Tanvi Nikhil Shroff

(213) 272-4022 • tnshroff16@gmail.com • linkedin.com/in/tanvi-shroff • tshroff16.github.io/portfolio-tanvishroff/

Data Scientist at University of Rochester with experience creating solutions to solve complex business problems through data science. Adept in fostering cross-functional collaboration and communicating technical concepts to non-technical audiences.

SKILLS:

Languages:	Python, R, SQL, Linux, C++
Technologies:	Tableau, Git, JIRA, Databricks, PySpark, MLflow, Jupyter Notebook, Excel, PowerPoint
Libraries and Frameworks:	NumPy, Pandas, Matplotlib, NLTK, scikit-learn, PyTorch
Machine Learning Algorithms:	Regression, Random Forest, Gradient Boosting, XG Boost, Clustering, Forecasting

WORK EXPERIENCE:

Data Scientist, *University of Rochester* **June 2023 – Present**

Tool Stack: Python, SQL, Tableau, Oracle SQL Server, Databricks, Cognos, Denodo, JIRA

- Built an end-to-end **student retention predictive model** to help university leadership identify at-risk students early, enabling targeted interventions and resource allocations.
- Developed a **Student Performance Dashboard** and **Course Enrollment & Job Placement Dashboard** for the university to utilize and develop key business strategies.
- Led initiatives to resolve data discrepancies and improve the consistency and **accuracy** of student records, reducing data validation time for other teams and increasing **efficiency** by **50%**.
- Designed an Oracle **database architecture** in collaboration with multiple teams, focusing on user authentication, authorization, and data security.
- Fostered strong **collaboration** with **cross-functional** teams to implement the predictive model, ensuring smooth integration with Cognos, Denodo, and Oracle MySQL.

Data Analyst, *Paychex* **September 2023 – December 2023**

Tool Stack: Python, Git, SQL, Excel, Microsoft SQL Server, Databricks

- Built and deployed a **random forest classification** model to predict sales for Paychex's payroll solution with an 85% recall.
- Developed strategies for revenue **optimization** and shared actionable recommendations with the stakeholders.
- **Optimized** complex **SQL** queries by eliminating view dependencies and streamlining joins reducing runtime by 50%.
- Established Git for data integration solutions, enabling seamless updates and collaboration across teams.
- **Collaborated** with Sales, Marketing, and Data Engineering teams to ensure seamless data integration and data accuracy.

Research Analyst, *IBM* **August 2021 – May 2022**

Tool Stack: Python, NLP

- Led a team of 4 to convert English to Indian Sign Language by generating a 3D avatar with hand signs.
- Leveraged speech-to-text API, phrase re-ordering and **NLTK** library to perform **tokenization** and **lemmatization** for text.
- Accelerated the inference latency by 500%, by caching the SiGML files containing HamNoSys notations.

Data Science Intern, *Women in Data Science* **May 2021 – July 2021**

Tool Stack: Excel, SQL, Python, NLP

- Performed exploratory data analysis for consumer forum complaints using **NLP** and derived actionable insights.
- Designed an engagement score model using sentiment analysis on twitter voice and text data.

EDUCATION:

University of Rochester, Rochester, New York **August 2022 – May 2024**

Master of Science, Computer Science (Data Science) GPA: 3.83 / 4.0

Savitribai Phule Pune University, Pune, India **August 2018 – May 2022**

Bachelor of Technology, Information Technology GPA: 3.80 / 4.0

Awarded the **'Cummins – Excellence Award'** as the Best Outgoing Student Award for 2021-2022

ACADEMIC PROJECTS:

Social Media Analytics, *University of Rochester*

- Identified potential target consumers for new products using **sentiment analysis** and **kmeans clustering**.
- Leveraged **topic modelling** using stemming & lemmatization to generate key business insights.

CITI Bikes Forecasting Model, *University of Rochester*

- Developed an end-to-end **pipeline** for a bike-sharing system using **Databricks**, **PySpark**, and real-time **structured streaming**.
- Built an **FB prophet forecasting** model to predict hourly net bike change for a station.
- Developed a real-time monitoring application utilizing **Git** and **MLflow** for model tracking and registry.

PUBLICATIONS:

- Shroff, Tanvi, *et al.* "Literature Review on Machine Translation Systems for Sign Language Generation." International Conference on Data Management, Analytics & Innovation. Springer, 2023