

# Sarthak Shandilya

Data Scientist | Python | SQL | Machine Learning | Data Analytics & Visualization | IBM Certified

📍 Sagar, MP 📞 +91 9893846001 📧 sarthakshandilya9@gmail.com 🔗 <https://sarthak-portfolio.streamlit.app>

[sarthxk20](#)

[sarthxk20](#)

## Summary

- Data Scientist with strong foundations in Python, SQL, and Machine Learning, experienced in building end-to-end data, ML, and NLP systems with an emphasis on scalable and explainable solutions.
- Designed and implemented applied ML and NLP pipelines, including an ATS-style resume screening system using TF-IDF and BERT embeddings and a SpaceX launch analysis workflow built on feature engineering and supervised learning.
- Hands-on experience with statistical modeling, deep learning architectures (ANN, CNN, RNN, LSTM), and transformer models (BERT, GPT), with a focus on model evaluation, reproducibility, and production deployment.

## Projects

### SpaceX Launch Dashboard

- Engineered an automated end-to-end data pipeline by integrating the SpaceX API with external datasets, enabling continuous data ingestion, cleaning, transformation, and feature engineering.
- Built and evaluated a Logistic Regression model for mission success prediction, improving model performance by 12% and achieving an F1-score of 0.89 through feature optimization.
- Developed and deployed an interactive Streamlit dashboard with Plotly visualizations to analyze mission success trends, rocket performance, and launchpad efficiency in real time.

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

### Resume Screening System (ATS Simulation)

- Built an end-to-end NLP-based resume screening pipeline simulating core ATS logic by evaluating candidate resumes against job descriptions using statistical and semantic text representations.
- Applied TF-IDF vectorization and BERT-based sentence embeddings to capture both keyword relevance and contextual similarity, enabling robust resume-job matching and explainable skill gap analysis.
- Deployed the system as a cloud-ready Streamlit application, emphasizing inference stability, dependency management, and reproducible machine learning workflows.

🔗 <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
- Excel, Power BI, Tableau, Jupyter Notebook, VS Code
- Hadoop, Apache Spark
- Azure Machine Learning (Azure ML)

### Data Science, ML & GenAI

- Data Cleaning, EDA, Statistical Analysis, Feature Engineering, Data Pipelines (ETL)
- Classification, Regression, Clustering, PCA, Random Forest, 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

## Education

### Dr. Harisingh Gour Central University

Sociology & History  
8.15

Bachelor of Arts  
2020 – 2023

- Built a strong foundation in research, analytical thinking, and data interpretation through social science coursework.
- Developed quantitative reasoning and hypothesis-driven analysis skills for data-focused problem solving.
- Transitioned into Data Science through the **IBM Data Science Professional Certificate**, gaining industry-aligned training in Python, SQL, data analysis, and machine learning.

## Certifications

### IBM Data Science Professional Certificate

IBM

October 2025

- 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.

🔗 <https://www.coursera.org/account/accomplishments/professional-cert/4UGBZJCVM3HZ>