

# Pranav Kumar Sasikumar

pk100815@gmail.com • +91 88383 81325 • linkedin.com/in/pranav-kumar-sasikumar • GitHub

## Professional Experience

---

### Machine Learning Data Associate II

Jul 2025 – Present

*Amazon – AGI Data Science, Chennai, India*

- Built and maintained scalable ML pipelines for large-scale multimodal datasets (text, image, audio, video), supporting machine learning training, validation, and evaluation workflows.
- Conducted exploratory data analysis, statistical error testing, and label quality assessment to identify noise, bias, and failure modes impacting model performance.
- Developed automated data validation and transformation frameworks to enable reproducible, production-ready ML pipelines and reliable evaluation across model iterations.

### Machine Learning Research & Teaching Assistant

Oct 2023 – Dec 2024

*University of Sheffield, UK*

- Designed, implemented, and evaluated regression, classification, ensemble, and generative models using statistically rigorous experimentation, cross-validation, and sensitivity analysis.
- Built reproducible Python-based ML pipelines (pandas, NumPy, scikit-learn) for data preprocessing, model training, and evaluation across structured and semi-structured datasets.
- Mentored and assisted undergraduate students on ML system design, experimental methodology, debugging, and performance optimisation.

### Custom Software Engineering Associate

Oct 2022 – Jul 2023

*Accenture Solutions Pvt. Ltd., Chennai*

- Designed and maintained Python based ETL/ELT pipelines across Azure, SQL Server, IBM Cloud, and Cleo MFT to support near real-time data processing.
- Implemented automated data validation, logging, and monitoring frameworks to improve reliability and consistency of production data workflows.
- Optimised SQL queries and batch processing jobs, reducing query latency by 27% and improving overall pipeline efficiency.

## Education

---

### MSc in Artificial Intelligence

Sep 2023 – Sep 2025

*University of Sheffield, UK*

### B.E. in Electrical and Electronics Engineering

Sep 2018 – Jun 2022

*Panimalar Engineering College, India*

## Technical Skills

---

**Programming & ML:** Python (Pandas, NumPy, PySpark), SQL, Scikit-learn, TensorFlow, PyTorch

**Data Science & Analytics:** Forecasting, A/B Testing, Segmentation, Hypothesis Testing, Feature Engineering

**Visualization & BI:** Tableau, Power BI, Matplotlib, Seaborn, Chart.js

**Data Engineering:** ETL Pipelines, Apache Spark, Hadoop, Data Quality Validation

**NLP & Text Analytics:** TF-IDF, Embeddings, Sentiment Analysis, Chatbots, Prompt Engineering

**Cloud & Deployment:** AWS, Azure, IBM Cloud, Docker, Kubernetes

## Projects

---

### LLM-Based Threat Detection & API Deployment [GitHub]

- Developed a Flask REST API with TF-IDF + tree-based ML models to detect SQLi/XSS attacks (98.4% accuracy).
- Integrated GPT-based assistant for natural-language explanations of alerts, improving interpretability for security analysts.
- Reduced false positives by 21% and improved triage efficiency.

### Zero-Dimensional Cardiovascular Modelling & Sensitivity Analysis [GitHub]

- Created 0D cardiovascular models in Python, reducing computational complexity by 30%.
- Applied Sobol & Morris sensitivity analysis to identify critical physiological parameters.

### Customer Churn & Credit Risk Prediction (Lloyds Banking Group – Forage)

- Built ML models (ROC-AUC 0.82) to identify at-risk customers, providing actionable insights for retention strategy.
- Performed feature importance analysis to derive actionable business insights for retention strategies.

### Topic Classification using Feedforward Neural Networks [GitHub]

- Built topic classification system on AG News dataset using GloVe embeddings; improved baseline accuracy by 15%.
- Designed image orientation detection pipeline using PCA + MLP across 10,000+ images, achieving 98% accuracy on high-resolution inputs.

## Certifications

---

- Google Data Analytics Professional Certificate