Anshul Gupta

Bristol | ans2852000@gmail.com | +44 790 945 24 01 | +91 810 773 99 59

Education

University of Bristol (UK), MS Data Science (Distinction) Sept 2024 – Nov 2025

- **Coursework:** Statistical Computing, Large-Scale Data Engineering, AI and Text Analytics, Visual Analytics

Experience

Data Engineer, Deloitte, Bengaluru Feb 2023 – Jun 2024

- Designed and delivered a reliable daily ELT pipeline ingesting 5–20M records via Kafka into Postgres in <1 hour, meeting strict data-integrity and SLA requirements and reducing load latency by 1–2 hours.
- Developed Python automation for job scheduling, pattern matching, and monitoring, eliminating hours of manual effort per week and improving operational reliability.
- Automated network-testing workflows by scripting command sequences and implementing state-machine-based response validation for a major U.S. telecom client.
- Resolved recurring production pipeline failures and refactored key components following SOLID principles, improving code quality, stability, and maintainability.

Projects

Automated Polymer Property Extraction from Research Articles (Jun 2025–Sept 2025)

- Built an end-to-end NLP system to extract polymer properties from 3,000+ full-text research papers, creating unified corpora and a consistent entity–relation annotation schema.
- Designed a scalable data-harvesting and preprocessing pipeline using regex and ontology-based annotation, enabling high-coverage, high-quality training data generation.
- Developed a character-aware BiLSTM-CRF model that outperformed BiLSTM and SciBERT transformer baselines, achieving 96% macro F1 (+30% improvement) while reducing compute requirements and inference time by ∼80%.
- Delivered a deployable FastAPI + Streamlit application integrating data ingestion, model inference, and structured database storage, with insights transferable to NLP tasks in healthcare, finance, and legal domains.

Stratifying Tumors and Normal Tissues via MIR100HG-Centered Multi-Omics Signatures
Group Administrator, Industry Project with Nottingham Trent University, Jan 2025–Apr 2025

- Reduced a 56,000-feature multi-omics dataset to a compact 10–20-feature MIR100HG signature using differential expression and network analysis, enabling clear cancer subtype separation.
- Identified robust tumour clusters by combining outlier removal (Isolation Forest) with PCA (PC1 explaining 42–76% variance) and validated feature importance using a Random Forest classifier (F1 > 90%).
- Demonstrated prognostic value by stratifying pancreatic and lung cancer patients into high- and low-risk groups through Cox regression and Kaplan–Meier survival analysis (significant log-rank p-values).
- Delivered a reproducible analysis pipeline across multiple datasets, providing an interpretable biomarker signature with applications in precision oncology and multi-omics research.

Automated Onboarding/Offboarding Architecture (Architecture Design Project)
180 Degrees Consulting — Global Digital & IT Specialist, 2025

- Designed a scalable, cost-efficient onboarding/offboarding architecture using Google Apps Script for validation, n8n for orchestration, and a Python microservice on AWS Lambda for automated provisioning.
- Defined an adaptable workflow integrating APIs (Google Workspace Admin, GitHub, Slack) with Postgres/S3 audit logs to support reliable, low-maintenance operations for volunteer-heavy teams.

Skills

Python, SQL, PostgreSQL, Redis, Kafka, Docker, AWS, CI/CD (Jenkins, GitHub Actions), Git, Linux, Bash, FastAPI, Django, PyTorch, Pandas, NumPy, REST APIs, Tableau