SUJIT TILAKRAJ THAKUR

Position: Data Scientist   **Contact**: +1-8064012452   **Portfolio** - https://sujit-data.github.io/Sujit-Portfolio/   **Email**: Sujittilakrajthakur@gmail.com
Projects: https://github.com/Sujit-Data?tab=repositories   **LinkedIn**: https://www.linkedin.com/in/sujit-thakur-ssbb-6b9692116/

Skills

Programming and Software: Python | SQL | R | Microsoft Power BI | Tableau | Big Query | Hadoop | AWS | Oracle | Git | GitHub | Spark | Airflow | NoSQL | Databricks | JavaScript | REST API | Spring boot | Docker | Computer Vision | AWS Sage maker | Streamlit | Snowflake | DataBricks | Dash | Angular

Forte: Machine Learning | Data Science | Statistical Analysis | Design of Experiment | Reinforcement learning | Supervised Learning | Unsupervised Learning | Data Analysis | Data visualization | Transformer Pipeline | Data mining | cross-functional team | Communication Mastery | Data Modelling | Project Management | Large Language Model | Natural language Processing | Classification | Regression | Clustering

Professional Experience

Data Scientist - Biogen Inc.   (December 2022 – Present)
- Developed supervised and unsupervised machine learning models for clinical contracts and drug discovery with over 90% accuracy
- Led R&D initiative to transform business, medical, and scientific questions into analytical ones and map out solutions, delivery, and impact.
- Spearheaded hypothesis-driven exploratory analyses, chose suitable ML algorithms, and developed sophisticated optimization engines to deliver impactful data solutions.
- Restructured data for clinical trial and bioscience data models, leading to a 6% improvement in on-time delivery of Clinical Contracts, and spearheaded initiatives resulting in a 40% efficiency gain within R&D Governance through Automation and Analytics projects, querying large data using SQL and large-scale distributed system like Oracle, Snowflake and Databricks.
- Demonstrated strong communication skills and leadership skills in international projects, optimizing vendor selection, creating Power BI and Tableau dashboards to communicate with both technical and non-technical audience, and contributing to $10 million in savings for Clinical Services by redesigning data models. Performed data task as full stack data scientist at biopharmaceutical company.

Jr. Data Scientist – Tech Mahindra ltd.   (February 2020 – July 2021)
- Elevated product analytics by 15% using statistical methods like Bayesian and ANOVA and hypothesis testing.
- Conceptualized and developed advanced machine learning models for R&D team and Product management team to extract significant business insights from data and identify the roots behind the patterns for imaging analytics.
- Spearheaded use of image prediction machine learning model for product failure.
- Directed and coordinated development and deployment of large language machine learning models.
- Developed Machine Learning models for predicting product failure and target metrics. Mentored my team of analysts on KPI development.
- Achieved a 25% improvement in operational efficiency, optimized product data and analysis by 30%, and reduced big data processing time by 40% through initiatives such as data quality frameworks and SQL optimizations. Documented the outcomes with utmost precision.

Graduate Assistant - Texas Tech University.   (August 2021 – May 2022)
- Developed machine learning models for PhD students to aid in their research and mentored undergraduate students in Engineering Economics.

Internships

Data Analyst Intern - Biogen Inc.   (July 2022 – December 2022)
- Implemented robust data solutions, leveraging Power BI, Python, and creating Tableau visualizations for cross-functional teams.
- Spearheaded a data pipeline project, resulting in a 25% reduction in engineers' time and enhanced operational efficiency.

Trainee – Analytics and Field Service – Mahindra and Mahindra ltd.   (February 2019 – February 2020)
- Developed Savings Dashboard for Field service Gujarat Team using Microsoft Power BI and Angular.
- Introduced a CI/CD pipeline for improved process efficiency.

Education

Masters (Postgraduate in Data Science) (Machine Learning and Artificial Intelligence) - University of Texas at Austin (Aug 2023 – Apr 2024), GPA – 4.33

Master of Science in Industrial Engineering - Texas Tech University (Aug 2021 – Dec 2022), GPA 4.0

Bachelor of Engineering in Mechanical Engineering - University of Mumbai (July 2014 – Jan 2019), GPA 3.7

Recent Honors and Projects

Projects: Fraud Detection, Sentiment Analysis NLP, HR Attrition Rate Project, German Credit Risk, e-commerce marketing analysis, Bank Loan Churn Prediction rate, Electric utility tariff project. **Honors and Awards**: Hackathon – Machine Learning Winner 2024, Best Data Analyst and Data Scientist Award 2023, Spot Award for Best Data Scientist 2021.