

HARSH PANDYA

☎ 812-778-5344 ✉ harshmpandya.00@gmail.com 🔗 [linkedin.com/in/harsh-pandya-ai/](https://www.linkedin.com/in/harsh-pandya-ai/) 📊 public.tableau.com/harshpandya

Education

Indiana University - Bloomington

Masters in Data Science - 3.77 GPA

Aug. 2022 – May 2024

Bloomington, IN

Mumbai University

Bachelor of Engineering in Information Technology - 3.87 GPA

May 2018 – July 2022

Mumbai, India

Experience

Data Scientist

Oct. 2022 – Present

Indiana University Athletics

Bloomington, IN

- Identified inefficiencies in U.S. Diving team's training methods and athlete recruitment process. Collaborated closely with the Olympic head coach and the **U.S. Olympic and Paralympic Committee** to boost training and recruiting efficiency by **40%** through building dashboards for identifying **KPIs** while **automating the reporting process**.
- Addressed the lack of daily players performance feedback and data availability by constructing a **Python** based data pipeline, to support the development of **data infrastructure**, scraping **1 million** data records with **Selenium** for daily training analysis, resulting in a **50%** increase in athlete performance rates in competitions.
- Recognized the team's success rate challenge, strategized, analyzed strengths through **statistical analysis**, thereby leading to the team's victory in both the **2023** and **2024 Big Ten championships** for IU.

Machine Learning Associate Intern

Feb. 2022 – May 2022

Infosys Springboard

Mumbai, India

- Utilized advanced deep learning techniques like **GANs** for image-to-image translation utilizing **TensorFlow** to address **color vision deficiency** in individuals conducting formal experiments to refine algorithms for enhanced accuracy.
- Resolved data generation challenges by communicating requirements and **implementing a research-based algorithm** using **Python** to generate simulated images, achieving a notable **30%** increase in data creation.

Machine Learning Intern

Apr. 2021 – Jul. 2021

AiBorne

Mumbai, India

- Led a **3-member** team and collaborated with manufacturing personnel to develop computer vision models leveraging **RCNN**, **Detectron2**, and **OCR** (Google Vision API) for accurately extracting car mileage information, culminating in a **60%** reduction in insurance claim processing time, maintaining a strong focus on customer and stakeholder needs.
- Enhanced accuracy** and **minimized false positives** (identified through **root cause analysis**) in damage detection models for cars through research-driven enhancements, employing preprocessing strategies. Accomplished **4%** increase in accuracy, enabling precise price estimation for damages, boosting the Customer Lifetime Value by **15%**.

Projects

Healthcare: Predict Patients Length of Stay | Python, Snowflake, AWS Sagemaker

Ongoing

- Planned to develop regression models, integrate predictions into Snowflake, **deploy scoring**, and **automate status emails** to enhance predictive analytics to enable optimized resource allocation and streamlined operations to build an end to end **retraining pipeline** by checking **Data** and **Model Drift** and redeploy the model.

Loan Default Prediction | Python, DVC, Docker, AWS (S3, ECR, EC2), Github Actions

March 2024

- Built and deployed an **XGBoost** binary classifier (0.8-F1 score) on Cloud Infrastructure, integrating Explainable AI techniques like **SHAP** and **Anchors**, to enhance transparency and interpretability for stakeholders and regulators, optimizing decision-making and improving regulatory compliance, aiming to increase **customer retention rate**.
- Designed workflow for ML lifecycle with DVC for Experimentation tracking, Docker, and Github Actions for **CI/CD**, reducing deployment time by **50%** by integrating the model into production, ensuring seamless deployment.

Telecom Data Analysis to Improve Service Quality | Python

September 2023

- Conducted Telecom Data Analysis by **designing experiments**, utilizing EDA techniques like outlier detection, **MICE imputation**, **Chi-square**, and **ANOVA** to identify challenges, establish success criteria, and improve service quality, aiming to reduce **customer churn rate** through targeted interventions.

Technical Skills

Competencies: Machine Learning, MLOps (ML Model-Interpretability, Evaluation, Deployment, Maintenance), Computer Vision, Generative AI, NLP, Predictive Modeling, Statistical Modeling, Regression Analysis, Stakeholder Management, LLMs, Data Mining, Cloud Computing Services(AWS), Quantitative and Qualitative Analysis.

Programming Languages: Python (Numpy, Pandas), R, SQL.

Visualizations: Matplotlib, Seaborn, Tableau, ZOHO, ArcGIS.

Tools/Frameworks: TensorFlow, PyTorch, FastAPI, Kubernetes, AWS, Snowflake, Containerization (Docker), Git, CI/CD.