Sahil Paresh Poonatar
Citizen of the United States of America
I bridge the gap between cluttered numbers and useful, effective knowledge

Merck email: sahil.paresh.poonatar@merck.com
Personal email: sahilpoonatar24@gmail.com
LinkedIn: https://www.linkedin.com/in/sahilpoonatar
Contact No: +91 8668911633

Education

Degree	Institute	GPA	Year
MSc. Computer Science and ML Specialization	Georgia Institute of Technology (Georgia Tech)	3.20/4	2021-
B.Tech Biotechnology	National Institute of Technology, Warangal (NITW)	7.74/10	2017-2021

Technical Skills

- Languages: Fluent: Python,SQL Comfortable: C++, JavaScript, HTML/CSS
- Tools: AWS RedShift, Excel, SQL Workbench, JupyterHub, RWDEx, Dataiku, Docker, Bash, Latex, Git
- Skills: Big Data Analytics, Machine Learning, Data Science, Data Visualization, Data Storytelling, Agent Based Modeling
- Basic : Artificial Intelligence, Cloud Computing, Web Development

Experience

- **Merck, Sharp and Dohme,** Pune Data Science Hub — **Payer Value and Access Team, CAS Data Science Team**
  Associate Data Scientist Intern — *June 2023 - present*
  - **ICER CE Assessment:** Project to understand if there is any association between ICER's Cost Effectiveness Assessment and US Commercial Health Plan Drugs Coverage and Management in preparation for Sotatercept's ICER Review. Ran multiple analyses on Commercial filled and rejection claim rates, Commercial plans formulary statuses and cost effectiveness proxies. Ran Random Forest, Logistic Regression and XGBoost models, taking into consideration multiple other variables to find the features most closely tied to coverage restrictiveness, obtaining an average F1 Score of 75%. No clear association was found, with annual treatment cost and specialty product status being the most significant features. Data sources used were ICER, IQVIA eLAAD, DRG Fingertip, Precision Hierarchy Bridge, Navlin and IPD Pricing.
  - **Market Access Landscape for LAG-3/mCRC:** Aided in the creation of the Pipeline LAG-3 therapy for mCRC Asset Access Strategy Development Framework by creating and adding content to the slide deck on Market Access Landscape, Early Economic Assessment, and pricing scenarios. Data source used was Clarivate DRG
  - **Competitive Pricing for GMAx:** Automated standardization between products' dosage strength variables from four different data sources to help compare prices between them. The automated code provided an average of 97% standardization conversion across data sources and reduced time spent on standardizing from days (previously had been done manually) to mere milliseconds. Data sources used were SSR, IQVIA, Analysource and POLI.

- **AISSMS Institute of Management**, Pune, India — **Business Analytics and Data Visualization**
  Guest Lecturer — *June 2023*
  - Took a semester's worth of classes (33 hours), introducing second year MBA students to Business Analytics, Data Visualization and Natural Language Processing. Familiarized students, wanting to have a specialization Data Analytics, to programming languages, softwares and tools such as Python, SQL, Jupyter Notebook, Tableau and Pandas, NumPy and NLTK libraries
  - Supervised and graded assignments and capstone project presentation, contributing to the success of numerous students who secured job offers as business analysts.

- **Zytex India**, Mumbai, India — **Probiotics Research Team**
  Data Analyst Intern — *May - July 2019 and February - May 2023*
  - Initially joined as a microbiology lab intern and worked on the Analysis of Bacillus Species from Soil for Probiotic Production. Introduced Google OpenRefine to the team to streamline data storage and wrangling process of 5000 bacteria species' data
  - Started and spearheaded the Bacillus Probiotics Data Science Team by conducting numerous classification machine learning techniques, such as Random Forest, Neural Networks, Boosting, Support Vector Machines and K-Nearest Neighbor on data containing the results of hemolytic, sporulation, and extracellular polymeric substance experiments on Bacillus isolate samples, to gauge whether a sample would be useful for the company's manufacturing processes, obtaining 92% accuracy.
  - Ran Clustering and Dimensionality Reduction algorithms to gain better insight into the potential relationships and patterns within the data that led to better understanding of bacteria selection for probiotics production.

- **Department of Biotechnology**, National Institute of Technology, Warangal — **Prof. Asim Bikas Das Ph.D.**
  Data Scientist Intern — January - April 2021
  - Worked on predicting Breast Cancer from family tree data
  - Used genome data from healthy people and breast cancer patients to identify genes more commonly susceptible to mutation, thus found in breast cancer patients
  - Utilized Support Vector Machines and Random Forest for classification

Projects

- **Green Analytics Interactive Tool - Tableau dashboard**

Built a global scale, interactive platform that provides insight on the relationships between various socio-economic factors and deforestation. Is meant to be used by Non-Profit organizations, government agencies and independent researchers to drive policy decisions. Used classification and clustering algorithms and Holt-Winter forecasting to provide greater insight into the relationships between deforestation and features like GDP, GNI, Inflation and Unemployment.

- **Breast CancerCare Hub - Tableau dashboard**

A dynamic hub for exploring Breast Cancer, acting as a one-stop-shop for all information on this cancer with a clear emphasis on user-friendliness and user-centric design, meant specifically for users not familiar with the disease to gain knowledge of its incidence, prevalence, and survival rates