

Shruti Houji

Data Scientist & Analyst

sghouji@iu.edu • 812-778-4724

[GitHub](#) • [LinkedIn](#) • [Portfolio](#) • Bloomington, IN 47408

Dedicated and tech-savvy data science graduate with hands-on internship experience. Proficient in data processing, advanced statistical analysis, data visualization, and machine learning. Eager to apply strong analytical skills to drive data-driven solutions.

Education

Master in Data Science

Indiana University, Bloomington, IN

Aug 2022 – May 2024

GPA - 3.770

Bachelor in Computer Engineering

Pune University, Pune, India

Jun 2016 – Apr 2020

GPA - 4

Skills

Programming Languages: C++, C#, Python, R, DAX

Data Analysis Tools: Microsoft Power BI, MS Excel VBA, Tableau, Matplotlib, Seaborn

Databases and Web Technologies: MySQL, NoSQL, Neo4j, BigQuery, Snowflake, HTML5, JavaScript

Cloud Platforms: AWS (S3, EC2, Lambda, SageMaker), Google Cloud Platform(GCP)

Other: Statistics, Data Structures, Databricks, PyTorch, Git/Github, Flask, Pandas, NumPy, Scikit-learn, Hadoop MapReduce

Work Experience

Danfoss Power Solutions, Cleveland, OH

May 2023 - Aug 2023

Data Science Intern

- Interpreted complex datasets, performed exploratory data analysis and statistical analysis to uncover trends, patterns, driving informed decision-making and strategic planning initiatives with a 5% increase in accuracy.
- Integrated diverse data sources into data warehouses, built data pipelines to extract, transform, and load (ETL) data and SQL databases enabling data modeling & processing to streamline workflow, reducing data processing time by 10%.
- Provided automation solutions for production planning reports using Python, R and Excel VBA macros, cutting creation time by 30 minutes.
- Crafted KPIs, 5 data visualizations, and productivity reports in Power BI, using Python scripts, and DAX language to deliver recurring Business Intelligence solutions, delivering insights and informing decisions.

Cognizant Technology Solutions, Pune, India

Dec 2020 – May 2022

Programmer Analyst – Developer

- Utilized Python, C#, Power BI, HTML, and JavaScript to clean input files for financial analysis and develop a web application.
- Collaborated with clients on the POC project to gather sprint module feedback, enhancing communication skills for business requirements and continuous improvement, resulting in a 10% increase in client satisfaction.
- Streamlined input feature decision-making by 5% through entrepreneurial leadership & collaborative teams.
- Boosted business efficiency by 20% with Python OCR scripts and software solutions for financial data processing using Datacap, .NET, Docker, Oracle, Git, GCP and agile methodologies within SDLC framework.
- Co-ordinated with a 10-person cross-functional team and stakeholders on Jira tasks, using organizational, problem-solving, and troubleshooting skills, presenting technical materials for imaging data analysis and validation tests.

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 infrastructure to reduce resource utilization by 15%.

Datawage Navigator

- Structured a robust MySQL database design allowing users to access job listings in specific locations within milliseconds.
- Integrated CRUD operations and dashboard page using Flask to get real-time updates with 25% 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 recommendations.

Real-time Intrusion Detection Systems for IOT Networks using ML

- Built a dataset by self-generated IoT node using Tshark script to get instantaneous 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 frontend notification to detect real-time attacks through websockets.

Home Credit Default Risk

- Engineered an advanced neural network model in PyTorch and predictive models, addressing financial instability, enhancing loan lending processes, achieving a Kaggle score of 76%.