AWANTIKA SRIVASTAVA

Machine Learning Engineer | Applied AI (Vision & GenAI)

+91-8920482037 | sawantika81@gmail.com | LINKEDIN | Github

PROFILE SUMMARY

Machine Learning Engineer with 2+ years of experience building production-grade Computer Vision and document understanding systems. Strong expertise in segmentation models, object detection, OCR pipelines, and structured data extraction from complex visual inputs. Hands-on experience deploying batch and near real-time inference systems on Linux environments. Skilled in designing reliable CV pipelines with precision-focused evaluation, robust preprocessing, and geometry-aware post-processing.

CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python, C++, SQL., Pandas, Numpy, Scikit-learn, Jupyter Notebook.
- **Statistics & Mathematics:** Statistical Modeling, Descriptive Statistics, Hypothesis Testing, Probability, Sampling, Scenario Analysis.
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, Model Evaluation Metrics (Accuracy, Precision, Recall, F1-score, ROC-AUC).
- **Deep Learning & AI:** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Model Fine-tuning, Time-series Forecasting, ARIMA, SRIMA, TensorFlow, PyTorch
- **Computer Vision:** Image Classification, Object Detection (YOLO, SSD, MobileNet, ResNet), Real-time Video Analytics, Image Pipelines, Semantic Segmentation, Image Consistency & Feature Extraction, OpenCV
- **Document Intelligence & OCR:** OCR Pipelines (PaddleOCR), Table Extraction, Layout Understanding, Symbol & Annotation Detection, PDF to Structured Data Conversion, Post-processing & Text Normalization
- **NLP:** Text preprocessing, Tokenization, Chunking, Sentiment Analysis, Topic Modeling (LSA, LDA), Transformer-based models, NLTK, TF-IDF
- **Deployment & Systems:** Linux-based Inference, Batch & Near Real-time Processing, REST APIs (Flask / FastAPI), Model Optimization, Evaluation Harness (Precision / Recall / F1), Docker, CI/CD
- **Cloud & Platforms:** Hugging Face Spaces, Replicate, AWS EC2, S3
- **Tools & Collaboration:** Git, GitHub, Documentation, Docker, CI/CD, REST APIs, FastAPI/Flask

EXPERIENCE

Machine Learning Engineer | PPS International Pvt. Ltd. **January 2024-Present**

- Designed and deployed segmentation and **object detection pipelines** for **real**-world production systems.
- Trained **CNN-based** models for structured element **detection** and spatial **feature extraction**.
- Built **OCR pipelines** for **extracting** structured information from **scanned** documents and noisy inputs.
- **Developed** post-processing modules to **convert** bounding **boxes** and model outputs into structured **JSON** formats.
- Deployed inference services **on Linux servers** supporting batch and near real-time processing.
- Created evaluation harness **for precision/recall** benchmarking and error analysis.
- Improved robustness across varied input quality **using preprocessing**, normalization, and threshold tuning.
- **Collaborated** with downstream engineering teams to integrate structured outputs into optimization systems.

PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- **Designed and deployed a** real-time computer vision–based ML system **for unsafe** driver behavior detection using **CNN-based SSD MobileNet** model.
- Trained and optimized models on large-scale video datasets, achieving **20–25 FPS** real-time processing with **<150 ms inference latency**.
- Implemented **end-to-end ML pipelines** for data ingestion, preprocessing, model training, evaluation, and production inference.
- Deployed optimized models using **TensorFlow Lite** on edge/production environments for continuous monitoring.
- Built a **Flask-based web dashboard** to visualize detections and automatically record **30-second event clips**, reducing manual review effort.

Chatbot Using LLM & RAG | Applied ML Project

- Built a Lightweight **LLM-Powered chatbot** using **Tiny LLaMA** to answer user queries over content.
- Implemented a **Retrieval-Augmented Generation (RAG) pipeline** to retrieve relevant resume sections for contextual question answering.
- Selected **Tiny LLaMA** to ensure **low memory footprint and fast inference**, making the solution suitable for resource-constrained environments.
- Applied **prompt engineering techniques** to improve response relevance and consistency.
- Deployed the chatbot as an interactive **Streamlit web application** for real-time user interaction.

YouTube Comments Sentiment Analyzer | link-https://youtube-ai-analyzer-ndzqo6r2mepjrtsdjmwaxl.streamlit.app/

- Deployed transformer-based **NLP models** as **production-ready** services with **REST APIs**.
- **Fine-tuned** and served a **DistilBERT**-based **sentiment** classification model for large-scale text inference.
- **Built** and deployed an interactive **streamlit web application** to perform real-time **sentiment analysis** on YouTube comments.
- **Processed** high-volume text **data** with sub-second inference **latency** for real-time sentiment analysis.

CERTIFICATION

- IBM Data Science & AI Certification
- AWS Generative AI with Large Language Models
- OpenCV Computer Vision Certification

EDUCATION

IMS Engineering College, Ghaziabad **September - 2020**

Bachelor of Technology