Rachit Sabharwal's CV

- Phone: +1 585 281 1928
- Email: rachit-sabharwal@outlook.com
- Location: 10211 Camden Garden Lane, Katy, Texas, 77494
- Website: www.rachitsabharwal.me
- GitHub: rachitest

Education

The University of California, Berkeley, Certificate in Software Development and Programming

- June 2023 – present
- Berkeley, CA

The University of Texas Health Science Center at Houston, Doctor of Philosophy in Biostatistics

- Aug 2022 – present
- Houston, TX
- Advanced Certificate in Data Science

The University of Texas Health Science Center at Houston, Master of Science in Biostatistics

- Jan 2020 – May 2022
- Houston, TX
- Thesis — BioRec: A Biomedical Recommendation System for Academic Conferences and Journals
- Certificate in Data Science

Experience

The University of Texas Health Science Center at Houston, School of Public Health, Biostatistics and Data Science - Graduate Research Assistant (Doctoral)

- Sept 2022 – Oct 2024
- Houston, TX
- Establish good DevOps practices, architect repository structures, author comprehensive documentation, and implement CI/CD pipelines to streamline development and testing processes
- Apply a mix of classical statistical models and advanced machine learning techniques for comprehensive data analysis on varied topics including

— vaccine non-response, serostatus, pediatric comorbidities arising from Covid-19, etc.
- Develop a weekly report generation pipeline incorporating CI/CD, ensuring seamless integration of new data and automated report updates
- Plan and enact a test suite for both code and data spanning the entire data engineering and data science life-cycle from data ingestion to report generation

The University of Texas Health Science Center at Houston, School of Public Health, Biostatistics and Data Science - Graduate Research Assistant (Master's)

- Feb 2020 – Aug 2022
- Houston, TX
- Maintained Covid-19 dashboards using Python and Tableau for the Texas Covid-19 Dashboard Project
- Created, deployed, and maintained accessible and responsive web apps on multiple platforms (Streamlit and Heroku) for academic conference recommender systems
- Built and serviced high content ETL pipelines using Python, R, and Cron to feed recommender systems and Covid-19 dashboards

Genentech, Research and Early Development, Development Sciences & Informatics - Informatics Intern

- May 2021 – Jan 2022
- San Francisco, CA
- Used Deep Transfer Learning via PyTorch and Raytune to create a neural network to predict adverse events for drugs
- Created a Knowledge Graph with Neo4j and a Graph Neural Network using NetworkX and PyTorch to generate gene expression signature-likes for drugs
- Developed a framework for the tokenization of internal documents for ingestion into text-mining application
- Compared work flow management/ETL frameworks (Airflow vs. Prefect vs. Luigi) for use with all NLP pipelines and presented the results

Procter & Gamble, Consumer & Market Knowledge - Advanced Analytics Co-Op

- Jan 2021 – May 2021

- Cincinnati, OH

- Created predictive models, analytics, and visualizations that facilitated a deep understanding of consumer and shopper behaviors

- Used parallel computing (Dask and Modin) to develop both predictive and explanatory models enabling insights into market trends and retailer behavior

- Upgraded teams nascent DevOps by implementing unit-testing via Pytest and Great Expectations, containerization via Poetry and Docker/Kubernetes, and agile via Jira and Confluence ## Bristol Myers Squibb, Data Engineering Intern

- June 2020 – Aug 2020

- San Francisco, CA

- Utilized Python, HTML, CSS, and Javascript in creating a multifeatured patent recommendation app to significantly improve scientists' workflow

- Designed and maintained both relational and graph databases in PostgreSQL and Neo4j

- Conducted literature reviews on research concerning Recommendation Systems, Information Retrieval Systems, and BERT and BERT variations (BioBERT, SciBERT, etc.)

Publications

Factors associated with elevated SARS-CoV-2 immune response in children and adolescents (10.3389/fped.2024.1393321)

- Aug 2024
- Messiah SE, Abbas R, Bergqvist E, Swartz MD, Talebi Y, **Sabharwal R**, Han H, Valerio-Shewmaker MA, DeSantis SM, Yaseen A, Gandhi HA, Amavisca XF, Ross JA, Padilla LN, Gonzalez MO, Wu L, Silberman MA, Lakey D, Shuford JA, Pont SJ, Boerwinkle E
- Frontiers in Pediatrics

Baseline characteristics of SARS-CoV-2 vaccine non-responders in a large population-based sample (10.1371/journal.pone.0303420)

- May 2024
- Yaseen A, DeSantis SM, **Sabharwal R**, Talebi Y, Swartz MD, Zhang S, Leon Novelo L, Pinzon-Gomez CL, Messiah SE, Valerio-Shewmaker M, Kohl HW 3rd, Ross J, Lakey D, Shuford JA, Pont SJ, Boerwinkle E
- PLoS One

An Interactive Online Dashboard with COVID-19 Trends and Data Analysis in Northeast and South Texas (https://research.ebsco.com/linkprocessor/plink?id=894625e1-7146-30bf-aa2c-9f5637dac41e)

- Apr 2024
- Zhang Z, **Sabharwal R**, Lee M, Zhang K, McGaha P, Crum M, Bauer C, Fisher-Hoch SP, McCormick JB, Reininger BM, Thomas S, Guajardo E, Pinon D, Yaseen A
- Texas Public Health Journal

Long-term immune response to SARS-CoV-2 infection and vaccination in children and adolescents (10.1038/s41390-023-02857-y)

- Oct 2023
- Messiah SE, Talebi Y, Swartz MD, **Sabharwal R**, Han H, Bergqvist E, Kohl HW 3rd, Valerio-Shewmaker M, DeSantis SM, Yaseen A, Kelder SH, Ross J, Padilla LN, Gonzalez MO, Wu L, Lakey D, Shuford JA, Pont SJ, Boerwinkle E
- Pediatric Research

Scholarly recommendation systems: a literature survey (10.1007/s10115-023-01901-x)

- June 2023
- Zhang Z, Patra BG, Yaseen A, Zhu J, **Sabharwal R**, Roberts K, Cao T, Wu H
- Knowledge and Information Systems

Data Cleaning for eCommerce: Standardizing Data Handling Practices for eCommerce Datasets

- May 2021
- **Sabharwal R**
- Procter & Gamble, Internal White Paper

Skills

- Languages: English (Native/Bilingual), Hindi (Native/Bilingual), French (Intermediate)

- Work Authorization: US Citizenship, Canadian Citizenship # Technical Skills

- Machine Learning: Scikit-learn, TidyModels, Raytune, Optuna, Pytorch, Tensorflow, Huggingface, JAX

- Programming Languages: Python, R, SAS, MATLAB, Javascript, C, Java, HTML, CSS

- Databases: RDBMS (PostgreSQL, SQLite, MySQL), NoSQL DBMS (MongoDB, Elasticsearch, Neo4J), BigQuery

- Cloud and Distributed Computing: AWS (AWS HPC), GCP, Azure, Spark, Hadoop, Slurm, On-Prem HPC

- DevOps: Git, GitHub, GitLab, Docker, GitHub/GitLab CI/CD, Jenkins, Kubernetes, Jira, Confluence

- Workflow Orchestration: Airflow, Prefect, Cron, Luigi

- Frameworks and Platforms: Streamlit, FastAPI, Django, Flask, Heroku, Replit, Great Expectations, PyTest

- Tooling: VSCode, RStudio, Quarto, Jupyter, PyCharm, CLion, IntelliJ IDEA, Confluence, Slack, Tableau, Power BI, Stata, DBeaver

- Operating Systems: Windows, Linux (Ubuntu, and Mint), MacOS

- General Computing: Microsoft Office, Google Workspace