Devanshu Gupta

623-290-3858 | dgupta77@asu.edu | linkedin.com/in/devanshu0gupta | github.com/devanshugupta

Summary

Experienced data scientist with 2.5+ years of experience in Python, SQL, Statistics, Exploratory Data Analysis, Gen AI, and scalable cloud architectures to deliver data-driven insights.

Education

Master of Science in Computer Science **Expected - 05/25**
Arizona State University, Tempe, AZ, USA *GPA: 3.8/4*
Coursework: Cloud Computing, Data Mining, Reinforcement Learning, Statistical ML, Planning & Learning Methods in AI

Bachelor of Technology in Computer Science & Engineering **08/17 – 05/21**
University Institute of Technology RGPV, IN *GPA: 8.4/10*

Experience

Tata Consultancy Services | Data Scientist | India **08/21 – 08/23**

- Developed predictive Machine learning models for customer churn using ensemble methods, resulting 30% reduction in customer attrition and saving the company approximately $5M annually
- Designed and deployed an advanced custom large language model (LLM) utilizing RAG architecture on Azure OpenAI services, overcoming long-context problem increasing content generation quality by 30%.
- Conducted Exploratory Data Analysis, Statistics, data visualization for demand forecasting using Python, R, SQL.
- Developed automated systems, preprocessing 6TB+ of operational and financial structured, unstructured data daily from multiple data sources to uncover key business trends through data analysis.
- Investigated user behavior analytics and constructed custom Power BI dashboards; identified significant trends leading to feature enhancements, augmenting user engagement by 30%
- Collaborated with cross-functional teams, documentation, unit testing, accelerating project delivery timelines by 15%.

Tata Consultancy Services | Data Science Intern (Research) | India **04/21 – 07/21**

- Developed Talent Experience Gap Analyzer and Productivity Prediction, a predictive model to identify KPIs of 10% low-performing employees to provide actionable insights for HR analytics.

Emphasis Corp | Machine Learning Intern | CA **03/20 – 09/20**

- Created customer facing bank chatbot using LLMs, Python, DialogFlow API (Google) reducing 30% query resolution time and enhancing engagement by automation.
- Engineered infinite playlist music recommendation system (20K songs) via Python, Tkinter GUI, and collaborative filtering to predict user preferences, listening patterns and enhance personalization.

Projects

Image Recognition as a Service using AWS, ASU AWS Cloud project | [Code] **09/24 – 12/24**

- Built efficient and scalable AWS cloud application IaaS resources, Flask framework handling 1000 concurrent requests.
- Leveraged AWS resources like EC2 for deploying machine learning model, S3 for object storage, and SQS for auto-scaling.
- Configured EC2 instances to dynamically scale and shut down in 30s, enhancing application elasticity and cost efficiency.

Image Search Engine | [Code] **09/23 – 12/23**

- Launched an open source project with custom kernels in Convolutional Nueral Networks (CNNs) enabling image classification, user relevance feedback model using SVM (10K images) with 95% accuracy.
- Improved nearest neighbour search time 1 hr to 30 secs by creating scalable vector databases via LSH, DBScan clustering.

AdPrompter : Generative AI for Ads | [Code] **09/23 – 12/23**

- Modeled autonomous system for generating 500+ images by fine-tuning GPT-3.5, Reinforment learning human feedback training custom LLMs, VLMs (Stable Diffusion XL) reducing time & cost for advertisements.

Healthcare Data Analytics **05/21 – 08/21**

- Scheduled healthcare.gov API data integration every 2hrs in Azure Data Warehouse, Power BI providing analytics.
- Automated 2 ETL data streaming pipelines for near real-time data analysis to help detect anomalies and provide alerts.

Technical Skills

Programming & Databases: Python, R, SQL, MySQL, MongoDB, NoSQL, PostgreSQL, Agile, Git.
Cloud Technologies: Azure (Databricks, Data Factory, ADLS), AWS (S3, Lambda, EC2, SQS), GCP, Talend DI, Docker.
Big Data & Machine learning: Hadoop, Spark, Hive, Kafka, Airflow, Gen AI, NLP, Pandas, Numpy, Tensorflow, keras, pyTorch, scikit-learn, Statistics, Data Visualization (Matplotlib, Seaborn, Power BI), Excel.