

Arpit Kumar

✉ arpikumar2004 | [in](#) Arpit Kumar | 📩 arpikumar@kgpian.iitkgp.ac.in | ☎ +91-9506035861

WORKING INTERESTS

Applied AI Research | Optimization Algorithms | Time Series Analysis | Deep Learning | Quantitative Research
EDUCATION

Indian Institute of Technology (IIT) Kharagpur, India

Major: Integrated B.Tech - M.Tech (Dual Degree) in Chemical Engineering

CGPA: 8.46/10

2022 - 2027

TECHNICAL SKILLS

Programming Languages: Python • C • C++ • MySQL • MATLAB • Html • CSS • JavaScript • LaTeX

Libraries: NumPy • Pandas • PyTorch • TensorFlow • Keras • Scikit-learn • Matplotlib • Seaborn • Plotly

Softwares & Technologies: Apache Airflow • Apache Kafka • MLflow • DVC • FastAPI • Flask • Docker

ACHIEVEMENTS

- Achieved **Expert** title on Codeforces (highest rating: **1612**) under handle: **_Badassium**, ranked 203 in Div 2 Round **1032**
- Achieved a **Silver** Medal in the Open IIT Data Analytics 2023 from Rajendra Prashad Hall of Residence in year 2023
- Achieved a **Silver** Medal in the Open IIT Case Study 2023 from Rajendra Prashad Hall of Residence in year 2023
- Achieved a **Bronze** Medal in the General-Championship ChemQuest 2024 from Rajendra Prashad Hall of Residence
- Achieved an outstanding All India Rank of **4478** in JEE Advanced 2022 among **1,50,000+** candidates in year 2022-2023
- Achieved an outstanding percentile of **99.2%tile** in JEE Mains - April 2022 among **10,00,000+** candidates in year 2022

PROJECTS

Multimodal Price Prediction using Text, Image, and Tabular Data | Amazon ML Challenge

Kharagpur, India

Developed an end-to-end multimodal ML pipeline for product price estimation

Oct 2025 - Oct 2025

- Built a multimodal price prediction system combining **BERT**-based text embeddings, **CLIP** image features, and tabular data with **UMAP**-based dimensionality reduction and ensemble stacking (**Linear**, **RF**, **LGBM**, **XGB**, **CatBoost**).
- Achieved **SMAPE-42.89%**, secured **top 0.5%** globally by building a Modular, OOP-driven, YAML-configured experimental ML pipeline for rapid experimentation, cross-validation, and meta-learning via a **Ridge**/**LightGBM** stacker.

Deep Learning - Based Text Summarization System | Course Project

IIT Kharagpur, India

Developed a Deep-learning based end-to-end pipeline to summarize large text

Jun 2025 - Jun 2025

- Developed summarization model processes **100K+** documents, cut average summary length by **75%** & retained **92%** info
- Automated processing and training pipelines, boost throughput **5x** from **5K** to **25K** doc/hr & streamlined model updates
- Improve summary quality **35%** by ROUGE-L, F1 gain in extractive baselines & reduced manual review workload by **60%**

COMPETITIONS

Volatility Curve Prediction by NK Securities 2025

Kharagpur, India

Optimised predictive models with advanced quantitative techniques & novel geometric ensembling strategy

May 2025 - May 2025

- Pioneered novel geometric ensembling strategy leveraging Euclidean geometry and data clustering for enhanced accuracy.
- Developed predictive models using **Parametric Curve Fitting**, **Denoising** Transformers, and **Iterative** Imputation.
- Demonstrated strong analytical thinking, integration of advanced machine learning, statistical, and geometric principles.

DTL Quant Challenge 2024 by DTL | Top 20 (National Level)

Kharagpur, India

DTL | Developed a dynamic alpha model integrating RSI, mean reversion, volatility, and trend signals

Sep 2024 - Sep 2024

- Designed alpha model by RSI, mean reversion, volatility, trend signals with adaptive weights for shifting market regimes.
- Achieved alpha scores of **2.42** (in-sample), **1.32** (out-of-sample), & **0.48** (live), validating robustness across testing stages.
- Ranked **19th** nationally in a live trading simulation, demonstrating strong predictive accuracy & innovative model design.

General Championship Data Analytics by Evva Health | Gold

IIT Kharagpur, India

Technology Students' Gymkhana, IIT Kharagpur | Developed a Social and Healthcare Risk Scorecard

Mar 2024 - Apr 2024

- Achieved **82.89%** accuracy by Voting ensemble **BERT**, Naive Bayes on patient-reported health data for risk classification
- Reduce manual data collection time by **60%** by deploying a **Streamlit** web app with real-time scoring & dynamic feedback
- Scrapped **1000+** entries by **BeautifulSoup/Selenium** and **Bifactor & MIRT** models to optimise healthcare resource in India

Open IIT Data Analytics Competition | Silver

IIT Kharagpur, India

Technology Students' Gymkhana, IIT Kharagpur | Prediction of Footfall on time-series data

Nov 2023 - Dec 2023

- Secured **2nd** rank among institute, predicting **city footfall** from prior years time series data using **time-series models**
- Scrapped data by **BeautifulSoup** and applied **EDA**, Time series analysis, and clustering techniques by **K-Mean Cluster**.
- Utilized forecasting models **FBProphet**, **Random Forest**, **LSTM**, and ensemble methods & achieved **accuracy** of 86.63%.

RELEVANT COURSEWORK

Finance : Entrepreneurial Essentials, Stochastic Processes in Finance, Mathematical Finance, Financial Institutions & Markets

Data Science : Machine Learning, Neural Network & Deep Learning, Regression and Time Series Modelling, Programming and Data Structures, Computer-Aided Process Engineering, High Performance Computing*

Mathematics : Mathematics I, Mathematics II, Probability and Statistics, Optimization Techniques, Advanced Mathematical Techniques, Transform Calculus, Introduction to Stochastic Processes, Partial Differential Equations