

# Aayush Pradhan

+1 (848) 437-1919 | [aayushpradhan29@gmail.com](mailto:aayushpradhan29@gmail.com) | [linkedin.com/in/aayush-pradhan-b26a00147](https://www.linkedin.com/in/aayush-pradhan-b26a00147) | [github.com/aayush2912](https://github.com/aayush2912)

## EDUCATION

### Rutgers University, New Brunswick, New Jersey, USA

Master of Science in Data Science, GPA 3.55

08/2023 – 05/2025

### NMIMS University, Mumbai, India

Master of Business Administration in Technology Management, GPA 3.59

07/2021 – 05/2022

### Mukesh Patel School of Technology Management & Engineering, Mumbai, India

Bachelor of Technology in Computer Science (Information Technology), GPA 3.59

07/2017 – 05/2021

## TECHNICAL SKILLS

**Programming Languages:** Python, Java, C++, R, SQL, JavaScript, Go (familiar)

**Core CS & SWE Skills:** Data Structures & Algorithms, OOP, Design Patterns, System Design, Distributed Systems

**Frameworks & Libraries:** TensorFlow, PyTorch, Scikit-learn, Flask, .NET, Pandas, NumPy, Matplotlib, Seaborn, Plotly, APIs

**Databases & Storage:** MySQL, PostgreSQL, SQL Server, Oracle SQL, MongoDB, DynamoDB, BigQuery, Redis

**Cloud Platforms:** AWS (EC2, S3, RDS, Lambda, SNS, SQS, Elastic Beanstalk), GCP (BigQuery, GKE, Dataproc), Azure (basic)

**DevOps & Infrastructure:** Docker, Kubernetes, Terraform, Git, GitHub Actions, Jenkins, CI/CD Pipelines

**Data Engineering & ETL:** ETL Pipelines (Sqoop, Spark, Python), Data Modelling, Feature Engineering, Data Warehousing (dbt, BigQuery)

**Machine Learning & AI:** Ensemble Models (XGBoost, LightGBM, Random Forest), LSTM, NLP (BERT, T5, spaCy), SHAP, MLOps

**Tools & Other:** Tableau, Power BI, Excel (Advanced), VBA Macros, Jupyter Notebooks, R Shiny

## PROFESSIONAL EXPERIENCE

### Machine Learning Engineer – Cap Bon Consulting, New York

09/2025 – Present

- Processed financial data using Python/SQL, improving model stability by 25% through advanced data cleaning.
- Engineered a modular ML pipeline using Python, SQL, and TensorFlow/PyTorch, reducing experiment runtime by 40%.
- Optimized large-scale time-series forecasting models (LSTM, XGBoost, Random Forest) by implementing distributed training and hyperparameter tuning with Optuna/GridSearchCV, cutting RMSE by 18% and increasing recall of profitable trades by 22%.

### Software Engineer – ANG InfoTech LLC, New Jersey

07/2024 – 05/2025

- Optimized ETL pipelines to import/export data from Oracle/MySQL into HDFS using Sqoop, performing data deduplication and cleansing that improved data quality and analytics readiness by 30%.
- Applied statistical modelling and simulation analysis, enabling data-driven insights to enhance decision-making for client projects.
- Leveraged AWS (RDS, S3, Lambda) for scalable data processing and cloud-based deployment resulting in cutting downtime by 30%.
- Drafted technical design specs, prototypes, and automated test suites, accelerating feature validation and ensuring compliance.

### Associate Solution Analyst – Deloitte Touche Tohmatsu Limited, India

06/2022 – 08/2023

- Conducted risk assessments and data validation on financial and operational datasets using SQL and Python, improving audit accuracy and efficiency by 45% across business units and strengthening decision support for auditors.
- Optimized data processing workflows, reducing report generation time by 20% while increasing consistency of financial insights.
- Designed Tableau dashboards to visualize audit KPIs, reducing reporting time by 40% and improving visibility of compliance findings.

## KEY PROJECTS

### NewsVision – Personalized News Recommender System [\[Link\]](#)

- Developed a hybrid recommender using Sentence-BERT and SVD, achieving 91.2% Recall@5 on real-time news data.
- Launched a Flask API with Redis caching and AWS EC2 + NGINX, reducing API response time by 40% and supporting 1K+ daily requests.

### QueryCraft - Conversion of Natural Language Questions to Schema-Specific SQL Queries

- Designed and employed a natural language to SQL query generation system using transformer-based models (BART, T5), achieving a 35% improvement in query accuracy across multiple complexity levels through execution-based evaluation logic.
- Constructed a multi-head architecture integrating spaCy, all-MiniLM-L6-v2, and custom schema classifiers to enhance semantic understanding and column mapping, trained on 80K+ WikiSQL queries for accurate, schema-aware SQL generation at scale.

### CrAlpto - AI-Based Cryptocurrency Analytics with Multi-Model Forecasting [\[Link\]](#)

- Designed a real-time cryptocurrency analytics platform leveraging Binance API and PostgreSQL-based ETL pipelines; added LSTM, XGBoost, and hybrid models, improving prediction accuracy by 18% and reducing latency by 150ms across 500+ daily users.

### OncoRisk AI — High-Fidelity Breast Cancer Malignancy Prediction Pipeline with End-to-End Cloud Deployment [\[Link\]](#)

- Developed an end-to-end malignancy prediction pipeline leveraging SHAP-driven dimensionality reduction, polynomial feature interaction modelling, and LightGBM for high-stakes classification, achieving 0.9977 ROC AUC on FNA cytology datasets.

### MediRisk-AI – Clinical Risk Prediction Platform [\[Link\]](#)

- Engineered an end-to-end ML pipeline with LightGBM/CatBoost models, SHAP, and a modular Flask + Streamlit application, incorporating batch/single prediction APIs, decision curve analysis, and reproducible deployment on cloud infrastructure.

## PUBLICATIONS

Solutions to Vulnerabilities and Threats in Software Defined Networking (SDN) [\[ResearchGate\]](#) [\[ScienceDirect\]](#)

04/2020