

Pranav Dhawan

Washington, DC | dhawanpranav02@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

George Washington University Washington, DC
Master of Science, Data Science (GPA 3.74) Expected May 2026
Relevant Coursework: *Machine Learning, Deep Learning, NLP, Data Mining, Cloud Computing, Time Series Analysis*

Manipal University Jaipur Rajasthan, India
Bachelor of Technology, Computer Science and Engineering May 2024
Relevant Coursework: *Algorithms & Data Structures, Database Management Systems*

PROFESSIONAL EXPERIENCE

Lumina Datamatics
Machine Learning Engineer Feb 2024 – Aug 2024

- Fine-tuned Computer Vision models to detect and extract complex equations from 10,000+ unstructured documents, achieving high precision in technical expression recognition and eliminating the need for manual post-processing.
- Reduced manual correction overhead by 16% and cut inference latency by 0.3ms per page by replacing LayoutParser with a custom YOLO-based document layout pipeline.
- Reduced legal document search query time from minutes to under 5 seconds by architecting a hybrid RAG system with optimized vector embeddings and FAISS indexing, improving search relevance for counsel teams.

HCL Technologies, Noida, India
Machine Learning Intern Jul 2023 – Sept 2023

- Identified the top 5 drivers of workforce performance with 87% prediction accuracy by building Logistic Regression and Random Forest models trained on activity data from 500+ employees, enabling targeted HR interventions.
- Engineered 12+ data-driven KPIs from raw employee activity logs (screen time, app usage) using SQL and Python, surfacing patterns that differentiated high-performance clusters from the broader population.
- Designed interactive Tableau dashboards to visualize productivity trends, enabling management to make data-driven workforce planning decisions.

Ernst & Young, Gurgaon, India
Summer Intern May 2023 – Jul 2023

- Consolidated Sales & HR KPI reporting into 4 Power BI dashboards—covering revenue trends and attrition rates—cutting cross-functional reporting turnaround and supporting strategic planning for senior leadership.
- Achieved 100% reporting accuracy across monthly business reviews by automating ETL workflows for 5+ data sources using Alteryx, eliminating manual data cleaning and transformation.

PROJECT EXPERIENCE

[Multimodal Financial Time Series Forecasting](#) Present

- Designing a GNN-based multimodal architecture fusing LSTM (time series), FinBERT (NLP sentiment), and graph networks to model inter-sector dependencies for stock prediction using the FinMultiTime dataset with attention-based late fusion, targeting sub-2% MAPE on held-out test data.

[Edge-Based PII Detection & Censoring System](#) Sept – Dec 2025

- Achieved 98.1% F1-score and 97.9% recall across 54 PII entity types by architecting and benchmarking 4 transformer architectures, selecting DeBERTa as the production model and deploying it via ONNX for <100ms on-device inference.
- Built an end-to-end production demo in Streamlit with Tesseract OCR for PDF/image ingestion, a full model-training pipeline, and SHAP/LIME explainability analysis to surface entity-level confidence scores.

PROFESSIONAL SKILLS

Programming Languages: Python, R

Libraries and Frameworks: pandas, numpy, scikit-learn, matplotlib, seaborn, PyTorch, TensorFlow, HuggingFace

Data Management: SQL, MySQL, AWS, Google Cloud Platform

Visualization: Power BI, Tableau, Streamlit