

SUBHRAJYOTY ROY

Research Fellow, Statistics

 [Google Scholar](#)

 roysubhra98@gmail.com

 (+91) 80139 76355

 github.com/subroy13

 Kolkata, India

 [/in/subroy13](#)

INTERESTS

- Robust and High-dimensional statistical inference.
- Exploratory Data Analysis.
- Theory of Machine Learning.
- Deep Learning.

SKILLS

Programming: R, Python, SQL, PHP, JavaScript (NestJS, NodeJS, React, NextJS).

Technologies: AWS (S3, Lambda, ECS, Sagemaker, Bedrock), Docker, MS-Office, \LaTeX

Theoretical: Statistics, Machine Learning, Mathematics, Algorithms, System Architecture Design.

EDUCATION

2021 - Current	Ph.D. in Statistics (External Research Fellow) Title of thesis: Robust Matrix Factorization using Minimum Distance Estimators: Theory and Applications. Supervisors: Ayanendranath Basu and Abhik Ghosh.	Indian Statistical Insititute, Kolkata
2019-2021	Masters of Statistics (M.Stat.) Grade: 93.8%, Specialization: Financial Statistics.	Indian Statistical Insititute, Kolkata
2016 - 2019	Bachelor of Statistics (B.Stat.) Grade: 93.6%.	Indian Statistical Insititute, Kolkata
2014 - 2016	Higher Secondary (WBCHSE) Grade: 93.4%, District Rank: 2nd.	Nimta High School, Kolkata
2014	Secondary (WBCSSE) Grade: 94.57%, State Rank: 20th.	Nimta High School, Kolkata

PUBLICATIONS

Articles

1. Roy, S., Ghosh, A. & Basu, A. (2024) Robust singular value decomposition with application to video surveillance background modelling. *Statistics & Computing*, 34(5), 178-.
2. Pyne, A., Roy, S., Ghosh, A., & Basu, A. (2024). Robust and efficient estimation in ordinal response models using the density power divergence. *Statistics*, 1-40.
3. Bhaduri, R., Roy, S., & Pal, S. K. (2022). Rough-Fuzzy CPD: a gradual change point detection algorithm. *Journal of Data, Information and Management*, 4(3-4), 243-266.
4. Ghatak, A., Singh Patel, S., Bonnerjee, S., & Roy, S. (2022). A generalized epidemiological model with dynamic and asymptomatic population. *Statistical Methods in Medical Research*, 31(11), 2137-2163.
5. Roy, S., Sengupta, D., Rudra, K., & Saha, U. S. (2020). Analysis of Pollution Patterns in Regions of Kolkata. *Calcutta Statistical Association Bulletin*, 72(2), 133-170.
6. Mukherjee, D., Dalal, A., Roy, S. (2019). Feasibility of Transparent Price Discovery in Tea through Auction in India. *Commodity Insights Yearbook (MCX)*, 44-52.

Preprints

1. Deb, S., Neves, C., & Roy, S. (2024). Nonparametric quantile regression for spatio-temporal processes. *arXiv preprint arXiv:2405.13783*.
2. Roy, S. (2024). Trustworthy Dimensionality Reduction. *arXiv preprint arXiv:2405.05868*. (Master's dissertation thesis).
3. Roy, S., Basu, A., & Ghosh, A. (2023). Robust Principal Component Analysis using Density Power Divergence. *arXiv preprint arXiv:2309.13531*. (currently in review at the *Journal of Machine Learning Research*).
4. Roy, S., Sarkar, A., Ghosh, A., & Basu, A. (2023). Breakdown Point Analysis of a General Class of Minimum Divergence Estimator. *arXiv preprint arXiv:2304.07466*. (currently in review at *Bernoulli*).
5. Bhaduri, R., Bonnerjee, S., & Roy, S. (2019). Onset detection: A new approach to QBH system. *arXiv preprint arXiv:1908.07409*.

INDUSTRY EXPERIENCE

July, 2022 - Current	Principal Information Researcher <ul style="list-style-type: none">• Continuous system improvement, research and new technology development across the company.• Implementation of several microservices powering the backend of SysCloud's application.• Implementation of chatbots, smart search navigation systems using Natural Language Processing and Generative AI models.• Building companywide automation and adoption to customized AI solutions using prompting, RAG, etc.	SysCloud Pvt. Technologies Ltd.
-------------------------	--	---------------------------------

July, 2021 - June, 2022	Data Scientist <ul style="list-style-type: none"> Developed models for optimized bidding strategy for Google Ads using reinforcement learning. Unified, scalable and robust system architecture development of backup, restore and export of any cloud, resulting in business expansion for 4 new clouds (Quickbooks, Hubspot, Salesforce and Slack). 	SysCloud Pvt. Technologies Ltd.
2020 May - June	Data Scientist Intern <ul style="list-style-type: none"> Understanding of impact of COVID-19 on Aviation industry with an analysis of historical tweets data. 	General Electric (Aviation)
2019 May - June	Data Scientist Intern <ul style="list-style-type: none"> Developed cost optimization techniques for backup and restore of G-suite and Office 365 files. Solving phishing email detection problem using combination of NLP and Deep learning. 	SysCloud Pvt. Technologies Ltd.

PROJECTS, TOOLS, PACKAGES

2023	callgrid-reader - A typescript utility package to read, parse and analyze callgrind files for profiling analysis of codes	NPM Link
2023	decompy - Python package for multiple robust matrix and tensor decomposition algorithms	PyPI Link
2022	roufcp - Python package implementing a gradual changepoint detection algorithm using rough-fuzzy set theory	PyPI Link
2021	rsvddpd - R package to perform Robust SVD using Density Power Divergence	CRAN Link
2020	COVID-19 Tracker: A Dashboard for prediction and control assistance Developed a dashboard to visualize Covid-19 situation in different states of India. It dynamically estimates the time varying reproduction rate and prediction of COVID-19 spread for various lockdown scenarios for district levels in India, based on a modified eSIR model. The project was funded by Indian Institute of Management, Vishakhapatnam.	
2019 Class Project	Analysis of Climber's challenges in scaling Mount Rainier Modelling of climber's success rate in climbing Mount Rainier based on weather modelling. Modelling of climber's team sizes for insights about mountaineering based local economy.	Report Link
2019 Class Project	An Introduction of Causal Inference using Direct Acyclic Graph Understanding of the setup and assumptions for causal inference. Estimating effect of intervention using frontdoor and backdoor criterion, simulations and applications to real world datasets to estimate causality.	Report Link

Note: I am also a contributor to the open source project [LangChain](#).

ACCOLADES

2024	AWS Cloud Practitioner Certification	Amazon Web Services (AWS)
2023	AIR 6 at National Eligibility Test (Mathematics)	Council of Scientific and Industrial Research (CSIR), Govt. of India
2022	J. K. Ghosh Memorial Gold Medal	Indian Statistical Institute, Kolkata
2020	Scaled score 324 at Graduate Record Examination (GRE - General)	Educational Testing Service (ETS)
2020	Nikhilesh Bhattacharya Memorial Gold Medal	Indian Statistical Institute, Kolkata
2020	D. Basu Memorial Gold Medal	Indian Statistical Institute, Kolkata
2019	Scaled score 890 (93 percentile) at GRE Mathematics Subject Test	Educational Testing Service (ETS)
2018	Top Quartile Listing at Simon Marais Mathematical Competition	Australian Mathematical Science Institute
2016	AIR 1 in Undergraduate Entrance Examination	Institute of Mathematics and Application, Bhubaneswar

LANGUAGES

- **English** - Professional Working proficiency
- **Hindi** - Professional Working proficiency
- **Japanese** - Elementary proficiency
- **Bengali** - Native proficiency

I certify that the information provided in this resume is true and accurate to the best of my knowledge

— Subhrajyoty Roy