Suriya G
Data Scientist

Namakkal, India | suriyagurumoorthi02@gmail.com | Linkedin | GitHub | Portfolio

EDUCATION

Vellore Institute of Technology Vellore, Tamil Nadu

MSc Computational Statistics and Data Analytics (Integrated) Sept. 2020 – May 2025

Programme Representative and Merit Scholar for 3 consecutive years CGPA 9.36/10.0

EXPERIENCE

Data Science Intern Aug. 2023

LetsGrowMore *(Virtual)*

- Worked on many initiatives at LetsGrowMore as an intern, learning data science and analysis. Finished many tasks at the firm and received a certificate.
- **Stock Price Prediction:** Built and forecasted LSTM-based models for accurate stock price prediction, essential for financial decision-making.   (GitHub Link)
- **EDA on Terrorism Dataset:** A thorough EDA revealed terrorist tendencies and patterns.   (GitHub Link)

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, Data Structures and Algorithm, Big Data Management

Soft Skills: Story Telling, Highly Adaptable and agile

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

PROJECTS

British Airways Customer Prediction | *Python* (GitHub Link)

- Demonstrated proficiency in data cleaning, Pre-processing, and machine learning model building techniques.
- Developed a robust predictive model using Random Forest, capable of forecasting passenger demand with high accuracy.
- Achieved an accuracy of 85 % using Random Forest, outperforming Decision Tree

Cycle Sales Analysis | *Excel, Tableau* (Tableau Link)

- Performed data cleansing in Excel to guarantee Tableau visualisations were accurate and reliable, providing a solid foundation for analysis.
- Visualized sales data across various dimensions, including brand, product category, customer , and sales region.
- Provided actionable insights into sales performance, customer preferences, and product positioning.

Analysis of vehicle Accidents | *MySQL* (GitHub Link)

- Fabricated a thorough analysis of accidents, uncovering hidden patterns and extracting meaningful insights.
- 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 using LLM | *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.

Youtube Summarizer | *Python, Generative AI, Gemini Pro* (GitHub Link)
- Developed a YouTube video summarization system using Google Gemini LLM model.
- Integrated Transcript API for accurate video transcriptions. Implemented query answering to enhance user interaction with the content.
- Streamlined YouTube content consumption through automated summarization.