Krishna Thakar

krsnathkr.github.io | (858)-408-5293 | krsnathkr@gmail.com | LinkedIn | GitHub

Education:

Southeast Missouri State University

Masters in Computer Science - 4.0 GPA — May 27'

Bachelors in CS w/ Minor: Data Science - 3.9 GPA — May 25'

Coursera Certifications:

Supervised Machine Learning: Regression and Classification - Deeplearning.AI <Link> — Feb 24'

Exploratory Data Analysis for Machine Learning – IBM <Link> — June 24'

Job Experience:

GA Web Developer – SEMO — Aug 25' – Present

- Developing reusable web components and custom page templates from scratch using HTML, CSS, JavaScript, and VTL.
- Maintaining and updating the **official university website - semo.edu**, using the Cascade CMS.

Machine Learning Researcher - SEMO <Research Link> — Jan 25' – April 25'

- Worked under the mentorship of Dr. Mohamed Abu Sheha and Dr. Emmanuel Thompson, comparing traditional ML models and deep learning models against fine-tuned RoBERTa for three-way sentiment classification.
- Processed ~7 million Yelp reviews, performing **text cleaning, tokenization, lemmatization, negation handling, vectorization**, and balancing sentiment classes into equal thirds.
- **Developed and evaluated multiple models** - including Logistic Regression, SVM, Naïve Bayes, Random Forest, BiLSTM, LSTM, CNN, RNN, GRU, and a fine tuned RoBERTa - using 5-fold cross validation, confusion matrices, and ROC curves to measure accuracy, precision, recall, F1, and ROC AUC.
- Demonstrated that RoBERTa achieved top performance (accuracy 0.80, AUC 0.93) through systematic cross-validation and in-depth analysis.

Projects:

StealthChess.AI – Secret Cheating Chess Assistant |YOLOv8, Stockfish, Meta Ray-Ban <Link> — April 25'

- Made a real-time chess AI that analyzes live streams from Meta Ray-Ban smart glasses, detecting boards and pieces with YOLOv8 and whispering Stockfish recommended moves into an earpiece.
- Improved detection accuracy by 30% by training custom **YOLOv8** models and implementing precision grid mapping for live, noisy video feeds.
- Designed and optimized a **full-stack ML pipeline** for real-time inference, including model deployment, live video preprocessing, object detection, and post-processing for actionable decision-making

ASA DataFest 2025 – Winner, Best Use of Statistical Analysis | R, RStudio: <Link> — April 25'

- Worked with a real-world dataset of 200,000+ U.S. office lease transactions (2018–2024) from Savills, focusing analysis on tech-sector leases larger than 10,000 sq. ft. to extract meaningful business insights, and framed meaningful research questions
- Conducted exploratory data analysis and built **Logistic Regression, Decision Tree, and Random Forest models** in R to predict tenant movement; identified space type, lease year, and square footage as top predictors.
- Presented findings and model outcomes on stage with a team of 4, earning **Best Use of Statistical Analysis** at ASA DataFest 2025 (18-team, 24-hour competition) for impactful insights and model interpretability.

Knowbl – Agentic RAG Assistant with Real-Time Web + PDF | LangChain, MCP, BeautifulSoup: <Link> — April 25'

- Built an agentic RAG pipeline combining real-time web search and document Q&A, using FAISS vector indexing and OpenAI embeddings to return citation-backed answers; handled 10+ queries in local tests with consistent <2s response time.
- Designed an **async ingestion and chunking pipeline** (10K tokens w/ 500 overlap) to preprocess and embed content into a vectorstore, enabling scalable retrieval from documents up to 100+ pages.
- Built an MCP-compatible backend with FastMCP for tool modularity and API-style interaction, supporting flexible web data retrieval and advanced filtering

Technical Skills:

- **Programming Languages**: Python, R, Java, SQL (PostgreSQL, MySQL), MongoDB
- **Frameworks & Libraries**: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly, Streamlit, OpenCV, NLTK, Hugging Face Transformers, TensorFlow, PyTorch, Keras, CUDA, LLM, FAISS, LangChain, Chroma, FastAPI, ETL
- **Tools and Platforms**: Snowflake, Databricks, AWS (EC2, S3), Azure, Docker, Kubernetes, Git, Linux, CI/CD pipelines, Airflow.