

AWANTIKA SRIVASTAVA

Machine Learning Engineer | AWS | OCR

+91-8920482037 | sawantika81@gmail.com | [LINKEDIN](#) | [Github](#)

PROFILE SUMMARY

Machine Learning Engineer with **2+ years of hands-on experience** in designing, training, and deploying **end-to-end machine learning and AI solutions**. Strong expertise in **Python-based ML development**, data pipelines, and production deployment on **AWS environments**. Experienced in building scalable ML systems, working with **MongoDB for data management**, and contributing to real-world AI applications including **computer vision and OCR-related workflows**. Adept at collaborating in remote, cross-functional teams and delivering production-ready ML solutions..

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 :** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, Computer Vision.
- **Deep Learning Frameworks:** TensorFlow, keras, Pytorch, TensorFlow Lite.
- **OCR & Document AI:** OCR Pipelines, Text Extraction, Image-to-Text Processing (Tesseract – Exposure).
- **NLP:** Text preprocessing, Tokenization, Sentiment Analysis, Topic Modeling (LSA, LDA), Transformer-based models.
- **LLMs:** Transformers, HuggingFace, DistilBERT, Prompt Engineering, LangChain, RAG.
- **Databases & Data Management:** MongoDB, Data Ingestion, Data Validation, Schema Design
- **Cloud & Deployment:** AWS (EC2, S3, Sagemaker – Exposure), Model Deployment, API-based Serving, CI/CD.
- **Tools & Collaboration:** Git, GitHub, Agile / SDLC, Documentation, Cross-functional Collaboration.

EXPERIENCE

Machine Learning Engineer | PPS International Pvt. Ltd.

January 2024-Present

- Designed and implemented **end-to-end pipelines** covering data ingestion, preprocessing, model training, **evaluation** and deployment.
- Developed and deployed **machine learning and deep learning models** using **Python, TensorFlow, and PyTorch** for real-world applications.
- Built **computer vision-based ML systems**, applying image preprocessing and CNN models to extract meaningful insights from visual data.
- Integrated ML pipelines with **MongoDB** to manage structured and unstructured data reliably.
- Deployed and tested ML models on **AWS environments**, ensuring scalability, stability, and performance.
- Analyzed and optimized deployed models to improve **efficiency, latency, and reliability**.
- Collaborated with cross-functional teams to understand project requirements and deliver impactful ML solutions.
- Documented ML workflows, experiments, and deployment processes to ensure clear knowledge transfer.

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 **TinyLLaMA** to answer user queries over content.
- Implemented a **Retrieval-Augmented Generation (RAG) pipeline** to retrieve relevant resume sections for contextual question answering.
- Selected **TinyLLaMA** 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-ndzqo6r2mepjtsdjmwxl.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 (Electrical and electronics engineering)