

AMAN PANDEY

508-373-8918 | amanpandey.ds@gmail.com | linkedin.com/in/amanpandeyy | github.com/aman-720

SUMMARY

MS Data Science student at ASU (GPA **4.0**/4.0, Dec 2026) with **4+ years** of production experience building ML pipelines, NLP systems, and data analytics platforms. Skilled in Python, PyTorch, deep learning, and cloud deployment (AWS, Docker, FastAPI). Seeking an AI/ML Engineering internship focused on Agentic AI, Retrieval-Augmented Generation (RAG), and large-scale ML systems.

EDUCATION

Arizona State University M.S., Data Science, Analytics and Engineering	GPA: 4.00	Tempe, AZ Jan 2025 – Dec 2026
Amity University B.Tech., Computer Science		Noida, India Jul 2016 – May 2020

EXPERIENCE

- Software Engineer** *My Next Film, New Delhi, India* Apr 2023 – Dec 2024
- Engineered a multilingual NLP pipeline supporting **114 languages** using seq2seq Transformers on AWS (EC2, S3, Lambda) with Google/Azure speech APIs, improving translation accuracy by **76%** and reducing manual review costs; delivered insights via Tableau dashboards for stakeholder reporting.
 - Developed a reviewer web app with automated task allocation, reducing project cycle time by **41%** and improving translation quality by **20%**; automated **400+** voice narrations using Amazon Polly, matching accent and timbre to character profiles across markets.
- Data Analyst** *Youth Buzz, Noida, India* Sep 2022 – Mar 2023
- Boosted Net Promoter Score by **+10** through analytics-driven strategy; automated survey reporting with zero-shot NLP classification and built Power BI dashboards to surface customer experience insights for leadership.
 - Built churn prediction models (logistic regression, **84% AUC**) on **50K+** customer records using Python (pandas, scikit-learn) and SQL; identified fee-driven attrition via RFM clustering, advising a strategy that cut attrition **50%** in fee-sensitive cohorts within one quarter.
- Software Developer** *Invesca Technology, Noida, India* Dec 2020 – Jul 2022
- Architected Celery/Redis distributed task queues processing **2M+ daily transactions**; reduced pipeline latency by **40%** and achieved **99.9% SLA** compliance across peak campaign windows handling **10x** normal traffic volume.
 - Built log analytics dashboards and multithreaded Python services that boosted backend throughput by **35%**; automated anomaly detection alerts, preventing overload incidents during peak operations.

PROJECTS

- GlucoCast** | *Generative Diffusion Framework-Based Blood Glucose Forecasting*
- Built a conditional diffusion model generating privacy-preserving synthetic CGM data conditioned on meals, insulin, and activity; outperformed LSTM/CNN baselines by **18% RMSE** on the OhioT1DM dataset.
- FinFusion** | *Deep Learning for S&P 500 Forecasting* | GitHub
- Benchmarked ARIMAX, LSTM, and Temporal Fusion Transformer across **450+ experiments** (11 phases) for S&P 500 return prediction; discovered gradient collapse in financial TFT and identified weekly resampling as optimal configuration achieving **59.1%** directional accuracy across 9-fold rolling evaluation (2016–2024).
- Pulse2Symphony** | *Biosignal-Conditioned Music Generation*
- Built a mobile app extracting HRV features (SDNN, RMSSD, LF/HF ratio) from smartphone camera PPG, classifying mood via CNN-LSTM into Russell's valence-arousal space, and generating personalized instrumental MIDI using an emotion-conditioned Transformer decoder with REMI tokenization.
- Traitlytics** | *LinkedIn Personality Predictor* | GitHub
- Built an NLP pipeline predicting Big-Five personality traits from LinkedIn text using BERT and RoBERTa with TF-IDF features; deployed batch and real-time REST endpoints on AWS (FastAPI, Docker, EC2).
- BasketIQ** | *Market Basket Analysis* | GitHub
- Analyzed **32.4M** Instacart transactions to mine Apriori association rules and segment users into **5 RFM-based** clusters using K-Means; visualized findings in Tableau to drive targeted marketing and retention strategies.

TECHNICAL SKILLS

Programming & Analytics: Python (pandas, NumPy, SciPy, scikit-learn), R, Java, SQL (PostgreSQL, BigQuery), PySpark, Snowflake, Git, Statistics, Data Modeling, Data Mining, Algorithms

Visualization & BI: Tableau, Power BI, Looker Studio, Google Analytics, Matplotlib, Seaborn, Advanced Excel

Machine Learning: Deep Learning (PyTorch, TensorFlow), NLP (BERT, RoBERTa), Transformers, LSTM, Large Language Models (LLMs), Feature Engineering, CNN, RNN, Diffusion Models, Reinforcement Learning, Statistical Inference

GenAI & LLM Ops: Agentic AI, Generative AI, LangChain, Retrieval-Augmented Generation (RAG), Model Context Protocol (MCP), Prompt Engineering

MLOps & Cloud: Docker, Kubernetes, CI/CD, FastAPI, Celery, Redis, AWS (EC2, S3, Lambda, API Gateway), Azure, ETL/ELT Pipelines, Distributed Systems, Anomaly Detection