GOPALIKA SHARMA

Boston, MA | (617) 223-7988 | [Personal Email](#) | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

OBJECTIVE

Seasoned project lead Data Scientist seeking a Data Science Manager role to leverage extensive experience in Generative AI, LLM, Machine learning, Statistical modeling, with a proven track record of successfully leading diverse projects and driving strategic initiatives.

WORK EXPERIENCE

Sanofi Pharmaceuticals, MA

✦ Technical Project Manager   January 2024 - Present

Data Speciality: Real World Data; e.g. CRM Text, Medical Inquiries
- Spearheading the development of a GPT-4 RAG engine to generate relevant insights for KOLs, enhancing product and marketing strategies.
- Managed the products life cycle in a GxP compliant agile manner
- Collaborate with cross-functional teams to define project scope, objectives, and deliverables.

Technical Skills: Atlassian Suite, Agile Methodologies, RAG, GPT-4, Pinecone, Snowflake, and more.

Key Achievement: Improved accuracy of AI-generated responses to specific queries using Generative AI.

✦ Lead Data Scientist   April 2023 - January 2024

Data Speciality: Real World Data; e.g. Raw text, Electronic Health Record, Insurance Claims and Patient Data
- Built an analytics platform in PowerBI to derive real world evidence and hypothesis from the characteristics and treatment journeys of patients in a broad cohort using the LightGBM model
- Built a ML model using G-estimation and SHAP to compare performance of biologics to determine efficacy of a treatment
- Created a standard for digital release process of data products, in adherence with quality and compliance team
- Spearheaded the establishment of OpenShift deployment for our web based products
- Lead a team of Junior data scientists and analysts, fostering a collaborative and innovative environment

Technical Skills: Argo WorkFlow, Github CI/CD, Docker, OpenShift, PowerBI, Snowflake, Atlassian Suite

Key Achievement: The analytical platform generates hypotheses which is used to build studies and conduct experiments which otherwise would have taken months

✦ Data Scientist   January 2022 - April 2023

Data Speciality: Real World Data; i.e. CRM free text, Social Media Data
- Developed an automated compliance engine for CRM free text using Natural Language Processing AI models.
- Created a reusable machine learning component in Python for anonymization, summarization, language detection and translation
- Facilitated the development of a COVID-19 social listening dashboard to analyze public sentiments about the vaccine through tweets.

Technical Skills: Python, Atlassian Suite, BERT, Github CI/CD, AWS, FastText, Spacy NER, and more.

Key Achievement: The compliance engine has reduced the assessment time required by Medical Governance Officers by 82%

Corbus Pharmaceuticals, MA   January 2021 - January 2022
✦ Data Scientist
Data Speciality: Text, Drug Trial Data
- Developed a custom streamlit application identifying various target areas as potential investment opportunities using ensemble learning
- Built a text analysis-NLP tool in Python to analyze research papers for better lead target identification

Technical Skills: Ensemble Learning, Streamlit, Python, Snowflake, and more.

Key Achievement: The predictions tool developed in Streamlit using Python reduced analysis time by 35%

Novartis, MA   January 2020- June 2020
✦ Data Scientist (Co-op)
Data Speciality: Clinical Trial Data, Drug Data
- Developed a machine learning model in Python using XGBoost to predict the feasibility of various pharmaceutical compounds for preclinical trials

Technical Skills: Numpy, Pandas, Sklearn, Seaborn, Matplotlib, R Studio, Jupyter Notebooks and more.

Key Achievement: The drug discovery team utilized the model to identify the best compounds for preclinical trials, thereby saving the company significant money and resources.