SUBHAJIT TARAFDAR

ACADEMIC QUALIFICATIONS

Year	Degree /Board	University /Institution	%/CGPA
2025*	Post Graduate Diploma in Business Analytics	IIM Calcutta, IIT Kharagpur, ISI Kolkata	-
2022	M.Tech. Biomedical Engineering	Indian Institute of Technology, Kharagpur	8.26/10
2018	B.E. Electrical Engineering	Jadavpur University	7.38/10
2013	CLASS XII	Uluberia High School	80.6 %
2011	CLASS X	Uluberia High School	82 %

KEY SKILLS/TOOLS Machine Learning, PyTorch, SQL, Programming, PySpark, Big Data, Python, Statistical Modeling, NLP

AWARDS AND ACHIEVEMENTS

Master' Thesis Project (Computer Vision)	■ Developed **multi-class classification;** used Chest **X-ray** images to identify normal, **covid-19** & **pneumonia** ■ **Normalized** the pixels, **resized** images, used **segmentation** and **augmentation** to **pre-process** the images ■ Implemented **ensemble** tech. on **VGG16, InceptionV3, DenseNet121** & **MobileNet,** on **ImageNet weights** ■ Achieved **interpretability** using **GRAD-CAM** algo; achieved **recall** of **0.89** and **0.93 accuracy** using **VGG16**
Case Competitions	■ Ranked **2nd** /1004 in "Prodalytics", **dashboarded** Zomato Bangalore data, organized by **XLRI, Jamshedpur** ■ Ranked **9th** /243; built **Credit Score Classification model** for banking industry; organized by **IIM, Udaipur** ■ **National Finalist** out of over **1800** teams; **Consulting case competition** conducted by **XLRI, Jamshedpur**
Certifications	■ Completed **Winter School** on **Deep Learning certification**; achieving **2nd** rank; conducted by **ISI**, **Kolkata**

ACADEMIC PROJECTS

Used Car Price Prediction (Regression)	■ Implemented **multiple-linear regression** models to predict price of used cars; utilizing**1500+ data points** ■ Used **VIF** to handle **multicollinearity**;effectively handled **leverage** points, **outlier**; performed **OLS** & **ridge** ■ Validated model assumptions via **residual** analysis; improved **adjusted R^2** from **0.6**(baseline MLR) to **0.78**
Market Risk Analysis (Time Series)	■ Modelled daily stock price volatility of **NIFTY50** data with **1600+** observations using a **GARCH(1,1)** model ■ Performed pre-processing, **ADF** test for stationarity, analyzed **ACF**, **PACF** plots of returns square & **residual** ■ Forecasted exchange rate via **ARIMAX**, interest rate differentials as **exogeneous**, achieving **MAPE** of **1.26%**
Heart Attack Prediction (Classification)	■ Developed models to predict heart attack occurrences with over **4.4L** historical data and **40** feature columns ■ Removed **duplicated** values, **imputed** missing values, detected **outliers**; applied **SMOTE** for data balancing ■ Implemented multiple feature selection techniques; developed classifiers via **LR**, **DT**, **RF** & **XG-Boost** model ■ Employed **F1** scores to compare models; **Ridge** provided **AUC 0.94;** got **f1-score** of **0.94** for random forest
Customer Need Identification (NLP)	■ **Scrapped** over **30K** mobile reviews from **Flipkart**; performed **topic modeling** & **sentiment classification** ■ Performed **tokenization**, **stemming** & **POS tag.**; used **LDA**; got **6** topics as **optimal** using **coherence score** ■ Trained **LSTM** using labelled dataset; achieved **90.4%** accuracy; got topic & sentiment level for each review
Traffic Flow Forecast (GNN)	■ Implemented traffic forecasting with **GNN** on **PeMDS7** dataset;capturing **spatial dependency** using graphs ■ Designed **graph convolution network,** with **LSTM** to predict the **next 3** timestamps from the **previous 12** ■ Improved results with **graph attention network** and **LSTM** for **9 timestamps**; reduced **rmse** value to **6.25**

ADDITIONAL PROJECTS

RAG Generated Financial Insights (Gen-AI)	■ Implemented **chatbot** using **Retrieval Augmented Generation** & **Langchain** for financial reports analysis ■ Loaded docs & created chunks via **TextSplitter**; created embeddings using **OpenAI** & stored in **Pinecone** ■ Used **MMR** for retrieving chunk; incorporated **memory** & created **retrieval chain** for efficient conversation
Movie Recommendation System (Pyspark)	■ Developed **hybrid** movie recommendation system using **45K+ movies** data, **270K+** users & 2**5M+ ratings** ■ Created **content** based recommendation system using **graph**; used **Adamic Adar** Index to check similarity ■ Implemented **user-based collaborative** filtering; used **cosine similarity** for recommending top **10** movies ■ Used **ALS** in **PySpark** to get user latent embeddings; got **13** optimum **latent** factors; got rmse value of **0.83**
Customer Segmentation (Clustering)	■ Analyzed Indian bank customer dataset with over **1L+** observations to identify distinct customer segments ■ Performed **feature engineering** and **RFM** analysis; explored **K-Means**, **Agglomerative**, **GMM** & **DBSCAN** ■ Identified **7** segments using **elbow** method; did cluster profiling; **silhouette** score **0.54** and **DB index 0.79**
Image Caption Generation (CNN & NLP)	■ Implemented **encoder-decoder** models on **8K+** images to train & generate automatic caption from images ■ Performed **resizing**, **normalization**; used transfer learning on **DenseNet121** to encode & LSTM to decode ■ Improved **ROUGE** value to **0.144** & **CIDEr** to **0.146** using **ResNet50** as encoder & **transformer** as decoder

POSITIONS OF RESPONSIBILITY & EXTRA CURRICULARS

Organizing Committee, Srijan (Tech-Mgmt Fest)	■ Led a team of **10+** members to successfully conduct **15+** events with footfall of **200+** participants city-wide ■ Coordinated with **8+** vendors to streamline supply logistics of equipment to successfully conduct the events ■ Arranged & provided hospitality to **200+** participants and chief guest to ensure smooth conduct of the fest
Sports	■ Captained departmental football team to semi-finals in **U-turn** | Member of winning football team in **Arena**

ELECTIVES : HCA, Financial Risk Managment , Bayesian Methods **INTERESTS : Football, E-Sports**