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: Data Visualization, Big Data, Machine Learning, Data Mining, ML Cloud, Network Optimization

JSS Science and Technology University | GPA 8.5/10 **Nov 2020**
Bachelor of Engineering, Industrial Engineering
Coursework: Statistical Quality Control, C Programming, Linear Algebra, Advanced Calculus, Operations Research, Metrology

SKILLS

Programming: Python, **Django**, Flask API, Javascript, SQL, **Jenkins**, GitLab, Linux, AWS, **Docker**, Apache Airflow**,** C Programming
Data Analytics: Pandas, Numpy, Matplotlib, Seaborn, PySpark, Databricks, Data Lakehouse, JMP, Tableau, Power BI.
Other Skills: JIRA, **Seagate Product Knowledge**, Lucid Charts, Web Scraping, Power Query, **Data Storage Systems**.

PROFESSIONAL EXPERIENCE

Seagate Technology | Engineering Data Analyst Intern **May 2024 – Aug 2024**
- Implemented ETL pipeline to process and analyze high-volume data for characterizing emerging memories (FeRAM).
- Calculated failure metrics, including Bit Error Rates (**BER**) across various address levels, and modeled failure rate (**CFR**) to determine how the error correction code (**ECC**) requirements relate to reliability thresholds.
- Designed a **Python**-based linter to automatically validate YAML configuration files by leveraging testers' domain expertise, reducing lengthy test re-runs.
- Developed unit tests using **Pytest** & integrated them into GitLab and **Jenkins** CI/CD pipelines, streamlining agile development.

Cline Center for Advanced Social Research | Research Assistant **Sep 2023 - Present**
- Engineered data pipelines to transform unstructured news articles into structured data, tracking police use of lethal force.
- Utilized **MongoDB** to ingest articles & developed automated pipelines to regularly convert them into structured **MySQL** databases
- Implemented scalable data ingestion workflows with Azure OpenAI Studio, optimizing analytics and automation processes.

RailTEC at Illinois | Summer Research Intern **May 2023 - Aug 2023**
- Implemented an automated railway track inspection process by leveraging Big Data Analytics and LRAIL, an AI-based system.
- Utilized Databricks, Azure Synapse, & **PySpark** to analyze periodic data for tracking and **forecasting** rail component **health**.
- Developed Flask-based **REST API** endpoints to automate real-time data delivery, enhancing analytics infrastructure efficiency.

Sapiens International Corporation | Software Developer **Mar 2021 - Nov 2022**
- Enhanced our Insurtech product using SQL & ETL, which helped companies administer legacy Insurance products.
- Developed reporting interfaces using SQL and Shell Scripting, including 'ECASS', a cache facility designed for client's actuarial team.
- Developed complex SQL data fix scripts and supported Agile processes using Jira to manage evolving requirements.

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.

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.

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.

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.
- Completed certification courses like Applied Machine Learning, Applied Plotting & Data Representation in Python, and more.
- 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,