Riyanshi Bohra

352-757-8079 | riyanshibohraa@gmail.com | Portfolio | LinkedIn | GitHub

Education

University of Arizona — Tucson, AZ
Master of Science (M.S.) in Data Science; ***GPA:*** *3.8/4.0* — *Aug 2023 – May 2025*
Coursework: Artificial Intelligence, Machine Learning (ML), Applied Natural Language Processing (NLP), Data Visualization, Data Mining & Warehousing

Manipal University Jaipur — Jaipur, India
Bachelor of Technology (B.Tech) in Information Technology; ***GPA:*** *3.72/4.0* — *Jul 2019 – May 2023*
Coursework: Deep Learning, Relational Databases (RDBMS), Data Structures & Algorithms, Operating Systems, Probability & Statistics, Object-Oriented Programming (OOP)

Professional Experience

Data Scientist (Research Assistant) — Jan 2024 – Present
Zuckerman College of Public Health, University of Arizona — *Tucson, AZ*

- Developed Python and SQL pipelines to transform **1B+** public health records spanning **15** years into a centralized reporting layer, accelerating insight generation for **3** state-funded reports impacting **2M+** students
- Built Tableau dashboards visualizing **30+** spatial and temporal trends across **200+** districts, driving **60%** increase in usage by non-technical stakeholders through improved data clarity
- Applied predictive modeling in R to analyze **1,200** school sites and segment infrastructure equity gaps, enabling **$750K** in redirected investment toward underserved zones

Data Science Intern (Applied ML) — Jul 2022 – Sep 2022
PwC (PricewaterhouseCoopers) — *Mumbai, India*

- Engineered ETL pipelines using Python, SQL, and Apache Spark, automating data ingestion and transformation of **15,000+** monthly manufacturing records, reducing cycle time by **45%**
- Trained machine learning models on **7+** years of production data to optimize operational workflows, increasing tablet output by **20%** annually (**400K → 480K units**) and achieving cost savings of **$50,000**/year
- Designed **5** interactive Power BI dashboards with A/B testing integration, supporting trend analysis and KPI monitoring across **7** production lines, saving **350+** man-hours/week

Technical Skills

Languages: Python (Pandas, NumPy, Scikit-learn, Matplotlib), SQL (MySQL, PostgreSQL), R, JavaScript
Frameworks & Libraries: Flask, FastAPI, React, Next.js, TensorFlow, PyTorch, SpaCy, Hugging Face Transformers
Data & BI Tools: Tableau, Power BI, Looker, Microsoft Excel (Pivot Tables, VBA), A/B Testing (SciPy, Statsmodels)
Generative AI: LangChain, LangGraph, VectorDB (Pinecone, Deep Lake), Retrieval-Augmented Generation (RAG)
Developer Tools & Cloud: Git/GitHub, Docker, VS Code, Jupyter Notebook, Cursor, RStudio, AWS, Azure

Project Experience

Biometric Predictors of Emotional States | *Python, TensorFlow, Hugging Face Transformers* — GitHub

- Engineered a hybrid deep learning pipeline combining GPT-2 for mood simulation and BERT for emotion classification, processing **50,000+** biometric-text pairs with **4x** improved detection accuracy
- Built an ensemble ML framework integrating LSTM and Gradient Boosting to analyze **15+** wearable health metrics, achieving a **95.7%** F1-score and cutting latency by **82%** compared to RNN baseline models

SafeDrive AI: Real-time Distracted Driving Detection | *Python, TensorFlow, Computer Vision* — GitHub

- Deployed a CNN-based classification pipeline using TensorFlow and Keras, analyzing live dash camera feeds to detect and classify **5** types of driver distractions with **92%** accuracy on a **50,000**-image dataset
- Integrated a real-time AI-powered alert system, using edge computing for on-device inference, reducing distraction incidents by **30%** across **200** monitored drivers in a **3-month** pilot study

DocTalk: AI-Powered Document Assistant | *LangChain, Whisper, Eleven Labs, Streamlit* — GitHub

- Developed a voice-interactive assistant using LangChain, OpenAI APIs, and Whisper, enabling natural language querying with **90%** transcription accuracy and sub-**2.5s** response time
- Implemented Pinecone vector search over **100+** internal documents, improving retrieval precision by **22%** using manual scoring and enabling scalable stakeholder-facing NLP insights