Suraj Kumar

Data Science and AI enthusiast with strong expertise in Machine Learning, Deep Learning, and Generative AI. Skilled in building end-to-end ML pipelines, developing interactive web applications, and deploying scalable AI solutions. Experienced in feature engineering, data preprocessing, and model interpretability, with hands-on projects in predictive analytics and recommendation systems.

PROJECTS

Breast Cancer Diagnosis Predictor Mar 2025 - Present

Python, Streamlit, Scikit-learn

- Developed an interactive Streamlit web app to predict whether a breast mass is benign or malignant using clinical cell measurements.
- Processed and modeled data from the Breast Cancer Wisconsin (Diagnostic) Dataset, implementing a complete ML workflow including preprocessing, model training, and evaluation.
- Enhanced interpretability with dynamic radar chart visualizations for clinical features.
- Serialized the trained ML model and scaler for seamless backend integration with the UI.
- Styled the application with custom CSS for a clean and intuitive user experience.
- Deployed the application live on Streamlit Cloud, enabling real-time access and interaction.

Laptop Recommendation System Mar 2025 - Present

Python, Streamlit, Scikit-learn

- Collected and curated 1000+ laptop specifications, engineering features like processor, RAM, GPU, storage, display, and OS for model training.
- Implemented a Content-Based Recommendation System using TF-IDF vectorization and Cosine Similarity to suggest top-3 laptops similar to user input.
- Developed an interactive Streamlit web app with input forms, prediction page, file upload, resource download, and contact modules.
- Optimized the preprocessing pipeline, cached the similarity matrix, and improved inference time by 40%.
- Deployed the system with a modular code structure and dataset support for easy scalability and maintainability.

DocuMind: AI Study Partner Mar 2025 - Present

Python, Streamlit, OpenAI GPT-4o, LangChain

- Developed an interactive Streamlit web app to predict whether a breast mass is benign or malignant using clinical cell measurements.
- Processed and modeled data from the Breast Cancer Wisconsin (Diagnostic) Dataset, implementing a complete ML workflow including preprocessing, model training, and evaluation.
- Enhanced interpretability with dynamic radar chart visualizations for clinical features.
- Serialized the trained ML model and scaler for seamless backend integration with the UI.
- Styled the application with custom CSS for a clean and intuitive user experience.
- Deployed the application live on Streamlit Cloud, enabling real-time access and interaction.

CONTACT

- +91-7616477447
- SurajkSharma7@outlook.com
- linkedin.com/in/surajksharma7

SKILLS

- Programming - Python, C++, SQL,
- Machine Learning
- MlOps
- Feature Engineering
- Deep Learning
- Generative AI
- Data Preprocessing
- Feature Selection
- Data Visualization
- Data Analysis
- Natural Language Processing

EDUCATION

University of Petroleum & Energy Studies
Mca- Aiml
July 2024 - July 2026

Bhagwan Mahavir University
B.Sc-Data Science
July 2021 - July 2024

CERTIFICATIONS

- Data Science Menorship Program by Campus -X

PUBLICATIONS

Elevating Large Language Models' Comprehensibility with SHAP Values

Submitted for Publish