

Suriya G

Data Scientist

Namakkal, India | suriyagurumoorthi02@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | +91 73393 10945

EDUCATION

Vellore Institute of Technology

MSc Computational Statistics and Data Analytics (Integrated)

Programme Representative and Merit Scholar for 4 consecutive years

Vellore, Tamil Nadu

Sept. 2020 – May 2025

CGPA 9.39/10.0

EXPERIENCE

Data Analyst Intern

May. 2023 – Jun. 2023

Hindustan Aeronautics Limited (HAL)

Bangalore, Karnataka

- Designed "**Helicopter Failure Analysis**" project, organising and cleaning data for meaningful dashboards.
- Optimally used Excel, Python, SQL, and Tableau to clean data, analyze it, and craft interactive dashboards.
- Strengthened soft skills in independent and team work, honing problem-solving and decision-making capabilities.

TECHNICAL SKILLS

Softwares: Python, SQL, Tableau, R, Excel

Hard Skills: Git, Generative AI, Large Language Models, Prompt Engineering, Data Visualisation, Machine Learning, Deep Learning.

Soft Skills: Story Telling, Highly Adaptable

Libraries: Pandas, NumPy, Matplotlib, Tensorflow, Scikit-Learn

PROJECTS

Credit Risk Classification | *Python, Machine Learning*

[\(GitHub Link\)](#)

- Implemented end-to-end credit risk classification, involving data transformation techniques and
- Conducted comprehensive hyperparameter tuning for multiple classification models, enhancing model performance and accuracy.
- Achieved 73 % precision with Naive Bayes, identifying it as the leading model for credit risk prediction and deployed using streamlit

Analysis of vehicle Accidents | *MySQL*

[\(GitHub Link\)](#)

- Fabricated a thorough analysis of accidents, uncovering hidden patterns and extracting meaningful insights from the dataset.
- Performed time-based analysis and found seasonal trends, peak accident times, and patterns.
- Demonstrated intermediate SQL proficiency by using JOIN, DECLARE, CASE, SET, WITH functions.

Customer Lifetime Value Prediction | *MySQL, Python*

[\(GitHub Link\)](#)

- Employed Python for data cleaning and customer segmentation based on RFM Scores.
- Leveraged SQL to pinpoint high-performing customers, optimizing marketing and retention efforts.
- Utilized ML models for predictive analysis, forecasting new customer behaviors based on purchase patterns.

Trip Itinerary Generator | *Python, Generative AI, Gemini Pro*

[\(GitHub Link\)](#)

- Designed an interactive system allowing users to input and to receive a personalized daily itinerary.
- Leveraged the Gemini Pro API for advanced natural language processing and understanding
- Incorporated real-time data sources for weather forecasts, foreign exchange rates, and flight fares using relevant prompt engineering.

Fine-Tuned Chatbot using Open Source LLM | *Python, Generative AI, OpenAI*

- Developed a chatbot using a customized university dataset, leveraging OpenAI API and LangChain framework to enhance functionality.
- Integrated Pinecone for efficient vector storage, managing a database of 250 vectors with 1536 dimensions to improve response accuracy.
- Used open source gpt-3.5-turbo-instruct model to optimize response.