Neeraj Jawahirani

+1(850)300-3851 | neerajjawahirani@gmail.com | linkedin.com/Neeraj-jawahirani | neerajj.netlify.app

EDUCATION

Florida State University — Tallahassee, FL
Masters of Science in Data Science - Computer Science, GPA: 3.9 — August 2023 – May 2025

University of Pune — Pune, India
Bachelors of Engineering in Information Technology, GPA: 9.1/10 — July 2018 – May 2022

TECHNICAL SKILLS

Programming: C, C++, Java, Advanced Java, Python (NumPy, Pandas, Matplotlib, Plotly, Seaborn), R
Machine Learning & AI: Scikit-learn, XGBoost, TensorFlow, PyTorch, NLTK, OpenCV, LangChain, Natural Language Processing (NLP), Generative AI, Large Language Models (LLMs)
Databases: SQL, SQLite, NoSQL, MongoDB, PostgreSQL, Amazon RDS, Hadoop
Tools & Cloud: Git, Microsoft Office, Power BI, Tableau, AWS (S3, Redshift), Snowflake, Microsoft Azure, Bitbucket

EXPERIENCE

COAPS - Florida State University (FSU) — Tallahassee, FL
Web Developer / Analyst — September 2024 – April 2025

- Designed and maintained ETL pipelines using SQL and Python to process data from 30+ marine research vessels, automating ingestion schedules and reducing manual handling. Conducted data profiling to flag missing values, detect outliers, and identify sensor anomalies in voyage logs.
- Refactored legacy PHP/SQL backend to modular architecture, reducing page load times by 40% and enabling cleaner database access for researchers. Set up GitHub Actions for CI/CD, cutting deployment time from hours to minutes and reducing build errors.
- Automated recurring data validation and reporting tasks using Python scripts, cutting manual effort by 25%. Introduced rule-based quality checks (e.g., range validation, null handling), reducing reporting inconsistencies and ensuring cleaner datasets for downstream analysis.

IBM- Techademy — Pune, India
Data Engineer Trainee — March 2022 – October 2022

- Completed IBM-Techademy Data Engineering program with 3 applied projects involving ETL design, SQL optimization, and data modeling. Scripted workflows using Python and SQL to process 500K+ raw records, applying validation logic to preserve data accuracy.
- Built a batch processing pipeline using Hadoop and PySpark to clean, aggregate, and load 200K+ records into AWS S3. Outlined the workflow to handle large datasets in chunks, minimizing memory usage and preparing data for downstream analysis on a cloud setup managing over 100GB.
- Used SQL and Python to clean, reshape, and optimize large datasets for analytics. Improved query performance through indexing, optimized joins and filters, and table restructuring, reducing execution time by 40% on MySQL under simulated high-load conditions.

Geniobits Private LTD — Pune, India
Web Data Analyst & Team Lead Intern — May 2021 – December 2021

- Led a data annotation team of 100+ interns to preprocess and label over 500K records using PySpark and Python scripts. Integrated Apache Airflow to schedule and monitor ETL jobs, improving data readiness speed by 40% for analytics in an edtech platform.
- Implemented and launched 5 dashboards on AWS using Power BI and Python, allowing users to filter and monitor over 100K learner records in real time. Reduced stakeholder reporting time by 50% and improved visibility into key learning performance metrics.
- Developed and served two ML models with PyTorch for classification and regression tasks, integrated via REST APIs into lightweight deployment pipelines. Raised model accuracy from 78% to 89% and cut deployment time by 60% using Docker containers.

PROJECTS

AI-Powered News Summarizer — January 2025 – February 2025

- Engineered a real-time news summarization system using LangChain and GPT-3.5 to process articles from 20+ online sources. Tuned prompts, extracted keywords, and cleaned inputs to rank and summarize the most relevant articles using a scoring-based method.
- Created a responsive Streamlit web app comparing extractive (TextRank) and generative (LLM) summaries. Achieved a 30% improvement in summary consistency based on user testing and delivered concise, real-time takeaways for faster content consumption.

Data-Driven Fitness Tracker with Predictive Modeling — November 2024 – December 2024

- Developed a fitness tracking application using Python, Pandas, and Scikit-learn to analyze 100+ user profiles. Trained a decision-tree model to recommend diet and workout plans, improving plan matching accuracy by 30% based on simulated user data.
- Configured a modular system with a mobile-friendly frontend, integrating Matplotlib for real-time metric visualizations. Achieved a 40% increase in model accuracy through validation and ensured high responsiveness across devices during testing.

Twitter Sentiment Analysis — January 2024 – February 2024

- Processed and analyzed over 1 million Twitter posts using a Python-based NLP pipeline with a Naive Bayes classifier. Assembled detailed visual dashboards using Matplotlib and Seaborn to track sentiment trends by hashtag and topic, reducing processing latency by 25%.
- Boosted classification accuracy from 70% to 84% through feature scaling, engineered variables, and hyperparameter tuning. Enabled near real-time sentiment updates to monitor public opinion shifts across multiple user segments.

Body Performance Analysis — October 2023 – December 2023

- Performed multiple regression analysis on 1,000+ records from the Korean Sports Promotion Foundation to study relationships between age, physical activity, and grip strength. Identified significant correlations that guided athletic training recommendations.
- Applied data visualization and linear modeling to isolate key physiological predictors of grip strength. Findings were used to support tailored training plans for 200+ athletes, improving exercise targeting by age group, fitness level, and activity intensity.

Distributed Storage System using IPFS — August 2021 – May 2022

- Set up and integrated a decentralized file storage system using Django and IPFS, managing over 2TB of encrypted data across 100+ nodes. Maintained high availability and enabled secure access for 500+ global users through content-addressable storage.
- Published a research paper titled "Distributed Storage System Using IPFS" in the International Journal of Research and Analytical Reviews (IJRAR), contributing valuable insights to the field (Paper ID: IJRAR1CNP041).

LEADERSHIP

Information Technology Engineering Student Association | Management Head — January 2021 - January 2022
Toastmaster DYPCOE | Founder and President — March 2020 - January 2022