VAIDEHI SHAH

vaidehigmn1502@gmail.com | (781) 698-5866 | linkedin.com/in/-vaidehi-shah/ | github.com/vaidehi-nshah | vaidehi-nshah.github.io/Portfolio/

EDUCATION

M.Sc in Applied Data Analytics	**Sep 2022 - Jan 2024**
Boston University	*Boston, MA*
B.Tech in Computer Science Engineering	**Jul 2018 - May 2022**
Pandit Deendayal Energy University	*Gujarat, India*

Certifications

Data Analysis with R, **Google Data Analytics**, Web Development for Everybody	**Coursera**
Relational Database and MySQL	**Stanford Lagunita Program**

WORK EXPERIENCE

Graduate Assistant — **Dec 2022 - Present**
Boston University — *Boston, MA*

- Enhanced University LMS, **cutting faculty workload by 2 hours weekly**. Integrated LTI tools while maintaining **100% resolution rate** with ServiceNow, demonstrating **strategic communication and decision-making skills** that improved system efficiency.
- Worked on extensive reports and DDA queries with the help of SQL and Excel functions. Researched generative AI for plagiarism detection, focusing on GPT technology.

Graduate Research Assistant — **May 2023 - Jul 2023**
Analysis of Mutual Funds by CuSum and ML (Dr. Eugene Pinsky) — *Boston University, MA*

- Constructed a comprehensive **Data Visualization pipeline** for 150 large-cap mutual funds, facilitating fund performance evaluation through the application of K-means clustering on CUSUM data after data processing.
- Examined trends in historical data to infer that Mutual Funds with AUM exceeding $10 billion and a low Turnover Ratio of approximately 0.2 outperformed the Index fund, showcasing **data-driven solutions in financial analysis**.

Digital Media Associate Manager — **Oct 2021 - Mar 2022**
BigDeal, GlobalVOX — *Gujarat, India*

- **Collaborated cross-functionally** with a PR team, utilizing customer and market analytics to optimize marketing channels for a Blockchain project launch, resulting in a remarkable **30% increase in user engagement**.
- Employed data-driven approach in documentation of comprehensive reports, leveraging user feedback data on digital platforms to enhance project visibility and digital marketing, and align with business processes and business strategy.

PROJECTS

Hospital Administration Data Warehouse — **Boston University,** *Spring 2023*

- Designed and built a PostgreSQL data warehouse for small clinics, improving data consistency 60% through referential integrity and indexing. Also achieved **15% faster query performance** via feature engineering in the ETL process.
- Engineered a 12-table schema in pgAdmin **using a Snowflake Dimensional Model**. Leveraged Python's Faker library to generate and collect data to populate the warehouse.

LAPD Crime Big Data Analysis — **Boston University,** *Fall 2023*

- Orchestrated the integration of LA crime dataset with PostgreSQL and Tableau, facilitating dynamic and **tailored visualizations for key business questions**, including demographic-based insights and location-based hotspot identification.
- Spearheaded **end-to-end development of data pipeline** on Google Cloud Platform by applying RandomForest Classifier and Regressor with PCA to predict Victim Age, Sex, and Descent.

Property Type Categorization — **Boston University,** *Spring 2023*

- Streamlined quantitative data analysis workflow, **reducing feature dimensionality by 40%** through effective feature selection methods, validation and testing led to a **12% improvement in model accuracy** as compared to training.
- Optimized a Machine Learning System to 88.16% accuracy by evaluating 30 models with 5 Classification algorithms and 5 Feature Selection methods.

Sentiment Analysis using BERT on Movie Reviews — **Boston University,** *Spring 2023*

- Achieved 88% accuracy by implementing BERT Base (Uncased) data model on a Movie Reviews dataset, demonstrating **effectiveness of Natural Language Processing techniques for sentiment analysis**.
- Optimized hyperparameters, fine-tuned architecture, and refined dataset over multiple epochs to identify an **6% enhancement** in model's performance.

TECHNICAL SKILLS

- **Languages**: Python for Data Science (PySpark, Seaborn, skLearn, SmyPy, Matplotlib, Pandas, NumPy), R Programming, SQL, HTML5/CSS3, C, C++
- **IDEs**: Apache, Google Colab, Jupyter Notebook, R Studio, VS Code, pgAdmin, Microsoft Workspace, Google Workspace
- **Cloud Data Tools**: AWS (EC2, S3, EMR), Apache Airflow, Docker / Docker-compose
- **Tools**: Tableau, PowerBI, Apache Spark, PostGRESQL, Git, JIRA, MS Office, Agile, SAS, Slack