

# Sohom Ghosh

✉ [sohom1ghos@gmail.com](mailto:sohom1ghos@gmail.com), ☎ +91 8001734384, [Website](#), [LinkedIn](#), [Google Scholar](#), [GitHub](#)

## SUMMARY

---

- Passionate **Data Scientist** responsible for improving the digital experience of millions of users and empowering business to take **data driven** decisions using **Artificial Intelligence**
- Co-author of 2 **books** on Natural Language Processing (NLP) and 4+ **papers** at Proceedings of International Conferences and Refereed Journals
- Interests: **Industrial Applications of Machine Learning and Deep Learning based NLP**

## WORK EXPERIENCES

---

JUN 2019 - PRESENT	<p>Senior Analyst - Data Science at <a href="#">FIDELITY INVESTMENTS</a>, Bengaluru, India <i>Artificial Intelligence, Center of Excellence</i></p> <p>Developed Direct Answer Algorithm to enhance search experience of users Built models to predict intents from profiles and web-activities of users Created a re-usable Python library for summarizing call transcripts Publications: 1 conference paper, 1 book</p>
JAN 2017 - JUN 2019	<p>Data Scientist at <a href="#">TIMES INTERNET</a>, Noida, India <i>TimesJobs, Central Growth and Analytics</i></p> <p>Developed a Recommendation System to suggest relevant skills for TimesJobs Presented insights for impacting business decisions by analyzing interest graphs &amp; behaviours of 450M+ 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] Built a predictive model using PySpark to target emails &amp; deployed it on Hadoop cluster Other projects: Sales Analytics (Upsell, Cross-sell, Churn Model), Digital Product Analytics, B2B Cross-walk Analytics, Fraud Analytics in affiliate marketing, Chatbot prototype development, Data Engineering, Deploying ML models in production Publications: 1 journal paper, 1 book</p>

## EDUCATION

---

2019	<p>M.Tech in SOFTWARE SYSTEMS (Specialization: Data Analytics) <a href="#">BITS, Pilani</a>, India Thesis: Is it a click bait? Let's predict using Machine Learning GPA: 9.31/10</p>
2016	<p>B.Tech in COMPUTER SCIENCE AND ENGINEERING <a href="#">Heritage Institute of Technology, Kolkata</a>, India Project Topics: Natural Language Processing, Data Mining, Predictive Analytics GPA: 8.22/10</p>
2012	<p>Senior Secondary Education: Central Board of Secondary Education (CBSE) PERCENTAGE: 80.80</p>
2010	<p>Secondary Education: Indian Certificate of Secondary Education (ICSE) PERCENTAGE: 91.57</p>

## SELECTED AWARDS AND PUBLICATIONS

---

### Achievements

- **Fidelity Investments**: Received Excellence in Action & You've Earned It (Team Awards), GS & PI (Peer to Peer Recognition) [2020]
- Hackathons: **Kaggle**: Bronze-1, **TechGig**: CodeGladiators Finalists [2018]
- **Times Internet** (Times Business Solutions): Received **Rock Star Award** for increasing click-through rate of emails using Machine Learning [Mar 2018]
- **GATE**: Computer Science & Information Technology 2016 (**96.97 percentile**), **Nation-wide Interactive Maths Olympiad 2008** (All India Rank : 337)
- Received the **Best Presentation Award** for presenting the short paper titled "*A Model of Opinion Mining to Compute Score from Curriculum Vitae*" at 22<sup>nd</sup> **West Bengal State Science and Technology Congress 2015** [28<sup>th</sup> Feb - 1<sup>st</sup> Mar 2015].

### 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) [\[link\]](#)

### Conference Proceedings

- A Chopra, S Agrawal and **S Ghosh**, "*Applying Transfer Learning for Improving Domain-Specific Search Experience Using Query to Question Similarity*", in proceedings of **3<sup>rd</sup> International Conference on Algorithms, Computing and Artificial Intelligence [ACAI-2020]**, Sanya, China (Accepted for publication by ACM ICPS)
- **S Ghosh**, A Mitra, P Basuchowdhuri and S K Saha, "*Analysis of Online Product Purchase and Predicting Items for Co-purchase*", in proceedings of **3<sup>rd</sup> International Conference on Advanced Computing, Networking and Informatics [ICACNI-2015]** Vol-1, Smart Innovation, Systems and Technologies Vol 43, pp 581-591, Springer Verlag [ISBN: 978-81-322-2538-6] [\[link\]](#) [\[ppt\]](#)
- S Gurung, **S Ghosh**, J Jaiswal and P Basuchowdhuri, "*Extraction and Analysis of Publication Data of Conferences*", in proceedings of **2<sup>nd</sup> IEEE International Conference on Advances in Computing and Communication Engineering [ICACCE-2015]**, pp 588-593, ISBN: 978-1-4799-1734-1/15 [\[link\]](#) [\[ppt\]](#)

## CERTIFICATIONS AND TRAINING

---

- Completed **Natural Language Processing Specialization** (collection of 4 courses) offered by deeplearning.ai over Coursera with >90% score in each course [Sep 2020] [\[link\]](#)
- Completed online course on **Machine Learning for Data Science and Analytics** offered by **Columbia University** over edX (Honor Code) with 81% [4<sup>th</sup> Mar 2016] [\[link\]](#)
- Completed online course on **Social & Economic Networks: Model & Analysis** offered by **Stanford University** over Coursera with 84% score [5<sup>th</sup> Jan to 2<sup>nd</sup> Mar 2014]
- Completed **International Summer Training on Big Data Analytics**, at **IIT- Kharagpur**, Softwares used: - Matlab, Hadoop, Map Reduce, R, MS Excel [30<sup>th</sup> Jun - 11<sup>th</sup> Jul 2014]
- Participating in various **technical sessions** (Cloud - AWS, Computer Vision etc.), **domain training** (Mutual Funds, Stocks etc.) and **soft-skill development workshops** (Creating Effective Presentations, Communicating with Impact etc.) being offered at **Fidelity Investments** [2019 - Present]
- Participated in various **technical** (Machine Learning, Python etc.) and **soft-skill training** sessions organized by **Times Internet** [2017 - 2019]
- Participated in **Knowledge Dissemination Program on Data Mining** offered by **IIT-Kharagpur** [18<sup>th</sup> - 23<sup>th</sup> April 2015]

## TECHNICAL SKILLS

---

- **Data Science & Analytics:**
  - Primary Technologies: Python, Spark, Hadoop, SQL, AWS, Excel
  - Other Technologies: Keras, Flask,  $\text{\LaTeX}$
  - Techniques: Regression, Random Forest, XGBoost, Clustering, Neural Nets & Deep Learning
  - Expertise: Natural Language Processing (NLP)
  - Domains: Financial Services, Consumer Internet based Digital Products
- **Relevant Courses completed:** 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 and Hindi
- **Hobbies:** Playing Harmonica, Reading Books and Trekking