

# Sohom Ghosh

✉sohom1ghosh@gmail.com | ☎+91 8001734384 | [Website](#) | [LinkedIn](#) | [Google Scholar](#) | [GitHub](#)

## SUMMARY

---

- Data Scientist (Professional Experience: 4.5+ years)** responsible for improving the digital lives & financial well-being of millions of users, Delivered **15+ impactful projects**
- Co-authored: **2 books** on **Natural Language Processing (NLP)**, **5+ papers** at proceedings of **International Conferences, Workshops & Refereed Journals (IEEE, Springer, ACM)**
- M.Tech (BITS, Pilani)**, Completed **9+ MOOCs & Certifications**, Received **5+ Awards/Recognitions**
- Research Interests: **Applications of Natural Language Processing in the FinTech Industry**

## WORK EXPERIENCES

---

- |                     |   |
|---------------------|---|
| JUN 2019 - PRESENT  | <div><b>Data Scientist at <a href="#">FIDELITY INVESTMENTS</a>, Bengaluru, India</b><br/><i>Artificial Intelligence, Center of Excellence</i><ul style="list-style-type: none"><li>•Developed a direct answer algorithm to enhance search experience of users</li><li>•Built models to predict intents from profiles &amp; web-activities of users</li><li>•Created a re-usable Python library for summarizing call transcripts</li><li>•Publications: 2 conference/workshop papers, <b>1 book</b>. Awards: 3</li></ul></div>   |
| JAN 2017 - JUN 2019 | <div><b>Data Scientist at <a href="#">TIMES INTERNET</a>, Noida, India</b><br/><i>Central Analytics Team</i><ul style="list-style-type: none"><li>•Developed a recommendation system to suggest relevant skills for <a href="#">TimesJobs</a></li><li>•Presented insights for impacting business decisions by analyzing interest graphs &amp; behaviours of 450+ million monthly visitors across 39+ digital products (<a href="#">Gaana</a>, <a href="#">Times of India</a>, <a href="#">Economic Times</a>, <a href="#">MX Player</a>, <a href="#">CricBuzz</a> etc.) [Data size: &gt;1TB]</li><li>•Built a predictive model using PySpark to target emails &amp; deployed it on Hadoop cluster thereby increasing their open rate</li><li>•Other projects: Sales Analytics (Upsell, Cross-sell, Churn Model development), Digital Product Analytics, B2B Cross-walk Analytics, Fraud Analytics in affiliate marketing (saved 17,000+ USD), Data Engineering, Deployed 3 models in production</li><li>•Publications: <b>1 journal paper</b>, <b>1 book</b>, <b>1 article</b>. Awards: 1</li></ul></div> |
| JUN 2015 - JAN 2017 | <div><b>Intern, Corporate Trainer (part-time), Analyst (Intern) &amp; Freelancer</b><br/><b><a href="#">Indian Statistical Institute</a>, <a href="#">Novel R&amp;D</a>, <a href="#">MathLogic</a> &amp; <a href="#">Upwork</a></b><ul style="list-style-type: none"><li>•Worked on Data Mining &amp; Predictive Analytics</li><li>•Trained working professionals on Text Analytics, Machine Learning &amp; Big Data</li><li>•Developed &amp; maintained Amazon Web Services cloud infrastructure while hosting Spark Hackathon for one of the world's largest corporate firm. Prepared dashboards &amp; reports</li></ul></div>  |

## EDUCATION

---

- |      |   |
|------|---|
| 2019 | <div><b>M.Tech in SOFTWARE SYSTEMS (Specialization: Data Analytics)</b><br/><b><a href="#">BITS, Pilani</a>, India (Work Integrated Learning Programme)</b><br/>Thesis: Is it a click bait? Let's predict using Machine Learning<br/>GPA: 9.31/10</div> |
| 2016 | <div><b>B.Tech in COMPUTER SCIENCE &amp; ENGINEERING</b><br/><b><a href="#">Heritage Institute of Technology, Kolkata</a>, India</b><br/>Thesis: Solving Real Life Problems using Machine Learning Techniques<br/>GPA: 8.22/10</div>                    |

## SELECTED PUBLICATIONS & ACHIEVEMENTS

---

### Books

- R Chopra, A M Godbole, N Sadvilkar, M B Shah, S Ghosh, D Gunning, [The Natural Language Processing Workshop](#), Packt Publishing, ISBN: 9781800208421 [Aug 2020] [\[code\]](#)
- S Ghosh, D Gunning, [Natural Language Processing Fundamentals](#), Packt Publishing, ISBN: 9781789954043 [Mar 2019] [\[code\]](#)

### Journals

- S Ghosh, “Identifying click baits using various machine learning and deep learning techniques”, in [International Journal of Information Technology](#), Springer, ISSN : 2511-2104 (Print), ISSN: 2511-2112 (Online), DOI 10.1007/s41870-020-00473-1, May 2020 [\[link\]](#) [\[ppt\]](#)
- A Mitra, S Ghosh, P Basuchowdhuri, M K Shekhawat and S K Saha, “Recommendation system based on product purchase analysis”, in [Innovations in System and Software Engineering \(A NASA journal\)](#), Springer London, Vol. 12, Issue 3, Sep 2016, pp 177–192 ISSN: 1614-5046 (Print) 1614-5054 (Online) [\[link\]](#)

### Conferences & Workshops

- S Ghosh and A Chopra, “Using Transformer based Ensemble Learning to Classify Scientific Articles”, in proceedings of [Workshop & Shared Task on Scope Detection of the Peer Review Articles \[SDPRA-2021\]](#) (Collocated with PAKDD-2021) (Accepted for publication by Springer LNAI) [\[code\]](#)
- A Chopra, S Agrawal and S Ghosh, “Applying Transfer Learning for Improving Domain-Specific Search Experience Using Query to Question Similarity”, in proceedings of [International Conference on Algorithms, Computing and Artificial Intelligence \[ACAI-2020\]](#), Sanya, China (Accepted for publication by ACM ICPS) [\[pre-print\]](#)

\* For more refer to the [Google Scholar Profile](#)

### Achievements

- [Fidelity Investments](#): Excellence in Action (2 times) & You’ve Earned It (Team Awards), GS & PI (Peer to Peer Recognition) [2020, 2021]
- [IndiaHikes](#): Green Getter [Jan 2021]
- Hackathons: [Kaggle](#): 1 Bronze (Competition) & 2 Bronze (Discussion), [TechGig](#): CodeGladiators-2018 Finalists, [Analytics Vidhya](#): 4 times in top 25
- [Times Internet](#) (Times Business Solutions): Rock Star Award [Mar 2018]

### TECHNICAL SKILLS

---

- Tools & Technologies: Python, SQL, Cloud (AWS)
- Libraries & Frameworks: Keras, Scikit-learn, Pandas, Numpy, XG-Boost, LightGBM, Flask, Spark, Hadoop
- Methods & Algorithms: Regression (Linear/Logistic), Decision Trees, Random Forest, Gradient Boosting Machine, XGBoost, Clustering, PCA, Neural Nets & Deep Learning
- Others:  $\text{\LaTeX}$ , MS Office (Word, Excel, Powerpoint)

- Expertise: **Natural Language Processing**
- Domains: **FinTech (Financial Services), Consumer Internet based Digital Products**

## SELECTED CERTIFICATIONS & COURSES

---

- Business English Level **8, 9 & 10 Learnship** [Nov 2020, Jan Feb 2021]
- **Introduction to Research**, NPTEL, Score: 87%, Topper (Elite + Silver) [Jan 2021] [\[link\]](#)
- **Natural Language Processing Specialization**, Coursera, Score>90% [Sep 2020] [\[link\]](#)
- **Machine Learning for Data Science and Analytics**, edX, Score: 81% [Mar 2016] [\[link\]](#)
- **Social & Economic Networks: Model & Analysis**, Coursera, Score:84% [Mar 2014] [\[link\]](#)
- Summer Training on Big Data Analytics, **IIT- Kharagpur** [Jun - Jul 2014]
- Knowledge Dissemination Program on Data Mining, **IIT- Kharagpur** [Apr 2015]
- Participating in various technical sessions (Cloud - AWS, Machine Learning Engineering, Computer Vision etc.), domain training (Mutual Funds, Stocks etc.) & soft-skill development workshops (Creating Effective Presentations, Communicating with Impact etc.) being organized by **Fidelity Investments** [2019 - Present]
- Participated in various technical (Machine Learning, Python etc.) & soft-skill training sessions organized by **Times Internet** [2017 - 2019]
- **Relevant (Post)/Graduate Level Courses:** Artificial Intelligence, Machine Learning, Statistics, Information Retrieval, Advanced Data Mining, Engineering Mathematics (Probability, Linear Algebra, Calculus), Discrete Mathematics (Graph Theory, Mathematical Logic, Combinatorics), Database Management Systems, Programming, Data Structures, Algorithms, Operations Research (Game Theory. Optimization etc.)

## PERSONAL DETAILS

---

- **Permanent Address:** Nabarun Nagar, opposite Rakhal Pirtala, Police Line Bazar, District: Purba Bardhaman, PO: Sripally, Burdwan - 713103, West Bengal, India
- **Languages:** English, Bengali & Hindi
- **Hobbies:** Playing Harmonica, Reading Books & Trekking