Raghav Sharma

📞 (571) 274-1930 || ✉ rs2190@georgetown.edu | Website | 🔗 LinkedIn | GitHub | 🏠 Washington D.C. |

EDUCATION

Master of Science, Data Science and Analytics, Georgetown University *August 2022- Expected May 2024*
Top 1% of cohort, Lead Teaching Assistant (DSAN-5200 Spring 2024) , Teaching Assistant (DSAN-5900 Spring 2023, DSAN-6300 Fall 2023) 4.0/4.0 GPA

Bachelor of Engineering, Electronics & Communication, Thapar Institute of Engineering and Technology *July 2015- May 2019*
Top 10% of cohort 3.3/4.0 GPA

WORK EXPERIENCE

Graduate Research Assistant | Lab for Globalization and Shared Prosperity, Georgetown University May 2023 – Present
- Incorporated NLP and supervised machine learning techniques to analyze SEC 10-K reports, creating a measure for data analysis.
- Implemented visually engaging data visualizations and communicated research findings through publications and websites, ensuring data collection and documentation for reproducibility and long-term viability.

Software Development Engineer (Data Engineering) | Sigmoid Analytics, Remote (India) recommendation August 2021- July 2022
- Managed the execution and release of Data Solutions projects, resulting in a $10M increase in funding and a contract extension, leading to a $5M profit for the firm.
- Crafted complex data pipelines, ensuring data quality and reliability for a rapidly growing business.
- Released firm projects on 'Pypi' after integration with third-party applications like Apache Superset, Redash, and Metabase.

Software Engineer | Digifuse, Remote (India) August 2020- July 2021
- Improved code quality and maintainability by implementing custom back-end solutions and refactoring existing code, resulting in a 65% increase in the SONAR score.
- Demonstrated proficiency in object-oriented programming (OOP) principles, enabling code flexibility and extensibility, while providing technical support and on-site software maintenance for end-users.

Software Engineer | HSBC Software Development India (HSDI), Pune, India July 2019- August 2020
- Advised the development of a Trade Transaction Posting utility API for the 'Global Trades and Receivable Finance' department, streamlining the capture of transaction details and reducing time and effort for users.
- Implemented the Awaiting Approval Requests (AWPR) functionality, enabling customers to perform transactions per limits of respective bank accounts as per HSBC policy.
- Delivered the Inbound Payments project, facilitating internal transfers and reducing processing time by 90%.

PROJECTS

Reddit Analysis | EC2, S3, Sagemaker, Azure, Spark, Python, NLP, ML Project Website September 2023- December 2023
- Analyzed 2 million r/TaylorSwift subreddit comments for insights into community sentiments and language patterns, while leveraging Spotify data to delve into Taylor Swift's presence and legacy on the platform.
- Conducted comprehensive analysis of Taylor Swift's music and its impact on audience by combining data from Spotify and Reddit, focusing on song popularity, sentiments, and emotions.
- Applied predictive modeling techniques to anticipate comment controversiality and popularity, providing valuable insights into the reception of Taylor Swift's music on digital platforms.

Visualizing UFO Sightings| Data Visualization, Plotly, Altair, Highcharts, Python Project Website March 2023- May 2023
- Conducted a comprehensive analysis of 80,000+ UFO sightings in the United States, utilizing the National UFO Reporting Center (NUFORC) dataset from 1973 to 2022.
- Applied advanced data visualization techniques, creating interactive charts and graphs to elucidate insights into temporal, geospatial, and weather-related factors influencing UFO sightings.
- Led the exploration of emerging patterns in UFO sightings data, revealing correlations with advancements in technology and media usage, delivering impactful observations into potential societal influences on reporting trends.

Soccer Match Summarization | Web Scraping, Spacy, NLTK, NLP, Streamlit, Python Project Website March 2023- May 2023
- Scraped 10K+ rows of data from goal.com for the UEFA Champions League 2020-21 season to extract essential information from soccer match reports.
- Developed a streamlit application that analyzes and summarizes the data, using extractive and abstractive summarization techniques.
- Generated concise summaries and encapsulated each game's essence, providing a quick and informative overview of the matches.

ETL Task file (Prefect) | ETL, Prefect, Python Project Website November 2021- December 2021
- Developed an ETL workflow and a task file in Python to help the firm integrate with a workflow management tool designed to orchestrate data stacks.
- Authored the module containing a collection of tasks for interacting with Firebolt databases via the firebolt-python-SDK library for generating ETL flows with Prefect.

SKILLS

Programming Languages: Python, R
Machine Learning models: Regression, Naïve Bayes, KNN, Decision Trees, SVM, Random Forest, Adaboost, K-Means, Hierarchical clustering
Data Tools: MySQL, PostgreSQL, MS Excel, AWS (EC2, S3), Azure BLOB storage, Amazon Redshift, Firebolt, Oracle SQL Developer