Samiksha Rajendra Thikekar

(She/Her)

+1 202-843-7726 | samikshathikekar8@gmail.com | LinkedIn | Portfolio

Education

Georgetown University — Washington, D.C.
M.S. in Data Science & Analytics, GPA: 4.0/4.0 — *Aug 2022 – May 2024*

- Relevant Courses: Fundamentals of Data Science, ML App Deployment, Big Data & Cloud Computing, Probabilistic Modeling & Statistical Computing, Data Visualization, Data Structures & Algorithms, Computational Linguistics, Statistical Learning, Database Systems & SQL, Applied Time Series Modelling & Forecasting.
- Awarded Georgetown University's "Returning Student Scholarship" for innovative data analysis, visualization, and storytelling.

University of Mumbai — Mumbai, MH, India
B.E. in Electronics & Telecommunication, CGPA: 7.75/10.00 — *Jul 2015 – May 2019*

- Relevant Courses: Operating Systems, Neural Networks, Structured Programming Approach, Object Oriented Programming.

Technical Skills

Programming/Scripting/Mark-up Language: Python, R, SQL, JavaScript, HTML, CSS.
Libraries: Pandas, Numpy, Sklearn, Matplotlib, Plotly, Tensorflow, Dplyr, Seaborn, ggplot2, Spark, Altair, NLTK, Pickle, Keras.
Other Software & Tools: Git, AWS, Azure, MySQL, NoSQL, VScode, Tableau, Power BI, Quarto, Flask.

Professional Experience

Graduate Data Analyst — Apr 2023 – May 2024
Georgetown University's Center for Global Health Practice and Impact — *Washington, D.C.*

- Conducted analysis for program and research data pertaining to the healthcare and health system needs assessment, providing critical insights into healthcare needs and system performance, enabling strategic planning for health interventions.
- Developed data management code in Python that reduced manual data entry errors by 25% and enhanced processing efficiency by 45%. Also, produced finalized analytic datasets using Excel and Google Sheets.
- Created advanced data visualizations with Toucan Toco, supported routine evaluations, presented findings to both technical and non-technical stakeholders, aiding in tracking progress and measuring outcomes.

Associate Systems Analyst — Jul 2019 – Jul 2021
National Stock Exchange of India (NSEIT LTD) — *Mumbai, MH, India*

- Engineered 100+ web pages using React.js, ensuring responsiveness across devices and integrated accessibility features.
- Implemented APIs using Node.js. Performed features like transactions and triggers across 70+ tables and stored procedures in the SQL Server database which optimized server performance by 40%.
- Engaged with clients and collaborated with team members using GIT, Zoom and Microsoft Teams to deliver projects on-time.
- Employed Logstash to clean, parse, and transform 100M+ unstructured internal gateways, router, and firewall logs.
- Designed data analysis processes to craft visualizations and 15+ analytics dashboards in Kibana, facilitating real-time data monitoring and insights for the client.

Academic Projects

US Natural Gas Industry Analysis | Time Series Models — Jan 2024 – May 2024

- Conducted comprehensive analysis of the U.S. natural gas industry using data from 2005 to 2022, focusing on prices, consumption, production, imports, exports, stocks, and CO2 emissions.
- Forecasted natural gas variables using univariate (ARIMA/SARIMA) models, and analyzed complex economic relationships and environmental impacts using sophisticated multivariate (ARIMAX/SARIMAX/VAR) models.
- Assessed the financial health of leading natural gas companies using financial (ARCH/GARCH) models, and applied deep learning methods to enhance predictive accuracy in time series forecasting.

Reddit Analysis | Big Data, AWS, Azure, PySpark, SparkNLP — Aug 2023 – Dec 2023

- Analyzed 3M+ rows using AWS EC2, S3, Sagemaker, and Microsoft Azure for EDA, NLP, and ML to enhance content curation, recommendation algorithms, and community engagement on entertainment subreddits.
- Uncovered user engagement insights using techniques such as TF-IDF, Named Entity Recognition, and sentiment analysis. Cross-referenced popular Reddit discussions on movies and anime with Rotten Tomatoes reviews to gauge sentiment alignment.
- Applied Random Forest and Decision Tree classifiers for post categorization and popularity prediction, and Linear Regression and Decision Tree Regressor for post score prediction, revealing insights into community dynamics and content preferences.

Diabetes Checkpoint Pro App | ML, Flask, Heroku, Python — Aug 2023 – Dec 2023

- Built a diabetes prediction model using the Diabetes Health Indicators Dataset (253k+ rows, 21 features) with advanced classification algorithms.
- Evaluated model performance and implemented grid search CV for hyperparameter tuning, achieving an 87.9% recall score.
- Developed and deployed a Flask app on Heroku, featuring an intuitive MCQ-based interface for diabetes predictions.

Global Coffee Market Analysis | Data Visualization, Python, R — Jan 2023 – May 2023

- Crafted innovative data visualizations (choropleths, interactive histograms, animated plots) to analyze global coffee production, consumption, imports & exports, preferred chains, and quality leveraging libraries like ggplot2, Plotly, Altair, and Matplotlib.
- Showcased key insights including the United States as the top consumer, Brazil as a leading exporter, and Ethiopia's specialty coffee niche.