

AWANTIKA SRIVASTAVA

Data Scientist | Machine Learning Engineer | NLP & RAG Specialist

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

PROFILE SUMMARY

Data Scientist with 2+ years of experience in applying machine learning, statistical analysis, and NLP techniques to solve real-world business problems. Strong background in Python-based data science workflows, supervised and unsupervised learning algorithms, and end-to-end ML lifecycle. Experienced in working with databases, cloud platforms, and analytics pipelines, with the ability to learn and adapt to new AI/ML tools quickly.

CORE TECHNICAL SKILLS

- Programming Languages & Tools:** Python, Numpy, Pandas, Matplotlib, Scikitlearn, C++, SQL, PySpark
- 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, Feature Engineering.
- Deep Learning & GenAI:** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, RAG, LLMs, Vector DB.
- Deep Learning Frameworks:** PyTorch, TensorFlow, Hugging Face, LangChain, FAISS.
- Computer Vision:** Image Classification & Preprocessing, Object detection (YOLO, SSD, MobileNet, ResNet), Video Analytics, OpenCV
- NLP:** Text preprocessing, Tokenization, Chunking, Transformers, Sentiment Analysis, Semantic Search, Transformer-based models, NLTK, SciPy, Spacy.
- MLOps & Deployment:** Flask, REST APIs, Model Serving, Git, Docker (basic), ML Pipelines
- Cloud & Platforms:** AWS (EC2, S3 – basic), Linux
- Collaboration & Delivery:** Client-facing PoCs, Cross-functional Collaboration, Documentation, Agile Delivery

EXPERIENCE

Data Scientist | PPS International Pvt. Ltd.

January 2024-Present

- Designed and deployed **ML/DL models for decision augmentation and process automation**, improving solution efficiency and reliability.
- Built **end-to-end ML pipelines** covering data ingestion, preprocessing, feature engineering, model training, evaluation, and deployment.
- Performed **exploratory data analysis (EDA)** and feature selection to improve model performance and stability.
- Supported **client-facing PoCs and automation use cases**, contributing to **solution-led delivery, GTM readiness, and business value creation**.
- Collaborated with **ML engineers, data engineers, and platform teams** to evaluate deployment strategies and production constraints.
- Implemented **REST APIs and model serving workflows** to enable real-time and batch inference.
- Ensured **production readiness** through modular, reusable code, Git-based version control, and clear technical documentation.
- Monitored model outputs and supported **iterative improvements based on stakeholder feedback and delivery requirements**.

PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- Designed and deployed a **real-time ML systems** for unsafe behavior detection using **CNN-based object detection** models.
- Implemented **end-to-end ML pipelines** from data ingestion to production inference.
- Deployed** optimized models using **TensorFlow Lite** for continuous, **low-latency inference**.
- Developed a **Flask-based** web interface to display detections and recorded video clips in **real time**.
- Deployed** the complete solution on **edge devices**, enabling reliable on-device inference and **real-time event recordings**.

Amazon Stock Price prediction | Applied ML Project

- Built batch-oriented ML workflows** for time-series forecasting using **LSTM** models.
- Implemented data **preprocessing**, feature engineering, and model **evaluation pipelines** on **large** historical datasets.
- Designed **sliding-window** based **sequence generation** and trained LSTM model 5-year stock price data.
- Evaluated model performance using appropriate regression **metrics** and **trend-based analysis** for short-term forecasting.

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 (Electrical and electronics engineering)