SUBHRAJYOTY ROY

Nationality: Indian
(+91) 80139 76355
roysubhra98@gmail.com

Google Scholar
github.com/subroy13
/in/subroy13
Portfolio Website

CURRENT POSITIONS

External Research Fellow pursuing Ph.D. in Statistics from **Indian Statistical Institute, Kolkata**
Principal Information Researcher at **SysCloud Inc.**

RESEARCH INTERESTS

· Robust and High-dimensional statistical inference.
· Statistical divergences and minimum distance estimation.
· Spatio-temporal data analysis.
· Robust Machine Learning and Neural Networks.
· Statistical applications in finance and environmental Sciences.

SKILLS

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

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

EDUCATION

2021 - May, 2025 (Expected)	**Ph.D. in Statistics (External Research Fellow)** **Title of thesis:** Robust Matrix Factorization using the Density Power Divergence and its Applications. **Supervisors:** Ayanendranath Basu and Abhik Ghosh.	**Indian Statistical Insititute, Kolkata**
2019-2021	**Masters of Statistics (M.Stat.)** Grade: 93.8%, Financial Statistics specialization.	**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**

EXPERIENCE

July, 2022 - Current	**Principal Information Researcher** · Continuous system improvement, research and new technology development across the company. · Implementing a centralized control flow system to reduce resource overconsumption, resulting in 90% reduction in API rate-limit errors. · Development of SysCloud Sky and Scout, the AI-powered chatbots for customer and internal usages respectively. · Building companywide automation and adoption to customized AI solutions using prompting, RAG, AI agents, etc.	**SysCloud Inc.**
July, 2021 - June, 2022	**Data Scientist** · 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 a wide variety of SaaS apps (Quickbooks, Hubspot, Salesforce, Slack, Box, etc.).	**SysCloud Inc.**
2020 May - June	**Data Scientist Intern** · 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** · Developed cost optimization techniques for backup and restore of G-suite and Office 365 workspaces. · Solving phishing email detection problem using combination of NLP and Deep learning.	**SysCloud Inc.**

PUBLICATIONS

Articles

1. **Roy, S.**, Basu, A., & Ghosh, A. (2024). Robust Principal Component Analysis using Density Power Divergence. *Journal of Machine Learning Research*, 25(324), 1-40. [Link].

2. **Roy, S.**, Ghosh, A., & Basu, A. (2024). Robust singular value decomposition with application to video surveillance background modelling. *Statistics and Computing*, 34(5), 178. [Link].

3. Pyne, A., **Roy, S.**, Ghosh, A., & Basu, A. (2024). Robust and efficient estimation in ordinal response models using the density power divergence. *Statistics*, 58(3), 481-520. [Link].

4. 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), 243-266. [Link].

5. 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. [Link].

6. **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. [Link].

7. Mukherjee, D., Dalal, A., & **Roy, S.** (2019). Feasibility of Transparent Price Discovery in Tea through Auction in India. *Commodity Insights Yearbook (MCX)*, 44-52. [Link].

Preprints

1. Deb, S., Neves, C., & **Roy, S.** (2024). Nonparametric quantile regression for spatio-temporal processes. *arXiv preprint* arXiv:2405.13783. [Link].
2. **Roy, S.** (2024). Trustworthy Dimensionality Reduction. *arXiv preprint* arXiv:2405.05868. (*Master's dissertation thesis*). [Link].
3. **Roy, S.**, Sarkar, A., Ghosh, A., & Basu, A. (2023). Asymptotic Breakdown Point Analysis for a General Class of Minimum Divergence Estimators. *arXiv preprint* arXiv:2304.07466. [Link].
4. Bhaduri, R., Bonnerjee, S., & **Roy, S.** (2019). Onset detection: A new approach to QBH system. *arXiv preprint* arXiv:1908.07409. (*Undergraduate group project*). [Link].

PROJECTS, TOOLS, PACKAGES

2023	`Multi PDF QueryBoot` **- A streamlit application to chat with PDF documents using RAG and Large language models**	**Github Link**
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 the Indian Insititute of Management, Vishakhapatnam.

2019
Class Project   **Analysis of Climber's challenges in scaling Mount Rainier**   **Report Link**
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.

2019
Class Project   **An Introduction of Causal Inference using Direct Acyclic Graph**   **Report Link**
Understanding of the setup and assumptions for causal inference. Estimating the effect of intervention using frontdoor and backdoor criterion, simulations and applications to real-world datasets to estimate causality.

2018   **Chatbot using Seq2Seq Attention Mechanism**   **Github Link**
It is a chatbot with seq2seq neural network with basic attention mechanism, trained with Cornell Movie Corpus dataset.

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 Insititute, Kolkata**
2020	**Scaled score 324 at Graduate Record Examination (GRE - General)**	**Educational Testing Service (ETS)**
2020	**Nikhilesh Bhattacharya Memorial Gold Medal**	**Indian Statistical Insititute, Kolkata**
2020	**D. Basu Memorial Gold Medal**	**Indian Statistical Insititute, 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 Insititute**
2016	**AIR 1 in Undergraduate Entrance Examination**	**Insititute of Mathematics and Application, Bhubaneshwar**

ACTIVITIES

- Contributor to the open source project LangChain.

- Maintain a substack newsletter StatWizard with 200+ subscribers.

- Served as a session chair at the IMS Asia-Pacifim Rim Meeting 2024 (**Venue:** University of Melbourne, Australia).

I certify that the information provided above is true and accurate to the best of my knowledge

— *Subhrajyoty Roy*