Shruti Houji

Data Scientist/Analyst/Engineer

sghouji@iu.edu • 812-778-4724
GitHub • LinkedIn • Bloomington, IN 47408

Education

Master in data science — Aug 2022 – May 2024
Indiana University, Bloomington, IN — GPA - 3.762

Bachelor in computer engineering — Jun 2016 – Apr 2020
Pune University, Pune, India — GPA - 4

Technical Proficiencies

C++ | Python | Java | R | .NET | Microsoft Power BI | Tableau | MS Excel VBA | SQL | HTML5 | CSS | JavaScript | Microsoft Azure | Databricks | MongoDB | Neo4j | Snowflake | Hadoop MapReduce | PySpark | NLP | Keras | GCP | Jira | Flask Framework | Statistics | Pandas | SciKit-Learn | Tensorflow | PyTorch | AWS S3 | AWS Lambda | AWS SageMaker | Git | Docker | LLM

Career Experience

Danfoss Power Solutions, Cleveland, OH — May 2023 - Aug 2023
Data Science Intern

- Revamped daily, weekly, and monthly productivity reports of the shop floor in Microsoft Power BI. Enable the team to conveniently check status of orders every morning by connecting to MySQL server data to provide data solutions.
- Developed and maintain data pipelines to extract, transform, and load (ETL) data from various sources into data warehouses and databases to ensure data processing and quality by integrating data from multiple sources.
- Reduced manual creation time of report by 30 minutes to automate reporting for the production planning team with the help of Python and Excel VBA macros.
- Crafted insightful KPIs and 5 data visualizations, measured efficiency in achieving target goals by employing different inbuilt visuals, Python scripts, and DAX language.

Cognizant Technology Solutions, Pune, India — Dec 2020 – May 2022
Programmer Analyst - Developer

- Cleaned input files for data analysis, utilizing Python libraries and Microsoft Power BI to create graphical visualizations.
- Coordinated with a cross-functional team of 10 people on Jira tasks involving text and barcode extraction, classification group creation, and validation tests. Leveraged .NET and Docker to efficiently carry out business analysis tasks.
- Expedited input feature decision-making by 5%, streamlining process and enabling business requirements.
- Increased business process efficiency by 75% by implementing OCR solutions to process financial data using Data cap application, .NET technology, Oracle, Python, and agile methodologies along with SDLC framework.

Key Projects

Skin Lesion Classification for melanoma detection

- Spearheaded an end-to-end skin cancer detection system leveraging machine learning and amazon services like S3, SageMaker, Lambda, IAM and API Gateway to streamline analysis achieving 10% faster processing.
- Trained a neural network model with MONAI Pytorch and AWS cloud infrastucture to reduce resource utilization by 20%.

Datawage Navigator

- Structured a robust MySQL database design allowing users to access job listings in specific locations within milliseconds.
- Integrated CRUD operations and a dasboard page using Flask to get real time updates with 10% decrease in data latency.

Energy Analytics and Prediction in USA

- Employed time series models such as ARIMA, VAR, and LSTM for forecasting optimal energy resources for the upcoming decade by applying data driven approaches.
- Created a four-page dashboard using Flask, encompassing imports, exports, production, and consumption of the energy resource "Natural Gas" promoting sustainability.

Real-time Intrusion Detection Systems for IOT Networks using ML

- Built a dataset by self-generated IoT node using Tshark script to get instananeous network logs of around 1 GB data.
- Applied data mining techniques to develop an IDS using a Random Forest model with an accuracy of 99% with fronend notification to detect real-time attacks through websockets.

Home Credit Default Risk-Classification

- Engineered an advanced neural network model using PyTorch and ML models and addressed the challenge of financial instability, thereby simplifying and enhancing the loan lending process achieving an impressive Kaggle score of 76%.