Pavan Kumar Kokkiligadda

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

Education

Rutgers University-New Brunswick **January 2024 – December 2025**
Master of Science in Computer Science *New Brunswick, New Jersey*

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

Experience

Rutgers University-New Brunswick **August 2024 – Present**
Research Assistant *New Brunswick, New Jersey*

- 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 for named entity recognition.
- Implemented unsupervised topic modeling using BERT transformer and BERTopic framework, achieving an average topic coherence score of 0.75; created domain-grouped topic visualizations using folium, which enhanced stakeholder insights.
- Tested and deployed fine-tuned BART models from Hugging Face for Zero-Shot news classification on 10 broad topics.

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 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.

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 using Flask web framework and hosted MySQL database on Azure.
- Implemented RSA and AES encryption to secure sensitive patient information during transmission and storage.

Vibration Analysis of HydroCyclone | *Masters Thesis Project* **January 2023 – May 2023**

- Collected time-series vibration profile data from an Industrial HydroCyclone using accelerometers and a Raspberry Pi.
- Implemented Fast Fourier Transform Algorithm for denoising vibration signals, achieving a 6.15 dB increase in SNR.
- Discovered a positive correlation between feed inlet pressure and vibration intensity, quantified through RMS values.

Technical Skills

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

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

Libraries: NumPy, Pandas, scikit-learn, PySpark, spaCy, NLTK, transformers, BERTopic, Gensim, TensorFlow, PyTorch, OpenCV, SciPy, Matplotlib, Seaborn, Folium, GeoPy, Newspaper3k

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.
- All India Rank 429 in GATE-2022 Mining Engineering.