VAIDEHI SHAH

vaidehis@bu.edu | +1 (781) 698-5866 | linkedin.com/in/-vaidehi-shah/

EDUCATION

M.Sc in Applied Data Analytics (GPA: 3.53/4.0) **Sep 2022 - Jan 2024**
Boston University *Boston, MA*

B.Tech in Computer Engineering (GPA: 3.82/4.0) **Jul 2018 - May 2022**
Pandit Deendayal Energy University *Gujarat, India*

Certifications

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

WORK EXPERIENCE

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

- Enhanced University's Blackboard LMS by integrating LTI tools like Zoom and InScribe, reducing faculty workload by up to 2 hours weekly and maintaining a 100% resolution rate through ServiceNow support
- Formulated research on generative AI for plagiarism detection and GPT technology for faculty support

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

- Developed an efficient data visualization pipeline for a 25-year dataset comprising 150 large-cap mutual funds sourced from WRDS and Yahoo Finance. Evaluated fund performance relative to the S&P-500 benchmark through K-means ++ Clustering on CuSum data
- Examined trends in historical data to infer Mutual Funds with AUM exceeding $10 billion and a low Turnover Ratio of approximately 0.2 outperformed Index fund

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

- Collaborated with a team of 5 members in the Public Relations department to optimize marketing channels, leading to a remarkable 30% increase in user engagement for the launch of Blockchain project
- Leveraged user feedback data to gauge project impact and strategically utilized it to maximize project visibility across digital platforms - LinkedIn, Telegram, Facebook, and Medium

PROJECTS

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

- Achieved 88% accuracy by implementing BERT Base (Uncased) 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

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

- Built a PostgreSQL-based data warehouse in pgAdmin for small clinics, ensuring 60% improved data consistency through referential integrity on foreign keys and indexing
- Designed a 12-table schema, used Python's Faker library for data generation, and implemented a Snowflake Dimensional Model
- Achieved a 15% query performance improvement through feature engineering in the ELT process

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

- Streamlined data analysis workflow, reducing feature dimensionality by 40% through effective feature selection methods leading to a 12% improvement in model accuracy as compared to initial dataset
- Developed and evaluated 30 ML models using 5 classification algorithms and 5 Feature Selection Methods to achieve an average accuracy of 88.16%, with Random Forest model outperforming other models across multiple attribute selection methods

Cardiovascular Risk Assessment **Boston University,** *Fall 2022*

- Utilized a UCI dataset comprising 300 instances and 12 features, including critical factors such as serum creatinine and ejection fraction, to drive insights for more effective cardiovascular risk assessment and patient care in healthcare sector
- Implemented a Random Forest modeling approach to predict likelihood of heart failure. Achieved a predictive accuracy with an AUC of 90.3%, surpassing performance of Multiple Logistic Regression by 1%

Cataract Prediction **Boston University,** *Fall 2022*

- Designed and implemented a Convolutional Neural Network (CNN) architecture to develop a Cataract Prediction model
- Employed hyperparameter experimentation, model fine-tuning, and optimization techniques to achieve a predictive accuracy rate of 88%

TECHNICAL SKILLS

- **Languages**: Python (PySpark, Seaborn, skLearn, SmyPy, Matplotlib, Pandas, NumPy), R Language, SQL, HTML5/CSS, C/C++
- **IDEs**: Apache, Google Colab, Jupyter Notebook, R Studio, VS Code
- **Tools**: AWS(EC2, S3, EMR), Apache Spark, PostGRESQL, GitHub, JIRA, MS Office