

ARSALAN ANWAR

New York, US · 929-578-1098 · arsalananwarofficial@gmail.com · github.com/arsalananwar11 · linkedin.com/in/arsalan-anwar-ai · arsalananwar.com

Work Experience

Data Science Intern Medidata Solutions

Feb 2025 - Jun 2025
New York, USA

- Developed an **automated validation framework** to evaluate model performance for **clinical form extraction** and **edit check generation**, significantly reducing manual QA efforts
- Partnered with **cross-functional teams** to map **historical edit check metadata** to study protocols using Snowflake, **improving data traceability**
- Co-designed a POC for the next-gen **ML Prediction Service (Coder+)** for **intelligent medical coding** using **RAG** architecture with **vector embedded MedDRA** and **WHODrug** data stored in **Mongo Vector DB**

Data Science Intern Medidata Solutions

May 2024 - Aug 2024
New York, USA

- Automated clinical trial consent workflows using **NLP** and **Large Language Models (LLMs)**, cutting setup time by **45%** and costs by **\$6 million** annually
- Boosted document processing efficiency by **20%** with a custom algorithm for multilingual text extraction, eliminating manual language annotation
- Achieved **98% accuracy** in form-field mapping using **AWS Bedrock** and Claude AI, reducing annotation time by **87%** (15min to 2min)

Data Scientist Course5 Intelligence

Feb 2023 - Aug 2023
Bangalore, IN

- Boosted average revenue per customer by **15%** through market basket analysis (**Apriori**), enabling targeted upselling strategies for **50K+** users
- Reduced customer attrition risk by **20%** using a CLV-Churn ensemble ML model, effectively identifying high-value churn risks in the top **5%** segment, enhancing customer retention strategies

Data Scientist West Pharmaceutical Services

Jan 2020 - Feb 2023
Bangalore, IN

- Designed a **XceptionNet** deep learning model to classify 15+ defect categories in rubber stoppers, achieving **92% accuracy** (**37%** improvement over manual processes). Containerized with **Docker** and deployed on **Azure**, reducing defect resolution time by **62%**
- Trained particulate classification models (**Azure ML**) for structured/unstructured data, achieving **97% accuracy** and cutting analysis time by **80%** via **Azure Function Apps**
- Architected a document parsing system using **BERT** on **5 TB** of data, improving keyword-to-blog mapping by **23%**
- Built robust **ELT pipelines** on **Azure Data Factory** for 6+ data sources, reducing data preparation time by **40%**

Academic Projects

Live Flash Auctioning System | AWS (EC2, DynamoDB), WebSockets, Flask, Python

[Demo](#) [GitHub](#)

- Designed a real-time auction platform using **AWS EC2 Auto Scaling** and **DynamoDB**, handling **10K+ concurrent sessions** with **<200ms latency**, while deploying a **collaborative filtering model** (Python) to recommend auctions, boosting user engagement by **35%**
- Automated bid validation and user notifications via **AWS SQS/SES**, achieving **98% accuracy** in transaction processing while personalizing alerts using **interest-based segmentation**

Bengaluru Water Demand Forecasting | Python, Flask, SARIMA, REST APIs, DevOps

[Demo](#)

- Saved **2.1 billion liters** (25-30%) via **time series models** deployed as **Flask API** (MAPE < 8%)
- Integrated **Prophet** and **SARIMA** for reservoir management optimization

Technical Skills

Programming: Python, SQL, C#, Java, C/C++, PySpark, Spark

Machine Learning Frameworks: TensorFlow, PyTorch, Keras, Scikit-learn, Numpy, OpenCV

NLP/LLMs: BERT, RAG, LangChain, Transformers

Cloud/MLOps: AWS, Azure, Docker, MLFlow, Azure Data Factory, Firebase, Pinecone, MongoDB

Data Visualization & Other Tools: Power BI, Tableau, Flask, Django, Postman, Git, Snowflake, Excel

Methodologies: MLOps, Agile, Big Data, ETL/ELT, Data Warehousing, Explainable AI, Statistics

Education

New York University

Master of Science in Computer Science

New York, USA
Sep 2023 - May 2025

Visvesvaraya Technological University

Bachelor of Engineering in Computer Science

Bangalore, India
Aug 2016 - Aug 2020