








# SUBHRAJYOTY ROY

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

 [Google Scholar](#)  
 [github.com/subroy13](https://github.com/subroy13)  
 [/in/subroy13](#)  
 [Portfolio Website](#)

## CURRENT POSITIONS

- *Ph.D. Candidate* in Statistics from **Indian Statistical Institute, Kolkata** specializing in robust statistics and minimum divergence estimation with a focus on developing scalable matrix factorization algorithms for high-dimensional data.
- *Principal Information Researcher* at **SysCloud Inc.** with a focus on developing tools for business automation using machine learning and generative AI solutions.

## 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.

## EDUCATION

2021 - Current	<b>Ph.D. in Statistics (External Research Fellow)</b> <b>Thesis:</b> Robust Matrix Factorization using the Density Power Divergence and its Applications. <b>Supervisors:</b> Prof. Ayanendranath Basu and Prof. Abhik Ghosh.	Indian Statistical Institute, Kolkata
2019-2021	<b>Masters of Statistics (M.Stat.)</b> Grade: 93.8%, Financial Statistics specialization.	Indian Statistical Institute, Kolkata
2016 - 2019	<b>Bachelor of Statistics (B.Stat.)</b> Grade: 93.6%.	Indian Statistical Institute, Kolkata
2014 - 2016	<b>Higher Secondary (WBCHSE)</b> Grade: 93.4%, District Rank: 2nd.	Nimta High School, Kolkata
2014	<b>Secondary (WBCSSE)</b> Grade: 94.57%, State Rank: 20th.	Nimta High School, Kolkata

## EXPERIENCE

July, 2022 - Current	<b>Principal Information Researcher</b> <ul style="list-style-type: none"><li>• Continuous system improvement, research and new technology development across the company.</li><li>• Implementing a centralized control flow system to reduce resource overconsumption, resulting in 90% reduction in API rate-limit errors.</li><li>• Development of SysCloud Sky and Scout, the AI-powered chatbots for customer and internal usage.</li><li>• Building company-wide automation and adoption to customized AI solutions using prompting, RAG, AI agents, etc.</li></ul>	SysCloud Inc.
May, 2024 -July, 2024	<b>Mentor</b> <ul style="list-style-type: none"><li>• Course instructor for <i>Mathematics for Artificial Intelligence</i> module. Received a consistent rating of 4.5+.</li></ul>	Great Learning Education Series Pvt. Ltd.
July, 2021 - June, 2022	<b>Data Scientist</b> <ul style="list-style-type: none"><li>• Developed models for optimized bidding strategy for Google Ads using reinforcement learning.</li><li>• Unified, scalable and robust system architecture development of backup, restore and export of any cloud, resulting in business expansion for various SaaS apps (Quickbooks, Hubspot, Salesforce, Slack, Box, etc.).</li></ul>	SysCloud Inc.
2020 May - June	<b>Data Scientist Intern</b> <ul style="list-style-type: none"><li>• Understanding of the impact of COVID-19 on the aviation industry with an analysis of historical tweets data.</li></ul>	General Electric (Aviation)

- Developed cost optimization techniques for backup and restore of G-suite and Office 365 workspaces.
- Solving phishing email detection problem using a combination of NLP and Deep learning.

## **PUBLICATIONS**

---

### **Journal 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*. (Minor revision from "Bernoulli"). [\[Link\]](#).
4. Bhaduri, R., Bonnerjee, S., & **Roy, S.** (2019). Onset detection: A new approach to QBH system. *arXiv preprint arXiv:1908.07409*. (Undergraduate group project). [\[Link\]](#).

### **Working Projects**

1. **Roy, S.**, Deb, S., Karmakar, S. & Roy, R. (2024+). Nonparametric Regression of Spatio-temporal Data using Infinite Dimensional Covariates.
2. **Roy, S.**, Basu, S., Ghosh, A. & Basu, A. (2024+). Generalized Alpha-Beta Divergence, Associated Entropy and its Properties.
3. Jana, S., **Roy, S.**, Basu, A. & Ghosh, A. (2024+). Asymptotic Breakdown Point Analysis of Minimum Density Power Divergence Estimator under Independent Non-homogeneous Observations.

## **SEMINARS, CONFERENCES & LECTURES**

---

1. Best Paper Competition at **IISA Conference, 2024** at **Cochin University of Science and Technology**.  
*Title:* Robust Principal Component Analysis using Density Power Divergence.
2. Contributory talk at **IMS Asia Pacific Rim Meeting, 2024** at **University of Melbourne**.  
*Title:* A novel and scalable background modelling algorithm from video surveillance data.
3. Invited guest lecture for Robust Statistics course (M.Stat.'22) at **Indian Statistical Institute, Kolkata**.  
*Title:* A review of Robust Location and Scatter estimators.
4. Prasanta Chandra Mahalanobis memorial lecture at **Indian Statistical Institute, Kolkata**.  
*Title:* Algorithmic Fairness of Statistical Decision Systems.

5. D. Basu memorial lecture at **Indian Statistical Institute, Kolkata**.  
*Title:* tSNE: A way to visualize multivariate data.

## PROJECTS

---

2023	<b>Multi PDF Querybot</b> It is a streamlit application to chat with PDF documents using RAG and large language models. It uses a locally hosted Chroma Vector database, SQLite database for metadata storage, and Anthropic Claude model for access to foundational LLM.	<a href="#">Github Link</a>
2020	<b>COVID-19 Tracker: A Dashboard for prediction and control assistance</b> Developed a dashboard to visualize the 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	<b>Analysis of Climber's challenges in scaling Mount Rainier</b> 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.	<a href="#">Report Link</a>
2019 Class Project	<b>An Introduction of Causal Inference using Direct Acyclic Graph</b> 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.	<a href="#">Report Link</a>
2018	<b>Chatbot using Seq2Seq Attention Mechanism</b> It is a chatbot created with a seq2seq neural network with a basic attention mechanism, trained with the Cornell Movie Corpus dataset.	<a href="#">Github Link</a>

## TOOLS AND PACKAGES

---

2024	<b>Langchain - Contributor</b> Largest framework for developing applications powered by large language models	<a href="#">Github Commits</a>
2023	<b>Multi PDF QueryBoot - Developer &amp; Maintainer</b> A streamlit application to chat with PDF documents using RAG and Large language models	<a href="#">Github Link</a>
2023	<b>callgrid-reader - Developer &amp; Maintainer</b> A typescript utility package to read, parse and analyze callgrind files for profiling analysis of codes	<a href="#">NPM Link</a>
2023	<b>decompy - Developer &amp; Maintainer</b> Python package for multiple robust matrix and tensor decomposition algorithms	<a href="#">PyPI Link</a>
2022	<b>roufcp - Developer &amp; Maintainer</b> Python package implementing a gradual changepoint detection algorithm using rough-fuzzy set theory	<a href="#">PyPI Link</a>
2021	<b>rsvddpd - Developer &amp; Maintainer</b> R package to perform Robust SVD using Density Power Divergence	<a href="#">CRAN Link</a>

## ACCOLADES

---

2024	<b>Shortlisted for Best Paper Competition</b>	IISA Conference 2024
2024	<b>AWS Cloud Practitioner Certification</b>	Amazon Web Services (AWS)
2023	<b>AIR 6 at National Eligibility Test</b>	Council of Scientific and Industrial Research (CSIR), Govt. of India
2022	<b>J. K. Ghosh Memorial Gold Medal</b>	Indian Statistical Institute, Kolkata
2020	<b>Scaled score 324 at GRE General Test</b>	Educational Testing Service (ETS)
2020	<b>Nikhilesh Bhattacharya Memorial Gold Medal</b>	Indian Statistical Insititute, Kolkata
2020	<b>D. Basu Memorial Gold Medal</b>	Indian Statistical Insititute, Kolkata
2019	<b>93 percentile at GRE Mathematics Subject Test</b>	Educational Testing Service (ETS)
2018	<b>1st Quartile at Simon Marais Mathematical Competition</b>	Australian Mathematical Science Insititute
2016	<b>AIR 1 in Undergraduate Entrance Exam</b>	Insititute of Mathematics and Application, Bhubaneswar

## SKILLS

---

### Programming Technologies Tools

R, Python, SQL, PHP, JavaScript (NestJS, NodeJS, React, NextJS).  
AWS (S3, Lambda, ECS, Sagemaker, Bedrock), Docker, Langchain, Langflow.  
MS-Office,  $\text{\LaTeX}$

## ACTIVITIES

---

- Maintain a substack newsletter [StatWizard](#) with 200+ subscribers.
- Served as a session chair at the IMS Asia-Pacific Rim Meeting 2024 (**Venue:** University of Melbourne, Australia).

## REFERENCES

---

- **Ayaendranath Basu.** Professor, Indian Statistical Institute, Kolkata. [ayanbasu@isical.ac.in](mailto:ayanbasu@isical.ac.in)
- **Abhik Ghosh.** Associate Professor, Indian Statistical Institute, Kolkata. [abhik.ghosh@isical.ac.in](mailto:abhik.ghosh@isical.ac.in)
- **Soudeep Deb.** Associate Professor, Indian Institute of Management, Bangalore. [soudeep@iimb.ac.in](mailto:soudeep@iimb.ac.in)
- **Vijay Krishna.** CEO, SysCloud Inc. [vijay@syscloudsoftware.com](mailto:vijay@syscloudsoftware.com)
- **Diganta Mukherjee.** Professor, Indian Statistical Institute, Kolkata. [diganta@isical.ac.in](mailto:diganta@isical.ac.in)

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

— Subhrajyoty Roy