SURYAPRAKASH UPPALAPATI

D A T A S C I E N T I S T

Fairfax, VA 22031 | suppala6@gmu.edu | https://www.linkedin.com/in/suryaprakash55/ | +1 5712699148

SUMMARY

Aspiring Data Scientist with a strong academic background in Computer Science and specialization in ML/AI. Experienced in developing ML models and data visualization using Python, SQL, TensorFlow, and AWS SageMaker. Eager to leverage expertise in machine learning and analytical problem-solving to contribute to innovative engineering projects.

EDUCATION

George Mason University **Fairfax, Virginia.**

M.S. in Computer Science Expected Graduation, **December 2026**.

- **Concentration:** Machine Learning **GPA:** 3.8
- **Related Coursework:** Introduction to Artificial Intelligence, Theory and Applications of Data Mining, Software Design and Design Architecture, Analysis of Algorithms, Computer Systems and System Programming, Mathematical Foundations of Computer.

PROFESSIONAL EXPERIENCE

Voluntary Data Science Research Assistant, George Mason University **August 2024 – February 2025**

- Developed and fine-tuned ML models (Random Forest, SVM), achieving 85% accuracy, with feature importance analysis and cross-validation.
- Implemented K-means clustering with silhouette analysis for optimal region segmentation, improving data-driven decision-making.
- Designed and automated an ETL pipeline to process 100GB+ spatial data, reducing data processing time by 20%.
- Created interactive choropleth maps using GeoPandas and Folium, enabling stakeholders to visualize patterns effectively.

Data Science Intern **Bengaluru, India**

Virgosys Software **November 2023 – May 2024**

- Built data pipelines using Python, SQL, and Apache Airflow, reducing data processing time by 30%.
- Developed and deployed automated machine learning models for data processing and anomaly detection, reducing data errors by 35% and improving pipeline efficiency by 25%.
- Led code reviews and model optimization for ML workflows, focusing on efficient Docker-based deployment.
- Implemented automated testing with pytest, decreasing the time to detect bugs by 25% and improving the reliability of the system.

PROJECTS

Customer Churn Prediction System

- Developed an end-to-end customer churn prediction model using XGBoost and Random Forest, achieving 87% accuracy and 0.92 AUC
- Engineered 20+ features from customer interaction data, identifying key behavioral indicators that preceded churn events
- Implemented SHAP value analysis for model interpretability, providing actionable insights that reduced churn rate by 18%
- Designed automated ML pipeline with cross-validation and hyperparameter tuning using Grid Search, improving model robustness.

Domain-Specific Language Model Fine-Tuning with Meta Llama 2

- Fine-tuned using LORA adaptation reducing model deployment costs by 40% while achieving 30% accuracy gain
- Implemented domain-specific prompt templates reducing hallucination rate by 45%
- Evaluated using ROUGE-L (0.85), BLEU (0.78) and domain-specific accuracy metrics
- Applied gradient checkpointing and accumulated batches for efficient training

House Price Prediction

- Developed ML model reducing property valuation time by 80% with 90% prediction accuracy, enabling data-driven decisions
- Engineered 15+ features incorporating market dynamics (price/sqft trends, location clusters) and macroeconomic indicators
- Built automated feature selection pipeline using RFE, improving model interpretability while maintaining accuracy
- Created interactive dashboards visualizing price trends and feature importance for stakeholder decision-making

TECHNICAL SKILLS

- **Machine Learning:** PyTorch, Scikit-learn
- **Statistics:** Regression (Linear, Logistic), Clustering, Time Series Analysis, A/B Testing
- **Programming:** Python (NumPy, Pandas), SQL (Joins, Window Functions)
- **Cloud/Tools:** AWS SageMaker, Git, Jupyter
- **Visualization:** Seaborn, Matplotlib, GeoPandas
- **Analytics:** Statistical Testing, A/B Testing, Hypothesis Testing, Time Series Analysis
- **Cloud & Tools:** AWS (SageMaker, S3, Lambda, RDS), Docker, Git, Apache Airflow, Jupyter, LangChain , LangChain

CERTIFICATION

- **DeepLearning.AI:** Supervised Machine Learning: Regression and Classification
- **Udacity Certification:** Introducing Generative AI with AWS
- **AWS Academy Graduate:** AWS Academy Machine Learning Foundations
- **AWS Academy Graduate:** AWS Academy Cloud Foundations