Yashanjali Chavan

| 8108535444 | yashanjali.ch2602@outlook.com | www.linkedin.com/in/yashanjalichavan | https://github.com/yasha-tw0seven |

EDUCATION

Syracuse University, School of Information Studies — **Syracuse, NY**
Master of Science in Applied Data Science | **GPA: 3.78/4.0** — May 2024

Pune University, Cummins College of Engineering for Women — **Pune, India**
Bachelor of Technology in Electronics and Telecommunication Engineering | **GPA: 3.2/4.0** — June 2022

EXPERIENCE

Research Analyst — **Syracuse, NY**
Syracuse University — July 2024 – Present

- Built a complete data pipeline to collect 20,000+ Reddit posts and comments from specific subreddits using **Python** APIs, tracking sentiment patterns and user behavior across different online communities
- Cut data processing time significantly by cleaning text data through normalization, tokenization, and lemmatization which reduced overall data size by **35%** and organized everything in **SQL** databases for easier analysis

Graduate Research Assistant — **Syracuse, NY**
Syracuse University School of Information Studies — March 2023 – August 2023

- Compared **UMAP** and **CNN** techniques for reducing data dimensions on 10K data points, successfully cutting dimensions by **40%** while maintaining data quality

Created autoencoder models in **Python** to handle high dimensional datasets, which improved our processing speed by **15%**

PROJECTS

University Data Warehouse & Business Intelligence: [Link]

- Built a comprehensive data warehouse using fact and **dimension modeling** in **Excel**, creating 8 dimensional tables to better organize academic data.
- Set up ETL pipelines with **dbt Cloud**, **Snowflake**, and **SQL** to automatically pull data from 3 different source systems into our warehouse.
- Created **Power BI** dashboards tracking enrollment numbers, student performance, and resource usage for university departments.

Microsoft Stock Price Forecasting: [Link]

- Developed a **sentiment analysis** tool to process 5K+ Microsoft-related tweets daily, reaching **78%** accuracy in classifying tweet sentiment.
- Built **classification models** using 2 years of historical stock data combined with sentiment scores which performed **15%** better than **regression models** for short term predictions.
- Found that stock prices averaged **12%** higher during periods of positive social media sentiment.

Healthcare Cost Prediction: [Link]

- Analyzed healthcare management data in **R** to identify key factors driving medical costs.
- Built a **Random Forest** model that predicts healthcare expenses with **79.5%** accuracy.
- Created an interactive **Shiny R** application that combines predictive analytics with comprehensible visualizations, helping healthcare managers make better spending decisions.

TECHNICAL SKILLS

Languages: Python, R
Database Tools: MySQL, ETL, Data Warehouse, Snowflake, Cassandra, Mongo DB, Neo4j, Elasticsearch
Frameworks and Libraries: NumPy, Pandas, Pyspark, Scikit-learn
Visualization and Business Intelligence Tools: Power BI, Tableau, R Studio, MS Excel (Regression, Pivot Tables, Vlookup), Google Analytics (GA-4)
Coursework: Advanced Big Data Management, Data Warehouse, Big Data Analytics, Business Analytics, Applied ML, Statistics
Certifications: Snowflake Hands-on Essentials Data Warehouse, AWS Certified Cloud Practitioner.