

BHARATH NAVEEN

Tucson, AZ | +1 520-355-9462 | bharathnaveen.bn@gmail.com | [linkedin.com/in/bharath-naveen](https://www.linkedin.com/in/bharath-naveen) | github.com/Bharath-Naveen

Data Scientist • Machine Learning Engineer • Data Analyst

PROFESSIONAL SUMMARY

Data and machine learning practitioner with 3+ years of enterprise experience at Tata Consultancy Services and a track record of shipping end-to-end data systems — from Power BI reporting and SQL automation to containerized, cloud-deployed ML pipelines. Currently completing an MS in Information Science (Machine Learning) at the University of Arizona (GPA 3.89, May 2026). Comfortable owning ambiguous problems across the full stack: exploratory analysis, feature engineering, model training and evaluation, LLM/NLP integration, and deployment on AWS with Docker and CI/CD. Equally at home turning messy data into stakeholder-ready insight and building production systems that other engineers depend on. STEM OPT eligible; open to relocating anywhere in the U.S.

TECHNICAL SKILLS

Languages: Python, SQL (MySQL, PostgreSQL), R

Machine Learning & AI: scikit-learn, XGBoost, LightGBM, Random Forest, Logistic Regression, Neural Networks, model evaluation (ROC-AUC, F1, precision/recall), feature engineering

NLP & LLMs: Applied NLP, NLTK, TF-IDF, statistical NLP, LLM integration, RAG pipelines, brand-impersonation / text classification

Data & Analytics: pandas, NumPy, exploratory data analysis, statistical modeling, data mining, Power BI, Plotly, Streamlit, data visualization, business intelligence

Cloud & Infrastructure: AWS (Lambda, DynamoDB, API Gateway, S3, EC2, ECR, ECS), Azure (Fundamentals certified), Docker, Git, CI/CD

Data Engineering & Databases: ETL pipelines, ER modeling, schema design & normalization, MySQL, PostgreSQL, DynamoDB (NoSQL), serverless / DAO patterns

Web & Automation: Playwright, BeautifulSoup, automated data collection, workflow automation

Practices: Data quality & integrity, security & data privacy, stakeholder reporting, cross-functional collaboration

PROFESSIONAL EXPERIENCE

System Engineer — Data & Analytics — Tata Consultancy Services *Apr 2023 – Aug 2024*

- Built and maintained 2 Power BI dashboards from scratch — one for bug-report tracking, one for new-feature delivery progress — used by Microsoft stakeholders to make release decisions and eliminating manual update cycles entirely.
- Automated SQL querying and data collection workflows in Python, saving ~3 days of manual effort per reporting cycle while improving data quality, consistency, and pipeline reliability across the team.
- Conducted exploratory data analysis on 300+ test cases and 180+ defect records per cycle, applying statistical techniques to identify high-risk application areas and surface actionable insights for Engineering leads.
- Collaborated with a cross-functional team of 8 across QA, Development, and Operations — aligning on data sources, communicating findings, and driving product-quality improvements across Microsoft Office applications.
- Structured test metrics into clear, stakeholder-ready visualizations for client presentations, ensuring results were accurate and directly supported business decisions.

Associate System Engineer — Tata Consultancy Services *Jul 2022 – Apr 2023*

- Led 2 analysts to coordinate test execution and on-time QA delivery, mentoring on analytical workflows, documentation practices, and data-quality standards.
- Implemented automated testing workflows that reduced manual data collection effort, improved consistency of results, and enabled faster defect-pattern analysis.
- Analyzed defect trends and generated actionable recommendations for developers and QA leads, contributing to measurable improvements in system reliability and issue-resolution speed.

ASE Trainee — Tata Consultancy Services *Apr 2021 – Jun 2022*

- Executed 300+ structured test cases across Apple-based environments, documenting results to support data-driven QA reporting and building foundations in structured problem-solving and enterprise-scale data validation.

KEY PROJECTS

Phishing Detection System — MS Capstone — University of Arizona *2024 – 2025*

- Designed and built a containerized 4-layer analysis pipeline from scratch: ML triage (Logistic Regression, Random Forest, XGBoost, LightGBM), live browser capture via Playwright, HTML/DOM feature analysis, and a deterministic Evidence Adjudication Layer for ambiguous cases.

- Selected LightGBM as the primary model based on F1/precision trade-offs across 50K+ URL samples; used domain-aware dataset splitting to eliminate data leakage and ensure real-world robustness.
- Integrated LLMs for brand-impersonation detection from webpage content and screenshots; engineered URL-based features including domain signals and hosting patterns.
- Built a real-time Streamlit dashboard showing per-model probability scores, feature diagnostics, and explainable verdict reasoning; containerized with Docker and deployed to AWS ECR/ECS.
- Debugged and resolved cold-start latency and browser-sandboxing issues that only surface in production — shipped, not just designed.

VIP Cohort Analytics Platform — Cloud Systems — University of Arizona2024 – 2025

- Built Lambda-based serverless ETL workflows with DynamoDB DAO access patterns and indexing strategies within a CI/CD multi-environment deployment pipeline.
- Mapped an end-to-end production-scale AWS analytics architecture (API Gateway, Lambda, DynamoDB, containerized ETL) across automated collection, processing, and storage layers to optimize scalable data-system design.

BN Motors — Relational Database & Analytics System — Personal Project2024

- Designed an end-to-end dealership relational database from scratch — full ER modeling, schema design, and normalization — then authored complex SQL (joins, aggregations, subqueries, window functions) for inventory, sales, and customer analytics.
- Built analytics-driven reporting views that translated raw relational data into actionable business metrics for decision-makers.

Bank Loan Default Prediction — Academic Project2024

- Performed exploratory data analysis on real-world financial data, identifying statistical correlations and data-quality issues; applied feature engineering on economic variables to reduce dimensionality.
- Trained and benchmarked multiple classification models (Logistic Regression, Random Forest, XGBoost) on imbalanced data, selecting the optimal approach on ROC-AUC, precision, and recall while balancing false-positive/false-negative trade-offs.

EDUCATION

MS, Information Science — Machine Learning — University of ArizonaJul 2024 – May 2026

GPA 3.89 | Coursework: Machine Learning, Neural Networks, Data Mining & Discovery, Applied NLP, Data Analysis & Visualization, SQL/NoSQL Databases, Foundations of Data Science, Data Science for the Public Interest

BE, Mechanical Engineering — SEA College of Engineering & Technology2016 – 2020

CERTIFICATIONS

Microsoft Certified: Azure Fundamentals • Microsoft Word Expert (Microsoft 365) • Microsoft PowerPoint (Microsoft 365) • Java Certification • Programming Foundations: Software Testing / QA

LANGUAGES

English (Native / Bilingual) • Kannada (Native / Bilingual) • Spanish (Elementary)