Krishna Thakar

krsnathkr.github.io | (858)-408-5293

krsnathkr@gmail.com | LinkedIn | GitHub

Education:

Undergraduate in Computer Science w/ Minor: Data Science — May 25'

Southeast Missouri State University — Cape Girardeau, MO

GPA – 3.9/4.0

Presentations:

Thakar, Krishna. Sentiment Analysis of Yelp Review Dataset: A Comparative Study of Machine Learning Methods. April 15, 2025.

This research investigates the effectiveness of various machine learning and deep learning models for sentiment classification on Yelp review data. The study compares traditional approaches like SVM and Random Forest with deep learning architecture such as CNN, BiLSTM, and RoBERTa. Results showed that RoBERTa outperformed other models in classification accuracy. The project was conducted under the mentorship of Dr. Mohamed Abu Sheha and Dr. Emmanuel Thompson and was presented at the 33rd Student Research Conference at Southeast Missouri State University.

Affiliation: Southeast Missouri State University

Thakar, Krishna. From War Machine to Peacemaker: Duality of AI in Geopolitics. April 16, 2024.

This paper explores the dual roles of AI in warfare and peacebuilding. It highlights how AI-powered autonomous weapon systems raise ethical concerns and how AI can support peace negotiations and conflict resolution. The conclusion emphasizes the need for responsible AI development and international collaboration to leverage AI for global peace. This research, conducted under the guidance of Dr. Reshmi Mitra, is currently in progress for submission and was presented at the Student Research Conference at Southeast Missouri State University.

Affiliation: Southeast Missouri State University

Thakar, Krishna. Capstone Project – SEMO Esports Website. May 2025.

This capstone project developed a full-stack web platform to address overcrowding and booking inefficiencies in Southeast Missouri State University's Esports arena, which previously had no digital system in place. Built using React, Node.js, Express, MongoDB, Firebase Authentication, and Tailwind CSS, the platform enabled real-time station reservations, availability tracking, and administrative control. Used by over 150 students, the system reduced peak-hour congestion by more than 70% and saved staff an estimated 5+ hours per week by replacing manual scheduling processes. The completed project was presented to the entire Computer Science department and executives from local companies, showcasing its practical impact and real-world value to campus operations.

Affiliation: Southeast Missouri State University

Tic-Tac-Toe Game

Presented an interactive Tic-Tac-Toe game in front of the class, showcasing its engaging two-player functionality and innovative features. Developed the game using HTML, CSS, and JavaScript, with a dynamic, responsive game board and real-time score tracking utilizing local Storage to persist scores across sessions. Incorporated a dark mode toggle to enhance user experience, allowing seamless switching between light and dark themes. Utilized modular code to efficiently manage player turns, validate winning combinations, and handle draw scenarios, ensuring smooth and intuitive gameplay. The presentation highlighted the project's design, functionality, and user-centric enhancements.

Job Experience:

Student Researcher - Southeast State Missouri University <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 from 2005 - 2022, 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 stratified cross validation, confusion matrices, and ROC curves to measure accuracy, precision, recall, F1, and ROC AUC.
- Demonstrated that RoBERTa achieved top performance (accuracy 0.80112, AUC 0.93237) through systematic cross-validation and in-depth analysis.

Resident Assistant - Southeast State Missouri University Jan 23' – May 25'

- Cultivating effective communication and mentoring skills by guiding 100+ residents through academic and personal challenges, offering tailored advice, and connecting them with campus resources.
- Demonstrating strong organizational abilities by planning and executing 10+ events annually, ensuring seamless coordination and active participation.
- Strengthening interpersonal skills by fostering a collaborative and inclusive environment, resolving conflicts, and providing ongoing support to residents.

Information Technology Staff – Southeast State Missouri University Sept 23' – Jan 25'

- Provided front-line technical support to 100+ students and faculty, resolving issues with software, printing, and system functionality at an on-campus IT help desk.
- Diagnosed and troubleshot Windows/macOS PCs, maintaining operational efficiency across 50+ public access computers and reducing downtime by 25%.
- Delivered prompt walk-up support with an average resolution time under 10 minutes for common technical problems.
- Documented incidents and resolutions in the university's ticketing system, improving tracking and future troubleshooting efforts.

Concession Stand Worker - Southeast State Missouri University Nov 21' – May 25'

- Served 200+ customers per shift during peak campus events, providing fast, friendly, and accurate service in a high-volume environment.
- Maintained inventory of food and beverage supplies, reducing stock shortages through proactive restocking and organization.
- Ensured food safety and cleanliness standards were consistently met, contributing to successful health inspections with 0 violations.

Projects:

Sales & Customer Data Pipeline | Azure Data Factory, Databricks, dbt, Apache Spark, ADLS Gen2: <Link> July 25'

- Engineered a sales & customer data pipeline using dbt and Spark with a Medallion architecture, reducing analytics reporting time by 40% and processing over 1M+ records.
- Developed PySpark transformations in Azure Databricks to cleanse, dedupe, and enrich data across bronze, silver, and gold layers, delivering sub-5-minute SLA processing and reducing compute costs by 30%.
- Modeled and versioned 15+ data assets with dbt on Databricks - incorporating snapshots, automated tests, and lineage documentation - to support reliable, self-service BI reporting.

Knowbl – Agentic RAG Assistant with Real-Time Web + PDF | LangChain, MCP, BeautifulSoup: <Link> June 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.
- Integrated Exa API to retrieve 5-10 web results per query, scraped content using BeautifulSoup, and embedded the text using OpenAI Embeddings into a FAISS vectorstore for fast retrieval.
- Built an MCP-compatible backend with FastMCP for tool modularity and API-style interaction, supporting flexible web data retrieval and advanced filtering
- Processed live and static data into 10K-token chunks with 500-token overlap, enabling accurate semantic retrieval from web and PDFs without noticeable lag.

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 technology-sector leases larger than 10,000 sq. ft. to extract meaningful business insights.
- Framed the research question: "What factors influence whether a tenant relocates to a new office or remains in the same space (through renewal, expansion, or restructure)?" Focused the scope on tech-sector leases larger than 10,000 sq. ft. to uncover actionable business insights.
- 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.

Student Research Conference 2025 – Presenter, Sentiment Classification Using ML & Deep Learning: <Link> April 25'

- Worked under the mentorship of Dr. Mohamed Abu Sheha and Dr. Emmanuel Thompson, comparing traditional ML models (SVM, Random Forest) and deep learning models (CNN, BiLSTM) against fine-tuned RoBERTa on a 20,000-sample Yelp review dataset for three-way sentiment classification.

- Improved training and evaluation time by ~30% in local runs by designing an efficient text preprocessing pipeline and applying TF-IDF, Word2Vec, and BERT embeddings selectively based on model type.
- Focused on model architecture trade-offs, evaluation metrics like AUC and F1, and demonstrated experience tuning both classical ML and deep learning models for large-scale NLP tasks.

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

- Engineered 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

SEMO Esports Website | React, Node.js, PostgreSQL, Express, Tailwind, Git: <Link> June 24'

- Identified a critical gap in Esports resource management by observing disorganized, overcrowded station usage - conceptualized and built SEMO's first-ever Esports web platform from scratch, digitizing operations that previously had no online presence.
- Developed and deployed a full stack booking system using AWS, React, Node.js, and PostgreSQL, enabling over 150+ students to reserve gaming stations in advance, reducing on-site congestion by over 70% during peak hours.
- Integrated Firebase Authentication and role-based access controls, streamlining user access and empowering staff to manage reservations, usage stats, and maintenance through a dedicated admin panel.
- Accelerated daily operations, replacing inefficient spreadsheets and manual logs - saving staff 5+ hours per week and increasing booking reliability and student satisfaction.

Echo Chamber | Python, Streamlit, NewsAPI, NLTK, Sklearn: <Link> June 24'

- Tackled the challenge of combating news bias and information silos by building The Echo Chamber, a Python- and Streamlit-based platform integrating sentiment analysis (98% accuracy) and TF-IDF-based article recommendations; resulted in a 90% user satisfaction rate and over 200 active users within the first month.
- The platform uses NewsAPI to fetch articles, NLTK for sentiment analysis, and a TF-IDF-based recommendation system in Python to provide users with accurate, personalized news suggestions.

Market Radar | Streamlit, Plotly, yfinance, Google Gemini API: <Link> June 24'

- A Streamlit application integrating **Google Gemini API** and **Yahoo Finance** to deliver real-time financial insights and dynamic stock visualizations, reducing analysis time by 40%. 50+ users to make informed investment decisions with features like key financial metrics, analyst ratings, and interactive stock trends.
- Optimized data fetching and visualization processes, improving system performance by 30% and enhancing user engagement through an intuitive interface.

Penny Planner | Streamlit, Plotly, OpenAI API: <Link> July 24'

- Identified a gap in personal finance tools by enabling users to upload and analyze their bank statements for real-time income, expense, and transaction insights—solving the lack of accessible financial clarity in everyday spending.
- Analyzed 100+ bank statements, generated interactive visualizations, and delivered AI-powered insights via OpenAI API, improving financial clarity and literacy for users by 90%, as reflected in satisfaction surveys.

Titanic Survival Predictor | Python, NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn: <Link> January 24'

- I solved the problem of predicting survival rates on the Titanic using machine learning algorithms, achieving an accuracy score of 83.24%.
- Conducted data cleaning, feature engineering, and model tuning to manage missing data and enhance accuracy and addressed issues of data imbalance and feature relevance, leading to a substantial improvement in model performance.

New York Housing Price Analysis | Python, Pandas, Scikit-learn, Matplotlib: <Link> June 24'

- Analyzed New York housing prices using Python, uncovering a 78% correlation between property size and price and identifying premium neighborhoods with 20% higher average prices through geospatial and statistical visualizations.
- Improved price prediction accuracy by 30% through advanced data preprocessing and exploratory analysis.
- Developed interactive visualizations with Plotly, resulting in a 25% increase in data interpretability and actionable insights.

Tic-tac-toe Game | JavaScript, HTML, CSS: <Link> April 24'

- The Tic-Tac-Toe game delivers an engaging gameplay experience with real-time feedback and interactive elements. Players can compete in a friendly match while tracking their scores across multiple rounds. The dark mode feature ensures accessibility and modern aesthetics, making the game enjoyable for a broad audience.

- Overcame challenges related to real-time score updates and ensured compatibility across multiple devices, maintaining smooth functionality on various screen sizes.

Multi-Port Client-Server Application | Flask, Python, Socket Programming, Requests: <Link> May 24'

- Developed a fault-tolerant distributed system with Python, ensuring seamless client-server communication across 3 server instances (ports 8080, 8081, 8082).
- Designed robust mechanisms for client registration, data sharing, and fault tolerance, automatically reassigning clients during server failures. Enabled over 90% reliability in message delivery across 6 clients, validated by comprehensive unit tests achieving 100% pass rate on fault handling scenarios.

Academic Excellence and Recognition:

- International Excellence Award: 50% scholarship at SEMO - 4 times in 4 years
- Dean's List – Spring 22', Fall 22', Spring 23', Fall 23', Spring 24' & Fall 24'
- President's List - Fall 22', Spring 23', Fall 23', Spring 24' & Fall 24'
- First Year Leadership Program Certificate
- Graphic Designer of International Student Association

Certifications:

Supervised Machine Learning: Regression and Classification - Deeplearning.AI Feb 24'

Exploratory Data Analysis for Machine Learning – IBM June 24'

Skills:

Python	NLTK	Databricks
R	Hugging Face	AWS (EC2, S3, PVS)
Java	Transformers	Azure
SQL (PostgreSQL, MySQL)	PyTorch	Docker
MongoDB	Keras	Kubernetes
Scikit-learn	CUDA	Git
Matplotlib	LSTM	Linux
Seaborn	BiLSTM	CI/CD pipeline
Plotly	CNN	Hadoop
Streamlit	RAG LLM	Apache Spark
OpenCV	Snowflake	Kafka