AWANTIKA SRIVASTAVA

Data Scientist | Machine Learning, LLMs & AWS

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

PROFILE SUMMARY

AI Engineer / Data Scientist with **2+ years of hands-on experience** in building, evaluating, and operationalizing **machine learning and data science solutions**. Strong foundation in **supervised and unsupervised learning**, feature engineering, and model evaluation, with growing exposure to **deep learning, LLM workflows, and AWS-based ML systems**. Experienced in collaborating with engineers and data scientists to deliver **production-ready AI components** using Python, scikit-learn, and cloud-native tools.

CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python, C++, SQL., Pandas, Numpy, Scikit-learn, Data Cleaning, Data Processing.
- **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), Transfer Learning, Model Fine-tuning., Time – Series forecasting, ARIMA, SARIMA.
- **Deep Learning Frameworks:** TensorFlow, keras, Pytorch, TensorFlow Lite.
- **NLP:** Text preprocessing, Tokenization, Chunking, Splitting, Sentiment Analysis, Topic Modeling (LSA, LDA).
- **LLMs:** Transformers, HuggingFace, DistilBERT, Prompt Engineering, LangChain, RAG, Vector DB, Embedding.
- **Explainability:** Model Interpretability, SHAP, LIME (Exposure)
- **Cloud & MLOps:** AWS Sagemaker, S3, Lambda, Bedrock (Exposure), CI/CD, Jenkins, GitHub Action.
- **Tools & Collaboration:** Git, GitHub, Docker, Agile / SDLC, Documentation, Cross-functional Collaboration.

EXPERIENCE

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

- Worked on **data acquisition**, **cleaning**, **enrichment and transformation** to support ML model development.
- Built and evaluated **supervised ML models** for classification and regression use cases using Python and scikit-learn.
- Applied **unsupervised learning techniques**, including clustering and anomaly detection, to identify patterns in unlabeled data.
- Supported development of **deep learning models (CNN, RNN/LSTM)** under guidance for real-world analytics applications.
- Designed and implemented **end-to-end ML pipelines**, covering feature engineering, training, evaluation, and deployment readiness.
- Assisted in deploying ML models on **AWS SageMaker and cloud-based environments**, ensuring scalability and reliability.
- Used existing **CI/CD pipelines** for training, versioning, and deployment of ML models.
- Monitored model **performance** using **dashboards** and **logs**, supporting **debugging** and iterative improvement.
- **Collaborated** closely with senior data scientists, ML engineers, and platform teams to ship production **AI** features.

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.

Amazon Stock Price Forecasting using LSTM | Applied ML Project

- Analyzed **5 years of historical Amazon stock price data** to identify long-term trends, short-term movements, and market volatility patterns.
- Built an **LSTM-based time series forecasting model** using a **14-day sliding window** to learn temporal dependencies and predict **next-day closing prices**.
- Performed **data cleaning, normalization, and sequence generation** to improve model stability and forecasting accuracy.
- Validated trends and model outputs through **time-series analysis and error metrics (RMSE, MAE)** to ensure reliable predictions.
- Visualized actual vs predicted prices using **candlestick charts**, enabling clear interpretation of market trends and forecast insights.

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)