Dilip Francies

LinkedIn GitHub dfranci@iu.edu +1-812-803-4479

Education

- **Indiana University, Bloomington** Indiana, US
  Master of Science in Data Science, GPA: 3.84 *August 2023 - May 2025*
  Courses: Deep Learning Systems, Image Processing, Applied Machine Learning, Elements of Artificial Intelligence, Scientific Visualization, Statistics, Cloud Computing, Applied Algorithms, Big Data Applications

Experience

- **Institutional Analytics, Indiana University**
  Intern - Research and Analytics *Jan 2024 - Present*
  - **Explainable AI**: Applied tree-based models and weighted Shapley values to pinpoint factors influencing student academic performance and retention, delivering actionable insights to IU's Vice President of Student Affairs.
  - **Bias Mitigation**: Reduced predictive model bias by rigorously analyzing demographic fairness metrics (equal odds, disparate impact, true/false positive rates) using IBM's AIF360.
  - **ML Experimentation Framework**: Developed scalable ML pipelines using Scikit-learn and PyTorch on BigRed-200 HPC clusters, accelerating model experimentation by 6x.
  - **Unsupervised Clustering**: Leveraged PySpark and variational deep embeddings to cluster 16,000 courses from 20M rows of institutional data, improving clustering robustness by a 7% silhouette score.
  - **Natural Language to SQL - LLM Finetuning**: Developing a NL2SQL proof-of-concept by constructing a semantic layer with database schema embeddings, leveraging PEFT on SQL Coder, CodeLlama, and Phi-3 models to optimize query generation accuracy via RAG techniques.
  - **Data Visualization - Winner**: Winner of the Data Visualization competition to represent IU at the Big Ten Data Visualization Championship. Expertise in developing multimodal data visualizations using Seaborn, Plotly, ggplot2, Matplotlib, Tableau, and Power BI. *BTAA Data Viz Gallery*
- **Department of History - Indiana University**
  Research Assistant *May 2024 - Present*
  - **Multi-Modal Integration**: Developing a multimodal ML pipeline leveraging customized CNN feature extraction on elephant ivory images and textual descriptions to enhance geographical origin prediction models.
  - **Analytical Machine Learning**: Developed XRF spectral analysis for ivory provenance mapping with 82% accuracy using Random Forest, ALS correction, and multivariate statistical modeling.
  - **Automated Web Scraping**: Built web scrapers for 100,000+ ivory samples across 12 museums using BS4 and Selenium, creating an open-source research database and mapping ivory trade route patterns. *ArchivesofIvory*
- **Centre of Bioinformatics Research - Indiana University**
  Research Assistant *May 2024 - September 2024*
  - **Multi-Modal ML**: Developed a PyTorch multi-modal ML framework for Alzheimer's prediction, integrating ResNet-50 3D MRI feature extraction and transformer-based SNP embedding, enabling precision multi-class neurological classification of gigabyte-scale genomic and imaging data.
  - **Model Interpretability**: Implemented large-scale CNN visualization using Grad-CAM on IU-BigRed200 Supercomputer, leveraging ResNet-101 and Vision Transformer for advanced feature extraction and interpretability analysis.
- **Hexaware Technologies, Mumbai, India**
  Senior Data Analyst - Machine Learning *November 2021 - July 2023*
  - **Facial Recognition**: Engineered a 128-dimensional feature vector-based facial recognition system using Siamese one-shot learning, achieving 99% accuracy across a 2000-employee enterprise authentication system.
  - **Computer Vision - Object Detection**: Developed a real-time object detection system using OpenCV, TensorFlow, and YOLO-V3 (Darknet-53), creating an employee proctoring system.
  - **Model Optimization**: Accelerated machine learning inference by 70% through INT8 quantization-aware training and FP16 optimization, while reducing model size by 75% via TensorFlow Lite conversion for edge device integration.
  - **Cloud Deployments & A/B Testing**: Engineered a computer vision model deployment pipeline, containerizing TensorFlow models in AWS ECS, reducing latency by 3x, and implementing A/B testing for performance optimization.
  - **Big Data Analytics**: Analyzed 200k+ records to identify customer investment patterns for targeted marketing through segmentation.

Projects

- **Organ Segmentation in 3-Dimensional CT Scans — Computer Vision**
  - **Medical Imaging**: Developed a MONAI SegResNet-powered 3D organ segmentation system for 104 CT scans, leveraging voxel-based deep learning to enable automated volumetric diagnostic analysis.
- **Vehicle Dynamic Model Estimation — Indy Autonomous Challenge, Las Vegas**
  - **Physics-Informed Neural Networks**: Developed a PINN with LSTM/GRU architectures using a bio-inspired genetic algorithm for parameter optimization for autonomous racing vehicle dynamics; validated on a real-time race car.

Skills Summary

- **Computer Vision**: ResNet-50/101, Vision Transformers, YOLO, Multi-Modal Feature Extraction, Image Segmentation, Facial Recognition, Neural Network Interpretability, Inference Optimization, Model Serving
- **ML/AI**: Scikit-learn, TensorFlow, PyTorch, XGBoost, Optuna, Grid Search, OpenCV, Spacy, Transformers
- **Certifications**: (Deep Learning Specialization),CNN Specialization Image Segmentation