SARTHAK ARORA

sarthakvarora@gmail.com

LinkedIn | GitHub | Twitter

EXPERIENCE

ML ENGINEER | *PREDLI AB, STOCKHOLM* **Aug '23 – Present**
- Using OpenAI, ChromaDB and Langchain to generate a retriever-based **DBQA** system for a Climate consultancy client.
- Creating a Quiz Generator from Transcripts with Topic Clustering and Evaluation using SerpAPI.
- Fine-tuning LLM Performance on a Tweet generation system using a **RLHF** model based on Tweet engagement.

ML ENGINEER | *GANGES LAB, COLLABORATIVE EARTH* **Jan '23 – Present**
- Using Machine Learning and **GEE** to classify and cluster different "Naalas" for the Ganges river from Satellite Images.
- Conducted **fieldwork** around the Ganges River which involved scoping the area of research and meeting stakeholders.
- Conducting **research** for a publication around "Naala" Identification, Taxonomy and Remediation.

RESEARCH SCIENTIST | *UNIVERSITY OF CALIFORNIA, BERKELEY* **Sept '22 – Present**
- Conducted a time-series **Anomaly Detection** analysis of Power Plant Emissions in San Diego to find areas of **Climate policy** improvements.
- Analysing Regressors for a **Wildfire Survey** conducted by MWPA to cater to the needs of people in Marin County.

DATA SCIENTIST | INDIFI TECHNOLOGIES PVT. LTD., GURGAON **Aug '22 – Jul '23**
- Created a model which predicts Default Risk with **96% precision** by analyzing CIBIL Trend using **LSTM** Model.
- Used **Fuzzy Logic** to match addresses for verification, leading to a reduction in manual checks by **~4 hours** daily.

RESEARCH SCHOLAR | *UNIVERSITY OF CALIFORNIA, BERKELEY* **Apr '22 – Jul '22**
- Conducting Research to develop a **wildfire risk assessment** tool under the mentorship of Dr. Thomas Azwell.
- Working on **Remote Sensing** and **GEE** datasets for the state of California to predict wildfires with an accuracy of **82%**.
- Gaining valuable insights in applying Data Preparation, Machine Learning, Model **Deployment** and GIS mapping.

ML ENGINEER | *AI FOR GOOD FOUNDATION, REMOTE* **Feb '22 – Present**
- Working on the **Climate SDG** project, which is backed by **Microsoft** and the **UN** Development Program.
- Creating a **Graph** based Visualisation for Climate Entities using **Named Entity Recognition** and **Neo4J**.
- Cleaning Scraped data using **LDA Model** for topic modeling, leading to a Recall value of **83.5% in noise detection**.

SENIOR EXECUTIVE ANALYST | *DISNEY+ HOTSTAR, MUMBAI* **Feb '21 – Aug '21**
- Reported daily figures for projects such as IPL, Burbank for Star and Disney. Reducing TAT by ~**3** hours in the process.
- Designed a Health Score metric to predict user churn by modeling variables, leading to a rise in conversion by **1.5 pp**.
- Analyzed campaigns over engagement, affluence, affinity on Conversion and CTR%, leading to **0.5 pp** rise in renewals.

EDUCATION

Post-graduate Program	Plaksha Technology Leaders Program	3.88/4	2022
B.Sc. (H) Statistics	Ramjas College, Delhi University	7.18/10	2018
Class XII	Delhi Public School, Kanpur	95.25%	2015

PROJECTS
- Delivered a talk at renowned Machine Learning Conference **ICML 2022**.
- Co-wrote a Literature Survey paper on **Sarcasm Detection** in **NLP**.
- Co-founded an app that uses AI/ML to model physiological indicators and provide actionable **Mental Health** insights.
- Created a Machine Learning model for **Drug Discovery** using **GANs**.
- Created an Image Classification model using **transfer learning** on **VGG-16**.
- Predicted future order dates and SKUs purchased to create a **Recommendation Engine** for an e-commerce platform.
- Calculated the likelihood of defaulters by predicting **credit scores** for different individuals.
- Created a real-time **face and features detector** on live videos using **HAAR and CNN** for Computer Vision.

SKILLS
- **Programming Languages:** Python, R, SQL, Shell Script, Cipher
- **Machine Learning:** TensorFlow, PyTorch, Scikit-Learn, Pandas, NumPy, Matplotlib, Seaborn, NLP, CV, RL, Graphs
- **Others:** A/B Testing, Algorithms, AWS, Beautiful Soup, CSS, Data Structures, ETL, Excel, HTML, GIS, Google Analytics, Google Earth Engine, Product Management, Prototyping, Selenium, Spark, Tableau

ACHIEVEMENTS
- Received a **Best Paper award** at **FAccT 2023** for the paper Queer In AI: A Case Study in Community-Led Participatory AI
- **Silver Medalist** for academic excellence at Plaksha University, TLP Batch of 2022.
- Winning Team for the Plaksha Entrepreneurial **Challenge Lab** where we pitched a Mental Health startup.
- In the top **5%** out of 14k teams on Kaggle for the Titanic Project and received **first** position in class.

LEADERSHIP POSITIONS
- Core-Organizer and Program Chair for leading Machine Learning Conference **NeurIPS 2022** as a part of QueerInAI.
- **Taught** subjects such as ML, Data Science, Research, and Leadership to students based out of Africa with **Belong**.

PUBLICATIONS

- Queer in AI - ACM'22
  Hetvi Jethwani, Arjun Subramonian, William Agnew, MaryLena Bleile, **Sarthak Arora**, Maria Ryskina, Jeffrey Xiong
  pdf | abstract | cite
- Queer In AI: A Case Study in Community-Led Participatory AI - FAccT'23 *Won best paper
  Organizers of Queer in AI
  pdf | abstract