PRANSHU GOYAL

(571) 618-4567 • NY • pg2592@nyu.edu • LinkedIn • GitHub • Portfolio

TECHNICAL SKILLS

Machine Learning and AI : PyTorch, Tensorflow, LangChain, Hugging Face, Unsloth, Pandas, Scikit-learn, QLoRA, CUDA
Big Data and Analytics : PySpark, Vector Database, Grafana, MongoDB, Kafka, PowerBI, Hadoop, Dask, Tableau, Airflow
Programming Languages and Frameworks : C, C++, Java, Python, R, SQL, HTML5, Angular, Flask, Streamlit, PostgreSQL
Cloud : AWS, SageMaker, Lambda, S3, Docker, Kubernetes, Jenkins, Azure, Google Cloud Platform, Slurm, EC2, Singularity

EDUCATION

M.S in Computer Science **September 2023 - May 2025**
New York University (Recipient of Merit-based scholarship)
Coursework: Machine Learning, Big Data, Deep Learning, Natural Language Processing, Artificial Intelligence, Data Science

B.E in Computer Engineering **July 2018 - June 2022**
Thapar Institute of Engineering and Technology, India
Relevant Coursework: Predictive Analytics using Statistics, Optimization Techniques, Probability and Statistics

WORK EXPERIENCE

Machine Learning Engineer Intern, LOCOMeX, New York **June 2024 - August 2024**

- Trained and deployed XGBoost, Random Forest models using AWS Lambda, ECS and Docker, optimizing performance and scalability, while **achieving 80% accuracy in predicting a company's social score** based on ITA-defined parameters.
- Built ETL data pipelines using AWS S3, API Gateway, and Lambda functions, **reducing data processing time by 35%** and improving data reliability, to streamline workflows for model deployment.
- Enhanced a semantic search model deployed on AWS by using **all-distilroberta-v1 embeddings**, improving supplier **recommendation accuracy by 40% and relevance by 30% in matching RFP requirements.**

Software Development Engineer, ION Trading, India **January 2022 - June 2023**

- Spearheaded the development of **50+ features** for server and frontend components in C++, Java and Angular within a microservice architecture, aimed at booking and processing trades, for banks like UBS, Scotia, Mitsubishi, Pierpont, TD bank.
- Engaged in the complete SDLC from requirement analysis to system design, coding, and quality assurance (unit testing and ATDD); ensuring on time and high quality project delivery for **15+ client requested paid enhancements.**
- Involved in the re-engineering of a legacy C++ server component, transitioning it to Java improving its performance.
- Contributed to the CI/CD pipeline by creating jobs with Jenkins and Docker, **achieving a 12% reduction** in release time.

Machine Learning Engineer Intern, PAYTM, India **July 2021 - August 2021**

- Boosted facial liveness detection **accuracy to 93.4%** by integrating MobileNetV2 and a self engineered CNN model.
- Created an OCR Flask application utilizing tesseract for accurate text extraction from user-uploaded IDs for online KYC.
- Implemented observability for services using NewRelic, Prometheus & Grafana, ensuring compliance with MLOps practices.

Data Analyst Intern, ONACTUATE, India **May 2021 - June 2021**

- Created over 25 interactive PowerBI data visualizations to enhance client decision-making through in-depth analysis.
- Leveraged Azure Data Factory for data integration, utilized Azure SQL Database for managing and querying large datasets, and deployed Power BI reports on Azure for scalable and secure data sharing.

PROJECTS

Sentiment Insight for Finance [code]

- Conducted a comparative study efficiently **fine-tuning LLaMA3-8b and Mistral-7b** LLMs using **LoRA, and 4-bit quantization**, boosting Mistral's accuracy from 50.3% to 84.3% and LLaMA's from 31.9% to 80.6%, with Mistral demonstrating superior capabilities in financial sentiment analysis.

Real-Time Forex Arbitrage and Price Prediction System [code]

- Architected a forex trading system with **Spark for data processing** and arbitrage calculations, and **deployed a multivariate bi-directional LSTM model** for accurate price prediction. Integrated **Flask for real-time serving and automated workflows with Apache Airflow**, enhancing strategy effectiveness and enabling real-time monitoring via a Looker Studio dashboard.

NYUAssistant [code]

- Developed an AI-driven query system for the NYU community, leveraging a RAG system with instructor embeddings, FAISS vector DB, and **FlashRank for enhanced data retrieval**. Engineered an **NLP pipeline using LangChain and Llama3-8b** LLM for precise answers, and integrated **Streamlit** to build an web application, boosting user engagement and functionality.

IndoFashion CLIP [code]

- **Fine-tuned CLIP** model on a specialized Indian fashion dataset to enhance multimodal machine learning capabilities, improving text-to-image retrieval accuracy and boosting performance by **87.3% for personalized fashion recommendations**.

PUBLICATIONS

- **Goyal, P.**, Agarwal, A., Singhal, K., Chinagundi, B., Singh Rana, P. (2023). Multilevel Deep Learning Model for Fabric Classification and Defect Detection.Lecture Notes in Networks and Systems, vol 653. **Springer**, Singapore.
- Singhal, K., Chawla, S., Agarwal, A., **Goyal, P**., Agarwal, A., Rana, P.S. (2023). Ensemble Model for Music Genre Classification. Advances in Intelligent Systems and Computing, vol 1442. **Springer**, Singapore.