

# PAVAN KUMAR KOKKILIGADDA

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

## EDUCATION

### Rutgers University-New Brunswick

Master of Science in Computer Science – CGPA: 3.7 / 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

## CERTIFICATIONS

• Databricks Certified Data Engineer Associate

Nov 2025

## EXPERIENCE

### Rutgers University-New Brunswick: ROPES Lab

Research Assistant / Supervisor: Prof. Robert Mieth

Sep 2025 - Present

New Brunswick, New Jersey

- Mapping energy and power resources across the U.S. East Coast by building an interactive map tool in Plotly and Dash.
- Integrating a dynamic right-hand panel that displays real-time statistics and census data for user-drawn lasso selections.

### 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 data pipeline to deduplicate 200,000 news articles across 600 news domains using shingling + MinHash LSH technique 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 visualizations on Geocoded GPEs using Plotly and Places 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.

### InsAnalytics

Data Analytics Intern

May 2021 - Aug 2021

Kolkata, India

- Built an epidemiological time-series 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

### AI Agent for Medical Practitioner's Licensing & Credentialing Automation | Rutgers Hackathon

Oct 2025 - Oct 2025

- Automated provider data extraction and credentialing PDF filling using hybrid OCR/LLM pipeline (AWS Textract, Gemini API).
- Built an end-to-end web app for CV/resume parsing, required-field validation, auto-filling and outputting submission-ready PDFs.

### Memory-Enhanced Modular Entity for Noticing, Tracking, and Organization | Course Project

Jan 2025 - Apr 2025

- Built MEMENTO, a context-aware memory assistant over live home-security camera streams.
- Architected a VLM+LLM pipeline that chunks video, captions clips with a VLM, stores as text embeddings in vector database, and uses an LLM to retrieve events by text, position, and time; developed iOS app for adding streams and querying past/live events.

### Text-to-SQL Supervised Fine-Tuned LLM Model | Course Project

Jan 2025 - Apr 2025

- Fine-tuned DeepSeek-R1-Distill-Llama-8B model on Spider 1.0 Text-to-SQL dataset, achieving 86% execution accuracy.
- Applied reinforcement learning (RLEF) using GRPO with execution-aware rewards to improve generated SQL query correctness.

### Vibration Analysis of HydroCyclone | Masters Thesis Project

Jan 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, R

**Tools/Frameworks:** Apache Spark, Databricks, Flask, ReactJS, Git/Github, AWS, Azure, Docker

**Libraries:** scikit-learn, PySpark, spaCy, Stanza, NLTK, transformers, BERTopic, Gensim, PyTorch, LangChain, OpenCV, YOLO, SciPy, Matplotlib, Seaborn, Plotly, Dash, GeoPy, GeoPandas, Newspaper3k, Trafilatura, langdetect, BeautifulSoup, Google Maps API

## 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.