Yash Krishna Bheke

yashbheke2000 | yash-bheke | Portfolio | yash.bheke2000@gmail.com | +1 5132765130

Professional Experience

Graduate Assistant (IS 8034: Big Data Integration) Sept 2025 - Present
University of Cincinnati Cincinnati, Ohio

- Migrated and optimized 15+ legacy data pipelines to Databricks platform using PySpark and SQL, improving processing efficiency by 31% for 20+ graduate students in Big Data Integration coursework
- Integrated AWS RDS and S3 with databricks, automating data ingestion and implementing scalable Unity Catalog solutions for reproducible ML experiments and collaborative model development.

Software Developer (Data Analytics and Machine Learning) Sept 2022 - Jul 2024
Accelya Mumbai, India

- Optimized the predictive analytics pipeline using XGBoost and dimensionality reduction techniques, accelerating model inference time by 97% (from 60+ min to < 2 min) enabling real-time analytics.
- Developed ensemble classification models (Random Forest, Gradient Boosting) to predict flight delays (92% confidence) based on historical data, and deployed model which served 5k+ predictions daily to support flight scheduling optimization and reduce operational costs by $1.2M annually.
- Designed a testing framework combining Bayesian inference and Multi-armed Bandit algorithms, along with Hypothesis testing (t-tests and chi-square) to identify 9 process improvements which led to 32% increase in user engagement ($p < 0.05$) and 16% reduction in financial reporting errors.
- Automated data pre-processing workflows using Python/Bash, eliminating 600+ hours of team's monthly manual work by implementing data cleaning, validation and EDA across 500+ GB datasets.

Projects

Airbnb Pricing Tool (Tech Stack: Python, Scikit-learn, SHAP, Google Colab)

- Preprocessed Airbnb listings dataset through text parsing (bathroom extraction), currency normalization, outlier removal, and engineered revenue-signal features for optimum pricing calculation.
- Optimized a Random Forest regressor (RandomizedSearchCV) to achieve R^2=0.58, MAE $49.69, RMSE $103.83; outperformed baseline by 35.9%. Identified top price drivers through permutation importance and SHAP to allow hosts to refine listings and access 90% prediction confidence bands.

Mood Disorder Predictor (Tech Stack: Python, SHAP, Matplotlib, Seaborn, Jupyter Notebook)

- Developed a privacy focused mood disorder prediction pipeline, achieving 96% accuracy and perfect recall on minority at risk classes via cross-validation enabling identification of vulnerable individuals.
- Visualized clinical drivers (Sleep, Exhaustion, Euphoria) for clinicians for early targeted intervention.

Education

University of Cincinnati, Carl H. Lindner College of Business August 2024 – May 2026
Master of Science, Information Systems, 3.97 GPA
Courses: Gen AI, Statistical Computing, Datamining for BI, Data Analysis, AI ML, Data Visualization

University of Mumbai August 2018 – May 2022
Bachelor of Engineering, Electronics and Telecommunications, 3.66 GPA

Technical Skills

- **Languages:** Python (Numpy, Pandas, TensorFlow, Keras, Spark), SQL, R, Bash, C++
- **Data Science & ML:** NLP, LLMs, PyTorch, RAG, A/B Testing, Time Series Analysis
- **Databases & Big Data:** Oracle, MySQL, PostgreSQL, Snowflake, Databricks, BigQuery
- **Data Engineering:** ETL Pipelines, Batch & real-time Processing, Feature Engineering
- **Data Visualization & Analysis Tools:** Tableau, PowerBI, Looker, Excel, Zoho
- **DevOps, Infrastructure and OS:** Docker, Git, GitHub, Containerization, CI/CD, Windows, Linux