Swajan Reddy Gaddampally

+1 832-807-5260 — swajan.edu@gmail.com — linkedin.com/in/swajanreddy

Education

University of Houston, Houston, TX *August 2023 – December 2024*
Master's in Statistics and Data Science, GPA: 3.85
Courses: Machine Learning, Big Data Analytics, Deep Learning, Data Visualization, Statistical Analysis, Spatial Statistics

Jawaharlal Nehru Technological University, Hyderabad, India *August 2019 – July 2023*
Bachelor's in Computer Science, GPA: 8.2
Coursework: Database Systems, Data Structures and Algorithms, OOPS Concepts, Computer Networks, Operating Systems, IoT

Skills

- **Data Science and Machine Learning:** Python, Pandas, NumPy, PyTorch, TensorFlow, SpaCy, NLTK, LangChain, NLP, Knowledge Graphs, Hugging Face, Scikit-Learn, RAGs
- **Big Data Analytics and Engineering:** PySpark, Apache Spark, MySQL, Azure, Apache Kafka, Apache Airflow, Hadoop, ETL Pipelines
- **Cloud ML Ops and Deployment:** AWS, Azure, GCP, Kubeflow, MLflow, Docker, Git, Kubernetes
- **Data Visualization and Business Analytics:** Power BI, Tableau, Regression Analysis, Hypothesis Testing, Excel
- **Web and App Development:** Flask, Django, HTML/CSS, JavaScript, Tkinter, Salesforce Developer, UiPath (RPA Developer), Selenium

Professional Experience

Data Engineer (Freelancer) *March 2022 – present*
ACEEC Innovation Labs, Remote

- Led the design and implementation of automated executive reporting solutions, integrating data from over 6 diverse sources, including customer transaction data (from CRM systems), financial reports (from ERP systems), and web analytics (from Google Analytics), into a Snowflake data warehouse.
- Developed and optimized ETL pipelines using Python and Apache Airflow to automate data ingestion and transformation from structured and unstructured data sources. Utilized AWS S3 for storage and Redshift for data warehousing, resulting in a 100% speed improvement in reporting cycles.
- Created and implemented Apache Airflow workflows to schedule and manage data pipeline tasks, ensuring reliability and scalability of ETL jobs. Optimized task dependencies and execution times to handle large volumes of data across multiple environments.
- Created and optimized complex SQL queries using Stored Procedures, Common Table Expressions (CTEs), and User-Defined Functions (UDFs) to handle large-scale data processing tasks, reducing report generation time by 50%.
- Developed Power BI dashboards for business intelligence reporting, enabling cross-functional teams to access real-time insights on customer behavior, sales performance, and product utilization, resulting in actionable insights that led to a 15% increase in operational efficiency.
- Designed and implemented predictive maintenance models using logistic regression and Random Forest classifiers to predict customer subscription responses based on transactional and demographic data, improving marketing targeting strategies.
- Collaborated with data scientists to define reporting requirements, creating data pipelines for real-time customer segmentation analysis, and integrating the outputs into Power BI dashboards for real-time decision-making.
- Utilized Hadoop and Apache Spark for processing large-scale datasets, optimizing jobs for performance and reducing data processing time by 25%.

Machine Learning Research Assistant *September 2023 – December 2024*
University of Houston, Houston, TX, USA

- Collaborated on HR management simulator development, improving effectiveness by 15%, and launched RAG model infrastructure, boosting text data processing efficiency by 80%.
- Designed and managed data pipelines using Azure Data Factory and Azure Databricks, resulting in interactive dashboards and improvement in decision-making by creating databases of student grades and scores.
- Leveraged Apache Spark and performed Exploratory Data Analysis (EDA) and developed insightful Power BI dashboards, improving team strategies by 25% and effectively communicating performance benchmarks.

– Optimized AI models using Python libraries (spaCy, NLTK) to analyze student feedback, provided advice leading to a 30% improvement in student satisfaction rates.

Data Science Intern *January 2022 – March 2022*
Defense Research and Development Organization, Hyderabad, TG, India

– Enhanced project outcomes by 20% through Python, Machine Learning, and Data Analysis. Analyzed over 10,000 multisensor data points using linear regression, decision tree, and k-means clustering, improving data-driven decision-making by 25%.

– Built and deployed an end-to-end ML Ops pipeline for target classification using AWS and Azure, achieving 86% classification accuracy, reducing model response lag by 5 seconds.

– Streamlined ML pipelines and deployment processes leveraging Docker for containerization and implementing CI/CD practices to ensure efficient, scalable, and reliable model deployment across environments.

– Refined predictive models and machine learning algorithms using Python, reducing error rates in analysis processes by 15% and enhancing data reliability.

Projects

– **Human Mimic Chatbot using Transformers:** Developed deep learning software for human-like conversations, achieving 80% user satisfaction.

– **Image Captioning:** Implemented Transformer Encoder-Decoder architecture with pretrained Vision Transformer (ViT) and GPT-2 on the Flickr30k dataset.

– **Database Management System for Indian Cricket Team:** Designed relational databases using SQL in Oracle 7g for seamless data manipulation.

– **Custom Object Detection with TensorFlow:** Integrated Mask R-CNN-based model with 89% accuracy into camera software systems.

Research Publications

– **Survey on Chatbot Classification and Technologies**, IRJET

– **Statistics and ML in Data Science and Effect in Businesses**

– **Human Mimic Chatbot**, WJARR