Shreel Shah

LinkedIn | 732-781-6589 | shreelshah.github.io | shreelshah12@gmail.com | GitHub

Education

Bachelor of Science — **Penn State University** — *State College, PA, USA* — **08/2019 - 12/2022**

- Major in Computer Science; Smeal Business Fundamentals Certificate

Skills

- Python | Java | JavaScript | NextJS | React | HTML | C | C# | Powershell | MSSQL | NoSQL | Node | Express | Flask | Git
- Azure | AWS | Snowflake | Apache Airflow | SSIS | ETL | Data Pipelines | OOP | Cloud Computing | CI/CD
- Microservices | Distributed Systems | API | Backend | Full-Stack | Data Science | AWS Certified Cloud Practitioner

Experience

Software Engineer — **Deloitte** — *Arlington, VA, USA* — **01/2023 - Present**

- Leading the design and development of multiple enterprise-level **microservice** applications for the NIH/NIAID using the latest technologies of **Azure**, **AWS, Python**, **Powershell, C#**, **.NET**, **SSIS**, and **SQL**.
- Developed an API **microservice** hosted in a Linux environment that utilizes OCR and machine learning tools to extract clinical trial lab data encompassed in PDFs and store it in structured data marts/warehouses via **ETL** for extensive research analysis at the NIH.
- Architecting and implementing scalable APIs for detecting critical cybersecurity vulnerability threats across the institute utilizing **Python**, **C#**, and **Azure cloud technologies** that serve hundreds of thousands of scan requests daily.
- Designing, implementing, and maintaining a concurrent data pipeline with an up-time of 99.8% to process transactional data from 40+ data sources using **SSIS**, **Airflow**, and **PySpark**. Migrating over 20+ separate workflows into automated **ETL** processes to improve runtime and operational efficiency by almost 75% and reducing manual workload by 30% weekly.
- **Continuous Integration/Deployment Pipeline Integration**, pull requests, code reviews, load/stress testing, unit/integration/e2e testing.

AI/ML Data Scientist — **Nittany AI Alliance** — *State College, PA, USA* — **01/2022 - 12/2022**

- Built an automated system using NLP, NLU, Computer Vision, and the public Twitter API to predict whether a tweet contained misinformation to construct a graph network for extended analysis focusing on the accuracy of social media content.
- Trained a random forest model on a sensory dataset to predict sleep duration using phone activity, light, screen status, charge status, and noise with an average mean prediction loss of +-0.3 hours.
- Designed and developed an Android mobile emulator using Flutter, the Android Sensor API, and the machine learning model trained on sleep data to create a reward-based system that incentivized users to improve their sleeping habits and patterns.

Software Engineer, Intern — **Deloitte** — *Arlington, VA, USA* — **06/2022 - 08/2022**

- Engineered a web application and dashboard that allowed the U.S. Customs and Border Patrol to visualize/forecast airport staffing Requirements for the post-COVID travel surge nationwide using **Python Flask**, AWS Forecast, **SQL**, and **HTML/CSS**.
- Developed and implemented an end-to-end pipeline for a key metrics dashboard utilizing **Python**, Amazon RDS, and Power BI. Ingested data from disparate sources and used the pipeline to process semi-structured data and feed the final views for the dashboard.
- Automated financial estimation and revenue calculations by converting Excel models to **Python** scripts which improved data management and performance tracking to save business analysts and project leadership about 20 hours/week.

Predictive Data Scientist, Intern — **Verizon** — *Basking Ridge, NJ, USA* — **06/2021 - 08/2021**

- Worked on an organization-wide data anomaly and analytical pipeline observability tool for the Verizon Data Engineering team.
- Worked on a recommendation engine (Deployed to Production) that conditionally displayed devices, accessories, and services on the Verizon product's website based on the user order history, search patterns, and ongoing promotions to increase sales.
- Leveraged data visualization and analytics tools (**Python** and Microsoft Power BI) to help create transition plans and roadmaps from 4G to 5G NR in several regions across the northeast US.
- Engineered an automated serverless forecasting application for the 5G Network Partnerships team to calculate and determine the overhead required for a network upgrade in a specific city/region based on geolocation, existing infrastructure, and other factors.
- Partnered with a Senior Solutions Architect to design and deliver a panel presentation to the organizational leadership team demonstrating Verizon Home 5G and Fiber Business Internet.

Projects

- **Personal Portfolio Website**: Utilized **Next.js**, **React.js, HTML, and CSS** to build an interactive web app portfolio highlighting experiences, passions, interests, and skillsets. The website is deployed on GitHub pages. **(06/2024)**
- **Nittany Marketplace**: Developed a full-stack online marketplace with buying and selling functionality for users. Implemented a database management system with **Python** and **Flask** on the backend and linking **HTML/CSS** for the interactive UI on the front end. **(04/2022)**
- **Dynamic Memory Allocator**: Implemented a heap memory allocator in **C** to support standard libraries such as malloc, realloc, and free. Optimized throughput and utilization through segregated lists, caches, and efficient block insertion policies. **(03/2022)**