

# AWANTIKA SRIVASTAVA

AI Engineer | Machine Learning | GenAI | Cloud AI

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

## PROFILE SUMMARY

AI Engineer with 2+ years of experience in designing and deploying machine learning and AI solutions. Proficient in Python, SQL, ML/DL, and GenAI (LLMs, RAG) with hands-on experience in end-to-end ML pipelines, APIs, and cloud deployment. Strong at converting business problems into scalable, data-driven solutions.

## CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python, C++, SQL, Numpy, Pandas, SciPy, scikit-learn, FastAPI, RESTful API
- **Statistics & Mathematics:** EDA, Statistical Modeling, Hypothesis Testing, Probability, Linear Optimization, Trend Analysis.
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, Hyperparameter Tuning.
- **Deep Learning & GenAI:** Neural Networks, CNN, RNN, LSTM, BERT, Transfer Learning, Generative AI, LLMs, Prompt Engineering, RAG, LangChain
- **Frameworks & Tool:** TensorFlow, Keras, Pytorch, Flask, Hugging Face, Git, GitHub.
- **MLOps & Model Lifecycle:** MLflow, Model Versioning, CI/CD for ML Pipelines, Model Monitoring, Performance Regression Tracking, Experiment Tracking, Production ML Workflows
- **NLP:** Text preprocessing, Tokenization, Chunking, Sentiment Analysis, Transformer-based models, NLU, NLTK, spaCy, TextBlob
- **Cloud & Platforms:** AWS (EC2, S3), Azure, Cloud-based ML workflows, IBM Cloud (exposure).
- **Databases & Tools:** SQL, MongoDB, DevOps Tools, Docker, SaaS, Git, GitHub
- **Professional Skills:** Business Acumen, Problem Solving, Stakeholder Communication, Agile Collaboration, AI Ethics.

## EXPERIENCE

AI Engineer / Machine Learning Engineer | PPS International Pvt. Ltd.

January 2024-Present

- Designed and developed end-to-end machine learning pipelines, covering data ingestion, preprocessing, feature engineering, model training.
- Built and optimized ML models (classification & regression) using Python and scikit-learn, ensuring robustness, interpretability, and performance.
- Applied statistical modeling and hypothesis-driven analysis to validate assumptions and support data-driven business decisions.
- Conducted exploratory data analysis (EDA) to identify trends, biases, anomalies, and data quality issues impacting downstream models.
- Developed RESTful ML inference services using FastAPI, enabling integration with production systems.
- Implemented GenAI and LLM-based workflows, including prompt engineering and retrieval-augmented generation (RAG) for NLP use cases.
- Deployed ML solutions on cloud environments (AWS) with a focus on scalability, reliability, and performance.
- Applied MLOps best practices, including CI/CD pipelines, model versioning, monitoring, and automated validation checks.

## PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- Built and deployed a production-grade machine learning system to detect unsafe driver behavior, translating real-world safety requirements into AI-driven solutions.
- Designed and owned end-to-end ML pipelines covering data ingestion, preprocessing, feature engineering, model training, evaluation, and inference for real-world deployment.
- Trained CNN-based computer vision models (SSD MobileNet) using TensorFlow, and optimized inference using TensorFlow Lite for low-latency environments.
- Achieved real-time inference performance (20–25 FPS, <150 ms latency), enabling near real-time decision-making in operational settings.
- Integrated model predictions into downstream systems via REST APIs, delivering actionable insights to support operational and safety-related decision

Chatbot using LLM & RAG | Applied ML Project

- Built a production-grade Generative AI (GenAI) application using Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) for context-aware question answering.
- Designed and implemented document ingestion, embedding generation, vector retrieval, and LLM inference orchestration pipelines.
- Applied prompt engineering, response evaluation, and reliability tuning to improve output accuracy, consistency, and relevance.
- Deployed the solution as an observable LLM inference service, aligning with modern GenAI and LLM production workflows.

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

- Developed an end-to-end NLP modeling system using BERT / DistilBERT for large-scale sentiment classification.
- Performed EDA, text preprocessing, feature engineering, model training, and statistical evaluation on real-world user-generated data.
- Conducted baseline comparisons and experiment tracking to optimize model performance.
- Deployed the model as a REST API-based inference service to deliver actionable sentiment insights to business stakeholders.

## CERTIFICATION

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

## EDUCATION

IMS Engineering College, Ghaziabad  
Bachelor of Technology (Electrical and electronics engineering)

September - 2020