

Harshith Keshavamurthy

+1 (617) 999-6861 | Boston, MA 02135 | hk17@bu.edu | www.linkedin.com/in/harshith-k-bu/

SUMMARY

Data practitioner who turns messy data into decisions and shipped systems: **SQL/pandas EDA, KPI** & experiment design, and **ML/LLM** solutions (regression/classification, **clustering, RAG/Graph-RAG**) with **FastAPI** services on Azure. Comfortable going end-to-end—**data modeling (SQL/Neo4j)**, feature engineering, model/**evaluation (precision@k, recall@k, A/B)**, dashboards (**Tableau/Power BI**), and clear stakeholder **storytelling**.

SKILLS

- **Statistics & Experimentation:** hypothesis testing, confidence intervals, **A/B testing** (power, SRM), metric design, causal inference (basics)
- **Machine Learning:** linear/logistic regression, regularization, tree-based (**Random Forest, XGBoost, LightGBM**), clustering (K-Means, DBSCAN), time series (ARIMA, Prophet basics), **cross-validation & hyperparameter tuning**, calibration, SHAP
- **Data & SQL:** advanced **SQL** (joins; window functions ROW_NUMBER, LAG; CTEs; CASE WHEN), data modeling (star schema), pandas, NumPy, **feature engineering, geospatial** (lat/long, haversine); **PostgreSQL/MySQL**
- **LLM/RAG & Retrieval:** Large Language Models (**LLMs**), prompt engineering, **Retrieval-Augmented Generation**, LangChain, FAISS/ANN, SentenceTransformers (BGE, MiniLM), cross-encoder reranking; **evaluation** (precision@k, recall@k, hallucination)
- **Software/MLOps & BI:** Python, scikit-learn, PyTorch (U-Net), FastAPI, Streamlit, Docker, **Azure DevOps** (CI/CD), **Tableau/Power BI**, Excel/Sheets (PivotTables, XLOOKUP, Power Query), Weights & Biases, TensorBoard, Git

EXPERIENCE

Tietoevry – AI/ML Engineer Intern

Boston, Massachusetts

June 2025 – August 2025

- Translated analyst questions into measurable metrics for a **Graph-RAG** platform; defined an **evaluation protocol** (precision@k, recall@k, hallucination), built labeled test sets and a reviewer rubric, and used **bootstrap confidence intervals** for reliable comparisons.
- Ran **experiments/ablations** across retrieval configs—**Cypher graph hops, BM25, SentenceTransformers** (BGE, MiniLM), cross-encoder reranking—and k-values; performed **model selection** with cross-validation, led error analysis, and documented quality-latency-cost trade-offs.
- Analyzed the knowledge graph with **Neo4j/Cypher** to quantify coverage and severity distributions, measure CVE↔CWE link completeness, and generate triage lists; engineered features (**severity level, linkage density, reference richness**) to explain performance shifts.
- Operationalized insights via a **FastAPI** service and **Streamlit analyst UI**; instrumented telemetry (query volume, SLA compliance, failure causes) and shipped **dashboards + runbooks** for clear stakeholder readouts..

Nineleaps – Data Analyst (UBER ext)

Bengaluru, Karnataka, India

August 2023 – May 2024

- Primary involved during the **Uber Bus** India expansion; laid the **foundation for Kolkata and Chennai** by defining **KPI frameworks** (utilization, occupancy, route efficiency), **building demand baselines**, and creating city-readiness checklists that enabled pilot launches and scale-up.
- Built **geospatial** demand-hotspot and **route-candidate** models using **clustering (DBSCAN, K-Means)** with features (time of day, stop density, catchment size, POI proximity); validated with field inputs and historical ridership to produce ranked recommendations.
- Implemented **an analysis pipeline** for scenario testing and daily refreshes (**Python, SQL, Google Apps Script**); automated data pulls and QA; delivered **Tableau** launch scorecards used by ops/PMs for go/no-go and capacity planning.
- Authored **SOPs/playbooks** and **data quality checks** to standardize new-city onboarding and handoff; recognized with a **Best Employee Award** within three months.
- **Tech:** SQL, Python, Tableau, DBSCAN, K-Means, geospatial (lat/long, haversine), Google Apps Script, Git

Nineleaps – Data Science Intern

Bengaluru, Karnataka, India

February 2023 – August 2023

- Completed **DA/DE rotation**; cleaned and analyzed **10M+ records**; built KPI dashboards with **SQL, Excel, Tableau** for stakeholders.
- Prototyped scalable ETL with **PySpark, Hive, Hadoop**; added data-quality checks and basic scheduling to cut manual effort by ~40%.
- Implemented data modeling and workflow orchestration patterns; documented pipelines and handoff guides for engineering.
- **Tech:** PySpark, Hive, Hadoop, SQL, Excel, Tableau, Python

INDUSTRY-LED PROJECTS

Build Fellowship – Student Consultant

Applied fellowship; built research, scoring model, and KPI gates; delivered a board-style market-entry pitch

Boston, Massachusetts

July 2025 – August 2025

- Built **data pipelines and market-entry scoring model** (15 features across 5 countries): created a source log and data dictionary, **engineered features** (internet penetration, affordability, English proficiency, edtech adoption, regulation), and applied 1–5/100-point normalization with range & missing-data QA to produce per-country scorecards and a leaderboard.
- Converted **SWOT/PESTEL** into quantitative variables (e.g., pricing → elasticity proxy; regulation → compliance-risk flag); documented SOPs for refresh/validation and defined **go/no-go KPI gates** (NPS, CAC, completion, job-advancement) to turn analysis into decisions.
- Presented a **board-style USA launch recommendation** informed by the model's outputs; **proposed two routes to market** (certification-aligned tracks; credit-eligible academies) with phased timeline and budget guardrails; approach was well received and elements were considered for rollout.

PROJECTS

Breast Ultrasound Segmentation (BUSI)

February 2025 – May 2025

- Built a semantic segmentation pipeline with **U-Net (ResNet50 encoder, scSE decoder, softmax2d, 1-channel input)** to segment benign, malignant, background.
- Trained with **Albumentations** augmentations and a **class-weighted Dice + Focal loss**, stratified splits, early stopping, and LR scheduling; logged per-class & overall Dice during validation.
- Added **DenseCRF** post-processing to refine masks and implemented a clean **validation/evaluation harness (Dice per class + overall)** with visualizations.
- Tracked experiments in **Weights & Biases and TensorBoard**; shipped a reproducible loader + **submission generator** for test inference.

Formula 1 Race Outcome Prediction API

August 2025 – November 2025

- Built a machine learning model using Python and Scikit-Learn to predict driver finishing positions in Formula 1 races, **achieving 85% accuracy across 500+ race simulations** by leveraging features such as starting grid position, lap counts, and season performance metrics.
- Evaluated advanced algorithms, including Gradient Boosting, Random Forest, and Logistic Regression, identifying **Gradient Boosting as the best-performing model with a 10% accuracy improvement** over baseline methods, while engineering 15+ interaction features and optimizing hyperparameters with GridSearchCV, **reducing cross-validation error by 12%**.
- Conducted feature importance analysis, identifying starting grid position and average season performance as the top predictors, contributing to 70% of the model's predictive power.

Uber ETL Data Transformation Project

June 2023 – August 2023

- Led the development of an end-to-end ETL pipeline, improving Uber's data ecosystem for operational insights.
- Automated data extraction and transformation workflows, **reducing manual tasks by 60% and increasing data accuracy by 25%**. Results: Improved decision-making for Uber's management, **leading to a 30% increase in operational efficiency**.

EDUCATION

Boston University

Boston, Massachusetts

M.S in Applied Data Analytics

GPA: 3.73/4.0 | Courses: Advanced Machine Learning and Neural Networks, Web Mining and Graph Analytics, Big Data Analytics, Analytics and Data Visualization, Machine Learning, Data Science with Python, Data Mining.

RNS Institute of Technology (RNSIT)

Bangalore, Karnataka, India

Bachelor of Engineering in Computer Science

GPA: 8.0/10.0 | Courses: Data Structures, Algorithms, Database Management, Machine Learning, Artificial Intelligence