AWANTIKA SRIVASTAVA

AI Engineer | Generative AI | LLMs | RAG | Agentic AI

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

PROFILE SUMMARY

AI Software Engineer with **2+ years of experience** building and deploying **production-grade AI-powered applications**. Strong expertise in **Large Language Models (LLMs), Agentic AI, Retrieval-Augmented Generation (RAG), and scalable backend systems**. Proven ability to design **end-to-end AI solutions**, integrate RESTful APIs, and apply **software engineering best practices** to deliver reliable, high-performance AI systems.

CORE TECHNICAL SKILLS

- **Programming Languages & Tools:** Python, C++, Data Structures & Algorithms, OOPs, RESTful API Development.
- **Statistics & Mathematics:** 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 & GenAI:** Neural Networks, CNN, RNN, LSTM, BERT, Transfer Learning, Generative AI, LLMs, Prompt Engineering, RAG, Agentic AI Pipelines, Semantic Search, Hugging Face, LangChain
- **Frameworks:** TensorFlow, Keras, Pytorch, Flask.
- **MLOps & Model Lifecycle:** MLflow, Model Versioning, CI/CD for ML Pipelines, Model Monitoring, Performance Regression Tracking, Experiment Tracking, Production ML Workflows
- **NLP**: Text preprocessing, Tokenization, Sentiment Analysis, Transformer-based models, NLU, OpenCV.
- **Cloud & Deployment:** AWS, Docker, Kubernetes, Scalable Inference Services, Production Deployment, API-based Model Serving.
- **Data & Analytics:** Exploratory Data Analysis (EDA), Data Preprocessing, Unstructured Text Data, Evaluation Metrics (Precision, Recall, F1-score), Business-Oriented Model Insights.

EXPERIENCE

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

- Designed and developed **machine learning and AI solutions** to solve real-world business problems.
- Performed **data collection, cleaning, feature engineering, and exploratory data analysis (EDA)** on structured and unstructured datasets.
- Built and optimized **predictive, classification, and NLP models** using Python and deep learning frameworks.
- Collaborated with business and product stakeholders to translate requirements into **analytical problem statements**.
- Deployed trained models as **REST API-based inference services** for real-time usage.
- Monitored model performance and continuously improved models based on analytical results and feedback.
- Communicated analytical insights and model outcomes in a **business-meaningful manner** to stakeholders.

PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- Built and deployed a **production-grade machine learning system** to detect unsafe driver behavior, translating real-world safety requirements into AI-driven solutions.
- Designed and owned **end-to-end ML pipelines** covering data ingestion, preprocessing, feature engineering, model training, evaluation, and inference for real-world deployment.
- Trained **CNN-based computer vision models (SSD MobileNet)** using **TensorFlow**, and optimized inference using **TensorFlow Lite** for low-latency environments.
- Achieved **real-time inference performance (20–25 FPS, <150 ms latency)**, enabling near real-time decision-making in operational settings.
- Integrated model predictions into **downstream systems via REST APIs**, delivering **actionable insights** to support operational and safety-related decision

Chatbot using LLM & RAG | Applied ML Project

- Built a **production-grade Generative AI (GenAI) application** using **Large Language Models (LLMs)** and **Retrieval-Augmented Generation (RAG)** for context-aware question answering.
- Designed and implemented **document ingestion, embedding generation, vector retrieval, and LLM inference orchestration pipelines**.
- Applied **prompt engineering, response evaluation, and reliability tuning** to improve output accuracy, consistency, and relevance.
- Deployed the solution as an **observable LLM inference service**, aligning with modern **GenAI and LLM production workflows**.

YouTube Comments Sentiment Analyzer | link - https://youtube-ai-analyzer-ndzqo6r2mepjrtsdjmwaxl.streamlit.app/

- Developed an **end-to-end NLP modeling system** using **BERT / DistilBERT** for large-scale sentiment classification.
- Performed **EDA, text preprocessing, feature engineering, model training, and statistical evaluation** on real-world user-generated data.
- Conducted **baseline comparisons and experiment tracking** to optimize model performance.
- Deployed the model as a **REST API-based inference service** to deliver actionable sentiment insights to business stakeholders.

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)