

PRADYUMNA BADA

pbada2@illinois.edu | (217) 200-3976 | [Portfolio](#) | [LinkedIn](#) | [Github](#) | Urbana, Illinois

EDUCATION

University of Illinois, Urbana - Champaign | GPA 3.8/4

May 2025

Master of Science, Advanced Analytics

Coursework: Machine Learning, Data Mining, Big Data, Deep Learning, **ML Cloud (AWS)**, Computer Vision

JSS Science and Technology University | GPA 8.5/10

Nov 2020

Bachelor of Engineering, Industrial Engineering

Coursework: Operations Research, Statistical Quality Control, Linear Algebra, Calculus, Energy Engineering, C Programming

SKILLS

Programming: **Python**, SQL, PostgreSQL, **Flask API**, Django, Docker, GitLab, Azure, AWS, GCP, NoSQL, **MongoDB**, Jenkins.

Data Science: Scikit-Learn, **PyTorch**, Sagemaker, Pandas, Prophet, Seaborn, LangChain, Hugging Face, **PySpark**, Airflow, Databricks.

Other Skills: Web Scraping, Prompt Engineering, Chain Of Thought Prompting, **Azure OpenAI Studio**, LucidChart, Agentic AI.

PROFESSIONAL EXPERIENCE

Cline Center for Advanced Social Research | Research Assistant

Sep 2023 - Present

- Leveraged **LLMs** to extract structured data from 1000+ news articles every day, specifically tracking police use of lethal force.
- Implemented a vector database (FAISS) for ranking, clustering and deduplication of articles using vector embeddings.
- Reduced manual processing time by **67%**, saving **80 hours/week** using **AI** workflows, enhancing accuracy & efficiency.

Seagate Technology | Data Analyst Intern

May 2024 - Aug 2024

- Implemented **ETL** pipelines to process and analyze high-volume data for characterizing emerging memories (FeRAM).
- Designed a Python-based workflow to extract data from a Postgres DB, transform it using **PySpark**, & generate automated reports.
- Developed predictive models for failure metrics to ensure reliability & Integrated CI/CD pipelines using GitLab & Jenkins.

RailTEC at Illinois | Summer Research Intern

May 2023 - Aug 2023

- Implemented a **Vision AI-based** tool for railway track inspection, potentially reducing time from nearly 2 weeks to just 3 hours.
- Utilized **Databricks**, Azure Synapse, & PySpark to analyze periodic data for tracking and **forecasting** rail component health.
- Developed Python-based data visualizations to provide actionable insights, optimizing decision-making processes.

Sapiens International Corporation | SQL Developer

Mar 2021 - Nov 2022

- Enhanced our Insurtech product using SQL & ETL, enabling £250k annual reduction in legacy insurance operational costs.
- Developed reporting interfaces using SQL and Shell Scripting, including 'ECASS', a cache facility designed for client's actuarial team.
- Crafted robust SQL scripts to resolve 50+ critical data issues, ensuring data accuracy across policy, claims, & customer datasets.

ACADEMIC PROJECTS

Predictive Modeling for Failure Prevention in Scania Truck Air Pressure System (APS) | Tools: Python, GitHub, MICE

Goal: Develop a cost-based model to predict and minimize failures, considering the costs of unnecessary checks & breakdowns.

- Performed EDA, & data imputation to develop various ML models, including LR, KNN, DT, RF & used Grid Search CV to compare it.

Natural Language Processing with Disaster Tweets | Tools: Pytorch, Pandas, RoBERTa.

Goal: Predicting disaster-related tweets to automate Twitter monitoring for disaster relief organizations and news agencies.

- Implemented various ML models & encoders, including BOW, tf-idf, GloVe, BERT, and RoBERTa to achieve 81.1% accuracy.

Stock Price Forecasting Using News Sentiment Analysis | Tools: Python, Streamlit, Prophet

Goal: Develop an interactive web app to forecast stock prices by integrating sentiment analysis of financial news.

- Used a pretrained model for sentiment analysis and Prophet for forecasting, delivering a dynamic dashboard with real-time plots.

Displacement and Strain Measurement using Computer Vision | Tools: Python, Pytorch, TorchVision.

Goal: To harness AI for the precise measurement of displacement and strain in objects under load, facilitating real-time monitoring

- Developing CNNs, and other DL models to detect displacement between pair of speckle images.

LEADERSHIP & INVOLVEMENT

- Ensured timely meal service for hundreds of students daily through food servicing, dish cleaning, and table sanitation.
- Led the development team in resolving high-priority production issue in the absence of senior team members.
- Organized multiple treks to the Great Himalayan Ranges for groups of more than 10.
- Collaborated on creative projects, including short films & travel vlogs. [Video1](#), [Video2](#), [Video3](#), [Video4](#),

CERTIFICATIONS

- [Prompt Engineering](#), [Transformer Architecture Behind LLMs](#), [Applied Plotting in Python](#), [Python Data Structures](#), [Databricks](#)