SUBHRAJYOTY ROY

Statistician, Data Scientist,
and Solution Architect Researcher

🌐 statwizard.in ✉ roysubhra98@gmail.com
📞 (+91) 80139 76355 ⬡ github.com/subroy13
📍 Kolkata, India 🔗 /in/subroy13

SUMMARY

I am a Solution Architect Researcher with a strong background in Statistics, Machine Learning, System Design and Algorithms development. My speciality lies in transforming business problems into abstraction, comprehending them, building innovative solutions, building robust and scalable systems and validating them with Proof-of-Concepts.

SKILLS

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

Technologies and Tools: 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)**	**Indian Statistical Insititute, Kolkata**

Broad area of thesis: Robust Matrix Factorization and High Dimensional Statistical Inference.
Supervisors: Prof. Ayanendranath Basu and Dr. Abhik Ghosh.

2019-2021	**Masters of Statistics (M.Stat.)**	**Indian Statistical Insititute, Kolkata**

Grade: 93.8%, Specialization: Financial Statistics.

2016 - 2019	**Bachelor of Statistics (B.Stat.)**	**Indian Statistical Insititute, Kolkata**

Grade: 93.6%.

2014 - 2016	**Higher Secondary (WBCHSE)**	**Nimta High School, Kolkata**

Grade: 93.4%, District Rank: 2nd.

2014	**Secondary (WBCSSE)**	**Nimta High School, Kolkata**

Grade: 94.57%, State Rank: 20th.

EXPERIENCE

July, 2022 - Current	**Principal Information Researcher**	**SysCloud Pvt. Technologies Ltd.**

- Continuous system improvement, research and new technology development for rapid automation across the company.
- Implementation of several microservices powering the backend of SysCloud's application.
- Implementation of smart search navigation system using Natural Language Processing.
- Research about interesting use-cases of Large Language Models and integration of artificial intelligent systems into the business.

July, 2021 - June, 2022	**Data Scientist**	**SysCloud Pvt. Technologies Ltd.**

- 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).

2020 May - June	**Data Scientist Intern**	**General Electric (Aviation)**

- Understanding of impact of COVID-19 on Aviation industry with an analysis of historical tweets data using Natural Language Processing.

2019 May - June	**Data Scientist Intern**	**SysCloud Pvt. Technologies Ltd.**

- Developed a new architecture for G-suite and Office 365 backup and restore with cost optimization and performance efficiency in mind.
- Solving phishing email detection problem using combination of Natural Language Processing and Deep learning.

ACCOLADES

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 |

PUBLICATIONS

1. Roy, S., Basu, A., & Ghosh, A. (2023). Robust Principal Component Analysis using Density Power Divergence. *arXiv preprint arXiv:2309.13531.*

2. Roy, S., Ghosh, A., & Basu, A. (2023). Analysis of the rSVDdpd Algorithm: A Robust Singular Value Decomposition Method using Density Power Divergence. *arXiv preprint arXiv:2307.10591.*

3. Roy, S., Sarkar, A., Ghosh, A., & Basu, A. (2023). Breakdown Point Analysis of the Minimum S-Divergence Estimator. *arXiv preprint arXiv:2304.07466.*

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(3-4), 243-266.

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.

6. Pyne, A., Roy, S., Ghosh, A., & Basu, A. (2022). Robust and efficient estimation in ordinal response models using the density power divergence. *arXiv preprint arXiv:2208.14011.*

7. Roy, S., Basu, A., & Ghosh, A. (2021). A New Robust Scalable Singular Value Decomposition Algorithm for Video Surveillance Background Modelling. *arXiv preprint arXiv:2109.10680.*

8. 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.

9. Dalal, A., Mukherjee, D., & Roy, S. (2020). The Information Content of Taster's Valuation in Tea Auctions of India. *arXiv preprint arXiv:2005.02814.*

10. Bhaduri, R., Bonnerjee, S., & Roy, S. (2019). Onset detection: A new approach to QBH system. *arXiv preprint arXiv:1908.07409.*

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` **- A Python package implementing several robust matrix and tensor decomposition algorithms** **PyPI Link**
2022	`roufcp` **- A Python package implementing rough-fuzzy set theory based gradual changepoint detection algorithm** **PyPI Link**
2021	`rsvddpd` **- An R package to perform a robust Singular Value Decomposition 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 Insititute of Management, Vishakhapatnam.
2019 Class Project	**Analysis of Climber's challengens 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 effect of interven- tion using frontdoor and backdoor criterion, simulations and applications to real world datasets to estimate causality.

LANGUAGES

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