

Preethi Sighavi

+1 408-461-0313 | preethisighavi@gmail.com | [linkedin.com/in/preethisighavi](https://www.linkedin.com/in/preethisighavi) | github.com/preethisighavi | [Portfolio](#)

SUMMARY

Master's in Applied Data Intelligence with 3+ years applying machine learning, deep learning, and data analysis to real-world, large-scale problems. Builds multimodal and vision-language ML pipelines, retrieval systems, multi-agent RAG applications, and production data workflows in Python, SQL, and JavaScript. Combines hands-on AI/ML research with a strong foundation in data engineering, analytics, and software quality.

EDUCATION

San Jose State University — M.S. in Applied Data Intelligence Aug 2024 – Dec 2026
San Jose, California, USA | Coursework: Machine Learning, Deep Learning, Data Visualization, Big Data
Sri Sairam College of Engineering (VTU) — B.E. Computer Science Aug 2017 – Aug 2021
Bangalore, India | Coursework: DSA, Software Engineering, Python Applications, AI, ML, Big Data

RESEARCH & PROJECTS

Zero-Shot Video Retrieval & Event Localization with Vision-Language Models Ongoing

- Designed a zero-shot natural-language video retrieval and event-localization system using vision-language models (**CLIP**, **BLIP**, **Flamingo-inspired VLMs**), enabling timestamp-level event search across video corpora from free-text queries.
- Fine-tuned **CLIP ViT-B/32** on 670 MSVD videos against 27,763 text queries, achieving **78.7% Recall@10**, **50.7% MRR**, **median rank 2**, and ~0.0025 s/query retrieval latency.
- Built an end-to-end multimodal pipeline — 8-frame sampling, BLIP captioning, caption normalization, and embedding generation — supporting reproducible retrieval evaluation.

PermitLens — Multi-Agent RAG Permit Assistant (IBM SkillsBuild) Feb 2026 – May 2026

- Architected and deployed a full-stack, **multi-agent RAG** system (Next.js, React, Node.js, Express) serving permit workflows across 15–20 U.S. cities with city-specific document checklists.
- Designed retrieval, completeness-assessment, guardrail, and response-generation agents to classify document gaps and surface blocker-first next steps with source-backed reasoning.
- Implemented production AI governance — **PII scrubbing**, official-source citation enforcement, confidence-based findings, and IBM watsonx.governance monitoring — to curb hallucinations and unsupported approval claims.

EXPERIENCE

Data Operations Analyst — Tata Consultancy Services Jul 2022 – Jul 2024 Bangalore, India

- Uncovered **data drift** between marketing spend and sales across millions of transaction records using **Python** and **SQL**, directly prompting stakeholders to overhaul ROI reporting and reallocate budgets.
- Built automated anomaly-detection and data-quality monitoring workflows that cut the reporting cycle from **5 days to 2** and raised accuracy from **~86% to ~95%**, removing manual validation.
- Rebuilt end-to-end data extraction and transformation pipelines with rigorous cleaning, preprocessing, and **EDA** to deliver trustworthy, audit-ready datasets for downstream modeling.
- Designed and owned **interactive dashboards** in Power BI, Tableau, and Excel tracking CPA, channel ROI, and spend performance, guiding Product, Engineering, and Marketing stakeholders through the insights.

Associate System Engineer — Tata Consultancy Services Nov 2021 – Jul 2022 Bangalore, India

- Queried and analyzed application datasets in MySQL across **6 projects** to surface reporting insights, validate data integrity, and support business decision-making.
- Authored test plans, cases, and scripts (**manual + Selenium**) and designed **20–25 functional and data-validation scenarios** per project, improving release quality and requirement traceability.
- Tracked defects in **BMC Remedy** with clear reproduction steps and resolved **20+ critical production incidents** with cross-functional teams; ran regression testing to catch issues before release.
- Maintained structured **documentation** — test cases, SRS artifacts, and Tableau dashboards on incident trends — improving traceability, audit readiness, and knowledge transfer.

Machine Learning Intern — AVM IT Solutions Jan 2020 – Jun 2020 Bangalore, India

- Built and benchmarked Logistic Regression, Random Forest, and Neural Network classifiers (scikit-learn) for biometric recognition; **Random Forest outperformed the baseline by 6.33%** and was recommended for deployment.
- Engineered features from **2,000+ biometric records** (NumPy, Pandas, OpenCV) and ran feature-relevance analysis to isolate predictive signals and reduce training noise.
- Visualized model performance and feature distributions (Matplotlib, Seaborn, Tableau) to communicate findings to technical and business stakeholders.

TECHNICAL SKILLS

Languages & Query: Python, SQL, PL/SQL, JavaScript (Node.js, React), MySQL

Machine Learning & Deep Learning: PyTorch, scikit-learn, XGBoost, Regression, Classification, Clustering, Time Series (ARIMA), Feature Engineering, Anomaly Detection, EDA

Computer Vision & Multimodal: CLIP, BLIP, Vision-Language Models, Multimodal Retrieval, Event Localization, OpenCV
Data Engineering & Big Data: PySpark, Kafka, Airflow, Snowflake, Docker, Pipeline Management

Web & Frameworks: Next.js, React, Node.js, Express

Data Analysis & Visualization: Power BI, Tableau, Looker Studio, Excel (Pivot Tables, VLOOKUPS), Matplotlib, Seaborn

Software Engineering & QA: Debugging, Manual & Automated Testing (Selenium), Test Plan/Case/Script Writing, Defect Tracking (BMC Remedy), Regression Testing, SDLC, Agile/Scrum

Core Competencies: Data Cleaning, Statistical Analysis, KPI Reporting, Forecasting, Stakeholder Communication