


Sarthak Shandilya

Data Scientist | Python | SQL | Applied Machine Learning | Analytics & Visualization

📍 Sagar, MP 📞 [+91 9893846001](tel:+919893846001) ✉ sarthakshandilya9@gmail.com 🔗 <https://sarthak-portfolio.streamlit.app>

🔄 [sarthxk20](#)  [sarthxk20](#)

Summary

- Data scientist with strong foundations in Python, SQL, and data analysis, experienced in working with real-world datasets to derive insights through statistical analysis, predictive modeling, and machine learning.
- Hands-on experience across the data science lifecycle, including feature engineering, supervised learning, ensemble methods, validation workflows, and metric-driven model evaluation with a focus on interpretability and reliability.

Projects

SpaceX Launch Dashboard

- Engineered an end-to-end data pipeline integrating the SpaceX API with external datasets, enabling automated data ingestion, cleaning, transformation, and feature engineering.
- Built, evaluated, and deployed predictive models and interactive dashboards, improving model performance by 12% (F1-score: 0.89) through feature optimization and delivering real-time insights via Streamlit and Plotly visualizations.

🔗 <https://spacex-launch-dashboard.streamlit.app>

Resume Screening System (ATS Simulation)

- Built an end-to-end NLP-based resume screening pipeline simulating ATS logic, using TF-IDF and BERT embeddings to evaluate resume–job relevance and enable explainable skill gap analysis.
- Deployed the system as a Streamlit application with an emphasis on reproducible modeling, consistent inference results, and transparent evaluation of resume–job matching outputs.

🔗 <https://resume-screening-ai.streamlit.app>

Skills

Technical Skills

- Python, SQL, Pandas, NumPy, Scikit-Learn
- TensorFlow, PyTorch
- Matplotlib, Seaborn, Plotly, Streamlit
- Git & GitHub, REST APIs
- MongoDB, Excel, Power BI, Tableau, Jupyter Notebook, VS Code
- Hadoop, Apache Spark
- Azure Machine Learning (Azure ML)

Data Science, ML & GenAI

- Data Cleaning, EDA, Statistics & Statistical Analysis, Feature Engineering, Data Pipelines (ETL)
- Classification, Regression, Clustering, PCA, Random Forests, XGBoost
- Model Evaluation (Precision, Recall, F1, ROC-AUC), Model Deployment
- Deep Learning: ANN, CNN, RNN, LSTM
- NLP: Tokenization, Embeddings, Text Classification
- GenAI & LLMs: Transformers (BERT, GPT), Large Language Models, Prompt Engineering, Text Generation

Certifications

Applied Data Science Lab WorldQuant University

- Applied machine learning techniques to real-world datasets with a focus on model evaluation, validation, and data-driven decision making using iterative experimentation, including managing unstructured data with MongoDB.

Data Science Professional Certificate IBM

- Completed an industry-aligned data science curriculum with hands-on labs and capstone projects covering Python, SQL, statistical analysis, machine learning, APIs, data visualization, and end-to-end model development.

Machine Learning & AI with Python HarvardX (edX)

- Implemented and evaluated supervised learning models in Python, with a strong focus on ensemble methods (Bagging, Random Forests, Gradient Boosting, AdaBoost), bias–variance tradeoffs, and metric-driven model selection.

Education

Bachelor of Arts — Sociology & History | CGPA: 8.15

Dr. Harisingh Gour Central University (2020–2023)