Dhyey Joshi

+1 (930) 215-4548 | dhyjoshi@iu.edu | linkedin.com/in/dhyey-joshi12 | github.com/dhyeyjoshi | dhyeyjoshi.github.io

Education

Indiana University Bloomington — Bloomington, USA
Master of Science in Data Science | *GPA: 3.6/4.0* — *August 2023 – May 2025*
Relevant Coursework: Applied Algorithms, Big Data Systems, Machine Learning, Cloud Computing

Technical Skills

Languages: Python, Scala, R, SQL, Java, Pytorch, JavaScript, TypeScript, HTML, CSS
Libraries: Keras, Scikit-learn, NumPy, Pandas, Matplotlib, TensorFlow, SciPy, Theano, ggplot, Statmodels
Web Technologies: React, Django, Flask, FastAPI, RestAPI, Material-UI, Tailwind CSS
Big Data and Cloud: Apache Hadoop, Spark (MLlib), Amazon S3, Google BigQuery, MongoDB, Amazon DynamoDB, Amazon Redshift, Snowflake, Apache Kafka, AWS Glue, MySQL, PostgreSQL, NoSQL
Tools: GitHub, Git, Postman, Figma, VS Code, JIRA, Apache
DevOps: Docker, AWS, CI/CD, Kubernetes, Terraform, ETL, Bash, Linux/Unix

Experience

Graduate Research Assistant — March 2024 – Present
Indiana University Bloomington | Python, TensorFlow, React, Flask, FastAPI, AWS (S3, Glue) — *Bloomington, USA*

- Led research under Professor Jiangmei Wu to enhance **generative AI models**, fine-tuning **GANs** and the **Stable Diffusion Model**, resulting in the generation of 10,000+ high-fidelity 3D origami designs.
- Developed and maintained a responsive **web app** with **React.js**, **Material UI**, **Flask**, and **FastAPI**, improving deployment efficiency by 50% through a **CI/CD pipeline** on **JetStream2 Cloud**.
- Managed data storage in **Amazon S3**, reducing query response time by 40% and improving data processing speed by 35%, handling 1,000+ user inputs and 500+ design prompts in **Amazon DynamoDB**.
- Utilized **AWS Glue for ETL processes** and monitor application performance with **AWS CloudWatch**, ensuring efficient data transformation, integration, and system health.

Data Engineer Intern — January 2023 – May 2023
Symbiosis Centre for Applied Artificial Intelligence | MLflow, Docker, AWS SageMaker — *Pune, India*

- Led predictive model development for banking telemarketing, enhancing accuracy by 15% with **scalable MLOps tools (MLflow, Airflow)**, and boosted training efficiency by 40% with **Pyenv and Poetry**.
- Implemented **MLflow pipeline** to build a **Docker image** from ML models, push to **AWS ECR**, and deploy via **AWS SageMaker** endpoints, also configuring **S3** for model artifact storage.

Data Engineer (Co-op) — May 2021 – December 2021
Exposys Data Labs | Python, SQL, ETL, Tableau, Amazon Redshift — *Bangalore, India*

- Spearheaded **ETL** and **credit risk analysis** initiatives using **Python** script and **SQL** for data manipulation.
- Achieved over **95% accuracy** in **data cleansing, processing, and analyzing** over **500,000 loan records**.
- Developed dynamic **Tableau dashboards** to visualize risk scores and loan approval rates, integrating **predictive modeling** for better risk management and decision-making.
- Managed **AWS cloud-based workflows** utilizing **Amazon Redshift** for scalable and efficient data processing, handling over **1TB of data daily** and improving analytics processing speed by 30%.

Projects

A2Z-AmazeComparator | *Flask, React, Amazon RDS, Docker, Tailwind CSS* — January 2024 – May 2024

- Developed the backend with **Flask** and the frontend with **React**, using **Docker** for scalable and isolated deployment, ensuring seamless integration.
- Analyzed sales data using **Amazon RDS** to offer **insights into product pricing, ratings, and reviews**.
- Integrated a **Collaborative Filtering Recommendation System** and visualized data through dynamic dashboards, improving user decision-making efficiency.

FireEye | *TensorFlow, OpenCV, IoT,* ***Open Source Contributor*** — May 2022 – December 2022

- Developed a **Pile Fire Detection System** using **transfer learning** in **CNNs**, analyzing **10,000+ thermal images**, achieving **95% accuracy** in early fire detection to prevent Amazon forest fire outbreaks.
- Deployed a **real-time monitoring solution** on **Raspberry Pi** using **TensorFlow**, integrating fine-tuned **ResNet** and **InceptionV3 models** for detection and alerting, reducing response time by **50%**.

Publications

- Face-Mask Detection Using Stacked 2D Convolutional Neural Network.
- Parkinson's Detection Using Speech Processing and Ensemble Learning Algorithms.