ANIRUDDHA PATHAK

OBJECTIVE

Highly motivated graduate student with strong knowledge of Bayesian statistics, Nonparametric methods, Statistical Computing and Machine Learning seeking an internship opportunity to explore the practical implementation of data analytics into real-life business problems.

EDUCATION

Doctor of Philosophy in Statistics Aug 2020 - May 2026(Expected)
Iowa State University, Ames, IA; CGPA: 3.83/4
Advisor : Prof. Somak Dutta

Master of Science in Statistics Aug 2018 - Jul 2020
Indian Institute of Technology, Kanpur, India; CGPA: 9.2/10

Bachelor of Science with Honors in Statistics Jul 2015 - Jun 2018
University of Calcutta; Ramakrishna Mission Residential College, Narendrapur; India.
Honors(Major) Subject Aggregate Marks: 84.25%.
Minor Subjects: Mathematics, Computer Science.

DOCTORAL RESEARCH

Bayesian Variable Selection for Multi-Environment Agricultural Trials Aug 2023 – Present
Ph.D. Research Project

· Currently working on developing a Bayesian variable selection method that leverages the genotype information to model genome-environment interaction yield data.

Regularized AMMI Model for Multi-Environment Agricultural Trials Aug 2022 – Aug 2023
Ph.D. Research Project

· Developed a hierarchical regularized Additive Main-effects and Multiplicative Interaction (AMMI) model for modeling genome-environment interaction yield data in the presence of missing values and predicting yield for untested environments.

WORK EXPERIENCE

Graduate Research Assistant Aug 2023 - Present
Iowa State University

· Working on a project titled *A Machine Learning Postprocessor to Mitigate QPF Errors for Improved Hydrometeorological Forecasting* in collaboration with the Dept. of the Earth, Atmosphere, and Climate at Iowa State University.
· Developing ensemble Machine Learning models that utilize 8 different model forecasts and various weather parameters to make more accurate predictions for the precipitation system location and its intensity.

Quantitative Analyst Intern Jun 2024 – Aug 2024
Wells Fargo *Charlotte, NC*

· Worked in the Commercial Loss Forecasting team in the Risk Modeling Group CoE.
· Developed Top-down commercial loan balance models for forecasting outstanding and commitment balances in the context of data segments from the Wealth Investment and Management line of business.
· Tested the performance of the models using Scenario Sensitivity Testing and Backtesting.

Quantitative Analyst Intern Jun 2023 – Aug 2023
Wells Fargo *Charlotte, NC*

- Worked in the Balance Sheet and Stress Test Validation team in the Corporate Model Risk CoE.
- Built benchmark models using XGB2 and ReLU-DNN methods from the PiML Python toolbox for forecasting the default rate in the auto loan portfolio.
- Investigated the importance of the feature variables with model-specific and post-hoc model-agnostic explainability tools.

Graduate Teaching Assistant Jan 2021 - Jun 2023
Iowa State University

- Worked as a lab instructor for undergraduate level Probability and Statistics courses - STAT 101, STAT 326, STAT 342.
- Worked as a grader for undergraduate and graduate level Probability and Statistics courses - STAT 231, STAT 322, STAT 542, STAT 587, and graduate-level Measure-theoretic Probability course - STAT 642.

OTHER RESEARCH EXPERIENCE

Some Nonparametric Tests of Randomness for Rectangular Binary Grids
ISU STAT 680 Course Project under Prof. Ranjan Maitra, ISU Aug 2021 – Dec 2021

- Hidden messages can be stored in the least-significant-bits (LSBs) of images, called *Steganography.*
- Developed two nonparametric tests based on the counts of $k_1 \times k_2$ blocks to detect the existence of any hidden structure in the LSBs.
- Assessed the performance of these tests on simulated datasets with empirical power analysis.
- Applied these tests on a set of 1800 digital camera images of different scenes over 18 devices from 5 camera models.

Parsimonious Classification using Higher Order Markov Chains
Masters Project at IIT Kanpur under Prof. Subhajit Dutta Jan 2020 – Jun 2020

- Worked on Classification problems with categorical feature variables, which can be found in Biology, Computer Science, Finance, etc.
- Formulated three different classifiers - Maximum likelihood estimates based, Conditional Probability Mixture probability model based, LASSO regularized logistic regression on counting statistics based.
- Applied the classifiers on simulated datasets to assess the performance with their error rates and time efficiency.

OTHER TERM PROJECTS

Analysis and Forecasting of Gold Price Data
Course Project for 'Time Series Analysis' under Prof. Amit Mitra, IIT Kanpur Aug 2019 – Nov 2019

- Data Preparation including handling missing values, estimation, and elimination of the trend and the seasonal component.
- Tested stationarity of the data, built a suitable model through ARIMA, and analyzed with forecasting.

Some Exploratory Issues in Statistical Computing
IASc - INSA – NASI Science Academies' Summer Research Fellowship Program 2019 May 2019 - Jul 2019

- Project Guide: Prof. Saurabh Ghosh, Human Genetics Unit, Indian Statistical Institute Kolkata
- Reviewed the efficiency of likelihood-based and k-Means clustering in the context of the Iris dataset.
- Studied the application of the EM algorithm for allele proportion estimation in blood type data.
- Applied the methodologies to analyze the dataset on the effect of physical and genetic properties on Coronary Artery Disease.

RELEVANT COURSEWORK

At Iowa State University -

- **Theory** : Measure Theory, Adv. Probability Theory, Adv. Theory of Statistical Inference.
- **Methods** : Linear Models and Design of Experiments, Adv. Statistical Modelling.
- **Other Electives** : Applied Modern Multivariate Statistical Learning, Nonparametric Methods, Adv. Statistical Computing, Mathematical Finance, Advanced Bayesian Methods.

At IIT Kanpur, India -

- **Statistics**: Probability Theory, Sampling Theory, Regression Analysis, Statistical Inference, Analysis of Variance, Time Series Analysis, Stochastic Process, Statistical Simulation and Data Analysis, Probabilistic Theory of Pattern Recognition, Multivariate Analysis, Nonparametric Inference, Bayesian Analysis, Statistical and AI Techniques in Data Mining.
- **Mathematics** : Linear Algebra, Real Analysis, Complex Analysis, Convex Optimization.
- **Computer Science** : Computer Programming and Data Structures.

SCHOLASTIC ACHIEVEMENTS

- Received *AG2PI Student Conference Travel Award 2024* to present part of doctoral research at Maize Genetics Meeting 2024 in Raleigh, NC.
- Received award for Best Narrative (Statistics) in the 2023 International Cherry Blossom Prediction Competition organized by George Mason University.
- Selected under IASc - INSA – NASI Science Academies' Summer Research Fellowship Program 2019.
- Secured All India Rank 33 in IIT Joint Admission Test for M.Sc. 2018 among 2926 candidates from all across India.
- Recipient of INSPIRE-SHE Scholarship awarded by the Department of Science and Technology, Government of India, for pursuing an undergraduate degree in sciences at a leading institute.

TECHNICAL SKILLS

Programming Languages: R, Python, Julia, C++, RCpp, SQL

Key Python Libraries: NumPy, Pandas, Scikit-learn, Statsmodels, PySpark, Matplotlib, Seaborn

Other Tools: JAGS, Stan, JMP, Minitab, LaTeX

TALKS AND POSTER PRESENTATIONS

Talks –

- Joint Statistical Meetings (JSM) 2024, Portland, OR — Aug 2024
- IMS International Conference on Statistics and Data Science (ICSDS) 2023, Lisbon, Portugal — Dec 2023
- CMStatistics 2023, Berlin, Germany — Dec 2023

Poster Presentations –

- Raymond F. Baker Plant Breeding Symposium 2024, Ames, IA — Mar 2024
- Maize Genetics Meeting 2024, Raleigh, NC — Mar 2024

EXTRA-CURRICULAR ACTIVITIES

- Member, Statistics Dept. Computational Advisory Committee, Iowa State University — Aug 2023 - Present
- Member, Statistics for Community(STATCOM), Iowa State University — Oct 2022 - May 2024
- Volunteer, National Service Scheme, India — Jul 2015 - May 2017