Sahil Dhingra

New Jersey(317-559-6121) | sahilnits@gmail.com | LinkedIn| GitHub

SUMMARY

Dedicated and accomplished professional with 14 years of distinguished experience in the pharmaceutical industry, specializing in Clinical, Preclinical, and Manufacturing Quality domains. Proven track record of ensuring regulatory compliance and quality through expert product management and data analytics. Recently graduated with a degree in Data Science, equipped with knowledge in Machine Learning, Natural Language Processing (NLP), and deep learning. A highly adaptable candidate seeking a challenging role in Data Science, Data Analytics, or Data Engineering. Excels in integrating domain-specific knowledge with cutting-edge data science techniques to derive actionable insights and drive informed decision-making.

EDUCATION

1. 2023 | Indiana University Bloomington | MS in Data Science | 4.0/4.0 GPA
2. 2009 | N.I.T Silchar | BTech in Computer Science & Engineering | 3.5/4.0 GPA

TECHNICAL SKILLS

Languages: Python, PL-SQL, SQL, Shell-scripting.

Libraries: Pandas, NumPy, Scikit-Learn, TensorFlow, Spacy, NLTK, Keras, Matplotlib, Seaborn, Plotly.

Databases and Visualization Tools: PostgreSQL, MS SQL Server, SQLite, NoSQL, Tableau. Looker

Concepts: Statistics, Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, A/B Testing.

Statistical Tests: ANOVA, t-Test, chi-squared test, Kruskal-Wallis Test.

Licenses & certifications: Generative AI for Everyone, Generative Deep Learning (TensorFlow), CNN, Supervised Machine Learning: Regression and Classification, OCI 2019 Cloud Operations Certified Associate.

Tools: Visio, Informatica MDM, Informatica PowerCenter, Informatica Data quality, Workload Automation, Latex, Aginity Workbench, Denodo, Quality Docs, Eclinical, Clintrial, eCTS, Impact, TrackWise, Streamlit, Git.

WORK EXPERIENCE

TrackWise (Manufacturing Quality) | Branchburg, NJ, USA   JAN 2022 - Current

- Technical Project Lead for Enterprise Quality management TrackWise systems for CAPA, Audit management, Complaint management, Supplier, Deviations.
- Responsible for design and development for data integration pipelines into and from TrackWise system.
- Led the transition and further development of TrackWise system interfaces from a different vendor to TCS.

X-LIMS (Pre-Clinical) | Indianapolis, IN, USA   MAR 2019 - DEC 2021

- Led the implementation of a custom cloud-based LIMS application, strategically replacing Waters LIMS.
- Key features include Sample Management, Complete Storage Management, Integrated Instrument Management, Dynamic Access Management, Multi-vendor/Lab Support, Sample Assay, PRF Results, Data Analytics, and Dynamic Reporting ensuring 21 CFR Part 11 compliance for regulatory adherence.
- Significantly reduced PRF form creation turnaround time and provided support for multiple labs.
- Enabled sample management for Covid-19 test samples.

Data Integration Platform (Clinical R&D) | Indianapolis, IN, USA   AUG 2013 - FEB 2019

- Developed master data management strategy and data architecture, focusing on a data delivery platform with dimensional models for Risk Monitoring.
- Implemented Master Data Management for site, investigators, and compounds.
- Orchestrated near-real-time data integration, covering eDC Results, Lab Results, and Connected Devices.
- Enabled data management functionalities such as Visit & Subject reconciliation.
- Developed custom system for de-identifying and scrambled sensitive data from diverse source systems (Clintrial, Eclinical, Impact, eCTS, and GLS, CLRM, Denodo) which facilitated comprehensive testing of the entire data warehouse with a large and robust dataset while adhering to FDA guidelines and ensuring HIPAA compliance.
- Conducted meticulous review and testing of study protocol documents, proposing strategic amendments.

Clintrial Integrations (Clinical R&D) | Delhi, India   DEC 2009 - AUG 2013

- Enhancement, development, and maintenance of critical applications, including Eclinical, Clintrial, Clintrial Integration Solution, Client Safety System Integration, GLS Integration, MedDRA Integration, WHO drug Integration, Impact and Standard Code Integration, Clintrial Extract to Climate, Integrated Review, Inform, and Grants Integration.

- Implemented Six Sigma methodologies for the investigator payment module, achieving an annual saving of $1.6 million by eliminating overpayments.

Data Science:

O'Neill School of Public and Env. Affairs (https://bradrfulton.com/collaborators/ras/**)**   01/2023–05/2023

NLP Engineer and Data Scientist | Discerning Organizational Identities
- Fine-tuned BERT Zero-Shot for Text Classification on the philanthropic data (Form-990 fillings) of foundation grantee network of the entire US non-profit sector to develop a custom trained model.
- Utilized K-Means clustering in scikit-learn to label a large unlabeled dataset to prepare for classification, supervised learning model training.
- Integrated Open API for efficient and experimental religion classification in religious organization data. Additionally, pioneered NLP Tree-of-Thought Prompt Engineering techniques for Large Language Models (LLMs) to perform binary and multi-class classification, providing generative recommendations within US non-profit sector Form-990 filings.

Data Engineering and Visualization | Foundation and Grantees Network
- Created a robust architecture with data cleaning and processing routines for technical and non-technical access on Form 990 data.
- Designed and created intuitive Tableau Dashboards for Grants Network, and respective profiles of Grantees and Foundations. This was crucial for Sponsor to get further funding for the project.

Projects:

Deep Learning: Playlist Generation
- Small Language Modelling to generate playlist from song sequence. Worked on various Deep Learning Models like LSTM (Uni and Bi-directional_ with n-gram), Self-Attention and Cross-Attention based model with N-dimensional and positional embedding.

NLP Topic Modelling - Amazon Reviews
- Implemented topic modeling with LDA to extract top 10 latent themes from a corpus of Amazon reviews and hosted the application on Streamlit.

Machine Learning: Home Credit Default Risk
- Employed advanced predictive models like Logistic Regression, Random Forest, XGBoost, and Neural Network to predict if a person will be able to repay the loan.
- Conducted a comprehensive comparison of multiple machine learning models using ANOVA.

Data Visualization K-12 Shooting Data Visualization
- Exploring the pervasive issue of injuries from gunfire at schools, a leading cause of fatalities globally and within the US, we conducted a comprehensive analysis through a series of visualizations.