PRAKHAR GUPTA

+1 857-230-3243 | gupta.prak@northeastern.edu | LinkedIn | GitHub | Boston, MA

EDUCATION

Northeastern University, Boston, MA — Expected Apr. 2024
Master of Science in Data Analytics Engineering | GPA: 3.82/4.0
Focused Coursework: Foundation of Data Analytics, Data Management, Data Mining, Computation & Visualization

Vellore Institute of Technology, Vellore, India — Jun. 2020
Bachelor of Technology in Electronics and Comm. Engineering
Focused Coursework: Object-Oriented Programming, Data Structure and Algorithms, Probability & Statistics

SKILLS

- Languages: Python, HTML, CSS, Typescript, C++, R
- Database: MS SQL Server, PostgreSQL, SQL, MySQL, Azure, AWS
- ML Libraries: Pandas, NumPy, matplotlib, scikit-learn, NLTK, SciPy, ggplot2
- Analysis Tools: Power BI, Grafana, Google Analytics, Microsoft Excel
- Frameworks: Angular, Django Framework, Microsoft Visual Studio 2020, Anaconda, Postman, Jupyter Notebook, GitHub

EXPERIENCE

DXC Technology, Bangalore, India — Jul. 2020 to Aug. 2022
Professional 1 Software Engineer

Development of Data Quality Tool:

- Designed, coded, and built a software tool for Data Quality checks on incoming data through the creation of data pipelines and ETL processes with the help of a team - The tool was able to successfully clean and analyze data from over 10,000 records in just 15 minutes, improving the accuracy of the data by 95%.
- Identified and removed over 500 duplicate records using the tool, reducing the data size by 10%.
- Developed and designed ADF pipelines to automate trigger for Azure Databricks notebook, which incorporated data warehousing techniques and big data technologies like Hadoop and Spark - The ADF pipelines successfully processed over 100 GB of big data in just 30 minutes, reducing the manual processing time by over 70%.
- Eliminated human errors by automating the pipeline, resulting in a 100% accuracy rate in processing the data.
- Conducted sessions with clients to understand business requirements and used data modeling and data visualization techniques to provide business intelligence using Google Charts & Microsoft Power BI.
- Maintained a CI/CD platform for the tool using GitHub version control system and deployed it to the Azure cloud computing platform, ensuring scalability and performance optimization.

Development of Snowflake Data Extraction Scheduler:

- Implemented a secure and interactive user interface on Microsoft Power Apps. The interface enabled users to trigger the extraction of data from databases such as MS SQL Server, and dynamically build Azure Logic Apps.
- Developed an Azure function in Python to extract data from multiple relational databases and performed data migration to the user-defined cloud platform like AWS S3 and Azure.
- Ensured data security and privacy by integrating the tool with Azure Active Directory, providing role-based access control and data governance capabilities, resulting in a 40% decrease in data breaches.
- Followed Agile Methodology to meet client requirements and improve the tool's performance and scalability, resulting in a 25% increase in tool performance and a 30% reduction in development time.

ACADEMIC PROJECTS

Music Recommendation System — Oct. 2022 to Nov. 2022

- Designed and implemented a machine learning-based recommendation system utilizing techniques like K-means clustering for songs and genres, data mining and predictive analytics.
- Developed and maintained data warehousing and data processing infrastructure, integrating SQL and R for data manipulation and analysis, reducing the data processing time by 30% through time series analysis.
- Conducted data analysis, visualization, and modeling to derive meaningful insights and track the performance of the recommendation algorithm using business intelligence techniques and metrics such as precision, recall and accuracy.

Tweets Sentiment Analysis — Nov. 2022 to Dec. 2022

- Designed and executed a sentiment analysis pipeline utilizing natural language processing (NLP) techniques to classify tweets into positive, negative, or neutral categories.
- Conducted exploratory data analysis & processing of tweets including text cleaning, tokenization, and feature extraction.
- Utilized data visualization tools such as Tableau to create interactive dashboards and deliver meaningful data insights, enabling informed data-driven decision making and continuous improvement of system through data storytelling.

ACHIEVEMENTS/CERTIFICATION

- Achieved DXC Recognition award for teamwork and delivery by DXC Technology, Dec 2021
- Achieved FY22 H1 Collaborators Award by DXC Technology, September 2021
- AZ-900: Microsoft Azure Fundamentals Certification
- Natural Language Processing Specialization by DeepLearning.AI, Coursera Online Course