

# Sai Karthik Kosuri

<https://www.linkedin.com/in/karthikkosuri9/> | +1-(680)-697-4150 | karthikjobasap@gmail.com | <https://karthik.herokuapp.com/>

## EDUCATION

**Syracuse University** - Syracuse, NY  
Master of Science in Computer Science  
Graduate Teaching Assistant

August 2022 - May 2024  
GPA: 3.8/4  
August 2022 - December 2023

**Vellore Institute of Technology** - Vellore, India

Bachelor of Technology, Computer Science ,  
Courses: Data Structures & Algorithms, Artificial Intelligence (AI), Computer Networking, Parallel & Distributed Computing

July 2017 - June 2021  
GPA: 8.49/10

## PROFESSIONAL EXPERIENCE

**Data Scientist**, University of Pennsylvania – Philadelphia, Pennsylvania, USA

June 2024 - Present

- **Developing a Retrieval-Augmented Generation (RAG) system for scientific claim validation under DARPA SciFY**, involving the construction of a scalable document datastore from scientific PDFs, automated parsing and OCR extraction, embedding generation using domain-specific LLMs, and integration into a reasoning pipeline for evidence-based validation.
- Developed and optimized a comprehensive audio and transcript analysis pipeline by applying LDA, LIWC, and n-gram techniques to textual data and generating correlation word clouds with DLATK. Implemented machine learning models, including SVM, RidgeCV, and Random Forest, achieving a 30% improvement in loneliness prediction
- Designed and implemented a multi-modal feature extraction pipeline integrating OpenSMILE & Librosa for extracting acoustic features, and leveraging Whisper and wav2vec for advanced vector embeddings. Performed comprehensive combined analysis of audio and textual data, enhancing the robustness of predictive models for loneliness detection
- **Contributed to the open-source dlatk** Python library by designing and implementing the `add_pipeline` feature for Hugging Face integration and the `exportConvokit` method for data format conversion, enhancing the library's versatility
- Developed and scaled innovative features for a data collection website with Express and Angular, empowering more than 5,000 users to share their social media information effortlessly
- Ingested and refined large-scale Facebook posts and user information raw data, architected a MySQL schema for efficient storage solutions, empowering the team to identify user problems and elevate operational efficiency by 80%
- Handled a wide range of Linux server tasks, including orchestrating upgrades, executing seamless migrations, establishing and managing AWS instances, configuring Jenkins jobs, and deploying websites such as [TextAnalyzer](#) and [WWBP Wiki](#)

**Data Analyst Intern**, Siemens Healthineers – Cary, North Carolina, USA

June 2023 - August 2023

- Devised an efficient data processing system through Power Apps by building advanced search, report automation and streamlined data entry which optimized resource use and budget, saving \$50K annually across operation department
- Successfully managed the transition of data visualization and analytics processes from Power BI to Qlik Sense, ensuring seamless integration and utilization of SaaS-based datasets, enhancing business intelligence capabilities by 30%
- Integrated Python and SQL using Teradata SQL library for advanced data extraction and manipulation from Teradata, leading to a 50% reduction in data processing errors and enhancing data quality

**Software Engineer**, Broadridge Finance Solutions – Hyderabad, India

July 2021 - July 2022

- Created a website adhering to design principles leveraging TypeScript and Angular for responsive frontend, combined with a Go backend to fetch real-time API data then stored in AWS S3 and employed GitLab for codebase version control
- Optimized application performance by implementing concurrent Go routines for fetching and processing AWS service data streams, reducing retrieval times by 30% and improving responsiveness
- Designed automation linking GitLab webhooks to Jira via AWS Lambda and API Gateway, enhancing issue tracking and logging with CloudWatch, reducing issue resolution times by 40%
- Automated AWS infrastructure deployment via Jenkins CI/CD pipelines with Groovy scripts and terraform repositories for Infrastructure as Code reducing infrastructure provisioning time by 70%
- Developed robust APIs for the company website using Node.js, harnessing object-oriented programming principles and strategic code optimization resulting in a 30% increase in user engagement
- Led a cross-functional team to revamp Terraform scripts, boosting infrastructure agility by 30%. Achieved seamless state file management via S3, docker and security with DynamoDB state locking, resulting in a 25% faster deployment cycle

**Software Engineer Intern**, Broadridge Finance Solutions – Hyderabad, India

April 2021 - July 2021

- Collaborated with DevOps team to introduce Splunk dashboard to generate valuable log statements bringing down root cause analysis time for a bug by 80%
- Tested code against test suites using test pyramid concepts to write integration, end to end and unit test cases to get a code coverage of at least 95% and facilitate effective technical documentation for reviews in Confluence
- Applied Agile software development methodologies to rapidly prototype and deploy AWS cloud services solutions, including EC2 and Lambda, ensuring iterative improvements and stakeholder feedback integration

## PROJECT

### Deployment of Real-Time Voting Application on AWS

January 2023 - March 2023

- Utilized AWS EKS for backend orchestration and EC2 for React frontend hosting, ensuring enhanced performance, security & user experience with effective access policies and ALBs. Established MongoDB stateful set on EKS for data management

### Covi source: Hospital Finder & Booking App <https://covisource-vn.netlify.app/>

January 2022 - April 2022

- Designed and developed a Covi source (responsive desktop application for Tab and Mobile): encompassing full-stack design that enabled users to locate hospitals based on specific search criteria and make online reservations during the pandemic
- Tech Stack: ReactJS, HTML, CSS, JavaScript, Redux, Bootstrap, Firebase, Express, Node, MongoDB

## SKILLS

- Programming Languages: Go, Python, C, C++, Java, HTML, CSS, TypeScript, JavaScript, Bash , Groovy
- Frameworks and Libraries: Terraform, Angular, Scikit learn, Pandas, OpenCV, CV2, NLTK, TensorFlow, Matplotlib, Apache Kafka, MapReduce, Hadoop ,Ansible
- Data Processing & Analytics: Databricks, Unity Catalog, Snowflake
- Tools and Technologies: Amazon Web Services, Azure, Redux, Bootstrap, Kubernetes, Docker, GIT, Express, JIRA, Node JS, Restful APIs, IDEA, Eclipse, MySQL, MongoDB, Redis Cache, Oracle, Webpack, Power Apps, Power BI & Qlik Sense

## CERTIFICATIONS

- [AWS Certified Solutions Architect Associate](#) Issued by Amazon Web Services Training and Certification