AWANTIKA SRIVASTAVA

AI Engineer | Machine Learning, LLMs & AWS

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

PROFILE SUMMARY

Data Scientist with 2+ years of experience designing and deploying scalable machine learning and deep learning systems for real-world business applications. Strong expertise in classical machine learning, large-scale data processing using PySpark and Spark SQL, and production-grade ML pipeline development on Databricks and AWS. Experienced in building real-time inference systems, LLM-powered applications, and end-to-end ML solutions delivering measurable performance and business impact.

CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python, C++, SQL., Pandas, Numpy, Scikit-learn, SciPy, PySpark
- **Statistics & Mathematics:** EDA, Statistical Modeling, Descriptive Statistics, Hypothesis Testing, Probability, Sampling, A/B Testing
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, LightGBM, Model Evaluation Metrics (Accuracy, Precision, Recall, F1-score, ROC-AUC).
- **Deep Learning:** ANN, CNN (1D/2D), RNN, LSTM, BERT, Transfer Learning, Self-Attention, multi-head Attention, Time-Series Forecasting, Tensorflow, PyTorch, Keras
- **NLP:** Text preprocessing, Tokenization, Chunking, Sentiment Analysis, Transformer-based models, NLTK, TextBlob, SpaCy
- **Generative AI & LLMs:** LLMs (GPT, LLaMA, Gemini), Prompt Engineering, RAG, RAG Pipeline, Vector Search, FAISS, Embeddings, LLM Agents, Multi-Agent Systems, LLM Evaluation, Confidence Grading, LangChain
- **Computer Vision:** Image Classification & Preprocessing, Object detection (YOLO, SSD, MobileNet, ResNet), Video Analytics, OpenCV
- **Large-Scale Processing:** PySpark Data Pipelines, SQL, MySQL
- **MLOps & Deployment:** MLflow, Real-Time Inference APIs, REST API Integration, Docker, AWS (EC2, S3), CI/CD Pipelines, Scalable ML Deployment
- **Model Optimization:** Quantization, Structured Pruning, Performance Optimization, Latency Reduction, SHAP, LIME, Model Monitoring
- **Tools & Version Control:** Git, GitHub, Jupyter, Streamlit

EXPERIENCE

AI/ML Engineer | PPS International Pvt. Ltd. **January 2024-Present**

- Designed and implemented end-to-end ML pipelines using **PySpark and Spark SQL** for large structured datasets, enabling scalable model training and inference.
- Built and optimized classical **ML models (XGBoost, Random Forest**, GLMs) improving **prediction** accuracy by **15–20%** across business use cases.
- Deployed scalable ML workflows on **Databricks and AWS**, ensuring modular, production-ready architecture.
- **Developed real-time inference APIs** integrating ML models into applications with **low-latency response**.
- Implemented **robust feature engineering** strategies aligned with complex business objectives.
- Managed experiment **tracking** and model lifecycle using **MLflow** and **Git**-based version control.
- Built **LLM-based automation** tools leveraging **RAG pipelines** and **prompt engineering** for enterprise use cases.
- Applied **SHAP-based model explainability** techniques to improve transparency and stakeholder trust.

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-ndzqo6r2mepjrtsdjmwaxl.streamlit.app/

- Developed **LSTM-based time-series forecasting** model trained on 5+ years historical stock data reducing **RMSE** by 18**%.**
- Engineered lag **features**, **rolling statistics**, and volatility indicators improving prediction stability.
- **Implemented** custom **BPTT training** loop with adaptive learning rate scheduling for **stable convergence**.
- Applied **regularization techniques** reducing overfitting gap by **20**%.

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 (Electricaland electronics engineering)