

PAVAN KUMAR KOKKILIGADDA

pavan0417.k@gmail.com | +1 469 207 6377 | linkedin.com/in/machaax | machaax.netlify.app

Education

Rutgers University-New Brunswick

Master of Science in Computer Science – CGPA: 3.65 / 4.0

Jan 2024 - Dec 2025

New Brunswick, New Jersey

Indian Institute of Technology, Kharagpur

B.Tech & M.Tech (Hons.) in Engineering – CGPA: 7.32 / 10.00

Aug 2018 - May 2023

Kharagpur, India

Experience

Rutgers University—New Brunswick : ComLab

Research Assistant / Supervisor: Prof. Matthew Weber

Aug 2024 - Present

New Brunswick, New Jersey

- Extracted news articles from U.S. outlets using Newspaper3K and Trafilatura on A100 GPUs of Rutgers HPC (Amarel).
- Developed a multi-stage PySpark pipeline to deduplicate 200,000 news articles across 600 domains using shingling + MinHash LSH and to extract and validate geopolitical entities (GPE) with spaCy and Stanza.
- Implemented BERTopic (BERT) unsupervised topic modeling framework, achieving an average coherence score of 0.75.
- Tested and deployed GPT models via the OpenAI API for article summarization and fine-tuned BART, BERT, and DeBERTa -based models from Hugging Face for zero-shot news classification across 10 topics using consensus labeling.
- Created domain-grouped topic and GPE visualizations using folium and Google V3 API, to enhance stakeholder insights.

Verzeo

Machine Learning Intern

May 2022 - Jun 2022

Bangalore, India

- Developed real-time intrusion-notification system using OpenCV + YOLO (Ultralytics), tracking, and face embeddings.
- Reduced false alarms via NMS and IoU/confidence thresholds; unknown faces triggered SMTP email alerts to users.
- Built a sentiment analysis model using SVM and TfidfVectorizer to classify restaurant reviews; achieved 91% accuracy.

InsAnalytics

Data Analytics Intern

May 2021 - Aug 2021

Kolkata, India

- Built an SEIR epidemiological model to forecast COVID-19 cases; estimated parameters and validated with backtesting.
- Developed ETL/EDA pipelines using Pandas, NumPy, SciPy to clean/merge time-series data, handle missing values and outliers, and perform hypothesis tests; delivered data visualizations and dashboards using Matplotlib, Seaborn, Plotly.

Projects

Text-to-SQL Ensemble Supervised Fine-Tuned LLM Model | Class Project

Feb 2025 - Apr 2025

- Supervised fine-tuned the DeepSeek-R1-Distill-Llama-70B model alongside an ensemble of 5 small models (Llama 3.2, Google Gemma3, Mistral, Qwen, and Nvidia-Nemotron) on the Bird SQL Workbench and Spider text-to-SQL datasets.
- Integrated RLHF, rejection sampling, and consensus-based ensemble verification to generate and refine SQL queries.
- Achieved 86% execution accuracy on the Spider benchmark and 60% execution accuracy on the Bird benchmark.

Online Railway Ticket Booking System | Class Project

Sep 2024 - Nov 2024

- Developed an online railway booking system allowing users to search train schedules and make reservations.
- Implemented functionalities like one-way and round-trip reservations, schedule queries, and user authentication.
- Used HTML, CSS, and JavaScript for the front-end, Java and JDBC for server-side processing, and MySQL for database management with Apache Tomcat as the web server.

Secure Patient Data Management System | Class Project

Sep 2024 - Nov 2024

- Developed a full-stack web application enabling patients, doctors, and nurses to securely share and manage health data with role-based access control (RBAC) using Flask web framework and hosted MySQL database on Azure.
- Implemented RSA and AES encryption to secure sensitive patient information during transmission and storage.

Technical Skills

Languages: Python, C/C++, Java, JavaScript, HTML, CSS, MySQL

Tools/Frameworks: Flask, ReactJS, JSP, JDBC, Apache Tomcat, MongoDB/Mongoose, Git/Github, Azure

Libraries: scikit-learn, PySpark, spaCy, Stanza, NLTK, transformers, BERTopic, Gensim, PyTorch, OpenCV, YOLO, SciPy, Matplotlib, Seaborn, Plotly, Folium, GeoPy, GeoPandas, Newspaper3k, Trafilatura, langdetect, Beautiful Soup

Achievements

- All India Rank 4 in All India Maths Science Talent Examination 2014-2015 - Gold Medal.
- Solved 300+ problems on LeetCode and 3-Star competitive programmer on CodeChef.