

Shubham Pareek

(646) 418-8701 | sp4553@columbia.edu | linkedin.com/in/shubhampareek1211

EDUCATION

Columbia University

New York City, NY

Masters in Science, Data Science

May 2027

Relevant Course Work: Applied Machine Learning, Applied Deep Learning, Probability and Statistics, Exploratory Data Analysis/Visual.

Additional Role: Teaching Assistant - Cloud Computing.

SRM Institute of Science and Technology

Chennai, India

Bachelors of Technology in Computer Science and Engineering, CGPA: 8.83/10.0

Jul 2023

WORK EXPERIENCE

Weatherhead East Asian Institute, Columbia University

NY, NY

Data Analyst Research Intern

Sep 2025 - Present

- Devised and deployed production-grade web scraping pipelines using Scrapy collecting Financial Institute Data Records, enabling automated data collection from Big Stock Exchanges.
- Performed exploratory data analysis on financial transactions using pandas and seaborn, identifying key investment patterns and geopolitical correlations for research publications.
- Built NLP classification pipeline operating BERT and spaCy to extract and categorise investment relationships from company filings, attaining 89% F1-score on entity relationship classification.

Cognizant Technology Solutions

Hyderabad, India

Programmer Analyst

Jan 2024 - Jul 2025

- Resolved critical production incidents for premium Google Cloud customers within a 2-hour SLA, leveraging GKE troubleshooting methodologies, logging data, thus reducing customer downtime and decreasing mean time resolution by 45%.
- Delivered technical solutions for complex customer requirements in GKE INFRA and compute optimisation, DevOps.
- Architected and deployed scalable cloud infrastructure solutions for enterprise clients operating Terraform (HCL) and Google Cloud Platform APIs, accelerating deployment timelines by 65%.
- Mentored new project Joinee by creating onboarding technical guides and conducting technical sessions to accelerate team ramp-up and knowledge sharing.

PROJECTS

Energy Consumption Forecasting System

Jun 2025 - Oct 2025

- Developed a multi-model forecasting system for hourly energy prediction using 145k+ observations, implementing ARIMA, XGBoost and LSTM Architectures.
- Engineered 25+ temporal and statistical features including lag variables, rolling statistics, and external weather data, improving baseline LSTM model RMSE by 28%.
- Deployed production-ready REST API operating FastAPI and Streamlit dashboard with Docker, facilitating real-time forecasting with 95% confidence intervals.

Next-Gen Intrusion Detection System

Jun 2023 - Nov 2023

- Engineered an intrusion detection system leveraging a Kaggle dataset with supervised learning algorithms, scikit-learn, and pandas in Python, achieving 95%+ accuracy in detecting network anomalies and malicious traffic patterns.
- Engineered an interactive data visualisation dashboard using Streamlit/Plotly to display real-time network traffic analysis and threat detection metrics, enabling security teams to identify attack patterns.
- Built a conversational RAG interface operating Streamlit and Llama 2 LLM for interactive cybersecurity analysis, allowing users to query threat intelligence and receive contextual security insights through natural language.

SKILLS & CERTIFICATIONS

Programming Languages: Python, C, C++, SQL, R, Bash.

NLP and LLM: Hugging Face, BERT, spaCy, NLTK, LangChain, RAG, Fine-tuning, CNN, RNN, LSTM, Transformers.

Machine Learning: Scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM, Keras, ARIMA, LSTM, Pandas, NumPy, Matplotlib.

MLOps & Tools: Git, Docker, FastAPI, Flask, Streamlit, Jupyter, VS Code, R Studio.

Certification: Associate Cloud Engineer - Google Cloud | Professional Cloud DevOps Engineer - Google Cloud.