

SUPRIYA SINGH

Data Analyst

supriyasingh8168@gmail.com | 8828518168 | Mumbai, India | [LinkedIn](#) | [GitHub](#)

Data professional with a year of on-site experience evaluating and categorizing high daily volumes at **95%+ accuracy**, auditing team output, and reporting daily progress. Builds Tableau dashboards and Excel reports, and works in Python and SQL. Self-driven project builder with live-deployed ML and RAG projects.

PROFESSIONAL EXPERIENCE

HERE Technologies

Feb 2025 – Feb 2026

Trainee | Full-time | Navi Mumbai, On-site

- Evaluated high volumes of ML-filtered news articles daily on the Neptune platform to identify Location Data Issues (LDI), with **95%+ categorization accuracy** (QIL), feeding HERE's map correction pipeline.
- Progressed into a **task-assignment and quality-review role**: allocated daily workloads and audited team categorizations for consistency.
- Reported daily progress in scrum standups and tracked tasks in Jira (assigned, completed, pending), flagging bottlenecks early.
- Used AI tools (ChatGPT, Claude, Gemini, Copilot) at work to speed up analysis, formulas, and reporting.

Self-Employed

Feb 2026 – Present

Freelance WordPress Developer | Remote

- Delivered **4+ client websites end-to-end**: custom themes, plugin integration, and deployment, including bilingual builds.
- Achieved **100/100 on-page SEO scores** across client sites alongside core speed optimizations.

KEY PROJECTS

Job Fraud Detection System + Live API

Python | Scikit-learn | Random Forest | TF-IDF | FastAPI | Docker

- Achieved **98.07% accuracy** detecting fraudulent job postings on 18,000+ records using Random Forest + TF-IDF NLP; applied SMOTE for class imbalance.
- Deployed as a live, containerized REST API (FastAPI + Docker) with real-time fraud probability scoring.

RAG PDF Chatbot: AI Document Query Assistant

LangChain | Groq LLaMA3 | TF-IDF | Streamlit

- Built an end-to-end RAG pipeline for context-aware Q&A on any PDF; deployed live on Streamlit Cloud with multi-turn conversation.

House Price Prediction

Python | Machine Learning

- Built a regression model with focus on data quality: handling missing values, outlier detection, and feature selection.

TECHNICAL SKILLS

Languages	Python (Proficient), SQL (Proficient), JavaScript (Decent)
Data & Quality	Data categorization/annotation, quality assurance (QIL), Pandas, Excel, Tableau, Jira, dashboards and reporting
ML / AI	Scikit-learn, NLP/TF-IDF, LangChain, RAG, Random Forest
AI Tools	ChatGPT, Claude, Gemini, Microsoft Copilot
Cloud & Tools	AWS (S3, Lambda), Microsoft Azure (AZ-900), Docker, Git/GitHub, MySQL, WordPress
Frameworks	FastAPI, Flask, Django, ReactJS, Streamlit

EDUCATION & CERTIFICATIONS

B.Sc. in Information Technology

University of Mumbai, K.B. College of Arts & Commerce
Graduated April 2024 | CGPA: 8.22

Certifications

- AWS Certified Solutions Architect – Associate
- Microsoft Azure AZ-900 (Fundamentals)
- Python Full Stack – PySpider
- UiPath: Explore Automation Development (Diploma of Completion, Oct 2026)