Shirsh Mall

+91 8527766680 | shirshmall10work@gmail.com | linkedin.com/in/shirsh10mall | github.com/shirsh10mall

Education

Indian Institute of Technology Delhi — New Delhi
Master of Science (MSc.) in Physics · IIT-JAM Rank: 97 · CGPA: 7.4 — *Aug 2021 – May 2023*

- Relevant Coursework: **Deep Learning for Mechanics, Probability and Statistics**

Dyal Singh College, University of Delhi — New Delhi
Bachelor Of Science (BSc.) in Physics Honours · CGPA: 8.5 — *Aug 2018 – July 2021*

- Relevant Coursework: Linear Algebra, Calculus

Projects

GitHub Portfolio Page

Patient Survival Prediction — Tools: Scikit-learn, Tensorflow-Keras, Pandas, Plotly, Shap, Streamlit, Heroku

- Analyzed patient's medical records collected by hospitals, performed EDA, trained various models for predicting whether patient will survive or not (**binary** classification).
- Achieved **0.88 AUC Score** and tackled overfitting by hyper-tuning Deep Neural Network model.
- Created model's web app using **streamlit**, deployed using **Heroku**, and analyzed **model's explainability** using **Shap**.
- In healthcare, this project aids swift survival prediction, aiding critical decision-making in crises for healthcare providers.

Revenue Forecasting for Dynamic Pricing Optimization — Tools: Scikit-learn, Tensorflow-Keras, Optuna, Pandas, SHAP

- Developed a **predictive revenue model** using XGBoost to optimise dynamic pricing, predicting revenue per user action for an e-commerce company.
- Conducted data pre-processing, employed tree-based algorithms, hyperparameter optimization, and feature selection and deployed **Web App** the model using Streamlit.
- Explored **ensemble models** with H2O's AutoML, along with Shap values for model explainability.

Image Captioning and Image Retrieval System — Tools: Selenium, TensorFlow, PyTorch, Hugging Face, FAISS, FastAPI

- Collected diverse datasets: **1000 captions** from ChatGPT and **web scraped** Google images, and 150K GCC **image-caption pairs**, after thorough **preprocessing**.
- Utilized **ViT-BERT** pre-trained model and fine-tuned to **generate image captions** and evaluated using **Rouge** score.
- **Implemented CLIP model** for **image retrieval** using PyTorch. Enhanced the retrieval capabilities by **fine-tuning** OpenAI's pre-trained CLIP model for improved performance.
- Created **web app** with user-friendly tabs for image captioning and retrieval using **FastAPI, HTML**, and **CSS**, incorporating **FAISS indexing** to boost retrieval **speed by 10x**.
- In e-commerce, this project enhances the shopping experience by enabling image and text-based product searches.

Semantic Search-based Recommendation and QnA System — Tools: BeautifulSoup, Selenium, NLTK, Hugging Face, FAISS, Gradio

- **Web-Scraped** 1000 Reuters news articles with their summaries and added CNN news data for more depth.
- Fine-tuned BART models for news **summarisation**, generated news article **embeddings**, enabling **10x faster semantic search** with **FAISS indexing**.
- Implemented **context-based question-answering** using pre-trained models with Wikipedia-search dynamic content extraction.
- Created user-friendly **web applications** and **deployed** using **Gradio**
- The project facilitates content discovery and user engagement for news platforms through semantic search and real-time Q&A.

Technical Skills

Programming Language: Python | **Database**: SQL | **Data Science**: Machine Learning · Natural Language Processing · Computer Vision
Tools/Framework: NumPy · Pandas · Scikit-Learn · TensorFlow-Keras · PyTorch · Hugging Face · Plotly · Seaborn · Matplotlib
Model Deployment: Streamlit · Flask · FastAPI · Gradio | **Additional Skills:** Explainable AI (Shap) · Web Scraping

Research Experience and Academic Coursework

Masters Thesis: Particle Physics Data Analysis using Deep Learning

- **Achieved enhanced signal efficiency**, accurate signal and background **events classification**, precise transverse momentum prediction using **ANNs** and implemented **CNNs** models for **Jet-Images classification**, after jet clustering
- Employed **data analysis**, utilised **data simulation** tools for event generation using theoretical learning and **transformed data into image** representations for deeper analysis.

Assignments - Implementation from Scratch

- Multivariate non-linear regression, Binary classification, One-vs-all and Softmax multi-class classification algorithms.
- ANN and CNN multi-class classification using MNIST dataset, RNN for time series dataset and Physics Informed-Neural Network.