AWANTIKA SRIVASTAVA

Deep Learning Engineer | Vision Transformers | ADAS

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

PROFILE SUMMARY

Deep Learning Engineer With strong expertise in neural network design, Vision Transformers, and model for real-time inference system. Experienced in building perception models, implementing self-attention mechanisms, and optimizing deep networks using quantization and pruning techniques. Skilled in PyTorch-based model development, accuracy-latency tradeoff optimization, and deployment-oriented deep learning engineering. Strong foundation inlinear algebra, gradient-based optimization, and deep learning architecture design.

CORE TECHNICAL SKILLS

- **Programming & Data Handling:** Python, Numpy, Pandas, Matplotlib, Scikit-learn, C++, SQL.
- **Statistics & Mathematics:** EDA, Statistical Modeling, Hypothesis Testing, Probability, Linear Optimization, Linear Algebra (Vectors, Matrices)
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, Hyperparameter Tuning, Feature Engineering, Time-Series Forecasting, ARIMA, SARIMA
- **Deep Learning:** Neural Networks, CNN, RNN, LSTM, BERT, Vision Transformers (ViT), Self-Attention, Cross-Attention, Multi-Head Attention, Sequence Modeling, TensorFlow, PyTorch, Keras
- **Computer Vision**: Image Classification & Preprocessing, Object detection (YOLO, SSD, MobileNet, ResNet), Feature Extraction, Multi-Modal Fusion Concepts
- **NLP**: Text preprocessing, Tokenization, Chunkins, Transformers, Sentiment Analysis, Semantic Search, Transformer-based models, NLTK, SciPy
- **Model Optimization:** Quantization, Pruning, Model Compression, Efficient Inference Optimization, Accuracy-Latency Tradeoff
- **Tools:** Git, GitHub, Docker, AWS, EC2
- **Deployment:** REST APIs, Real-Time Inference Pipelines

EXPERIENCE

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

- Designed and implemented **Deep Learning architectures in PyTorch**, including custom **neural network layers** and **loss functions** for optimized model performance.
- Developed **Transformer-based models** utilizing **Self-Attention** and **Multi-Head Attention** mechanisms for advanced feature representation.
- Applied **Gradient Descent optimization** and **Backpropagation** techniques to ensure stable convergence and efficient training.
- Performed extensive **Hyperparameter Tuning** and architecture experimentation to improve accuracy and generalization.
- Implemented **Model Compression techniques** such as **Quantization** and **Structured Pruning** to reduce model size and enhance inference efficiency.
- Optimized models for **Latency, Throughput, and Memory Efficiency** to enable real-time inference deployment scenarios.
- Followed **Modular Software Engineering practices**, maintained **Git-based version control**, and collaborated with cross-functional teams for production integration

PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- Designed and **deployed a real-time computer vision**–based ML system to detect unsafe driver behaviors from continuous video streams.
- Trained and **optimized CNN-based** object detection **models (SSD MobileNet** architecture) to perform real-time inference on video data.
- **Implemented end-to-end ML pipelines** covering data ingestion, preprocessing, model training, evaluation, and production inference.
- Achieved **20–25 FPS** real-time processing with **<150 ms** inference latency by optimizing models for **deployment**.
- Built and integrated **a Flask-based** web **interface** to **visualize** detections and automatically recorded 30-second event clips, reducing manual **review** effort.

Amazon Stock Price prediction | Applied ML Project

- Built **LSTM-based time-series model** on 5+ years stock data, achieving **18% lower RMSE** vs baseline.
- Engineered **lag, rolling average, volatility features**, reducing **MAE by ~14%**.
- Implemented **BPTT with adaptive learning rate**, improving training stability.
- Applied **dropout & regularization**, cutting overfitting gap by **~20%** and enabling reliable forecasting.

Vision Transformer – Object Detection & BEV Mapping | Applied DL Project

- Implemented **Vision Transformer (ViT)** with **Patch Embeddings, Positional Encoding, Multi-Head Self-Attention**, achieving **6% higher feature accuracy** over CNN baseline.
- Experimented with **Cross-Attention** for spatial fusion and simulated **BEV-style feature projection** for structured perception modeling.
- Performed **hyperparameter tuning** (embedding size, heads, depth), improving convergence speed by **22%**.
- Applied **Structured Pruning & Quantization**, boosting inference throughput by **~30%** with <2% accuracy drop.
- Optimized models along the **accuracy-latency Pareto curve** for real-time deployment trade-offs.

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)