandas, Pyspark, Matplotlib, Seaborn, Pytorch, Tensorflow, Keras
- **Machine Learning:** Logistic Regression, SVM, Random Forest, LSTM, BERT
- **Others:** React, Flask, Javascript, HTML/CSS, GitHub, Gitlab, Jupyter Notebook

Experience

- **Engineering Analyst - Goldman Sachs, Bengaluru, India** — *July 2022 - July 2024*
  - Designed and maintained **ETL** pipelines leveraging **Kafka** and **Elasticsearch**, ensuring efficient data ingestion and optimized loading into **HDFS** for large-scale data processing.
  - Monitored, supported, and resolved issues and alerts in a complex automated data processing framework, handling **230 data groups** with **18,150 datasets**, ensuring seamless data flow and integrity.
  - Developed a characteristic framework using **PySpark** to implement 32 custom constraints on daily acquired financial data, enabling enhanced security customization for large-scale investments.
  - Onboarded **12 new datasets** by integrating **web scraping techniques** and file-based data acquisition, efficiently processing and storing structured data in **HDFS**.
  - Introduced **manual compaction feature for Parquet paths**, enhancing **storage efficiency** and adding **flexibility in data organization** within the framework, optimizing retrieval speeds and reducing storage overhead.
  - Integrated **CI/CD pipelines** into the data engineering workflow, automating builds, tests, and deployments using **GitHub** and **Jenkins**, significantly improving deployment reliability and reducing release turnaround time.
  - Executed **weekly production releases** with precision, ensuring smooth deployments by leveraging **GitHub for version control** and **real-time monitoring** to track changes and mitigate risks.
- **Summer Analyst, Intern - Goldman Sachs, Bengaluru, India** — *June 2021 - Aug 2024*
  - Designed and developed a new UI from scratch using **ReactJS** and **Redux** to facilitate adding, updating, and maintaining securities in the firm's database, enhancing user experience and operational efficiency.
  - Implemented **REST APIs** to enable seamless **database interactions**, improving **data consistency and system reliability**.
  - Built and tested **dynamic front-end components**, ensuring **robust database connectivity** and a **smooth, user-friendly interface**.

Projects

- **Twitter Sentiment Analysis | ML, LSTM, BERT, LDA, Flask** — *Feb'25 - Apr'25*
  Built a sentiment analysis system comparing traditional ML models (**Logistic Regression, SVM, Random Forest**) with deep learning approaches (**LSTM, BERT**) using **TF-IDF** and **Word2Vec**. Deployed a **Flask** web app for real-time sentiment classification and integrated topic modeling with **LDA** to surface key complaint themes.
- **Virtual Calendar Application (access on request) | Java, Swing** — *Feb'25 - Apr'25*
  Developed in **Java** using **MVC** design patterns, supporting creation and management of multiple calendars with single and recurring events. Enabled event editing, import/export via CSV, and multi-mode interaction through **GUI, text, and script-based execution**.
- **Guardians Gambit: A 2D Reinforcement Learning based game | Pygame, Q-Learning, Tiled** — *Oct'24 - Dec'24*
  Designing an **adversarial reinforcement learning** game where **thief and guard agents** are trained using **Q-learning** and **Deep Q-Networks (DQN)** to develop strategic behaviors for a simulated heist.
- **Natural Disaster Analysis using Satellite Images and Real-time Data | Python, U-Net, NLP** — *Jan'22 - May'22*
  Developed a disaster impact assessment system by integrating **segmented satellite images** with **real-time Twitter data**, leveraging **U-Net** for image segmentation and **NLP** for sentiment-based region classification.
- **Exploratory Data Analysis on Post-Disaster Heterogeneous Dataset | SnowFlake, Apache Superset** — *Dec'21 - Feb'22*
  Conducted statistical and visual analysis of disaster data to evaluate correlations between climate change, natural disasters, and economic impact, providing insights for risk management and preparedness.

Publications

[1] Srinivasan K, Ganesan S P, Mandyam S, Suresh S. (2023). **A Correlation Analysis and Visualization of Climate Change using Post-Disaster Heterogeneous Datasets**. *Global NEST Journal*, Vol. 25, Issue 7, pp. 155-165. DOI: 0.30955/gnj.004382