

CONTACT INFORMATION	<p>B2-224, Kameng Department of Mathematics, Indian Institute of Technology, Guwahati, Assam - 781039 Mob : +919044725608 Whatsapp : +919140803463 Founding Member @ LabTenX</p>	<p>📧 Blog (statML) : http://www.statml.in 📄 Github : https://github.com/yuvrajiro ✉️ LinkedIn : https://www.linkedin.com/in/yuvrajiro/ ✉️ Primary E-mail : rahul.goswami.131997@gmail.com ✉️ Academic E-mail : rahul.goswami@iitg.ac.in ✉️ Academic Page : Rahul Goswami ✉️ Secondary E-mail : yuvrajiro@gmail.com</p>
AFFILIATION	<ul style="list-style-type: none"> Indian Institute of Technology, Guwahati (206123106) Sorbonne University Abu Dhabi Institute of Actuaries, India (IAI-32892) 	
TL:DR;	Senior Research Fellow @ IIT Guwahati, Sorbonne Abu Dhabi for Innovation and Research Institute (SAFIR) Affiliate, Top Rated Plus Freelancer @ Upwork, Founder @ LabTenX	
EDUCATION	<p>MITx Micromaster Programme, Online. 2024–Current</p> <ul style="list-style-type: none"> Micromaster in Finance Course Cleared Two Courses 15.415.1x: Foundations of Modern Finance I : Certificate and 15.415.2x: Foundations of Modern Finance II : Certificate Currently Pursuing Financial Reporting Course. <p>Indian Institute of Technology, Guwahati. 2021–Current</p> <ul style="list-style-type: none"> Senior Research Fellow at Department of Mathematics GPA in Coursework : 8.76/10 – via 42 credits In my coursework, I took 6 Courses: 1. Probability Theory, 2. Real Analysis, 3. Linear Algebra, 4. Advanced Topics in ML, 5. Advanced Statistical Algorithm, 6. Statistical Foundation for Data Science Course Repo for Advanced Statistical Algorithm : https://github.com/yuvrajiro/MA691 <p>Banaras Hindu University, Varanasi. 2018–2020</p> <ul style="list-style-type: none"> M.Sc. Statistics and Computing from the Center for Interdisciplinary Mathematical Sciences, with a GPA of 8.4/10. I completed 84 credits during my Master's degree, including courses in Bayesian statistics, computational statistics, regression, probability, and survival analysis. My interest in Bayesian inference was sparked during my Master's degree when I took my first course in Bayesian statistics. This technique has since become a key part of my approach to data analysis. I have consolidated some of my R programming codes on my website, which you can access here: https://yuvrajiro.github.io/Mini-R-Programs/. These codes demonstrate my skills in R programming and statistics. One of my notable projects is my credit card fraud detection project, which I completed using random forest. You can view the project here. While Random Forest is a subtle algorithm nowadays, coding it from scratch in Julia was quite challenging and provided key insights into coding tree-based algorithms. <p>Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur. 2015–2018</p> <ul style="list-style-type: none"> B.Sc. Statistics and Computer Science (2yrs. Mathematics) Department of Mathematics and Statistics For my Bachelor's, I majored in Statistics and Computer Science, with two years of Mathematics courses. I graduated with First Division honors. During my Bachelor's degree, I also completed the CS1 Certification Course by the Institute of Actuaries India, which is recognized by the Institute and Faculty of Actuaries (IFoA) in London. This certification demonstrates my knowledge and skills in actuarial science. 	

TECHNICAL SKILLS

- **Programming Languages:**
 - **Python:** Expert – Advanced capabilities in ML, data analysis, and software development.
 - **R:** Expert – Proficient in statistical analysis and visualization.
 - **Julia:** Intermediate – Experienced in high-performance numerical computing.
 - **C:** Basic – Familiar with low-level programming and algorithms.
 - **Lean** Intermediate - LEAN System prover
- **Machine Learning Frameworks and Libraries:**
 - **PyTorch, TensorFlow** – Expert in deep learning model development.
 - **scikit-learn, scikit-survival** – Proficient in traditional ML and survival analysis.
 - **Hugging Face Transformers** – Skilled in NLP and state-of-the-art language models.
 - **OpenAI GPT, Meta's LLaMA** – Familiar with cutting-edge large language models and their applications.
- **Data Analysis and Visualization:**
 - **ggplot2, dplyr** – Expert in data visualization and manipulation.
- **Tools and Technologies:**
 - Experienced with **Docker, Git**

WORK EXPERIENCE, FREELANCING, AND INTERNSHIP

- Math Expert @ ByteDance** March 2025 – Current
- Quality Reviewer of Phd Level Math Question
- AI Trainer @ Turing** August 2024 – Feb 2025
- Helping top companies to build AGI
 - Phd. Math Expert
- Media and Data Science Research Intern @ Adobe** May 2024 – August 2024
- Conducted research on the memorization, reasoning, and counting capabilities of large language models (LLMs).
 - Developed quantitative metrics to evaluate and measure LLM performance in these specific areas.
 - Used these metrics to guide the redesign and optimization of Adobe's in-house LLMs, enhancing their performance and efficiency.
- Freelanced NLP Expert @ AI2** March 2024 – May 2024
- Collaborated with the Allen Institute for AI on a project centered around natural language processing (NLP), focusing on leveraging advanced techniques to handle complex text data.
 - Applied sophisticated NLP methods to extract and analyze critical information from a large corpus of historical research papers, involving tasks such as entity recognition, relationship extraction, and summarization.
 - Contributed to the development of systems aimed at making historical research more accessible and useful by utilizing cutting-edge NLP technologies to improve information retrieval, context understanding, and knowledge synthesis.
- Research Collaborator @ Huntington's Disease Project** Dec 2022 – Present
- Collaborated with researchers from Iowa University on a project focused on studying Huntington's Disease.
 - Applied expertise in Linear Mixed-Effect models to analyze data and identify covariates affecting the Year to Onset of the disease.
 - Contributed to data analysis, interpretation of results, and discussions with the research team, furthering our understanding of the disease and its underlying factors.
- Machine Learning Engineer @ Ready Tensor** Mar 2021 – Dec 2021
- Contributed to the early-phase startup, Ready Tensor, by spearheading the implementation of advanced Machine Learning (ML) models on their platform. Collaborated closely with the team to design and develop ready-to-dockerize versions of ML models, ensuring ease of deployment and scalability.

- Demonstrated expertise in improving model performance, leading to improved accuracy and efficiency compared to previous iterations. Played a pivotal role in advancing the platform's capabilities by integrating cutting-edge ML algorithms.
- Specific projects included:
 - **rt_forecasting_TSMixer** (Public): Developed advanced time-series forecasting models using the TSMixer architecture. Python, Apache License 2.0.
 - **rt_forecasting_PatchMixer** (Public): Implemented the PatchMixer model for time series forecasting. Python, Apache License 2.0.
 - **rt_mc_class_catboost** (Public): Implemented a CatBoost model for multiclass classification.
 - **rt_adaboost_binary_classifier** (Public): Implemented an AdaBoost algorithm for binary classification. Python, MIT License.
 - **rt_xgboost_binary_classifier** (Public): Developed an XGBoost model for binary classification for Ready Tensor. Python, MIT License.

Subject Matter Expert @ Chegg Inc.

June 2018 – Feb 2021

- Helped more than 1000 students with their work on Statistics and Probability.
- Provided personalized assistance, explanations, and guidance on complex concepts, resulting in better understanding and academic performance for students.

PACKAGES

Shape Penalized Decision Forests (imbalanced-spdf)

- **Description:** Shape Penalized Decision Forests, for training ensemble classifiers tailored for imbalanced datasets. This package provides two primary implementations: SPBoDF and SPBoDF
- **PyPI Link:** <https://pypi.org/project/imbalanced-spdf/>
- **Documentation Link:** <https://imbalanced-spdf.readthedocs.io/en/latest/>
- **Project Overview and Future Plan:** Shape Penalized Decision Forests, for training ensemble classifiers tailored for imbalanced datasets. This package provides two primary implementations: SPBoDF (Shape Penalized Boosting Decision Forest): A boosting ensemble method that builds multiple trees sequentially, adjusting the weights of samples to focus on harder-to-classify instances, and SPBaDF (Shape Penalized Bagging Decision Forest): A bagging ensemble method that builds multiple trees independently on bootstrap samples of the dataset, improving robustness and reducing overfitting. Both implementations use the concept of Surface-to-Volume Regularization (SVR) to penalize irregular decision boundaries, thus improving generalization and addressing challenges associated with imbalanced datasets.

Combined Regression Strategy Survival (cobsurv)

- **Description:** cobsurv is a product of a broad research project, PBSA, which stands for Proximity-Based Survival Analysis. These currently contain combined regression strategy-based models proposed by me.
- **PyPI Link:** <https://pypi.org/project/cobsurv/>
- **Documentation Link:** <https://cobsurv.readthedocs.io/en/latest/>
- **Project Overview and Future Plan:** The aim of the project is to provide production-ready models for survival analysis. Future plans include implementing dockerized versions of survival models with Fast API integration. The API will be built inside the dockerized implementation, making it accessible for doctors to analyze survival data without the need for deep knowledge of survival models. The ultimate goal is to create a platform for doctors.

Fast Kaplan Meier Estimator (fastkme)

- **Description:** A faster Kaplan-Meier Estimator, specially created for working with a nonparametric Estimator.
- **Github Link:** <https://github.com/yuvrajiro/fastkme>
- **A blog:** <https://www.statml.in/2024-03-11-faster-kme/>
- **Project Overview:** This is a tiny package which computes faster Kaplan-Meier also comes with a weighted Kaplan-Meier curve.

RESEARCH
PAPER

- **Shape Penalized Decision Forests for Imbalanced Data Classification (Published):** Classification trees often yield fragmented minority boundaries under imbalanced data . The authors propose a surface-to-volume ratio (SVR) regularization that penalizes decision-set complexity, optimized via a greedy breadth-first splitting algorithm analogous to CART . Under mild smoothness and margin conditions, On twelve benchmarks, SVR-Trees improve true positive rate, F-measure, and geometric mean over oversampling and alternative trees, and extend naturally to multiclass settings. <https://ieeexplore.ieee.org/abstract/document/11003049>
- **MART: Moving Average Randomized Tree (Accepted @ Springer Machine Learning):** A randomized CAGR-based split method for predicting future trends in the stock market. [To be Published, Preprint not available](#)
- **Survival: A Different Approach (Under Review):** Developed and implemented novel data-dependent techniques to aggregate survival trees, enhancing the accuracy of survival estimates for patients. Explored innovative approaches for survival analysis, contributing to advancements in the field of statistical machine learning. Currently under review for publication.
- **Concordance-based Survival Cobra with Regression Type Weak Learners (Under Review):** Developed and proposed a novel survival analysis method, utilizing concordance-based techniques in combination with regression type weak learners. Explored innovative approaches for improving survival analysis accuracy and predictive performance. Paper available at: <https://arxiv.org/abs/2209.11919>
- **Integrated Brier Score based Survival Cobra - A Regression Based Approach (Under Review):** Proposed and developed an innovative survival analysis method using an integrated Brier score-based approach combined with regression techniques. Investigated the potential of this method to enhance survival prediction accuracy and evaluated its performance through empirical analysis. Paper available at: <https://arxiv.org/abs/2210.12006>

TEACHING
ASSISTANCE

- **Coursera DA112: Introduction to R Programming** Worked as a Teaching Assistant for the global course “DA112: Introduction to R Programming” on Coursera. Assisted in delivering course content, addressing student inquiries, and providing support to learners worldwide, enhancing their understanding of R programming.
- **Financial Engineering:** Collaborated with Dr. Arabin Kumar Dey in instructing the Financial Risk Analysis course at the Indian Institute of Technology Guwahati. Provided guidance and support to Bachelor’s Degree students, helping them comprehend and apply concepts related to financial risk analysis. Contributed to fostering a comprehensive understanding of financial engineering principles among the students.
- **Advanced Statistical Algorithm:** Assisted Dr. Arabin Kumar Dey in teaching the Advanced Statistical Algorithm course at the Indian Institute of Technology Guwahati. Interacted with students, addressed their queries, and provided guidance in understanding complex statistical algorithms. Conducted sessions on introductory Python with a focus on statistics to enhance students’ programming skills. Contributed to the effective delivery of the course and provided valuable support to students pursuing their Bachelor’s Degree.
- **R Programming:** Assisted Dr. Arabin Kumar Dey in teaching R programming to Undergraduate students at IIT Guwahati. Provided guidance and support to students, helping them understand and apply fundamental concepts of R programming. Contributed to creating a positive learning environment and fostering students’ programming skills.
- **Applied Statistics:** Assisted Dr. Palash Ghosh in to conduct bachelor level course at IIT Guwahati

CERTIFICATE
COURSES AND
ACHIEVE-
MENTS

- **Qualified Graduate Aptitude Test in Engineering (GATE) (Statistics)** Achieved All India Rank 63 out of thousands of candidates in the GATE exam, a highly competitive national-level examination conducted by premier institutes of India for admission to graduate programs in various subjects. Demonstrated proficiency in statistics and strong academic performance, which contributed to my admission to the Indian Institute of Technology, Guwahati.
- **Qualified IIT-JAM** Secured All India Rank 306 in the subject Mathematical Sciences in the Joint Admission Test for Masters (IIT-JAM), a prestigious national-level exam for admission into Masters programs at Indian Institutes of Technology (IITs). Demonstrated strong mathematical and analytical skills, contributing to my academic success and admission to a competitive Master’s program.

- **Fuzzy Logic and Neural Networks, NPTEL** Completed the Fuzzy Logic and Neural Networks course as part of my Master's program. Achieved successful qualification in the proctored exam conducted by the Indian Institute of Technology. The course deepened my understanding of fuzzy logic systems and neural network architectures, enabling me to apply these concepts to complex problem domains. Certificate available upon request.
- **Operations Research, NPTEL** Successfully completed the Operations Research course offered by NPTEL during my Master's program. Achieved a top ranking of 1% in the proctored exam conducted by the Indian Institute of Technology. The course enhanced my understanding of optimization techniques and their applications in real-world problem-solving. Certificate available at: [Operations Research NPTEL Certificate Link](#).
- **Andrew NG's Machine Learning Course** Completed the renowned Machine Learning Course offered by Andrew NG on Coursera. The course covered fundamental concepts and techniques in machine learning, providing a strong foundation in the field and enhancing my ability to develop and implement machine learning algorithms effectively. Certificate available at: [Coursera Certificate Link](#).
- **Practical Machine Learning With TensorFlow – NPTEL** Successfully completed the Massive Open Online Course (MOOC) on Practical Machine Learning With TensorFlow organized by NPTEL. The course provided comprehensive insights into practical applications of neural networks and deep learning techniques, contributing to my proficiency in machine learning and enhancing my practical skills in developing advanced models.
- **CS1 from Institute of Actuaries India** Successfully completed Core Statistics 1, a rigorous and comprehensive course, at par with the Institute and Faculty of Actuaries (IFoA) standards. Acquired in-depth knowledge and skills in fundamental statistical concepts, enhancing my analytical abilities and contributing to a solid foundation in statistics.