

# Rishab Khatokar R

📍 Tucson, Arizona    ✉ [rishab.khatokar@gmail.com](mailto:rishab.khatokar@gmail.com)    📞 +1(520)205-0384  
🌐 [in linkedin.com/in/rishab-khatokar-90286b344](https://www.linkedin.com/in/rishab-khatokar-90286b344)    🌐 <https://rish742.github.io>

## ABOUT ME

Aspiring Data Scientist passionate about using data-driven approaches to solve real-world problems. Graduate student in Information Science (Machine Learning) with hands-on experience in statistical modeling, deep learning, and predictive analytics. Skilled in Python, SQL, and machine learning frameworks like Scikit-learn and TensorFlow. Experienced in building end-to-end data pipelines, performing feature engineering, and deploying models on cloud platforms (AWS, Azure). Strong communicator with a growth mindset, blending analytical rigor with practical deployment skills.

## EDUCATION

### Master of Science in Information Science (Machine Learning)

2023 - 2025

University of Arizona, Tucson, USA

GPA: 3.78

- *Related coursework:* Applied Natural Language Processing, Data Analysis & Visualization, Data Warehousing, Artificial Intelligence, and Neural Networks

### Bachelor of Engineering in Electronics and Communication

2018 - 2022

Dayananda Sagar University, Bengaluru, India

CGPA: 7.42

- *Related coursework:* VLSI, Embedded Systems, Probability and Statistics

## TECHNICAL SKILLS

- **Programming Languages:** Python, R, C, C++, Java, SQL
- **Machine Learning & Data Science:** XGBoost, Scikit-Learn, Pandas, NumPy, TensorFlow, MATLAB
- **Cloud, MLOps & DevOps:** Microsoft Azure, AWS, GitHub Actions, CI/CD, Infrastructure as Code(IaC)
- **Data Engineering & Databases:** Data Warehousing, SQL, NoSQL, Vector Databases (Weaviate)
- **Tools & Technologies:** Tableau, CUDA, Excel, Jupyter notebooks

## PROJECTS

### AI Polypharmacy Risk Predictor – Capstone Project

- Integrated structured healthcare datasets (Synthea, DrugBank, RxNorm) and applied BioBERT for semantic alignment and drug interaction context extraction.
- Developed a machine learning pipeline for adverse interaction prediction, achieving 85% baseline accuracy, later improved to 90% through advanced feature engineering.
- Deployed the interactive application using Streamlit, enabling real-time drug risk visualization and analysis for end users.

### AI Telematics System

- Designed and implemented a real-time ML system for driver drowsiness detection using a synthetic telematics dataset, achieving 94% simulated detection accuracy based on eye closure duration and motion sensor patterns.
- Integrated Kafka-based streaming from Raspberry Pi devices, improving data throughput by 50% over traditional polling methods.
- Built and deployed a full-stack edge-to-cloud pipeline (Streamlit, Supabase), reducing data processing time by 45% end-to-end. Demonstrated scalable infrastructure with cloud integration and real-time visualization, maintaining latency during testing.

### Sequence Labeling for Scientific Quantity Recognition (MeasEval - SemEval 2021)

- Developed a Bi-LSTM-CRF model with pre-trained GloVe embeddings for sequence labeling, extracting quantities and measurements from scientific texts.
- Achieved an F1-score of 0.82 with 86% precision on the MeasEval dataset, leveraging deep learning, NLP, and TensorFlow to enhance scientific information extraction and data mining.

### Fine-Tuning RoBERTa for Commonsense Validation: SemEval-2020 Task

- Fine-tuned RoBERTa with Hugging Face Transformers achieving 92.9% accuracy (Subtask A) and 92.8% accuracy (Subtask B).
- Conducted hyperparameter tuning and evaluated commonsense reasoning on large-scale datasets.

## EXPERIENCE & LEADERSHIP

### IEEE Club - MDC Chair at Dayananda Sagar University

- Led technical initiatives and workshops to promote innovation in electronics and communication engineering.
- Organized seminars, hackathons, and research discussions to improve peer learning and collaboration.

## CERTIFICATIONS

- Career Essentials in GitHub Professional Certificate – GitHub (Issued Mar 2025)
- Data Quality: Core Concepts – LinkedIn (Issued Mar 2025)
- DevOps Professional Certificate – PagerDuty & LinkedIn (Issued Mar 2025)
- Vector Databases Professional Certificate – Weaviate (Issued Mar 2025)