Soyab Karki

662-352-1863 | sk2078@msstate.edu | linkedin.com/in/soyab-karki/ | github.com/SoyabKarki

EDUCATION

Mississippi State University   Starkville, MS
Bachelor of Science in Computer Science, Minor in Data Science and Mathematics   *Aug. 2023 – May 2027*
- **GPA:** 4.0

TECHNICAL SKILLS

Languages: Python, C/C++, JavaScript, Typescript, SQL, HTML/CSS
Libraries/Frameworks: Pandas, NumPy, Scikit-Learn, PyTorch, FastAPI, React.js, Express.js
Data Tools: Airflow, Kafka, Spark, AWS, PostgreSQL, Redis, Prometheus, Grafana, Tableau
Developer Tools: Git, Github, Docker, Kubernetes, Node.js

EXPERIENCE

Undergraduate Research Assistant   Aug. 2024 – Present
Mississippi State University   *Starkville, MS*
- Designed and implemented ETL pipelines for 300GB+ consumer review datasets using batch processing and Parquet storage, improving data processing speed and storage consumption by 10x.
- Automated filtering of 2,500+ research papers using OpenAI API and custom keyword matching in Python, reducing manual review time from days to minutes.
- Created publication-ready maps in ArcGIS Pro by visualizing spatial datasets for research projects, helping improve the clarity and impact of academic papers.

Student Worker   Jan. 2024 – May 2025
High Performance Computing Collaboratory   *Starkville, MS*
- Parallelized 10+ algorithmic problems using MPI in C++, demonstrating 5x performance gains over sequential code.
- Performed maintenance and assisted in building a $100,000 supercomputing cluster for the SC24 competition team, ensuring system stability and readiness for high-performance benchmarking tasks.

Data Scientist Intern   May 2024 – July 2024
United States Department of Agriculture - Agricultural Research Service   *Starkville, MS*
- Extracted and validated 1.5 years of meteorological data from Google Earth Engine using Python APIs and quality-control filters, enabling robust training datasets for machine learning models.
- Collaborated with 5 researchers to implement Gradient Boosting and Logistic Regression models in Python, improving baseline accuracy by 25% and providing insights to reduce crop contamination risk in field studies.
- Presented research findings at 3 regional conferences, clearly communicating complex ML workflows to academic and industry audiences, and receiving commendations for clarity and practical impact.

PROJECTS

Reddit Data Pipeline | *Python, Airflow, PostgreSQL, Docker, Prometheus, Grafana*   Aug. 2025
- Built a containerized ETL pipeline that pulls stock discussions from 3 Reddit forums, runs FinBERT sentiment analysis, enriches with news/stock data, and stores in PostgreSQL with materialized views for fast queries.
- Implemented Prometheus + Grafana dashboards for real-time pipeline metrics, improving system observability.

Resume Radar | *React.js, FastAPI, Chrome Extensions API, Redis, Docker*   July 2025
- Built a Chrome extension with FastAPI backend and Redis caching that analyzes resumes against job descriptions using OpenAI API, reducing keyword identification and system latency by 99.9%.

LEADERSHIP

Data Science Club at MSU | *President*   Feb. 2024 – Present
- Recruited 50+ members and raised $3,000+ to run datathons and workshops.

ACM SIG-HPC | *Founding Member*   Sep. 2023 – Present
- Built mini HPC clusters with Raspberry Pi + MPI to demonstrate parallel processing concepts.