AWANTIKA SRIVASTAVA

Data Scientist | Machine Learning, Data Science & NLP

+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, C, C++, SQL.
- **Statistics & Mathematics:** Statistical Modeling, Hypothesis Testing, A/B Testing, Confidence intervals, Bias, Probability, Linear Optimization, Trend Analysis, Risk Modeling.
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, LightGBM, Hyperparameter Tuning
- **Deep Learning:** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, Time-series Forecasting, ARIMA, SARIMA, Backpropagation, Gradient Descent
- **Computer Vision**: Image Classification & Preprocessing, Object detection (YOLO, SSD, MobileNet, ResNet), Video Analytics, OpenCV
- **NLP**: Text preprocessing & Embedding, Tokenization, Chunking, Sentiment Analysis, NER, Topic Modeling, Transformer-based models.
- **LLM Architectures & Models:** Self-Attention, Positional Embeddings, Token Embeddings, GPT, LLaMA, Fine-Tuning, Prompt Engineering
- **Retrieval-Augmented Generation (RAG**): FAISS, ChromaDB, Embedding Search, OpenAI Embedding,
- **Cloud & Data Science Platforms:** AWS, Sagemaker , EC2
- **Tools & SDLC:** Git, GitHub, Confluence, Structured Software Development.

EXPERIENCE

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

- Designed and implemented **machine learning and NLP models** using **scikit-learn, XGBoost, and transformer-based architectures.**
- Built and fine-tuned **LLMs (BERT-based and LLaMA-style models)** for **text generation, summarization, and question answering** tasks.
- Developed **Retrieval-Augmented Generation (RAG) pipelines** using **FAISS and vector embeddings** to improve contextual retrieval.
- Applied **NLP techniques** including **NER, topic modeling, and text embeddings** on large unstructured datasets.
- Implemented **REST APIs** for **real-time model inference** and production deployment.
- Worked extensively with **dimensionality reduction and feature selection techniques** to improve model performance.
- Optimized models using **gradient descent, backpropagation**, and performance tuning strategies.
- Collaborated with cross-functional teams and documented **model architectures, experiments, and results**.

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**.

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/

- 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)