SHUBHAM SHAILESH TAMHANE

shubhamtamhane2000@gmail.com | LinkedIn | Github | Portfolio | +1 585-363-9693

Education

University of Rochester   **Aug 2022 – Dec 2023**
Master of Science in Data Science – GPA: 3.96/4   *Rochester, NY*

University of Mumbai   **Aug 2018 – May 2022**
Bachelor of Engineering in Information Technology – GPA: 3.73/4   *Mumbai, India*
Relevant Courses: Statistics, Data Mining, Time Series, NLP, Machine Learning, AI, Data Structures, Big Data, DBMS

Experience

Data Analyst   **Feb 2024 – Jul 2024**
Indiana University, School of Public Health   *Bloomington, IN*
- Incorporated **Microsoft SQL Server** in combination with **Python** for in-depth analysis of healthcare datasets, exceeding **5M** records, to derive insights from Electronic Health Record (**EHR**) data
- Leveraged **T-SQL** to clean and preprocess extensive healthcare datasets, ensuring data integrity for a study on patient outcomes, which was conducted in compliance with **HIPAA** regulations
- Performed comprehensive **survival analysis** to assess disease diagnosis over 10 years of clinical data, integrating demographic and treatment variables to provide detailed reports
- Developed polynomial regression models in **R**, analyzing **25 years** of previous birth rates and disease spread while factoring in seasonality and other pertinent social variables in USA
- Conducted a **meta-analysis** of over 25 studies, including results from **A/B testing**, and visualized the findings using forest plots to summarize review results

Data Science Intern   **Jun 2023 – Dec 2023**
Regeneron Pharmaceuticals   *Tarrytown, NY*
- Orchestrated ETL pipelines in **Dataiku** for time series forecasting on a quarterly interval for a inventory management system stored in **Redshift** with **30,000+** products
- Employed a python-dash application invoking **MLOps** workflow to provide real-time estimations, customer analysis and model maintenance options to end users and stakeholders
- Implemented **ARIMA** and **ETS** family of models in JMP for their clear interpretability, achieving **80%** forecasting accuracy, thereby reducing material waste by **30%** each quarter
- Utilized **Seeq** to design interactive dashboards and alarms by fetching data from **OSI PI**, thus optimizing workflows and reducing task turnaround time for end users by **20%**
- Investigated predictive maintenance techniques to estimate Remaining Useful Life (**RUL**) of Mass Flow Controllers

Data Analyst   **Sep 2023 – Dec 2023**
Zalliant   *Amsterdam, NY*
- Led a team of **3** analysts to systematically code and analyze video data of cow activities using **Excel** and **Boris**, focusing on identifying multiple behavioral patterns within the same footage
- Designed a **Tableau** dashboard to display **key performance indicators (KPIs)** on cow behavior using data from **Azure Blob**, reducing decision-making time with real-time insights
- Engineered optimization strategies resulting in a **99.6%** reduction in sensor data collection frequency, thereby extending sensor battery life by **3** months

Software Development Engineer   **Sep 2022 – Jun 2023**
URMC - Center for Advanced Brain Imaging and Neurophysiology   *Rochester, NY*
- Instituted a verification system using pydicom to confirm over **10,000 DICOM** files within **60** seconds, while utilizing **MongoDB** to store metadata for data management and retrieval.
- Built a Flask web service with JWT authentication and caching, reducing JSON file load times by 91.6%
- Deployed the microservice using **Docker containers** for isolated execution, utilizing **Docker Compose** for orchestration across **3** operating systems.

Projects

Hospital Data Analysis using PowerBI | Dashboard Demo   **Jan 2024 – May 2024**
- Created an interactive hospital data dashboard in **PowerBI**, integrating over **500,000 records** from diverse sources.
- Engineered data models and used **DAX** for transformation, incorporating flexible filters for strategic insights.

News Data Engineering using Airflow, Snowflake and AWS | Project Link   **Oct 2023 – Dec 2023**
- Developed an automated data pipeline using Python and **Airflow** on an **EC2** instance, accessed via **Putty**, to fetch news articles about Apple Vision Pro from a specified API and transfer it to Amazon S3.
- Adopted AWS Glue and CLI commands to transfer data from S3 to Snowflake for simplifying the ETL pipeline

Technical Skills

Programming Languages: Python, SQL, R, C, C++, Java, DAX, Athena, Bash
Databases and Visualization: MySQL, MongoDB, SQL Server, Tableau, PowerBI, Excel, Redshift, S3
Cloud Platforms: **AWS Certified Cloud Practitioner**, Azure, Databricks, Snowflake, Dataiku, Seeq
Tools and Frameworks: Git, Docker, Pandas, Numpy, Matplotlib, Jira, Agile, Power Query