Aayush Panchal

Columbus, OH | m: 614-353-8167 | aayushpanchal1111@gmail.com | linkedin.com/in/atpp

Professional Experience

Ohio Health | Columbus, OH, June 2025 – August 2025 | **DATA SCIENCE INTERN**

- Designed and deployed an end-to-end AI pipeline combining Wav2Vec2 embeddings with Llama-70B LLM, enabling sentiment and topic extraction from more than 50,000 clinical audio files to accelerate insight generation.
- Automated ETL workflows using Azure Databricks Jobs API, creating parameterized pipelines that ensured repeatable, scalable, and reliable processing of large clinical datasets.
- Applied MLflow for systematic experiment tracking, enabling reproducibility and providing audit-ready records of models and results.
- Developed PySpark UDFs to integrate ML inference within distributed processing pipelines, reducing manual review time by 75% and significantly increasing throughput for clinical stakeholders.
- Partnered with clinical and IT teams to optimize model efficiency and infrastructure, achieving a 40% reduction in compute costs while maintaining quality and compliance standards.

Mollissoft | El Segundo, CA, May 2023 – February 2024 | **DATA ENGINEER**

- Devised and maintained scalable ETL pipelines in Python and SQL which improved the freshness of e-commerce analytics data by 40% and enhanced the accuracy of client reporting.
- Re-engineered MySQL database schemas and optimized complex queries which led to decreasing dashboard load times by 50% and improving the usability of data tools for business teams.
- Implemented monitoring frameworks on AWS that reduced downtime by 30% to ensure availability and reliability for client-facing platforms.
- Containerized data workflows using Docker and integrated them with CI/CD pipelines, which shortened release cycles by 35% and minimized disruptions during updates.
- Introduced standardized Git and GitHub practices for version control across the engineering team, reducing deployment errors by 20% and improving overall team productivity.

WallFort Inc. | Mumbai, India, May 2022 – August 2022 | **DATA ANALYST INTERN**

- Applied statistical models in R to estimate loan repayment probabilities under uncertainty, strengthening decision-making for credit assessments and financial risk management.
- Designed AI-enhanced Tableau dashboards and automated reports to surface actionable business and operational insights for internal teams.
- Built predictive models combined with A/B testing, enabling targeted marketing campaigns that improved customer retention by 15%.

Projects

Audio Sentiment Analysis Project | Ohio Health | Columbus, OH, June 2025 – August 2025

- Orchestrated Databricks pipelines to ingest 40,000+ WAV files, extracting embeddings and parsing topics with Wav2Vec2 + LLM integration.
- Accountable for implementing MLflow model registry to standardize and reproduce text-analytics experiments.
- Optimized compute efficiency with Delta Lake pruning and autoscaling**,** cutting costs by 40% while maintaining reliability**.**

Patient Follow-Up Analysis Project | Ohio Health | Columbus, OH, June 2025 – July 2025

- Constructed LLM-powered pipelines in Unity Catalog to extract insights from 20,000+ clinical records.
- Employed AI-driven pattern recognition to prioritize high-risk patients for follow-up care.

Education and Credentials

Master of Science (M.S.) in Statistics, 2026 (Expected); *The Ohio State University, College of Arts & Science, OH*
Bachelor of Science (B.S.) in Computer Science and Mathematics, (Statistics Concentration), 2022; *UMASS Amherst, MA*

Technical Skills

Programming Languages: Python, Java, SQL, C++, R, HTML, JavaScript, Scala
Frameworks: TensorFlow, PyTorch, PySpark, LangChain
Tools: Git, GitHub, Docker, Kubernetes, MLflow, Databricks, Delta Lake, AWS SageMaker, Microsoft Azure, Tableau, Power BI
Libraries: scikit-learn, XGBoost, SpaCy, Matplotlib