Pranav Hegde

602-561-9009 • pranavh4@gmail.com • linkedin.com/in/pranavh4 • github.com/pranavh4 • pranavh4.github.io

EDUCATION

M.S. Computer Science **May 2025**
Arizona State University, Tempe, AZ 4.0/4.0
- Distributed Systems, NLP, Cloud Computing, Data Processing at Scale, GenAI, Data Intensive Systems for ML, Data Mining

B.E. Computer Science and Engineering **July 2021**
Ramaiah Institute of Technology, Bangalore, India 9.55/10
- Data Structures, Algorithms, Computer Networks, Operating Systems, DBMS, Machine Learning, Deep Learning, Web Dev

TECHNICAL SKILLS

Languages and Build Systems: Java, Python, JavaScript/Typescript, HTML/CSS, C/C++, Maven, Gradle, Bazel, Jenkins
Libraries: Debezium, Dropwizard, Flask, PyTorch, Huggingface, Pandas, Numpy, Scikit-learn, Keras, NLTK, spaCy, Tensorflow, gRPC, JUnit, Selenium, Logback, jQuery, Rasa, BeautifulSoup, LlamaIndex, OpenCV, PyTesseract, React.js, Bootstrap
Cloud Native & Paradigms: GCP (Compute Engine, Pub/Sub, Dataflow, Cloud Storage, Cloud SQL, Cloud Functions, OAuth, Firebase), AWS (Lambda, S3, EC2, Cognito, API Gateway), Linux/Unix, Docker, Kubernetes, Git, Grafana, Prometheus, Keycloak, Ansible
Data: Apache (NiFi, Beam, Kafka), PostgreSQL, MySQL, BigQuery, DynamoDB, MongoDB, Neo4j, KSQL, PromQL, Redash, Airbyte

PROFESSIONAL EXPERIENCE

VECTARA **Tempe, USA**
Software Engineer Intern **January 2025 – May 2025**
- Enhanced Vectara's RAG pipeline by evaluating and integrating code-specific embedding models using Hugging Face Transformers. implemented plug-and-play support for OpenAI compatible embedders in the platform.
- Designed and executed a zero-downtime migration strategy consolidating per-customer MariaDB instances into a unified master schema using Flyway and SQL scripts.

CERMATI **Bangalore, India**
Senior Software Engineer – Data Platform **March 2021 - July 2023**
- Built a real-time data pipeline using Java, Debezium and Google Pub/Sub to stream changes from PostgreSQL databases, enabling instant data integration for analytics; contributed bug fixes to the Debezium and Pub/Sub open-source libraries.
- Owned and scaled the company's legacy event tracking system, leveraging Java, Google Pub/Sub, Apache Beam, and Dataflow to ingest over 100,000 events per second into BigQuery.
- Built and maintained ETL pipelines to ingest data from over 10 diverse sources—including PostgreSQL, MySQL, MongoDB, Mixpanel, and TikTok—into Google BigQuery and Cloud Storage.
- Led onboarding and mentorship of four new team members, streamlining their integration into the data platform team and delegating responsibilities to accelerate productivity.

SAMSUNG R&D INSTITUTE **Bangalore, India**
Samsung Prism Developer **March 2020 – August 2020**
- Trained a Random Forest Model using scikit-learn for classifying user-phone contact relationships with an accuracy of 73%.
- Created a robust Android Application using Java and Android Studio, containing the trained Model, enabling real-time and accurate relationship classification based on users' past interactions with their contacts.

GRAPHENE **Bangalore, India**
Summer Intern **June 2019 – July 2019**
- Designed and constructed a rule-based engine leveraging constituency parsing techniques using NLTK for automated extraction of product aspects, related phrases, and sentiments from online product reviews with an accuracy of 90%.
- Utilized Neo4J to craft a comprehensive knowledge graph, enhancing visualization and analysis of the extracted data.

PROJECTS

BelieVid, *Smart India Hackathon Winner* **July 2020 – August 2020**
- Led a team in developing a cross-platform solution (web, mobile, browser plugin) for detecting deepfakes, photoshopped images, and other media forgeries using cutting-edge CNN architectures. Spearheaded the implementation of forensic tools such as JPEG ghost detection and error-level analysis to assist in digital evidence verification.

JARVIS: Intelligent Network Intrusion Detection & Prevention System **January 2020 – June 2020**
- Designed and implemented a modular NIDPS using Snort and a Random Forest ML model trained on the CICIDS2017 dataset to dynamically detect and mitigate 15+ types of network attacks in real time.
- Published in the International Conference on Advances in Electronics, Computers and Communications (ICAECC).