Kruthika Srinivas Vasisht

Data Engineer | ML, Computer Vision, SQL/ETL, Dashboards & Business Intelligence
857-961-7680 | vasishtkruthika@gmail.com | LinkedIn | Github | Portfolio

Education

Northeastern University — Boston, MA
Master of Science in Data Analytics and Engineering — *Dec. 2025*

Visvesvaraya Technological University — Bangalore, India
Bachelor of Engineering in Computer Science and Engineering — *Aug. 2022*

Technical Skills

Data Engineering & Warehousing: SQL (MySQL/SQLite), Python ETL pipelines, data modeling, REST APIs, log parsing
ML & AI: Python (pandas, NumPy, scikit-learn), TensorFlow, YOLO object detection, feature engineering, experimentation
Analytics & Visualization: Plotly/Dash, Flask, Seaborn, Matplotlib, Tableau, PowerBI, Bootstrap UI, KPI dashboards
Platforms & Tools: GCP (Cloud Run, Storage), Azure DevOps, Firebase, Git/GitHub, Jupyter, medical device systems

Experience

Data Science Engineer | Medical Device Systems Engineering — Jan. 2025 – Dec. 2025
SpectraWAVE, Inc. | Acquired by Philips — *Bedford, MA*

- Owned end-to-end data platform for unstructured medical device logs. Engineered automated **ETL pipelines scaling to 1,000+ logs/month** across Dropbox, SharePoint, Azure DevOps with duplicate detection and JSON state management
- Established comprehensive **data validation framework with testing protocols** across ingestion-to-visualization pipeline, improving data accuracy across 10,000+ records
- Designed and deployed **FDA-regulated complaint automation pipeline** from Azure DevOps to GreenLight Guru QMS, generating structured PDFs with 15+ complaint fields (regulatory assessment, risk review, root cause, corrective actions)
- Built **multi-dashboard analytics platform** covering OCT image quality, field service, manufacturing performance, & startup diagnostics, enabling 5 cross-functional teams to make data-driven decisions previously requiring manual log analysis
- Created a **complaint analytics dashboard** (Rate/Distribution, Closure Stats, Workflow, Failure Mode) integrating ADO API, Knack API, and GreenLight Guru API for end-to-end complaint lifecycle visibility
- Automated Angio-OCT temporal-delay calibration using **YOLO object detection**, curating hundreds of manually annotated medical imaging frames to train model achieving **0.81+ confidence scores** and reducing manual effort by 80%
- Quantified links between manufacturing metrics and field outcomes, delivering service KPIs (uptime, response time, downtime cost) that informed stakeholder reviews and improved field reliability by 10%

Information Technology (Trainee) — Apr. 2022 – Jan. 2023
NTT Data — *Bangalore, India*

- Architected **Python/SQL ETL pipeline** with scheduling and query optimization, reducing dashboard load times by **40%**
- Designed **responsive UI using Bootstrap** and usage telemetry tracking (filters, latency metrics), increasing user engagement by **25%** for internal analytics platform
- Refactored data models and built **API-driven extracts**, improving dashboard reliability for 5+ departments

Signal processing & ML Research Intern — Dec. 2020 – Jun. 2021
Indian Institute of Science (IISc) — *Bangalore, India*

- Constructed voice pathology detection pipeline analyzing **jitter and shimmer irregularities** across stuttering conditions using **PRAAT signal processing**, improving medical assessment accuracy by **20%**
- Engineered data collection and preprocessing workflows for **10,000+ children's speech samples** from multiple sources, standardizing metadata schemas to identify voice maturation patterns across age and gender demographics

Projects

Massachusetts Mortgage Lending Analysis Dashboard Using Power BI — Dec. 2025

- Built interactive 3-page Power BI dashboard analyzing **825,000+ mortgage applications** across Massachusetts (2020-2023), implementing DAX measures for KPIs including approval rates, risk metrics, and lending volume trends
- Engineered **Python ETL pipeline** to preprocess and merge 5 datasets, creating calculated fields for risk flags, income bands, and denial categorization for comprehensive borrower risk analysis
- Implemented AI-driven analytics using Power BI's Key Influencers and Decomposition Tree visuals

Predicting Term Deposit Subscription Using Direct Marketing Campaign Data [Github] — Dec. 2025

- Optimized ensemble ML models (Logistic Regression, Neural Networks, Naive Bayes) achieving **87.75% F1-score** and **0.91 ROC-AUC** on **45,211 customer records** for marketing campaign optimization
- Applied **SMOTE oversampling** and feature engineering (interaction terms, outlier detection) to handle **88.3% class imbalance**, improving minority class recall to **80.87%**