(+1) 716 306-9840
New York, United States
sanjuktabaruah5@gmail.com

Sanjukta Baruah

Master's Of Science In Data Science from University At Buffalo

Website:sbaruah.com
GitHub: Sanjukta2212
LinkedIn: sanjuktabaruah

WORK EXPERIENCE 4 YEARS

Graduate Teaching Assistant (Part Time) **Jan 2024 — May 2024**
University At Buffalo ***Buffalo, New York, US***
Data Models and Query Language (CSE 4/560) | Class Strength: 253 students

- **Teaching and Tutoring:** Conducted sessions and provided one-on-one mentoring to over 50 students, enhancing their understanding of SQL, NoSQL, T-SQL, ER modeling, and normalization through hands-on technical demonstrations.
- **Technical Skills:** Developed and evaluated projects involving database schema design, data queries, and data integration, supplemented with product demonstrations to illustrate practical applications.
- **Software and Tools:** Guided students in using PostgreSQL, PGAdmin, Tableau, and Power BI for practical assignments, including technical support to troubleshoot and resolve issues.
- **Collaborative Development:** Collaborated with professors and GTAs to create and deliver a balanced curriculum on data management systems, incorporating technical demonstrations to facilitate learning.
- **Assessment and Classroom Management:** Administered and graded weekly quizzes, mid-terms, and final exams, maintaining high academic standards and integrity.
- **Computer Science Fundamentals:** Ensured the curriculum covers core computer science principles, including data modeling and data transformation.

Engagement Ambassador (Part Time) **Sept 2023 — Dec 2023**
Ruffalo Noel Levitz ***Buffalo, New York, US***

- **Identifying Prospective Donors:** Utilized storytelling techniques to identify potential donors aligned with the university's mission of helping students with their needs.
- **Securing Pledges:** Secured a total of 726 pledges, with 662 being $15 per month. Additionally, achieved 38 one-time pledges, including 26 pledges of $200 and 12 pledges of $500.
- **Effective Communication and Relationship Cultivation:** Played a pivotal role in meeting fundraising targets through personalized communication strategies, directly benefiting various university programs and building lasting donor relationships.

Technical Lead: Cloud Engineer (Full Time) **May 2019 — Nov 2022**
Cognizant Technology Solutions ***Pune, Maharashtra, India***
Azure | Automation | Migration | Extract, Transform, and Load Pipeline | Data Governance | Documentation | Active Directory | SQL | Python

- **Azure Cloud Solutions and Automation:** Proven expertise in administrating and deploying Microsoft Azure Cloud Solutions.
- **Technical Support:** Provided technical support for Azure environments, ensuring efficient operation and issue resolution.
- **KPI Management:** Managed KPI calculation and validation, driving process improvements with interactive dashboards. Delivered results to secure 100% customer retention and 15% quarterly revenue growth.
- **ELT Pipeline Optimization:** Orchestrated the Agile migration of 530TB of user and application-specific data from On-premises SSMS to Azure, leveraging Azure Databricks, Azure Data Lake Storage Gen2 for data storage, Synapse Analytics, Python, and SQL for ETL processes, followed by comprehensive insights development in Power BI for data visualization, enabling real-time updates and centralized business insights.
- **Azure Dashboards Deployment:** Deployed Azure Dashboards for data analysis, resulting in a 20% reduction in time spent on data preparation, reporting, and presentation tasks, while achieving a 95% adherence to SLAs.
- **Leadership and Team Coordination:** Led 15-day sprints with 5 user stories per member, optimizing ETL pipelines in Azure with SQL and Python, reducing storage costs by 40%.
- **Operations research** Led POC for nine Azure services, boosting efficiency by 20% and cutting costs by 15%. Implemented Privileged Identity Management, reducing security incidents by 60% for over 2000 users.
- **Azure Process Documentation:** Documented 39 Azure processes (SOPs) in Microsoft Office Software and chaired 23 team knowledge-sharing sessions, reaching 53 employees and increasing total productivity by 2.65 FTEs.
- **Azure Data Services Experience:** Experience with Azure Data Factory: browsing, exporting, and writing data pipelines in ADF.
- **SQL in Azure Databricks:** skilled in writing SQL in Azure Databricks and working with Azure Data Warehouse and Azure Synapse.
- **Post-Deployment Automation:** Automated post-deployment issue resolution, reducing resolution times by 50% and enhancing operational efficiency and client satisfaction using Dynamics 365 and Power Automation.
- **Risk Management:** Focused on resolving security recommendations to ensure ISO27001:2013 compliance, emphasizing data protection, governance, and privacy. Maintained Azure Security Center threshold at a minimum of 87% across all resources.
- **Automated User Creation:** Automated guest user account creation, reducing manual effort by 95% from 120 to 6 hours per month.
- **Infrastructure Planning:** Conducted client meetings to understand application configurations and translated business requirements into non-technical terms. Subsequently, curated cost-efficient infrastructure with optimal resources for applications.
- **Research and Innovation:** Conducted research on new Azure services and implemented improvements to enhance infrastructure capabilities.

(+1) 716 306-9840
New York, United States
sanjuktabaruah5@gmail.com

Sanjukta Baruah

Master's Of Science In Data Science from University At Buffalo

Website:sbaruah.com
GitHub: Sanjukta2212
LinkedIn: sanjuktabaruah

SKILLS

Visual Analytics Tools	Tableau, PowerBI, Microsoft Excel, VLOOKUP, PIVOT Tables, SSRS, SSIS, Microsoft Dynamics
Data Modelling Tools	Azure Data Factory, Azure Databricks, Azure Data Lake Services, Azure Synapse Analytics, Virtual Machines, Azure Storage Services, Azure Data Warehouse, Azure DevOps, Azure Cognitive Services, RESTAPI
Languages & Libraries	Python, SQL, VBA, R, Powershell, NumPy, Pandas, Matplotlib, Plotly, Seaborn, GGplot, Altair, Geoplotlib, PyTorch, NLTK, Matplotlib, Keras, PySpark, Scikit-Learn
Big Data	Data Mining, Apache Spark, Kafka, Tableau, Hadoop, Airflow, Snowflake, Splunk
Proficiencies	Docker, Git, Jenkins, JIRA, SSIS, SSRS, Machine Learning

EDUCATION

- **Master of Science in Data Science | GPA: 3.6/4.0**
  State University of New York at Buffalo, New York, United States **Jan 2023 - May 2024**
  Relevant Courses: Numerical Mathematics and Computing, Probability theory, Statistical Learning and Data Mining, Machine Learning, Data Structure and Algorithm, Applied Statistics, Data Model Query Language, Data intensive Computing, Deep Learning
- **Bachelor of Technology in Electronics and Communication Engineering | GPA: 7.81/10.0**
  Dibrugarh University Institute Of Engineering And Technology, Assam, India **Aug 2015 - May 2019**
  Relevant Courses: Probability and Random Processes, C++ and Object oriented Programming, Digital Circuit and Logic Design, Microprocessor and Microcontroller, Control Systems

ACADEMIC PROJECTS

Cancer Recurrence Prediction
- Traditional methods struggle to predict cancer recurrence and accurately identify key genetic mutations.
- Implemented the Genomic Impact Transformer (GIT) model with multi-head self-attention, applied PCA for dimensionality reduction, and created gene embeddings for functional impact analysis.
- Achieved 69.5% precision and 57.1% recall, PCA explained 75% of data variance, and improved gene similarity detection by 83.29%.

EconoMetrics Banking ER Analytics Link
- Created ER diagram and synthetic data using Faker, enhancing realism with real data for "Statewise" table and skewing distributions.
- Improved PostgreSQL performance with data normalization and advanced indexing, cutting response times by over 30%.
- Utilized Tableau to visualize transaction, demographic, and loan data, analyzing and informing strategic business decisions.

Seamless On-Prem to Azure Data Journey with Auto-Sync Pipeline Link
- Migrated AdventureWorks database to Azure SQL, ensuring data integrity.
- Developed and tested backup strategies for databases, ensuring seamless failover with geo-replication.
- Integrated Microsoft Entra Directory for secure access management.

Analyzing Car Accidents in the USA (2016-2021): Identifying Contributing Factors and Predicting Severity Link
- Cleaned and conducted EDA on 7.7M US accident records (2016-2021), addressing data imbalance with various sampling techniques.
- Addressed data leakage caused due to random sampling after PCA, leading to a notable increase of approximately 200% in recall, precision, and accuracy.
- Achieved 86% testing and 87% validation accuracy with optimized Neural Network. Also developed a Streamlit website for near real-time accident severity prediction, delivering results in under 2 seconds.

Climate Change Time-Series Analysis
- Developed a forecasting model to predict New York City's temperature trends using AR, MA, and ARMA models.
- Identified a potential temperature increase of up to 2°C over 30 years, informing environmental policy planning.
- Tools used: Python (Pandas, NumPy, StatsModels), Jupyter Notebooks.

CERTIFICATION

Azure Administrator Associate (AZ-104) Link
Azure AI Fundamentals (AI-900) Link
Generative AI (DeepLearning.AI) Link
CNN(DeepLearning.AI) Link
Introduction to Genomic Technologies (Johns Hopkins University) Link