AWANTIKA SRIVASTAVA

AI Engineer | Azure AI| LLM

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

PROFILE SUMMARY

Artificial Intelligence Engineer with 2+ years of experience building enterprise AI components using LLMs, Retrieval-Augmented Generation (RAG), and agentic workflows. Strong expertise in Azure-hosted AI services, API integrations, and modular AI solution design. Experienced in building reusable AI components, optimizing agent behavior, validating data flows, and deploying AI systems using CI/CD pipelines. Proficient in Python and Azure AI ecosystem fundamentals.

CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python, C++, SQL., REST API Development
- **Statistics & Mathematics:** EDA, 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), Feature Engineering
- **Deep Learning:** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, Time-series Forecasting, ARIMA, SRIMA.
- **Deep Learning Frameworks:** TensorFlow, keras, Pytorch, TensorFlow Lite.
- **NLP:** Text preprocessing, Tokenization, Chunking, Sentiment Analysis, Topic Modeling (LSA, LDA), Transformer-based models, Computer Vision, NLTK, TF-IDF
- **Generative AI & Agents:** LLM Agents & Agentic Workflows, RAG, Prompt Engineering, ReAct-style Reasoning Agents, Embedding Models & Vector Search, GPT, LLaMA, Claude Models
- **Cloud:** Azure OpenAI, Azure Cognitive Services, Azure DevOps, Understanding of Copilot, style agent architecture, Familiarity with Power Platform & Dataverse integration concepts
- **Tools & Collaboration:** Git, GitHub, CI/CD Pipelines, Agile , Documentation, Docker.

EXPERIENCE

AI/ML 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.

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