Maria Nikitha Suresh

marianikitha.suresh@gmail.com | +1 520 333 0012 | linkedin.com/in/maria-nikitha-588994242 |
github.com/marianikitha01| maria-nikitha-suresh.web.app

WORK EXPERIENCE

Machine learning Developer, Arizona Simulation Technology & Education Center   **Aug 2024 - Dec 2024**
- Achieved **95% accuracy** in classifying surgical sutures as **Good, Tight, or Loose** by engineering a YOLOv8 model in **Python/PyTorch.**
- Converted the model to ONNX and deployed on an NVIDIA T4 GPU, cutting per-frame latency from 120 ms to 30 ms.
- Implemented **precision, recall and F1-score** for model evaluation, optimizing for medical imaging challenges.
- Automated image preprocessing pipelines using Python (**OpenCV, NumPy**) to improve detection consistency and presented findings to research teams, showcasing **data-driven insights** for improving surgical training.

Data Analyst & Software Development Engineer, Protecto - Data Privacy Startup   **Jun 2022 - Jul 2023**
- Built privacy-risk dashboards in **Tableau** that let compliance teams visualise anomaly trends across millions of log events.
- Designed and implemented frontend features using Angular for internal **privacy risk analysis** tools, improving user experience and engagement.
- Utilized **statistical modeling** to detect anomalies and privacy risks, ensuring compliance with industry standards.
- Collaborated cross-functionally to implement data-driven BI solutions for enterprise data workflows.

Data Science Intern, Alpha Tech Academy - EdTech Startup   **Jul 2021 - Jan 2022**
- Analyzed enrollment and performance data using **pandas and NumPy** to identify engagement and outcome trends.
- Developed **Microsoft PowerBI** dashboards monitoring course completion rates, average scores, and student satisfaction.
- Built a **dropout-risk prediction model** in scikit-learn with **91%** accuracy to enable targeted support interventions.
- Automated end-to-end data preprocessing pipelines (data ingestion, cleaning and feature engineering), cutting report generation time by **40%**.

EDUCATION

University of Arizona, United States   **Aug 2023 - May 2025**
Master of Science : Data Science
Christ University, India   **June 2018 - May 2022**
Bachelor of Technology: Information Technology

PROJECTS

Chikankari Design Generator using Generative AI | Capstone Project   **Link**
- Built a generative pipeline using Stable Diffusion v1.5 fine-tuned with Advanced LoRA on 4000+ Chikankari embroidery images to synthesize authentic pattern designs.
- Developed training workflows with **Hugging Face Diffusers**, including data augmentation, checkpointing, and LoRA adaptation.
- Evaluated model performance using **FID and CLIPScor**e to assess realism and prompt alignment.
- Designed and implemented a **user interface with Streamlit** that allows users to input custom prompts and generate embroidery patterns within 5 seconds, making the tool accessible for non-technical users.

Early Readmission Prediction using Random Forest, SMOTE, and AWS-Based ETL   **Link**
- Developed a machine learning model to predict early hospital readmissions using a **Random Forest Classifier**, achieving a training accuracy of 96%.
- Implemented an ETL pipeline on AWS to automate data extraction, transformation, and loading processes. Utilized **AWS Glue** for data cataloging and **ETL operations.**
- Performed data preprocessing and oversampling techniques, specifically using **SMOTE**, to address class imbalance in the dataset, enhancing model performance.

Transforming E-Commerce Data with Azure: End-to-End Pipeline Implementation   **Link**
- Designed and deployed a scalable, end-to-end ETL pipeline using **Azure Data Factory, Azure Databricks, and Azure Data Lake Storage** to support high-volume E-commerce data processing.
- Ingested and integrated data from multiple sources into Azure Data Lake, enabling centralized and governed data access.
- Utilized Apache Spark on **Azure Databricks** for processing large-scale data, following Bronze, Silver, and Gold architecture.
- Leveraged Delta Lake for optimized storage, data versioning, and reliability. Implemented data quality checks and performance optimizations to enhance efficiency.

SKILLS: Deep Learning, Machine Learning, Data Engineering, Generative AI (GenAI), NLP, Data Analysis & Visualization.

TOOLS: Python, TensorFlow, PyTorch, Keras, Scikit-learn, OpenCV, NumPy, Pandas, SQL, R, C++, Git, AWS (Lambda, S3, Glue, Athena), Azure (Data Factory, Databricks, Data Lake Storage), Apache Spark, Transformers, Micrososft PowerBI, Tableau, GPT, BERT, NLTK, MySQL, MariaDB, MongoDB, R, Java, Bash, PHP, Relational Database.