Syed Affan Daimi

https://sulphatet.github.io/ · Hyderabad, Telangana · saffand03@gmail.com

Education	**Muffakham Jah College of Engineering and Technology, Affiliated to Osmania University**, Hyderabad, India — June 2020 - June 2024

Muffakham Jah College of Engineering and Technology, Affiliated to Osmania University, Hyderabad, India
Bachelor of Engineering, Computer Science and Engineering
CGPA: 8/10
June 2020 - June 2024

Indian Institute of Technology Madras, Chennai, India — *June 2020-Present*
Bachelor of Science In Data Science and Programming
Diploma in Programming — *Graduated 2024*
Diploma in Data Science — *Graduated 2023*
Projects: Modelling E-commerce Shopper's Behaviour (Machine Learning Practice), Goods Quality Prediction and Creditors' Returns Forecasting (Business Data Management), Full Stack Application of Library Management System (Modern Application Development)
GPA: overall 8/10; Project-work only 9.25/10; Data Science-only 8.6/10
Microsoft Azure Professional Certifications: AI Fundamentals [AI-900], Data Fundamentals [DP-900], Power Platform Fundamentals [PL-900]

Publications

Daimi, Syed Affan, and Asma Iqbal. "Ensemble Machine Learning to Predict and Feature Engineering to Identify Factors for Academic-Success." *Proceedings of the Fourth International Conference on Advances in Computer Engineering and Communication Systems (ICACECS 2023).* Vol. 18. Springer Nature, 2023.

Iqbal, Asma, **Syed Affan Daimi**, and K. Manjunatha Chari. "Performance Efficient and Fault Tolerant Approximate Adder." *Journal of Electronic Testing* 39.5 (2023): 571-582.

Daimi, Syed Affan, Askari, Arslaan Ahmed et. al. "Leveraging Machine Learning Models For Proactive Disaster Forecasting" Presented at *1st International Conference on AIML-Applications for Engineering & Technology (ICAET 2025).*

Under Review

Daimi, Syed Affan, Jindal, Garima, and Karlapalem, Kamal. "SpinTrix: Dynamic Network Visualizations." Manuscript under review, 2025.

Preprints

Daimi, Syed Affan. "A Scalable Data-Driven Framework for Systematic Analysis of SEC 10-K Filings Using Large Language Models." *arXiv preprint* arXiv:2409.17581 (2024).

Research Experience

Data Science and Analytics Center, International Institute of Information Technology
Associate Researcher with Dr. Kamal Karlapalem — *June 2024-Present*
Working on Knowledge Graphs, Network Visualization. Assisted GenePoweRx, a genomics startup, in automation and data mining, along with exploring LLM based solutions.

- Designed and implemented SpinTrix, an interactive D3.js dashboard that visualizes temporal graphs using spiral layouts, enabling users to track node changes and evolving community structures across all time slices, with the aim of reducing user's mental load and enhance pattern recognition of the spatial and temporal features of the graph.
- Developed persistent community selection and cross-time comparison features, allowing users to select and visually compare communities across time slices reducing visual clutter in large dynamic networks.
- Overhauled an automated report generation system by integrating doctors' input and genetic data with a knowledge base, streamlining the process to deliver personalized patient recommendations reducing manual *input time by 2X* and improving accuracy of reports as an external assistant at GenePoweRx.
- Led ideation and initial development of a RAG system to convert doctor audio notes into personalized lifestyle advice; designed the database and led ongoing data collection.

VIsual learninG and InteLligence (VIGIL) lab, Indian Institute of Technology Hyderabad

Research Intern with Dr. C Krishna Mohan   May 2023 – Aug 2023

Contributed to *Object Detection for Self-Driving Cars* project.

Designed an object detection system using 19,800 synchronized frames to detect six key urban objects, improving autonomous vehicle perception for Indian urban environment.

Voluntary Experience	*Teacher/Section Leader at* **Stanford Code In Place 2025**: Lead weekly virtual sessions for a cohort of 15-20 international students, delivering Stanford's CS106A curriculum. Facilitated interactive discussions, provided personalized feedback, and supported learners in mastering Python fundamentals.

NPTEL Program Outreach at **ACM CompEd 2023**

Open Source Contributions:

- **Vizro** *Maintained by* **McKinsey:** Modified CSS elements to improve data presentation, contributed a tutorial to user documentation.
- **HuggingFace x scikit-learn Sprint**: Overhauled existing scikit-learn tutorials by creating improved interactive versions using Gradio, and hosted them on Hugging Face Spaces.

Medium: Write technical blogs, mostly tutorials and guided projects for new frameworks, gaining 500+ views monthly.

Industry Experience	**Razor Group GmbH** November 2023 – May 2024

Data Science Intern

- **Forecasting, Demand Planning and Portfolio Steering:** Developed ML clusters to segment the product portfolio based on long-term sales trends, enhancing demand forecast accuracy by 20% and improving portfolio management.
- Improved explainability of black-box forecasting models using SHAP and model plots, identifying irrelevant features and enabling better model sensitivity and cross-learning.
- **Competitor Intelligence Database:** Built a database with 500,000+ competitor products identified and periodically tracked by writing custom python scripts uploaded on AWS Lambda.
- Partnered with operations to design and implement the database, aligning it with business requirements for actionable insights.
- **Demand Plan Reporting:** Lead reporting to identify inefficiencies in AWS Forecast, supporting the transition to custom ML solutions.

fellowship.ai   Oct 2022 – Jan 2023

Data Science Intern

Developed an image similarity tool using CLIP embeddings and K-Nearest Neighbours for competitor product analysis. Integrated the tool with a web scraper and multi-modal search for Nike/Adidas shoes, handling 5,000+ items and improving product suggestions. Optimized the pipeline by batching, storing embeddings, and applying dimensionality reduction (PaCMAP, t-SNE), reducing search time from 5-10 minutes to 30-45 seconds.

Projects	**Goods Quality Prediction and Creditors' Returns Forecasting for Pharmaceutical Company** (Capstone project under the guidance of Dr. Aaditya Chandel and Dr. Ashwin J. Baliga) Collected and cleaned Goods Returns data to create a regression model predicting retrieval returns for faulty medicines, improving inventory management decisions. Conducted statistical tests to assess creditor trustworthiness, revealing non-business creditors' greater tendency to default.

Verdure Curated a dataset of 85k images from 45k using image augmentation to train a PyTorch model for detecting plant health and developed a Streamlit interface. Collaborated in a 4-person team to integrate IoT components and create an Android app for plant care. Achieved 3rd place in Innovatia Panoply 2023 and 1st place in Envisage 2023, competing against 200+ participants.