

SKILLS

- *Programming*: Python (Pandas, Seaborn, Pytorch, Spacy, NLTK, Scikit-learn, Keras), R, SQL, PySpark, Docker, Java, JS, Tableau
- *Enterprise* - AWS, Linux, Streamlit, Hugging Face, Open-AI Codex, GitHub, Neo4j, MS Office Suite, Bloomberg, Jira, Confluence.
- *Databases* - Globaldata, Pitchbook, SEC Filings, Reactome, Human Protein Atlas, PubMed, clinicaltrials.gov & US Patent -REST
- Interests* - Generative AI, AI Ethics, research in NLP & genetics, diversity, and inclusion, deal-flow-automation
- Certifications* - Google Certified, Certified Investment Management Analyst®, Security Industry Essentials (FINRA)
- Publications* - SmFe1-yCoyO3 perovskite-type oxides for soot oxidation, Reducing Green House Effect Caused by Soot via Oxidation

EXPERIENCE (Selected)**Data Scientist / Algorithm Engineer**

January 2022 - November 2022

Pivotal Life Sciences, San Francisco, CA

- Developed automated NLP-based graph knowledge management to combine databases like Reactome, PubMed, OpenTargets, and DrugBank via a decoder-based sentence-transformer-based model to connect public and private companies with clinical trials, research, IP, and funding rounds, resulting in \$2M in savings then migrated legacy database to AWS S3/RDS data lake.
- Collaborated with partners on improving analysis of 100K PubMed abstracts (used - SageMaker and LLM), over 750 human diseases, and 47k genes, molecules, and therapeutics - providing faster (3 months to 3 days) & accurate (+12.6% accuracy) insights.
- Led the due diligence pipeline development to evaluate deals in collaboration with partners, principals, and associates. The pipeline optimized the data retrieval process resulting in 20x efficiency—mentored a team of four members and two clients.
- Performed due diligence on investments at the intersection of biotech, AI, ML, and RWE platforms deals over \$15M with a gross 1.6x MOIC for the fund and conducted hands-on technical sessions to onboard the stakeholders to the data pipeline.
- During the market downturn of 2022, combined Pitchbook, Internal CRM data, and \$500M funds of funds strategy portfolio using NLP-based similarity detection to find overlapping company names. The insights led to a classification model that found companies with cash higher than enterprise value saving over \$300K in consulting, three months, and securing three key deals.

Chief Data Engineer

September 2021- January 2022

ToughLeaf, New York, NY

- Led development of the construction bid display platform's backend by designing data pipelines and dynamic data models with AWS and NLP solutions to convert complex technical jargon into simplified hashtags.
- Improved database quality by detecting and removing duplicate entries, developed a TF-IDF ensemble with bi and trigram model (NLP) that identifies crucial terms, filtered by decision tree analysis to extract & convert 10138 technical public data entries.
- Collaborated with cross-functional teams to draft RFP for Empire State - Minority-Owned Business initiative to secure over \$6M of grant funding and a three-year contract to work with the state to develop and scale toughleaf.com.

Lead Data Scientist

December 2020 - September 2021

Columbia University Irving Medical Center (NY Presbyterian Children's Hospital), New York, NY

- Authored publication of an innovative clinical NLP-focused recommendation algorithm by collaborating with a social worker, two nurses, and a pediatrician, resulting in 2x cases reported to NY Administrative Childcare Services.
- Supervised, mentored, and trained new data scientists to improve the classification model to detect child abuse and neglect cases, risk factor analysis (ICD 9/10) for drug order notes, demographics, financial data, and triage nurse notes.
- Designed a classification detection model to alert clinicians of at-risk child patients by segmenting historically overlooked children using risk factor and pattern recognition; the model scanned & labeled over 33961 medical records detecting child maltreatment.
- Assisted in building ETL pipeline utilizing Epic & PySpark in mining over 106GB of HIPAA-protected unstructured data, restricted text with cleaning and formatting techniques, pattern recognition, and simplified data engineering pipeline with NLP.

Tutor and Associate at Columbia University

- Assisted faculty during the pandemic with information and technology support for 'Courseworks' (online course infrastructure), video broadcasting of lectures, and scheduling hybrid classes for inter-university-wide assignments.
- Tutored twenty-one student-athletes in physics, economics, statistical calculus, thermodynamics and chemistry, and biochemistry, taught underprivileged 9th and 10th grader women about climate change and AI, & taught incarcerated adults about AI and ML.

Process Engineer

July 2016 - April 2020

York Winery, Sula Winery & Delta Finochem, India

- Reduced operational costs of the winemaking process by \$38K by building MATLAB-based simulation of bioreactors. The simulation also helped conserve wine color, quality, and taste, simultaneously reducing the batches.
- Assisted in increasing the plant's output that manufactured active pharmaceutical ingredients (APIs - e.g., fenofibrate), fertilizers, and ammonia by increasing the recycling rate and decreasing process time by 33% per day, ultimately saving \$20K of logistic costs.

EDUCATION* **Executive Education Scholar, CIMA® Program****University of Chicago Booth School of Business, Chicago, IL*** **Master of Science in Chemical Engineering, Concentration in Data Sciences****Columbia University, New York, NY**• **Courses**: Advanced Mathematical Modeling, Natural Language Processing, Genomic Technologies, Corporate Finance and LBOs• **Honors**: Columbia Business School Build Lab Fellow, University of Illinois Sustainability Fellow (SISE)* **Bachelor of Technology in Chemical Engineering, Minor in Business Management****Manipal University, Manipal, India**

• Dean's award for best technical research in environmental engineering. Multiple prizes for winning national chess tournaments