



EDUCATION

| Degree/Certificate                         | Institute/Board                                | CGPA/Percentage | Year         |
|--------------------------------------------|------------------------------------------------|-----------------|--------------|
| Ph.D. Mathematics (Thesis Submitted)       | Indian Institute of Technology, Guwahati       | 8.76/10         | 2021–2026    |
| MITx MicroMasters in Finance (In Progress) | Massachusetts Institute of Technology (Online) | 4/5             | 2024–Present |
| M.Sc. Statistics and Computing             | Banaras Hindu University                       | 8.4/10          | 2018–2020    |
| B.Sc. Statistics & Computer Science        | DDU Gorakhpur University                       | First Division  | 2015–2018    |

EXPERIENCE

- Adobe** May 2024 – Aug. 2024  
*Ph.D. Media & Data Science Research Intern, [Certificate](#)* Noida, India
  - Researched memorization, reasoning, and counting capabilities of large language models (LLMs).
  - Designed quantitative metrics to evaluate the above capabilities and guided redesign of in-house LLMs.
- ByteDance** Mar. 2025 – May 2025  
*Math Expert* Remote
  - Quality reviewer of Ph.D.-level mathematics questions.
- Turing** Aug. 2024 – Feb. 2025  
*AI Trainer & Ph.D. Math Expert* Remote
  - Contributed to training data and evaluations for building advanced AI systems.
- Allen Institute for AI (AI2)** Mar. 2024 – May 2024  
*Freelanced NLP Expert* Remote
  - Applied advanced NLP (NER, relation extraction, summarization) to large corpora of historical research papers.
  - Improved information retrieval and knowledge synthesis pipelines for historical research.
- Huntington’s Disease Project (with University of Iowa collaborators)** Dec. 2022 – Present  
*Research Collaborator* Remote
  - Used linear mixed-effects models to study covariates affecting year-to-onset.
  - Led data analysis and interpretation to support scientific discussions and findings.
- Ready Tensor** Mar. 2021 – Dec. 2021  
*Machine Learning Engineer* Remote
  - Built production-ready, Dockerizable ML model templates for the Ready Tensor platform.
  - Implemented and optimized forecasting/classification models, improving accuracy and efficiency.
  - Select public repos: [TSMixer](#), [PatchMixer](#), [Catboost](#), [Adaboost Binary Classifier](#), [XGboost](#).
- Chegg Inc.** Jun. 2018 – Feb. 2021  
*Subject Matter Expert (Statistics & Probability), [Certificate](#)* Remote
  - Assisted 1000+ students with statistics and probability, improving understanding and outcomes.

SELECTED PROJECTS & PACKAGES

- imbalanced-spdf: Shape Penalized Decision Forests** 2025  
*PyPI package for imbalanced classification (SPBoDF/SPBaDF) with SVR regularization* [PyPI](#) | [Docs](#)
  - Ensembles penalize irregular decision boundaries via surface-to-volume regularization to improve generalization.
- memory-esn: Long-Memory Echo State Networks** 2026  
*Pip-installable reservoir-computing library (fESN, wESN) with Cython acceleration* [PyPI](#) | [Docs](#) | [GitHub](#)
  - Dual-reservoir design pairing a vanilla reservoir for short-term dynamics with a memory reservoir using fractional differencing and wavelet smoothing for long-range dependence.
  - Cython-accelerated kernels with pure-NumPy fallbacks, leakage-free time-series splitting utilities, examples, and a test suite. Companion software to the dengue forecasting paper.
- MART: Moving Average Randomized Trees** 2026  
*Reference implementation of the Springer Machine Learning paper* [GitHub](#)
  - Tree ensemble embedding multiple exponential-moving-average spans directly in the split criterion, removing the need for hand-built lag features; evaluated on 19 financial instruments.
  - Cython-accelerated MACD computation for fast training; MIT licensed.

- **fastkme: Fast Kaplan–Meier Estimator** 2024  
*Efficient nonparametric estimator with weighted KM curves* [GitHub](#) | [Blog](#)  
 – Lightweight package for accelerated KM estimation; includes weighted curve support.
- **Mini R Programs and Master’s Project on Credit Card Fraud Detection** 2020  
*Academic project* [Report](#)  
 – Implemented Random Forest training logic in Julia to deepen understanding of tree ensembles.

## PUBLICATIONS

Published • Accepted / in press • Under review / preprint

- **[Accepted, in press – NeurIPS 2026] CCTimeBoost: Learning Time-Varying Relative Risk with Case-Control Boosted Trees** (*sole author*) - A tree-based method for learning non-proportional risk for patients with explainability.
- **[Accepted, in press – TMLR] Deep Discrete-Time Survival Analysis with Guaranteed Monotonicity** (*sole author*) - Discrete-time neural survival model with monotonicity guarantees on the predicted survival curve; formerly circulated as KMNet. Code: [GitHub](#)
- **[Accepted, in press – ACML 2026] KESurv: A Kernel Ensemble Method for Patient-Specific Survival Prediction** (*sole author*) - Kernel ensemble approach to individualized survival prediction.
- **[Published – IEEE] Shape Penalized Decision Forests for Imbalanced Data Classification** - Improves TPR, F1, and G-mean across 12 benchmarks; natural multiclass extension. [Paper](#) | [PyPI](#)
- **[Published – Springer Machine Learning] MART: Moving Average Randomized Tree** (*equal first co-author with V. K. Giri*) - CAGR-based randomized split method for trend prediction in stock markets. [Paper](#)
- **[Under review – TMLR] Long-Memory Reservoir Computing for Data-Scarce Dengue Forecasting** (*first author*, with S. Paul, P. Ghosh and T. Chakraborty) - Long- and short-memory echo state reservoirs with fractional differencing and wavelet smoothing for forecasting under limited training data. [arXiv:2607.11272](#) | [Docs](#)
- **[Under review – NEJSDS] EnTransformer: A Deep Generative Transformer for Multivariate Probabilistic Forecasting** (*first co-author*, with R. Pathak, M. Panja, P. Ghosh and T. Chakraborty) - Generative transformer for multivariate probabilistic and spatiotemporal forecasting; New England Journal of Statistics in Data Science. [arXiv:2603.07108](#) | Code: [Genformer](#) | [Docs](#)
- **[Preprint] Concordance-based Survival COBRA with Regression-Type Weak Learners** - [arXiv:2209.11919](#).
- **[Preprint] Integrated Brier Score based Survival COBRA - A Regression-Based Approach** - [arXiv:2210.12006](#).

## TECHNICAL SKILLS

- **Programming:** Python (Expert), R (Expert), Cython (Intermediate), Julia (Intermediate), C (Basic), Lean (Intermediate)
- **ML/AI Frameworks:** PyTorch, TensorFlow, JAX, scikit-learn, scikit-survival, Hugging Face Transformers, LLMs (OpenAI GPT, LLaMA)
- **Performance:** Cython extensions with BLAS-backed kernels and NumPy fallbacks, profiling and optimization of training loops
- **Data & Visualization:** pandas, NumPy, SciPy, matplotlib, ggplot2, dplyr
- **Tools:** Docker, Git, PyPI packaging, Sphinx/ReadTheDocs documentation, pytest

## KEY COURSES TAKEN

- **Mathematics/Statistics:** Probability Theory, Real Analysis, Linear Algebra, Regression, Bayesian Statistics, Survival Analysis
- **Data Science/ML:** Advanced Topics in ML, Advanced Statistical Algorithms, Statistical Foundations for Data Science, Computational Statistics

## POSITIONS OF RESPONSIBILITY & TEACHING

- **Head Teaching Assistant (Lead for four TAs)**, Coursera AI Program DA112 *Global, Term-based*
- **Teaching Assistant**, IIT Guwahati - Financial Risk Analysis (with Dr. Arabin Kumar Dey) *Term-based*
- **Teaching Assistant**, IIT Guwahati - Advanced Statistical Algorithms (with Dr. Arabin Kumar Dey) *Term-based*
- **Teaching Assistant**, IIT Guwahati - R Programming (UG) *Term-based*
- **Teaching Assistant**, Applied Statistics (with Dr. Palash Ghosh) *Term-based*

## ACHIEVEMENTS & CERTIFICATIONS

- **GATE (Statistics)** - AIR 63, India *Certificate*
- **IIT-JAM (Mathematical Sciences)** - AIR 306, India *Certificate*
- **NPTEL - Operations Research (Top 1%)**, IIT-led MOOC *Certificate*
- **NPTEL - Practical Machine Learning with TensorFlow**, MOOC *Certificate*
- **NPTEL - Fuzzy Logic and Neural Networks**, MOOC *Certificate*
- **Andrew Ng’s Machine Learning (Coursera)**, MOOC *Certificate*
- **Institute of Actuaries of India - CS1**, Professional *Certificate*
- **Top Rated Plus Freelancer**, Upwork *Profile Link*