

AMAAN ALI KHAN BARKATH

+1-315-575-8003 | amaanali0477@gmail.com | linkedin.com/in/amaan-ali-khan/ | github.com/ak0477

EDUCATION

Syracuse University

Master of Science in Applied Data Science, GPA: 3.714

Syracuse, New York

Aug 2024 – May 2026

- **Relevant Coursework:** Machine Learning, Deep Learning, Building Human-centered AI Applications, Responsible AI

SRM Institute of Science and Technology

Bachelor of Technology, Computer Science and Engineering, GPA: 3.44

Chennai, India

Sep 2020 – May 2024

- **Relevant Coursework:** Statistics, Linear Algebra, Probability Theory, Data Structures and Algorithms, Database Systems

RELEVANT SKILLS

Programming & Tools: Python (pandas, NumPy, scikit-learn, TensorFlow, Keras), SQL, R, Git, Jupyter Notebooks, Bash

Machine Learning & AI: Supervised & Unsupervised Learning, Time Series Forecasting (Prophet, ARIMA, SARIMAX), Regression, Decision Trees, Neural Networks, NLP Fundamentals

Data Engineering & Databases: SQL, MySQL, Data Pipelines, ETL, Data Cleaning, Data Modeling

Visualization & BI Tools: Power BI, Tableau, Excel (PivotTables, VBA, Solver), Looker

Statistical Methods: A/B Testing, Hypothesis Testing, Regression Analysis, Predictive Modeling, MAPE/RMSE Evaluation

Cloud & Infrastructure: AWS (S3, EC2), Docker, CI/CD Basics

EXPERIENCE

Data Science Intern, The Fishin' Company, Remote, Pennsylvania

June 2025 – Dec 2025

- Reduced potential inventory waste by \$150K annually as measured by achieving 21% MAPE forecast accuracy, by architecting an end-to-end weekly demand forecasting pipeline using Prophet across 12 locations, 31 species, and 259 SKUs.
- Enabled supply chain team to potentially reduce stockouts by 27% as predicted by inventory tracking metrics, by designing interactive Power BI dashboards tracking forecast accuracy and demand trends.
- Validated optimal model selection as measured by cross-validation benchmarks, by experimenting with ARIMAX and SARIMAX models and conducting hyperparameter tuning against Prophet baseline.
- Delivered scalable data solutions aligned with business goals as demonstrated by weekly leadership presentations, by collaborating with cross-functional stakeholders to translate requirements into actionable insights.

Data Analyst Intern, Zero Meat, Chennai, India

May 2023 – Aug 2023

- Increased marketing ROI by 18% as measured by campaign attribution analysis, by developing customer segmentation models using K-Means clustering in Python on 20K+ transaction records to identify 4 high-value cohorts for targeted outreach.
- Reduced executive reporting time by 40% as measured by weekly hours saved, by building automated Power BI dashboards integrating sales, marketing, and customer behavior data with scheduled refresh pipelines.
- Improved promotional conversion rates by 12% as measured by A/B test statistical significance ($p < 0.05$), by designing and analyzing controlled experiments on campaign creatives and providing data-driven recommendations to marketing leadership.
- Saved 8+ hours of manual work weekly as measured by process efficiency gains, by engineering Python scripts to automate ETL workflows for weekly sales reporting and data validation.

ACADEMIC PROJECTS

RAGception – Research Paper Discovery Bot | Python, OpenAI API, SQLite, Streamlit, arXiv API

Dec 2024

- Automated discovery of 100+ RAG/LLM research papers weekly as measured by ingestion logs, by engineering an arXiv API integration pipeline that fetches, filters, and stores papers based on configurable keyword queries.
- Enabled semantic search across 500+ paper chunks with sub-second response time as measured by query latency, by implementing OpenAI text-embedding-ada-002 embeddings stored in SQLite with cosine similarity retrieval.
- Improved search relevance by 30% as measured by user feedback, by designing a hybrid search system combining vector similarity with keyword matching for better recall on technical queries.
- Deployed production web application on Streamlit Cloud serving research teams, by collaborating with 3 teammates on modular architecture integrating data pipeline, summarization, knowledge graph, and UI components.

FormAI – Real-Time Exercise Form Assessment | Python, MediaPipe, BiLSTM, OpenCV, NumPy

Dec 2024

- Achieved 80% classification accuracy distinguishing good vs bad squat form as measured by F1 score of 0.80 on validation set, by training a Bidirectional LSTM model on temporal sequences of 8 biomechanical features extracted from 500+ exercise videos.
- Enabled real-time pose analysis at 10 FPS as measured by inference throughput, by building a computer vision pipeline using MediaPipe Pose to extract and normalize 33 body landmarks from video frames.
- Classified 22 distinct exercise types with 49% accuracy under severe class imbalance as measured by held-out validation metrics, by engineering temporal sequence modeling with BiLSTM on pose-derived feature vectors.
- Improved user safety and feedback interpretability as measured by 100% correct classification on unseen real-world videos, by implementing rule-based threshold checks on knee angle, hip depth, and torso tilt that flag specific form errors.