

Harshit Joshi

(+91) 7042565768 | harshith.j.2003@gmail.com | New Delhi

[in](#) Harshit | [harshit1912003](#) | [Twitter](#) HarshitJos2003 | [GitHub](#) harshit1912003.github.io

OVERVIEW

Mathematics graduate with experience in developing end to end AI/ML models and implementing optimization solutions across diverse projects. Eager to leverage this strong analytical foundation for pioneering research in the field of stochastic modeling and optimization.

EXPERIENCE

Research Intern

Dr. Aparna Mehra, HOD, Mathematics, IIT Delhi

Dec 2024 – Jan 2025
New Delhi

- Implemented Cook et al.'s RDM-based DEA framework for non-homogeneous and negative-data DMUs.
- Reproduced and validated results from Bansal (2023) on Non-Homogeneous DEA with Negative Data.

Machine Learning Intern

Swachh.io

May 2024 – Jul 2024
New Delhi

- Developed YOLOv8 ANPR + emission detection pipeline achieving 89.5% precision, 80.7% recall, and mAP50 88.4%.
- Scraped and analyzed ~8.5k schools using Python (BeautifulSoup, pandas) for filtration efficiency analysis.

EDUCATION

Indian Institute of Technology Delhi

M.Sc. in Mathematics

CGPA: 7.8/10

New Delhi

Jul 2023 – May 2025

Hans Raj College, University of Delhi

B.Sc. in Mathematics

CGPA: 8.8/10

New Delhi

Aug 2020 – Jun 2023

Khaitan Public School

CBSE (Schooling)

Class XII: 95.33%, Class X: 96.20%

RESEARCH OUTPUTS

[S]=IN SUBMISSION [T]=THESIS

[S] Novel DEA Models for Non-Homogeneous DMUs with Negative Data

(Submitted – EJOR)

Co-Author

Developed new DEA models for non-homogeneous DMUs with negative data using RDM-based convex and non-convex (FDH) formulations. Proposed RDM-based EAT and CEAT models to mitigate overfitting of the efficient frontier, validated via statistical testing. Implemented peer-benchmarking methodology for granular efficiency analysis and performance improvement.

[T] Non-Convex Data Envelopment Analysis

(Awarded JRC Best Thesis, MSc)

Supervisor: Dr. Aparna Mehra Co-Author: Ananya Sharma

Comprehensive study of classical (CCR, BCC, SBM, Additive) and non-convex (FDH) DEA models with economic interpretation and validation. Extended EAT method for frontier analysis under Free Disposal assumptions using Monte-Carlo evaluation. Developed open-source Python library (Gurobi backend) with GUI for DEA/FDH and non-homogeneous data; available at [GitHub](#).

PROJECTS

• Chessboard FEN Estimation using Novel Features

CNNs, Feature Engineering, Regression Analysis



Developed a Deep CNN to predict Stockfish evaluations from FEN positions, achieving MAE of 152 centipawns; introduced regression based board state features improving accuracy by almost 100 cp; deployed model via Streamlit app. Accessible at [mtl782-chessapp.streamlit.app](#),

• Text Guided Image Clustering

Vision-Language Models, BLIP, ViLT, SBERT, DBSCAN



Combined BLIP, ViLT outputs and their SBERT embeddings with deep CNN features (ResNet, EfficientNet) for Food-101 clustering; achieved ARI of 0.8041 using DBSCAN, demonstrating strong alignment between semantic and visual features.

• Clustering using ToMATo Algorithm

Topological Data Analysis, Persistent Homology



Applied Topological Mode Analysis Tool (ToMATo) for high-dimensional clustering using topological persistence; achieved ARI 0.78 on MNIST vs 0.52 with k-Means, showing robustness to noise and density irregularities.

• Multi-Category Text Classification

C-(Bi)LSTM, Meta-Embeddings, Self-Attention, Transformers



Implemented C-(Bi)LSTM and Transformer-based models with dynamic meta-embeddings (CBOW + Word2Vec); integrated self-attention to enhance contextual representation and boost multi-label classification accuracy.

• N-FLP Robust Linear Regression

Robust Estimation, Econometrics, Monte Carlo Testing



Implemented the Normal-Filtered Log-Pareto (N-FLP) robust regression model for heavy-tailed data; applied to IPUMS 2019 Mincer wage estimation for outlier detection and improved fit stability.

• S&P 500 Diversification

Portfolio Optimization, Hierarchical Clustering, QCQP



Formulated a novel QCQP-based diversification framework using hierarchical clustering; reduced 501→9 constraints, outperforming HRP with Sharpe Ratio 1.15 vs 0.91 over 10-year backtest.

• Axelrod's Tournament

Game Theory, Iterated Prisoner's Dilemma Simulation



Simulated Prisoner's Dilemma strategies (Tit-for-Tat, Random, Friedman, Joss, Harrington); built a tournament framework with cumulative payoff tracking and visualized comparative outcomes.

SKILLS

- **Languages & Databases:** Python, R, Julia, C/C++, SQL
- **Machine Learning:** PyTorch, TensorFlow, HuggingFace, scikit, cvxpy, numpy

EXTRA CURRICULAR ACTIVITIES

- **Co-Author:** In Search of the Perfect Story (Quill Club Writers, 2018)
- **Volunteer:** Mathematics Society, IITD (2023–2024)
- **Certifications:** Operations Research (1, 2 & 3) – NTU; ML in Production; OpenCV Bootcamp

RELEVANT COURSEWORK

MTL505 (Computer Programming), MTL502 (Linear Algebra), MTL508, 32357616 (Mathematical Programming), ELL784 (Introduction to Machine Learning), MTL601 (Probability and Statistics), MTL732, 32357614 (Financial Mathematics), HSL613 (Econometrics), AIL721 (Deep Learning), MTL782 (Data Mining)