Sunil Sharma +91-9039805553 | 9.sunilsharma@gmail.com
mrxsierra.github.io | linkedin.com/in/sunilsharma97 | github.com/mrxsierra | kaggle.com/mrxsierra

SUMMARY

Results-driven ***Python Developer*** and ***Data Scientist*** with 1+ years of hands-on experience.

- Expertise in ***machine learning***, ***data analysis***, ***automation***, and end-to-end project delivery.
- Adept at leveraging advanced skills in ***SQL***, ***database design***, and ***backend development***.
- Demonstrated success in national-level hackathons and open-source projects.
- Strong focus on code quality, compliance, and real-world impact.
- Excellent communication, documentation, and problem-solving abilities.

Comfortable in both independent and collaborative environments.

SKILLS SUMMARY

- ***Programming Languages:*** Python, SQL, JavaScript, HTML, CSS
- ***Libraries/Frameworks:*** scikit-learn, XGBoost, LightGBM, Pandas, NumPy, Selenium, BeautifulSoup, Openpyxl, Pytest, Flask, Django, FastAPI
- ***Data Analysis & Visualization:*** pandas, numpy, matplotlib, seaborn, Plotly.js
- ***Machine Learning & Data Science:*** scikit-learn, XGBoost, LightGBM, SHAP
- ***Automation & Scripting:*** Python, ETL, Selenium, pytest
- ***Web Scraping & Data Extraction:*** Selenium, BeautifulSoup
- ***Database Design & SQL:*** SQLite, MySQL, PostgreSQL
- ***Tools:*** Jupyter Notebook, Git, GitHub, Docker, VSCode, Markdown
- ***Cloud Platforms:*** AWS, Azure, GCP (Basics)
- ***Operating Systems:*** Windows, Linux
- ***Other:*** Technical Documentation, CI/CD, Agile, Kubernetes (Basics), AI & Agentic Tools (Mistralai, Gemini, Langchain, Langgraph, Huggingface - PoC & experimental projects)

WORK EXPERIENCE

Hackathon Finalist – GSTN Predictive Binary Classification Challenge *(Aug 2024 – Oct 2024)*
Goods and Services Tax Network (Part-time) | Delhi, India
Role: Machine Learning Engineer / Data Scientist

- Led ***predictive model development*** for GST analytics in a national-level competition organized by the Government of India.
- Developed a robust and interpretable ***binary classification solution*** using a large, anonymized dataset (900,000+ records, 21 features).
- Achieved a Top 7% ranking (Top 17 out of 220+ teams), demonstrating expertise in handling severe class imbalance and missing data.
- Designed and implemented a complete ***machine learning pipeline***, including EDA, feature engineering, model selection, and hyperparameter tuning.
- Evaluated and reported on ***model performance***, achieving >97% accuracy and strong F1/MCC scores.
- Automated ***model training*** and ensured reproducibility while adhering to strict compliance standards for anonymization and data handling.
- Presented the solution, technical approach, and key insights to a panel of senior data scientists and domain experts.
- Awarded a finalist certificate and prize in recognition of technical excellence and solution transparency.
- Leveraged this experience to enhance skills in advanced ***ML modeling***, ***data preprocessing***, ***model explainability*** (SHAP), and technical communication.
- **Project Repository**: github.com/mrxsierra/gstn_dsp_pbc
- **Portfolio Link**: mrxsierra.github.io/projects/gstn-pbc/

PROJECTS

GSTN Hackathon: Predictive Binary Classification *(Aug 2024 – Oct 2024)*

Role: Machine Learning Engineer / Data Scientist

- Developed a ***high-performance***, ***interpretable ML pipeline*** for ***binary classification*** on large, imbalanced, anonymized datasets.
- Achieved >97% accuracy and strong F1/MCC using advanced modeling, feature engineering, and hyperparameter tuning.
- Ensured strict compliance, reproducibility, and transparent reporting; recognized as a top finalist.
- **Project Link**: mrxsierra.github.io/projects/gstn-pbc/#gstn-hackathon-predictive-binary-classification-project

Naukri Webscraper *(2024)*

Role: Python Developer / Automation Engineer

- Built a ***Selenium-powered tool*** to automate job search and data extraction from Naukri.com.
- Implemented skill-based filtering, CSV export, and robust automated testing with pytest.
- **Project Link**: mrxsierra.github.io/projects/naukri-webscraper/

Test Management Site *(2024)*

Role: Frontend Developer / Web Application Prototyper

- Designed and developed a vanilla ***JavaScript*** web application for creating, managing, and taking tests with result tracking.
- Integrated localStorage, third-party libraries, and CI/CD deployment via GitHub Actions and Pages.
- **Project Link**: mrxsierra.github.io/projects/test-site/

Examination Management System Database **(***Apr 2024 – Mar 2025)*

Role: Backend Developer / Database Engineer

- Architected and deployed a ***modular***, ***production-grade database system*** supporting multiple RDBMS (SQLite, MySQL, PostgreSQL).
- Automated schema management and data operations using ***Python***.
- Project Link: mrxsierra.github.io/projects/ems-db/

Paraxcel (*Feb 2025 – Mar 2025)*

Role: Python Application Developer

- Engineered a toolkit for advanced ***Excel data extraction*** and transformation with a user-friendly GUI (Pandas, Openpyxl, Tkinter).
- **Project Link**: mrxsierra.github.io/projects/paraxcel/

S3 Faker *(2024)*

Role: Data Engineer / Python Developer

- Built a high-performance ***fake data generator*** with ***AWS S3*** integration, supporting CSV, JSON, and Parquet formats.
- **Project Link**: mrxsierra.github.io/projects/s3-faker/

EDUCATION

Bachelor of Science (B.Sc.), Computer Science *(2015 – 2018)*

PMB Gujarati Science College, DAVV | Indore

12th Higher Secondary (PCM) – [59%] *(2014 – 2015)*

Sarafa Vidhya Niketan | Indore

10th High School – [73%] *(2012 – 2013)*

Sarafa Vidhya Niketan | Indore

CERTIFICATIONS

GSTN Hackathon – Top Finalist (Predictive ML) - *Goods and Services Tax Network*

- ***National-level hackathon:*** for ***AI/ML***-based GST analytics (900K+ records).
- **Certificate**: mrxsierra.github.io/cert/GSTN_Team_137.jpg

CS50P – Introduction to Programming with Python *HarvardX*

- ***Topics***: functions, conditionals, loops, debugging, exception handling, unit testing, ***OOP***.
- ***Certificate***: https://certificates.cs50.io/7683d52e-e848-45b6-b667-18dba0dac744.pdf?size=letter

CS50x – Introduction to Computer Science *HarvardX*

- **Topics**: ***algorithms***, ***data structures***, abstraction, encapsulation, resource management, security. Learned C, ***Python***, ***SQL***, ***and JavaScript***.
- Certificate: https://certificates.cs50.io/bf9109ae-3bed-49ce-8d31-10c361360b81.pdf?size=letter

CS50 SQL – Introduction to Databases with SQL *HarvardX*

- ***Topics***: ***relational databases***, ***SQL querying***, normalization, indexing, database integration with ***Python*** and ***Java***. Covered ***PostgreSQL*** and ***MySQL***.
- ***Certificate***: https://certificates.cs50.io/b8f3c962-95c1-4530-a6e3-af4125df66c1.pdf?size=letter

AI/ML for Geodata Analysis *IIRS/ISRO*

- ***Topics***: ***GIS processing***, ***image classification***, ***deep learning techniques***, and ***cloud computing*** using Google Earth Engine.
- ***Certificate***: mrxsierra.github.io/cert/IIRS.jpg

ADDITIONAL INFORMATION

Technical Blogger: Data science, machine learning, and database design.

- **Blog Portfolio**: mrxsierra.github.io/blog/
- ***Notable Posts***:
  - *Navigating the Nuances: A Developer's Guide to SQL Dialects (SQLite, MySQL, PostgreSQL) (mrxsierra.github.io/blog/2025/05/07/navigating-the-nuances-a-developers-guide-to-sql-dialects-sqlite-mysql-postgresql/)*
  - *Beyond the Schema: A Practical Guide to Querying and Interacting with SQLite, MySQL, & PostgreSQL (mrxsierra.github.io/blog/2025/05/07/beyond-the-schema-a-practical-guide-to-querying-and-interacting-with-sqlite-mysql--postgresql/)*
- Strong documentation and communication skills.
- Passionate about continuous learning and ***automation***.