Suraj Jaiswal
linkedin.com/in/surajjaiswal007
Portfolio: jaiswalsuraj487.github.io

Email : jaiswalsuraj487@gmail.com
Mobile : +91 7878189018

Education

- **Indian Institute of Technology Gandhinagar (IIT)** 2022 – 2024
  Master of Computer Science *CPI: 8.5*
- **Gujarat Technological University** 2018 – 2022
  Bachelor of Computer Science *CPI: 8.8*

Experience

- **Tiger Analytics** Chennai, India
  Data Analyst — *Full-time* *Dec 2024 – Present*
  - **Medical Note-Taking Application**
    - Developed product AI copilot for medical note-taking of Doctor-Patient conversation
    - ***Fine-tuned*** large language models (LLMs) using OpenAI API on medical guidelines and medical codes to generate outputs for healthcare
    - ***Medical Code Automation***: Developed a system that captures Doctor-Patient conversations to generate accurate ICD-10, CPT, SNOMED, and HCC codes using embedding-based similarity search and retrieval-augmented generation (RAG)
    - Medical Chatbot: Leveraged ***Graph-RAG*** for real-time insights on ***Athena EHR*** data
    - Applied few-shot ***prompt tuning*** to generate medical guideline-aligned and consistent LLM outputs
    - Wrote ***SQL*** queries for PostgreSQL and managed security tasks using AWS services(S3, ECR, EC2)
  - **Medical Q&A Application**
    - Created and optimized backend ML pipeline for product Medaura: Medical Q&A
    - Utilized ***HyDE + RAG*** using the PubMed database to get answers using medical data
  - **AI-Generated Bill of Materials Automation**
    - Designed a multi-stage GenAI pipeline leveraging ***AWS Bedrock*** (Claude 3.5 Sonnet) for entity extraction and structured Bill of Materials generation from complex catalog data
    - Engineered modular Python logic for rule-based parsing and pattern recognition across geometric layouts (Linear, U-shape, Hexagonal, etc.)
    - Implemented versioned deployment with MLflow PyFunc and automated inference & validation pipelines using Pandas, NumPy, and Boto3 to enhance scalability and reliability
    - Reduced manual BOM creation time by approx 80% and demonstrated measurable ROI from GenAI powered automation

- **NeuroReef Labs** Austin, Texas
  NLP Engineer *Full-time, Remote* *Oct 2023 – Oct 2024*
  - **Auto Code Evaluator**
    - Built an LLM-powered code evaluation engine integrating OpenAI API and ***prompt engineering*** for automated grading of regression and classification tasks
    - Enhanced **Streamlit** UI and implemented multithreading for parallel ZIP submission handling, improving throughput by 3×
    - Designed evaluation templates and scoring logic for reproducible, scalable assessment workflows
    - Streamlined model evaluation workflow, cutting manual review time by over 60%

- **Transportation Preference Choice Modeling**
  - Engineered an end-to-end discrete choice modeling framework analyzing 1,800+ traveler preferences using Multinomial Logit (MNL) models and behavioral analytics
  - Built preprocessing and feature engineering pipelines in **xLogit** and **Pandas**, optimizing wide-to-long data transformations and model estimation speed
  - Designed modular **YAML**-based experiment configurations enabling scalable simulations and reproducible hyperparameter tuning in discrete choice models
  - Developed interpretable utility estimation workflows to measure choice sensitivities and substitution probabilities under different scenario conditions
  - Improved forecasting accuracy by 25% and reduced manual modeling effort by 60% through fully automated elasticity and adoption analysis pipelines

Research Publications

- **Space to Policy: Scalable Brick Kiln Detection and Automatic Compliance Monitoring with Geospatial Data** (link)
  *Accepted at **ACM** Journal on Computing and Sustainable Societies*

- **Eye in Sky: Detection and Compliance Monitoring of Brick Kilns using Satellite Imagery** (link)
  *Accepted at **ACM** Compass on Computing and Sustainable Societies poster track*

- **Towards Scalable Identification of Brick Kilns from Satellite Imagery with Active Learning** (link)
  *Accepted at **NeurIPS** Active Learning in the Real World*

  - Led end-to-end development of a deep learning pipeline to detect ***30,000+*** brick kilns across 5 Indian states using moderate-resolution satellite imagery and oriented object detection models - ***YOLO-OBB***
  - Built a custom geospatial annotation tool using Leafmap and Esri Wayback Imagery; manually labeled 1,600+ kilns across 15,000+ km^2 in 4 airsheds based on air quality and policy relevance
  - Conducted model selection by evaluating YOLOv8 and YOLO11 OBB models across 5 configurations; chose YOLO11m-OBB for optimal performance (Weighted mAP@50 = 0.71)
  - Performed out-of-region generalization testing using "Leave-One-Region-Out" experiments to analyze model robustness and reduce exclusion errors
  - Applied semi-automated iterative labeling (precision = 58%) to validate 15K+ predictions and fine-tune models, increasing detection precision to 82% in key regions like Uttar Pradesh
  - Generated a comprehensive, hand-validated dataset of 30,638 brick kilns, classified by kiln type (CFCBK, FCBK, Zigzag) and ***geo-located across 5 states covering 520,000 km^2***

Skills

- **Programming Languages**: Python, SQL, Java, C
- **Technologies**: Langchain, Langsmith, OpenAI, Claude, FastAPI, Ultralytics YOLO, OpenCV, Tensorflow, Pytorch, MLflow, JAX, Geopandas, Raytune, Scikit-learn, Numpy, Pandas, Matplotlib
- **Tools**: AWS, AWS Bedrock, AWS Boto3, Hugging Face, Streamlit, Docker, Git, Visual Studio Code, Excel Automation, Notion, Jira

Achievements

- **Won The Third AI Engine Hackathon for Google Drive, ThirdAI Corp**
  Among 70+ teams. Built a Neural DB engine enabling intelligent query search on Google Drive
- **Achieved 97 Percentile in the Graduate Aptitude Test in Engineering (GATE) 2021**