

# Nyasha Ojha

76172 13258 | [nyashao2910@gmail.com](mailto:nyashao2910@gmail.com) | [linkedin.com/in/nyasha-ojha](https://www.linkedin.com/in/nyasha-ojha) | [github.com/Nyasha-Ojha](https://github.com/Nyasha-Ojha)

## Professional Summary

Computer Science undergrad building data systems with an engineer's mindset rather than a tool-collector's approach. Experienced in designing data pipelines, structuring raw datasets, and applying ML and deep learning models on well-prepared data. Strong in Python and SQL, with a clear understanding that reliable data engineering is the foundation for scalable analytics and ML-driven products.

## Technical Skills

- **Programming Languages:** Python, Java, SQL
- **Data Engineering & Processing:** ETL Pipelines, Apache Spark, Delta Lake, Incremental Data Processing, Data Modelling
- **Database & Storage:** PostgreSQL, MySQL, MongoDB, Spark SQL, Delta Tables
- **Machine Learning & Data Analysis:** Scikit-learn, Tensorflow, Pandas, Numpy
- **Cloud & Data Platforms:** AWS (IAM, EC2, S3, Lambda, Glue, Glue Crawlers, Athena), Databricks
- **DevOps & Tools:** GitHub, Github Actions (CI/CD)
- **Languages:** English, Hindi, Telugu, Bengali

## Open Source Contribution

### CaseLawAI – Open Source LegalTech Project

Jr Machine Learning Engineer | Omdena (USA Chapter) | *May 2024 – July 2024*

**Project Goal:** Streamlined legal research to detect patterns in employment discrimination cases using deep learning.

- Designed and implemented RNN and ANN modules for sequential contextual analysis and classification of discrimination types from legal case files, as part of a global 4-member DL subteam.
- Built an end-to-end deep learning pipeline: extracted case images, applied CNNs on images, performed OCR-based text extraction, then used RNN-LSTM for context retention and a final ANN classifier, improving model accuracy by 18% and
- Collaborated with 57 global members to optimize preprocessing with regex, benchmark algorithms, and integrate ML pipelines, hence reducing manual case review time by ~30%. leveraging Python, Scikit-learn, Pandas, NLTK, and Deep Learning frameworks.

## Projects

### Enterprise FMCG Analytics Platform

Individual Project | FMCG, ETL, Data Engineering

- Developed an end-to-end ETL pipeline from parent (Atlon) and acquired child (Sports Bar) companies, using Python, Spark SQL, Delta Lake and Medallion architecture to process 93K+ fact records across bronze, silver, and gold layers.
- Implemented historical backfill (5 months) and incremental updates via AWS S3 integration and Genie AI orchestration, handling data quality issues such as duplicates, nulls, and inconsistencies.
- Engineered dimension/fact tables, date dimension scripting, and denormalized views for BI dashboards, enabling analytics for revenue forecasting and top-product insights in a simulated enterprise acquisition scenario, hence building a scalable data platform.

### Urban Planning Using Satellite Imagery

Team Lead- Academic Project | Geospatial AI, Data Validation | *Sep 2024 - May 2025*

- Led a 10-member engineering team, to build a sustainable urbanisation plan, overseeing 4 models (population forecasting, land use/land cover, water resource allocation, road & traffic network analysis) and ensured data quality, regression checks, and output consistency across 556 satellite datasets and 3,000+ annual samples.
- Focused on population forecasting model, creating structured test plans, validating numerical outputs, and improving prediction reliability by ~20% through iterative modeling and data verification.
- Managed cross-functional collaboration via GitHub and Notion, tracked defects, monitored inconsistencies, and coordinated multiple model iterations to ensure consistent and accurate analytics outputs.

## Education

### Bachelor of Technology - Computer Science and Engineering

Vellore Institute of Technology

*Expected May 2026*

CGPA: 8.2

## Certifications

- **Fundamentals of Data Engineering | Coursera:** Mastered ETL pipelines, data warehousing, and orchestration
- **IBM GenAI | IBM:** Built generative AI applications using modern frameworks
- **Cloud Computing | NPTEL:** Implemented scalable cloud architectures and services

## Awards and Achievements

- **Technical Lead:** Led a 27-member team to successfully deliver 2 projects, increasing community engagement by 40%
- **Debate Team Core Member:** Represented school in inter- and intra-school competitive events

**Hobbies & Interests:** Certified drummer (3 years, Rock School of London), state-level youth boxing participant, and an avid follower of Formula 1 racing