

# Atharv Abhijeet Bagde

atharv.bagde1798@gmail.com • (+1) 812-803-9770  
linkedin.com/in/atharva-bagde-77970016b/ • Bloomington, IN  
atharvabagde.com/ • github.com/atharvabagde

## Qualifications Summary

Highly analytical and detail-oriented professional, currently pursuing a master's degree in data science from Indiana University; eager to leverage academic knowledge and expertise to support organizational growth and success as a **Data Scientist**.

### Data Analysis & Interpretation

- ◆ Proven expertise in analyzing complex datasets, extracting actionable insights, and identifying trends and patterns to inform decision-making processes.

### Statistical Modeling

- ◆ Adept at applying statistical techniques and models to solve real-world problems, including hypothesis testing, regression analysis, and predictive modeling.

### Deep Learning

- ◆ Skilled in developing and employing deep learning models and machine learning techniques for image segmentation, generation, and reconstruction.

### Coding & Programming

- ◆ Proficient in programming languages, such as Python and R with experience in data manipulation and analysis using relevant libraries and frameworks.

### Database Management

- ◆ Experienced in SQL for querying, extracting, and loading large data from relational databases, while ensuring data integrity and accessibility.

### Cloud Computing

- ◆ Certified developer of cloud-based pipelines with the capability of serverless data processing, distributed training, and model deployment on GCP and AWS.

## Education

### Master of Science in Data Science (CGPA: 3.4)

Courses: Exploratory Data Analysis, Database Concepts, Machine Learning, Financial Econometrics, Network Science  
Indiana University, Bloomington, IN, Aug 2022 to May 2024

### Bachelor of Engineering in Electronics and Telecommunication (CGPA: 3.78)

Mumbai University, Mumbai, IN, Aug 2016 to May 2020

## Technical Proficiencies

<b>Programming:</b>	Python, MATLAB, R, MySQL, TensorFlow, PyTorch,
<b>Frameworks:</b>	PyTorch, Tensorflow, Keras, ONNX, DeepSpeed, Scikit Learn, REST, FLASK, JAX, Huggingface
<b>Data Analytics:</b>	Matplotlib, Tableau, PowerBI, Seaborn, Gephi, ggplot2, OpenCV, SAS, Alteryx, QuPath
<b>Machine Learning:</b>	Regression, SVM, k-means, Xg-boost, Random Forest, k-NN, Graph Neural Networks
<b>Statistics:</b>	Bayesian Statistics, Probability Distributions, Multivariate Analysis, Time Series Analysis
<b>Methodologies:</b>	JIRA, Azure Boards, Asana, Agile, Kanban
<b>MLOPs Tools:</b>	DVC, Git, Kubeflow, MLFlow, Docker, CI/CD pipeline

## Professional Experience

### INDIANA UNIVERSITY BLOOMINGTON, Bloomington, IN Research Assistant

Oct 2022 to Aug 2023

Teaming up with researchers across universities on a **NIH** funded initiative – **HuBMAP** with the aim of constructing a 3D model of human body on a cellular level.

- Engaged with renowned academic professionals, including Dr. Katy Borner, Dr. Chao Chen, and Dr. Ali Eturk to enhance the performance of a **3D vascular segmentation** by experimenting with novel loss functions and metrics.

- Managed more than 100 GB of CT scans of the large intestine, extracting, and harnessing pertinent **intensity and texture-derived** characteristics for **tumor segmentation**, attaining an outstanding DICE score of 86.8 through meticulous fine-tuning of a **Vision Transformer** architecture.
- Spearheaded the creation of a cutting-edge image processing application, achieving an impressive 84% DICE score for the detection and segmentation of **FTU cell regions** within the **glomerulus** tissue structure.

**QUANTIPHI ANALYTICS SOLUTIONS PVT. LTD., Mumbai, IN**  
**Machine Learning Engineer**

**July 2020 to June 2022**

Collaborated with cross-functional teams, including machine learning engineers to develop and implement machine learning models using cutting-edge algorithms and techniques.

- Conceptualized and implemented a **digitization** pipeline for **ICU** notes, showcasing the capability to interpret graphical and tabular handwritten data via GCP **AutoML** and Faster **R-CNN** model architecture, resulting in 83% accuracy rate for graphical data and an 87% accuracy rate for tabular data.
- Spearheaded the development of a parameterized 3D **Style GAN** tailored for generating lifelike **CT organ images**, integrating age, gender, and modality as input parameters, yielding an impressive FID score of 7.8.
- Enhanced the **DICOM to NIFTI** conversion script by building a batch processing pipeline on GCP Dataflow using **Apache Beam** on the backend, leading to a 6x speedup and monthly savings of \$10K+ in cloud computing costs.
- Built a **3D U-Net** model using **PyTorch** and GCP **Vertex AI** to accurately localize and detect renal stones, in a CT image, with F1-score of 0.62, MAP score of 32.45 and a deviation of  $\pm 0.4$ mm for stone diameter.
- **Mentored** over 10 new **team members** with essential guidance through an ML foundational **training program** at Quantiphi, consistently facilitating **counseling** and **doubt-solving** sessions.
- Designed a **real-time object detection** system for surveillance videos employing the **YOLOv4** deep learning model that identified and tracked individuals, vehicles, and targeted objects promptly, boasting an impressive mAP score of 0.87.
- Consulted on 4 projects in the capacity of a **technical team leader**, overseeing **solution workflow** design, task **delegation**, and maintaining consistent **communication** with clients for **requirements gathering** and **feedback collection**.
- Created a dynamic data **visualization dashboard** in **Tableau** for tracking sales **KPIs**, seamlessly importing and analyzing data from Google Cloud Platform's **Big Query**, precipitating a remarkable 22% surge in sales revenue through data-informed decision-making and improved sales strategies.

**Academic Projects**

**Exploring US Philanthropic Giving Networks**

- Constructed **bipartite graph databases** and **embeddings** to analyze relationships between **1M+** US foundations and grantees by building **PySpark** data processing pipeline to optimally process the extensive **IRS tax data** and integrate it into **Neo4j** and **Node2Vec** frameworks.
- Developed **GraphSAGE** and **GAT** models in **PyTorch** and **NetworkX** to predict optimal foundation-grantee pairs based on features like geography, cause area, size, and relationships.

**Studying Microbiome Network in the Human Body**

- Compiled a comprehensive dataset comprising 100k+ news articles, 10k+ research publications, 500+ books, and online resources, utilizing precise data manipulation techniques and leveraging **Transformer-based NLP models** for targeted information extraction.
- Built a **graph** data structure that featured **microbiomes** as nodes and meticulously captured their intricate interactions with other microbiomes, food sources, and enzymes, thus constructing a comprehensive **knowledge network**.
- Applied **Graph Attention Networks (GAT)** to conduct an in-depth analysis of the graph's inherent properties and proficiently identified distinct microbial communities, enhancing the understanding of complex ecological relationships.