Shrivats Sudhir

+1 9086990054 | shrivats.sudhir@gmail.com | GitHub | LinkedIn

Education

University of Wisconsin, Madison

M.Sc. in Statistics and Data Science — ***Cumulative GPA: 3.83/4.00*** — *December 2024*

B.Sc. in Mathematics: Data and Risk Analysis; B.Sc. in Statistics — ***Cumulative GPA: 3.67/4.00*** — *August 2023*

Selected Coursework: Statistical Modeling, Financial Statistics, Machine Learning, Deep Learning, Applied Regression Analysis, Stochastic Modeling and Prediction, Linear Randomized Algorithms, Cloud Computing.

Extracurricular: Data Science @ UW-Madison, ML+X: A Machine Learning Community, Mathematics Directed Research Program, Peer-tutor at Mathematics Learning Center, Teach For India (TFI) volunteer.

Work Experience

Quantitative Risk Analyst Intern — *June 2024 – September 2024*

Risk Methodology Group, Nomura Services India Private Limited, Mumbai

- Developed dynamic staleness detection algorithm using **Python**, reducing false positives by over 90%.
- Presented the algorithm to cross-functional teams, including stakeholders in Tokyo and UK offices.
- Fully automated monthly report generation using **Python** and **PowerBI**, reducing labor hours by over 88%.
- Authored documentation for 11 key End-User Computing (EUC) based on **Excel VBA**, **PowerBI**, **Python**, and **R** to enhance accessibility and maintainability.

Research Experience

Research Intern — *January 2024 – May 2024*

Center for Sustainable Employment, Azim Premji University, Bangalore

- Developed a high-threaded parallel scraping algorithm using **Google Cloud VMs** and **Selenium** in **Python** to dynamically filter and retrieve 10.8 million rows ($\approx$ 1.7 GB) of publicly available data.

Research Assistant and Co-Author — *January 2023 – December 2023*

Directed Research Program, University of Wisconsin, Madison

- Utilized Statistical High-Performance Computing clusters with **Slurm** workload manager to implement large-scale Markov Chain Monte Carlo simulations using **JAGS** in **R**.
- Presented research findings to Ph.D. scholars, faculty members, and graduate students.

Project Portfolio

Clustering Spotify Podcasts with NLP-Driven Insights | *GitHub*

- Scraped $\approx$ *284,481* episode details from *818* podcasts using **Selenium** and Spotify API pipeline.
- Preprocessed and tokenized podcast descriptions with **NLTK**, including lemmatization and stopword removal.
- Developed metrics to quantify directional, overlap, diversity similarities, and engineered recommendation system.
- Deployed a **Dash** app for podcast clustering and personalized recommendations.

Predicting Flight Delays and Cancellations: An Integrated Analysis of Airport Data and Weather Data | *GitHub*

- Automated scraping of *23 GB* airport and *30 GB* weather data using **Selenium**.
- Utilized reverse geocoding, Haversine-based, and UTC-normalized alignments to join datasets.
- Trained random forest models, achieving $\approx$ *25 min* test RMSE for delays and $\approx$ *98%* test accuracy for cancellations.
- Developed scalable workflows on Google Cloud, and deployed interactive web-app using **Dash**.

Technical Skills

Programming: Python, R, SQL, Bash, JAGS.
Developer Tools: Git/Github, Linux Bash, Emacs, Pycharm, Visual Studio Code, Rstudio, Jupyter.
Computing: High-Performance Computing Cluster (Slurm), High-Throughput Computing (HTCondor), Google Cloud, Amazon Web Services (S3, Redshift, Glue, Canvas, EMR).
Miscellaneous: LaTeX, Power BI, Microsoft Office, Mathematica, Sagemath.