

ARSALAN SHAIKH

Data Scientist with experience in MLOps, medical imaging, financial risk analysis, and AI-driven platforms, focused on delivering reproducible models with real-world impact.

✉ skarsalan8983@gmail.com

🌐 [skarsalan07](#)

in [arsalanshaikh123](#)

🔗 [portfolio](#)

EDUCATION

B.tech in Ai & Data Science

DBATU University, Lonere

📅 2021- 2025

7.5 CGPA

INTERNSHIPS

Machine Learning intern

📅 Oct - Dec 2024

Infosys Springboard

📍 Remote

- Developed and optimized a CNN model for Iris Eye Tumor Detection, achieving 92% accuracy.
- Processed and augmented 10K+ images, improving model reliability and reducing false positives by 18%.
- Built a scalable, real-time diagnosis pipeline with full project documentation and CI/CD integration.

PROJECTS

Aimers- AI Interview Platform 🔗 (Jan-Mar 2025)

AI platform helping students assess knowledge and improve interview skills with personalized feedback.

- Developed a BERT-based evaluation system achieving 90%+ accuracy against expert scores.
- Optimized FastAPI backend, reducing answer evaluation time from 1.2s to 300ms.
- Improved recommendation relevance by 40% using TF-IDF + cosine similarity and statistical analysis.
- tracked experiments with MLflow and managed datasets with DVC for reproducibility.

VeriScore Ai - Loan Risk Detector 🔗 (Apr-Jun 2025)

Predictive system assessing loan risk by analyzing financial metrics to reduce defaults & improve approval accuracy.

- Analyzed 50K+ loan applications found 70% defaults had DTI > 20%.
- Engineered 11+ features (e.g., payment-to-income ratio) to improve insights and model performance.
- Set up MLflow + DVC for reproducible, multi-algorithm tuning identified final XGBoost model (AUC 0.91) that reduced false approvals by 22%.
- Developed a Power BI dashboard and prediction model, deployed on Render via FastAPI

Movie Recommendation System 🔗 (Sep-Oct 2024)

Personalized suggestions using IMDB data. Built a TF-IDF + cosine similarity engine with 85% accuracy and integrated thumbnails to boost engagement by 30%.

TECHNICAL STACK

Data Science / Machine Learning / Deep Learning

Regression, Classification, Clustering, Ensemble Methods , PCA, CNN, RNN, LSTM, GRU, Model Tuning, Feature Engineering, Feature Selection & Extraction

NLP

Text Preprocessing, Tokenization, Word2Vec, BERT, NLTK, spaCy, Text Classification, Similarity Analysis

Data Analysis & Visualization

EDA, Data Visualization (Matplotlib, Seaborn, Power BI, Tableau), SQL (MySQL), NoSQL (MongoDB)

Mathematics for ML & DL

Algebra, Probability, Statistics, Calculus, Matrices

Programming & Frameworks

Python (NumPy, Pandas, SciPy, Scikit-Learn, TensorFlow, Keras, PyTorch, FastAPI), HTML, CSS, Flask, Figma

MLOps & Deployment

CI/CD, Docker, MLflow, DVC, Render

CERTIFICATIONS

Career Essentials in Generative Ai

Microsoft and LinkedIn

Generative AI Professional

Oracle

ACHIEVEMENTS

- Led the team at Quasar Hackathon, advancing to finals by rapidly prototyping solutions under high-pressure conditions.

SOFT SKILLS

- **Fast Learner** • **Leadership** • **Optimistic**
- **Problem Solving** • **Communication Skills**

INTERESTS

- **AI/ML** • **Software Engineer** • **Developer**