

# Shirsh Mall

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

## Education

### Indian Institute of Technology Delhi

Master of Science (MSc.) in Physics · IIT-JAM Rank: 97 · CGPA: 7.4

New Delhi

Aug 2021 – May 2023

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

### Dyal Singh College, University of Delhi

Bachelor Of Science (BSc.) in Physics Honours · CGPA: 8.5

New Delhi

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-Cloud

- Analyzed **90k** medical records, conducted comprehensive **exploratory data analysis**, and developed a **Binary classification** model predicting patient will survive or not.
- Employed a variety of **ML** and **DNN** models, including multiple **custom stacked ensemble models**.
- Mitigated overfitting through hyperparameter tuning using **custom objective function**, resulting in an impressive **0.88 AUC Score**
- Developed a model **web app**, deployed seamlessly using **Streamlit**, and conducted model **explainability** analysis using **Shap**.
- In healthcare, this project aids swift survival prediction, aiding critical decision-making in crises for healthcare providers.

### Image Captioning and Image Retrieval System [↗](#)

Tools: Selenium, TensorFlow, PyTorch, Hugging Face, FAISS, FastAPI, Docker

- Collected diverse datasets: **1000** captions from ChatGPT and **web scraped** Google images, and **150K** GCC image-caption pairs.
- 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**, followed by **containerization** using **Docker**.
- This project can enhance e-commerce by utilizing image captioning for product tagging, creating descriptions, and improving product text-to-image similarity searches system internally.

### 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.

### Revenue Forecasting for Dynamic Pricing Optimization [↗](#)

Tools: Scikit-learn, Tensorflow, Optuna, Pandas, SHAP, AutoML

- Conducted **data pre-processing** and **EDA**, utilising **ML** and **ANN** algorithms to optimise dynamic pricing and forecast revenue per user action using e-commerce dataset.
- Developed a **predictive regression model** using XGBoost, followed by feature selection.
- Deployed a user-friendly **web app** for the model using **Streamlit**.
- Explored **ensemble models** with H2O's AutoML setup, along with **Shap** values for **model explainability**.

## Technical Skills

**Programming Language:** Python | **Database:** MySQL | **Data Science:** Machine Learning · Deep Learning · Natural Language Processing

**Frameworks:** Scikit-Learn · TensorFlow · PyTorch · Hugging Face | **Model Deployment:** Streamlit · Flask · FastAPI · Gradio · Docker

**Additional Skills:** Explainable AI (Shap) · Elementary Tableau · Web Scraping

## Research Experience and Academic Coursework

### Masters Thesis: Particle Physics Data Analysis using Deep Learning [↗](#)

- Achieved accurate **signal-background events classification** and constructed **predictive ANN models** for key physical variables.
- **Transformed data into image** for deeper analysis and implemented **CNN models** for **Jet-Images classification**, after jet clustering
- Employed **data analysis**, utilised **data simulation** tools for event generation using theoretical learning.
- This project facilitates identifying signal events, **reducing repetitive initial theoretical calculations**, aiding physicists in research.

### 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.