Pavan Kumar Kokkiligadda

🌐 Portfolio-Page 📞 469-207-6377 ✉ pavan0417.k@gmail.com in linkedin.com/in/machaax

Education

Rutgers University-New Brunswick   **January 2024 – December 2025**
Master of Science in Computer Science – CGPA: 3.67 / 4.0   *New Brunswick, New Jersey*

Indian Institute of Technology, Kharagpur   **August 2018 – May 2023**
B.Tech & M.Tech (Hons.) in Mining Engineering – CGPA: 7.32 / 10.00   *Kharagpur, India*

Experience

ComLab : Rutgers University-New Brunswick ⬈   **August 2024 – Present**
Research Assistant — Supervisor: Prof. Matthew Weber   *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 data analysis pipeline using PySpark to extract and identify geopolitical entities from thousands of news articles across 600+ domains, utilizing spaCy and Stanza for named entity recognition.
- 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   **May 2022 – June 2022**
Machine Learning Intern   *Bangalore, India*

- Developed a Theft prevention system that alerts users of intrusion into their house/property using computer vision.
- The system distinguishes family members as safe and sends an email with captured images if an intruder is detected.
- Built Sentiment Analysis machine learning model using SVM and TfidfVectorizer to classify restaurant reviews as positive or negative, achieving an accuracy of 82%.

InsAnalytics   **May 2021 – August 2021**
Data Analytics Intern   *Kolkata, India*

- Implemented SEIR model (susceptible, exposed, infectious, recovered) to forecast COVID-19 cases Trends.
- Played an essential role in collecting, cleaning, preparing data for further analysis, and creating visualizations.

Projects

Text-to-SQL Ensemble Supervised Fine-Tuned LLM Model | *Class Project*   **February 2025 – April 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-Nemetron) 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 72% execution accuracy on the Bird benchmark.

Online Railway Ticket Booking System | *Class Project*   **September 2024 – November 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*   **September 2024 – November 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
Libraries: Pandas, scikit-learn, PySpark, spaCy, Stanza, NLTK, transformers, BERTopic, Gensim, PyTorch, OpenCV, SciPy, Matplotlib, Seaborn, 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 ⬈.