Shweta Smriti Tripathi

New York | +1(332)250-6048 | sst2166@columbia.edu | Linkedin | Github

EDUCATION

Columbia University in the city of New York, Aug 2025 – Dec 2026 | New York, NY
Master of Science in Data Science
- Coursework: Applied Machine Learning, Applied Deep Learning, Exploratory Data Analytics, Probability & Statistics.

Kalinga Institute of Industrial Technology (GPA 3.8/4.0), Jun 2018 – Jul 2022 | Odisha, India
Bachelor of Technology in Computer Science and Engineering
- Courseworks: Artificial Intelligence, ML, DBMS, Software Engineering, Data Structures & Algorithms.
- Published and presented a peer-reviewed paper on AI-based fault diagnosis in coal mine power systems at ICOM 2020, achieving a 70% increase in anomaly detection accuracy using expert systems and SCADA analytics.

SKILLS

Programming & Frameworks: Python, Java, SQL, React, REST APIs, PyTorch, TensorFlow.
Data Platforms & MLOps: Snowflake, Azure Data Factory, AWS, PySpark, MLflow, Kubernetes.
Machine Learning: LLMs, LSTMs, Transformers, GNNs, Reinforcement Learning, CNNs, Hugging Face.
Experimentation & Methods: A/B Testing, Causal Inference, Time-Series Forecasting, Statistical Modeling.
Analytics & Visualization: Computer Vision, Power BI, Tableau, Pandas, NumPy, Matplotlib, Seaborn, Plotly, ETL.

PROFESSIONAL EXPERIENCE

Providence Global Center (Client: Advent Health), Jul 2024 – Jul 2025 | Hyderabad, India
Sr. Business Intelligence Data Engineer
- Spearheaded an 8-member analytics team to deliver **20+ executive dashboards** across 19 HITRUST domains, enabling **C-suite** invest/divest decisions and strengthening regulatory oversight.
- Developed and evaluated ML pipelines (XGBoost, LightGBM, CNN/LSTM/Transformers) over 10M+ logs, **improving anomaly-detection** accuracy to 87% and **reducing incident exposure** by 75%.
- Optimized 50+ **SQL/Snowflake pipelines** (5M+ rows/day), **cutting runtime** by 65% and **raising reporting efficiency** by 40% with near real-time validation.
- Integrated **PySpark + MLflow** into ETL, **decreasing deployment errors** by 40% and shortening rollout from 3 days to less than 12 hours.

DELL Technologies, *Business Intelligence Analyst* Jun 2022 – Jul 2024 | Bangalore, India
- Produced 30+ interactive dashboards (**Power BI/Tableau**) across 4M+ **Salesforce/Teradata/D&B records**, surfacing churn patterns and compliance risks that informed product investment.
- Conducted EDA on large **SQL/Teradata/SSIS** workflows, raising **data quality** by 60% and eliminating recurring errors.
- Automated preprocessing with **Python/PySpark + RPA**, handling 10M+ transactions, reducing prep time by 75% and **supporting real-time forecasting**.
- Led 10+ **cross-functional initiatives**, **prioritized** critical-path risks, and recommended fixes that **shortened delivery** by 30%; earned **Dell's Inspire Award** for streamlining reporting (+50% efficiency).

DELL Technologies, *Data Analyst Intern* May 2021 – Jul 2021 | Bangalore, India
- Constructed a SQL/Teradata **data-observability & anomaly-alerting** layer across 30+ pipelines (schema/row-count/latency/KPI drift), cutting detection time by over 90 percent and preventing bad refreshes from reaching executive dashboards.

HighRadius, *Machine Learning Intern* Jan 2021 – Mar 2021 | Hyderabad, India
- Designed a real-time order management dashboard (ReactJS, MySQL, and Java Servlets, integrated ML models and LLMs) to predict client payment dates, improving **forecast accuracy by 35%** and reducing manual invoicing by 25%.

PROJECTS

Brain Tumor Detection Using YOLOv10, (*OpenCV, Medical Imaging, CNNs, PyTorch*)
- Implemented a CNN-based MRI detection workflow, achieving 82% accuracy and **cutting review time by 40%**.

Time Series Forecasting for Energy Consumption, (*LSTM/Transformers, Scikit-learn, Python*)
- Built LSTM/Transformer forecasts on **5M+ records**, achieving **85% accuracy** and **reducing lead time by 30%** for planning.

Semantic Book Recommender System, (*LLMs, NLP, SpaCy, Scikit-learn, BERT,Hugging Face*)
- Trained a text-based recommender with greater than **80% accuracy**, **reducing manual search by 70%** and **boosting discovery by 70%**.

Cloud-Based IoT & Predictive Analytics for Smart Irrigation, (*Azure, Reinforcement Learning, Python, IoT Sensors*)
- Deployed an RL-driven IoT irrigation system, **lowering water use by 30%** and **increasing yields by 20%**.