Karan Badlani

Boston, MA | (617)-992-1174 | badlani.k@northeastern.edu | linkedin.com/in/karan-badlani

Education

Northeastern University, Boston MA **December 2025**

Master Of Science in Data Science; GPA: 3.6

Coursework: Database Management, Statistical Supervised Machine Learning, Large Language Models, Machine Learning Operations and Best Practices, Unsupervised Machine Learning models and Data Mining

Skills

Programming: Python (Pandas, Scikit-learn, Keras, TensorFlow, PyTorch, Spark, NLTK, Plotly), R, Java, SQL, NoSQL

Data Science: Predictive Analytics, NLP, Classification, Regression, Time Series Forecasting (ARIMA), PCA, Clustering (K-means, DBSCAN), Neural Networks, Feature Engineering, Hypothesis Testing, A/B Tests, GPT 4

Databases and Tools: MySQL, MongoDB, Docker, GitHub, Git, Apache Airflow, Postman, DBT, FastAPI

Data Visualization: Power BI, Tableau

Cloud: AWS (S3, Glue, SageMaker), Azure (ML Studio, ADF, Synapse Analytics, Databricks), Snowflake

Professional Experience

Data Science Analyst | MFS Investment Management, US **January 2025 – July 2025**

- Developed PCA-guided K-means clustering to **segment buying units** and surface hidden "Rising Stars clients", discovery of 22 % untapped prospects boosted **sales-conversion** by 11.5% within two quarters
- Engineered **statistical model** (Prophet) benchmarked against ARIMA/Naïve Bayes/LSTM on **Sales forecasting**; deployment reduced **forecast error** from $7.5 B to $1.65 B with ±3.7 % confidence
- Productionized the Sales pipeline in **Azure ML Studio** with ADF-powered Snowflake loads; **CI/CD MLOps** cut manual effort by 60 % and delivered near-real-time Tableau dashboards to Strategist and leadership team

Research and Teaching Assistant | Khoury College of Computer Sciences, US **May 2024 – December 2024**

- Led and mentored 100+ students in **data modeling**, EDA, and ML pipeline integration using **AutoML**, while designing course content and labs on Airflow, Docker, and **GCP** to elevate hands-on learning and project accuracy

Data Scientist | InfoCepts, India **December 2022 – June 2023**

- Improved bike rental demand prediction by 35% using ML models in Azure ML and **Databricks**, enabling smarter resource planning and 15% boost in **customer segmentation**
- Built 10+ scalable data pipelines and automated API ingestion using Logic Apps with Data engineering team, enhancing **KPI** accuracy by 20% and accelerating insight delivery via 7+ Power BI dashboards
- Engineered **end-to-end ETL** workflows with Azure Data Factory and Synapse, processing 1TB+ data and reducing operational costs by 9% through optimized space and analytics efficiency

Research Scientist | Shri Ramdeobaba College of Engineering, India January **2022 – June 2022**

- Designed and deployed end-to-end air quality analysis pipeline using **AWS** (S3, Glue, Redshift), improving classification model accuracy by 22% (Random Forest, XGBoost) and enabling informed KPIs via Streamlit dashboards

Academic Projects

Predictive Analytics on GCP with statistical validation | GCP, Docker, Airflow, Databricks, GitHub Actions

- Developed and deployed predictive analytics pipeline on Google Cloud Platform using Docker, Apache Airflow, and Databricks, reducing customer churn by 15% through real-time predictions and data-driven intervention
- Validated hypothesis testing (chi-squared, A/B tests) and automated ML workflows in a containerized environment to improve model accuracy by 12%, ensuring scalability and reproducibility across the CI/CD pipeline (GitHub Actions)

Predicting Purchase Intentions | Statistical Modeling – Python, Random Forest, SVM, Tableau (BI Platform)

- Executed EDA and feature engineering on a dataset of over 1M observations, using Random Forest to identify the most significant features and SVM to capture non-linear relationships in the data, resulting in 89% variance explanation
- Leveraged random under-sampling to address class imbalance, implementing 5+ classification algorithms, including Random Forest and SVM, and prepared KPIs and predictive insights in Tableau for business decision-making

Research and Publications

- Authored "Classifier accuracy in Time-series forecasting of air quality" at IBMEC, with the paper published in NeuroQuantology and indexed in SCOPUS with focus on predictive modeling and neural network
- Published "Pneumonia Detection through CNN Image Classification", presented at ICICA22, featured in IJNGC journal with Web of Science indexing showcasing expertise in computer vision and deep learning