

AWANTIKA SRIVASTAVA

Data Scientist | Machine Learning Engineer | NLP & RAG Specialist

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

PROFILE SUMMARY

Data Science professional with 2+ years of experience in data preparation, exploratory data analysis (EDA), feature engineering, and predictive modeling using Python and SQL. Experienced in working with structured and unstructured datasets, building analytical pipelines, and developing dashboards for performance monitoring. Strong exposure to process-oriented data analysis, anomaly detection, and model validation.

CORE TECHNICAL SKILLS

- **Programming & Data Handling:** Python, Numpy, Pandas, Matplotlib, Scikitlearn, Data Cleaning, Data Transformation, C++, SQL.
- **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, Time-Series Forecasting, ARIMA, SARIMA
- **Deep Learning & GenAI:** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, RAG, LLMs, TensorFlow, PyTorch
- **Computer Vision:** Image Classification & Preprocessing, Object detection (YOLO, SSD, MobileNet, ResNet), Video Analytics, OpenCV
- **NLP:** Text preprocessing, Tokenization, Chunkins, Transformers, Sentiment Analysis, Semantic Search, Transformer-based models, NLTK, SciPy
- **Data Visualization:** Power BI, Tableau, Plotly, Matplotlib, Seaborn, Dashboard Development
- **Process & Event Data Concepts:** Event Logs, Case IDs, Timestamp Analysis, Process Metrics, Data Quality Checks
- **ETL & Data Pipelines:** Data Preprocessing Pipelines, Data Integration, Data Consistency Validation
- **Enterprise Exposure:** SAP Data Exposure (conceptual), Cross-functional Collaboration, Business Reporting

EXPERIENCE

Junior 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 computer vision**-based ML system to detect unsafe driver behaviors from continuous video streams.
- Trained and **optimized CNN-based object detection models (SSD MobileNet architecture)** to perform real-time inference on video data.
- **Implemented end-to-end ML pipelines** covering data ingestion, preprocessing, model training, evaluation, and production inference.
- Achieved **20-25 FPS** real-time processing with **<150 ms** inference latency by optimizing models for **deployment**.
- Built and integrated a **Flask-based web interface** to **visualize** detections and automatically recorded 30-second event clips, reducing manual **review** effort.

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-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

Bachelor of Technology (Electrical and electronics engineering)

September - 2020