

# ARSALAN SHAIKH

[skarsalan8983@gmail.com](mailto:skarsalan8983@gmail.com) | +91-8983665317  
[linkedin.com/in/arsalanshaikh123](https://linkedin.com/in/arsalanshaikh123) | [github.com/skarsalan07](https://github.com/skarsalan07)

## PROFILE

Entry-level Data Scientist with hands-on experience in ML and AI solutions, including medical imaging, financial risk analysis, and an AI interview platform. Delivered strong results in model accuracy and performance across real-world projects. Skilled in full ML workflows and translating complex data into actionable insights.

## EXPERIENCE

**Deep Learning Intern (Remote)** | Infosys Springboard Oct 2024 - Dec 2024

- Achieved **92% accuracy** on the Iris Eye Tumour Detection CNN model that built and optimized.
- Augmented and normalized **10K+ images** to improve model robustness.
- Reduced **false positive rate** by 18% via hyperparameter tuning, built scalable pipeline for real-time diagnosis.
- Managed end-to-end project using CI/CD and maintained a clean, documented GitHub repository.

## PROJECTS

**Aimers - AI Interview Platform** [github.com/skarsalan07/Aimers](https://github.com/skarsalan07/Aimers)

- Built a BERT-based interview evaluation system with **90%+ accuracy**.
- Reduced answer evaluation time from **1.2s to 300ms** using FastAPI backend optimization.
- Enhanced recommendation relevance by **40%** using **TF-IDF + cosine similarity** and statistical analysis.
- Delivered a full-stack pipeline with modular software architecture and GitHub collaboration.

**Loan Risk Analysis & Prediction System** [github.com/skarsalan07/loan-project](https://github.com/skarsalan07/loan-project)

- 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.
- Trained an **XGBoost model**, achieved **91% AUC**, reducing false approvals by **22%**.
- Built interactive **Power BI dashboards** to visualize default trends backend deployed using FastAPI and Render.

## SKILLS

- Machine Learning & Deep Learning** : Regression, Classification, Clustering, Ensemble Methods (Voting, Stacking, Bagging, Boosting), PCA, CNN, RNN, LSTM, GRU, Model Tuning
- Natural Language Processing** : Text Preprocessing, Tokenization, Word2Vec, BERT, NLTK, spaCy, Text Classification, Similarity Analysis
- Data Analysis & Visualization** : EDA, Feature Engineering, SQL (MySQL), MongoDB, Power BI, Tableau, Matplotlib, Seaborn
- Programming & MLOps** : Python (Pandas, NumPy, Scikit-learn, TensorFlow, Keras, PyTorch, FastAPI), R, C++, Git, GitHub (CI/CD), Docker, MLflow, DVC, API Deployment

## EDUCATION

**B.Tech in Artificial Intelligence & Data Science** 2021 – 2025  
Marathwada Institute of Technology, Aurangabad **CGPA: 7.5**