

Experience

- The Goldman Sachs Group, Inc.** Bangalore, India
Associate, SRE and Alloy Streaming - ETL platform Mar. 2021 – Present
 - Captured timeliness and completeness metrics using prometheus on legacy systems including job scheduler, database queues, ETL platforms, etc.
 - Integrated the retry logic for data pipelines by merging extract, transform and load jobs into a single process and eliminated the need for any staging store in-between the jobs.
 - Supported the sourcing of data from rest sources involving complex authentication mechanisms.
Analyst, DocKnight - self service document digitization platform June. 2018 – Feb. 2021
 - Open-sourced the core docknight library which is used to generate document model of pdf documents. The document model deconstructs a document into pages, groups (a visual paragraph within a page), lines, and segments (similarly-styled text within a line).
 - Trained Random Forest Classifier to extract net asset value from 6000+ capital statements every quarter. Reduced the risk of an incorrect valuation getting approved manually. Lessened manual effort equivalent to one headcount.
- Amazon.com Inc.** Delhi, India
Intern, ATSD Invoice PDF Generation Jul. 2017 – Aug. 2017
 - Modeled the system to generate PDF documents from XSL-FO files by utilizing the technologies Apache FOP and Apache Velocity.
 - Constructed the generic invoice template in XSL-FO for Amazon ATS dedicated carriers.
Intern, Bidding Re-Architecture June. 2017 – Jul. 2017
 - Reduced the latency of transportation bidding services by replacing calls to DynamoDB APIs instead of SQLPlus. Performed end-to-end testing of the whole system in a beta environment.

Publications

- Harsh Gupta, Luciano Del Corro, Samuel Broscheit, Johannes Hoffart, Eliot Brenner; **Unsupervised Multi-View Post-OCR Error Correction With Language Models**; Empirical Methods in Natural Language Processing (EMNLP); November, 2021

Education

Qualification	Institute	Score	Duration
B.Tech	IIT Mandi, Himachal Pradesh	8.8/10	Aug 2014 - June 2018
Class XII	Greenway Modern School, Delhi	92.2%	Apr 2012 - Mar 2013
Class X	Greenway Modern School, Delhi	9.8/10	Apr 2010 - Mar 2011

Awards and Honors

- Represented IIT Mandi at the ACM ICPC India Regionals in 2015, 2016, and 2017, with the best rank being 7 out of 160 at the Kolkata/Kanpur Regionals, 2017.
- Ranked 2nd of 400 teams from all the IITs, NITs, BITs, and IISC, in GoldmanSachs Quantify, 2017.
- JEE Advanced 2014 Rank 3096 (out of 1.26 lakh), JEE Main 2014 Rank 7564 (13.56 lakh).

Activities

- Recruited interns from top-tier colleges by conducting in-person interviews.
- Instructed competitive programming skills to first-year students to develop coding environment of the IIT Mandi.
- Volunteered as Teaching Assistant for the course on Data Structures and Algorithms.
- Mentored two teams in Goldman Sachs NIC Datathon, 2020.

Skills

- Languages:* JAVA, Python, C++, C.
- Libraries:* STL, Pandas, Numpy, Spacy
- Web Design:* React, Redux, Django Framework.
- Others:* SQL, Git, Machine Learning, REST.