

Sohom Ghosh

Data Scientist (Exp: 3+ yrs, NLP, Machine Learning, Big Data), **M.Tech** (BITS, Pilani), **Kaggler & Researcher**

Mobile:	+91 8001734384
Email:	sohomlghosh@gmail.com
GitHub:	github.com/sohomghosh
LinkedIn:	in.linkedin.com/in/sohomghosh
Website:	sites.google.com/view/sohomghosh

SUMMARY

- **Passionate Data Detective** with **hands-on (industrial) experience** in solving **practical problems** using **data-driven approaches**
- **Publications** in various international conferences & journals (Springer, IEEE), **Co-Author** of the book NLP Fundamentals (Packt Publishing) [\[link\]](#)
- Receiver of **Rockstar Award** from Times Internet (TBS) for increasing click-through rate of emails by strategically targeting them using machine learning
- **Certifications:** Deep Learning with Python (Udemy), Social and Economic Networks (Coursera), Machine Learning for Data Science and Analytics (edX), Data Mining & Big Data Analytics (IIT-Kgp)

TECHNICAL SKILLS

PRIMARY TECHNOLOGIES	Python, SQL, AWS
OTHER TECHNOLOGIES	Keras, Spark, Hadoop, Excel
TECHNIQUES	Regression, Random Forest, Neural Net, XGBoost, Clustering, Deep Learning
EXPERTISE	Natural Language Processing (NLP)
DOMAINS	Internet, Consulting, Financial Services

WORK EXPERIENCE

JUNE 2019 – Present

Fidelity Investments, Bangalore [\[link\]](#)
Senior Analyst - Data Sciences

- **Project 1:** Working at the **Artificial Intelligence Center of Excellence** for improving users' search experience during Personal Investments. Developed **Machine Learning based models** to predict **intent**s from users' profiles & web-activities.
- **Project 2:** Building reusable libraries for simplifying various steps of Data Science projects (like Text cleaning, Memory reduction etc.).

JANUARY 2017 – JUNE 2019

Times Internet (Times of India Group), Noida [\[link\]](#)
Data Scientist (Manager-1)

- **Project 1:** Presenting **insights** for impacting business decisions by analyzing **interest graphs & be-**

haviours of **450M+** monthly visitors across **39+ digital products** (**Gaana, Times of India, Economic Times, MX Player, CricBuzz** etc.) [Data size: >1TB]
• **Project 2: Recommendation System** to suggest relevant skills during a job search on TimesJobs
• **Project 3:** Built predictive model using **PySpark** to **target emails**. Deployed it on **Hadoop cluster**. Outcome: **increased open rate** of emails by **5 %**
• **Project 4:** Developing **deep learning** based models to detect articles which tend to be **click-baits**
• **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**
⇒ Tools used: **Python, Spark, Hive, Sqoop, Hadoop, SQL, Flask, MongoDB, Tableau, AWS**

APRIL 2016 – JANUARY 2017

FN MathLogic Consulting Services, Gurgaon [\[link\]](#)
Analyst (Intern, Technical Consulting)

- Provided assistance on Machine Learning, R, Spark & Text Analytics to various clients (MNCs) and prepared dashboards for them using Excel & ShinyR
- Worked on NLP as a part-time intern (Apr - Jun'16)

RESEARCH PUBLICATIONS

- ⇒ **#Journals: 2, #Conferences: 3, #Book: 1**
- **Recommendation System based on Product Purchase Analysis**
[ISSE, Springer London] (NASA journal) [\[link\]](#)
[ICACNI, Springer] (International conference) [\[link\]](#)
 - Analyzed the Amazon Co-purchase Network
 - Examined dynamic buying patterns & developed algorithms to recommend products
 - **Extraction & Analysis of Publication Data of Conferences** *[IEEE ICACCE-2015, pg 588-593]* [\[link\]](#)
 - **Natural Language Processing Fundamentals** *[Packt Publishing] (Book Chapters 1-4)* [\[link\]](#)

INDEPENDENT PROJECTS

- **Kaggle:** GA Customer Revenue Prediction [\[link\]](#), Microsoft Malware Prediction (Top 15%) [\[link\]](#)
- **Hackathon Profiles:** Kaggle [\[link\]](#), Analytics Vidhya - Best rank: 95, CodeGladiators'18 Finalists
* (Other projects are showcased on GitHub) [\[link\]](#)

EDUCATION

- | | |
|------|--|
| 2019 | M.Tech (Software Systems) [WILP] ,
Specialization: Data Analytics
BITS Pilani, 9.31/10 |
| 2016 | B.Tech (Computer Science & Engg.)
Heritage Institute of Technology, 8.22/10 |