AWANTIKA SRIVASTAVA

Data Scientist | NLP | LLMs

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

PROFILE SUMMARY

Data Scientist with **2+ years of hands-on experience** in **NLP and applied LLMs**. Strong background in building **NLP pipelines, RAG systems, prompt engineering, and model evaluation** for large-scale unstructured text data. Experienced in working with **open-source and locally hosted LLMs**, optimizing models for **on-premise and restricted environments**, and translating **business problems into data-driven AI solutions** in procurement and sourcing-related domains.

CORE TECHNICAL SKILLS

- **Programming Languages & Tools:** Python, NumPy, Pandas, 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.
- **Deep Learning & AI:** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, PyTorch, TensorFlow
- **Computer Vision:** Image Classification & Preprocessing, Object detection (YOLO, SSD, MobileNet, ResNet), Video Analytics, OpenCV
- **NLP:** POS Tagging, Named Entity Recognition (NER), Text Classification, Sentiment Analysis, Semantic, Similarity, Word Embeddings (Word2Vec), NLTK, Information Extraction, Information Retrieval, Knowledge Representation Concepts
- **LLMs & RAG:** Applied LLMs, SaaS LLMs, Open-Source LLMs, Locally Hosted LLMs, Domain-Specific LLMs, Prompt Engineering, RAG, Contextual Retrieval
- **Embeddings & Vector Search:** Text Embeddings, Vector Similarity Search, Embedding-Based Retrieval
- **Open-Source Models & Fine-Tuning:** LLaMA, LoRA Fine-Tuning, QLoRA Fine-Tuning
- **Deployment & Integration:** API-Based Deployment, Batch Pipelines, Internal Tools Integration, On-Premise Deployment
- **MLOps & Monitoring:** Model Monitoring, Model Traceability, Model Maintainability, Deployment Best Practices

EXPERIENCE

Data Scientist | PPS International Pvt. Ltd.   **January 2024-Present**

- Developed and optimized **CNN-based deep learning** models for real-time **computer vision** safety systems.
- Implemented the **complete AI/ML lifecycle**, including data collection, preprocessing, model training, evaluation, and deployment.
- Deployed optimized models using **TensorFlow Lite** on **edge devices**, achieving low-latency inference under constrained CPU and memory environments.
- Designed and maintained **end-to-end inference pipelines** for continuous video stream processing.
- Built a **Flask-based web application** to visualize AI detections and automatically recorded video events for monitoring and analysis.
- Optimized inference workflows to ensure reliable performance on **low-resource hardware platforms**.
- Worked independently and collaboratively to deliver production-ready AI solutions with minimal supervision.

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