Abhishek Gajul

NJ, 07029 | (347) 378-5021| abhishek.g@mymailshub.com | LinkedIn

Professional Summary

- AI/ML Engineer with 5+ years of experience in designing, deploying, and scaling end-to-end machine learning solutions across NLP, Computer Vision, Generative AI (LLMs), and advanced analytics using deep learning and statistical modeling.
- Expert in developing supervised and unsupervised models using Logistic Regression, Random Forest, XGBoost, SVM, KMeans, DBSCAN, and ensemble techniques, optimized for high performance and scalability.
- Expert in fine-tuning and deploying deep learning architectures including CNNs, LSTMs, Transformers, and large language models such as BERT, GPT, LLaMA, utilizing frameworks like TensorFlow, PyTorch, Hugging Face, and Keras.
- Proficient in Python programming and data engineering with expertise in Pandas, NumPy, Scikit-learn, OpenCV, NLTK, SpaCy, Matplotlib, Seaborn, and tools for feature engineering, data visualization, and NLP pipeline development
- Skilled in ML infrastructure management and deployment using AWS (SageMaker, EC2, S3, Lambda), Google Cloud (Vertex AI, BigQuery), Azure DevOps, Docker, Kubernetes, Terraform, CI/CD pipelines, and MLOps best practices for production-grade AI systems.

Education

Master of Science in Computer Science | New Jersey institute of Technology

Skills

Language/ IDE's: Python, MATLAB, Jupyter Notebook, Google Colab, VS Code, SSMS
Machine Learning: Linear, Logistic Regression, Decision Trees, Random Forests, NumPy, SVM, A/B Testing, Naive Bayes
Deep Learning: CNN, RNN, LSTM, NLP, Large Language Model (LLM), LangChain, Hugging Face Transformers (BERT, GPT-3)
Cloud/Visualizations: AWS (EC2, SQS, SNS, Code Deploy, CloudWatch, API Gateway), GCP (Vertex AI, Google Cloud Storage), Tableau, Power BI
Statistical Techniques: Hypothesis Testing, Data Visualization, Data Modelling, A/B testing, Model Evaluation
Packages and Frameworks: NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, Keras, NLTK, XGBoost, PyTorch
Database and Tools: SQLServer, MySQL, PostgreSQL, Redis, Neo4j
AI apps: Large Language Models (LLMs), Transformers, Computer Vision, Generative AI, Prompting/Prompt Engineering

Work Experience

Johnson& Johnson, New Brunswick, NJ **Jan 2025-Present**

AI/ML Developer

- Designed and implemented advanced deep learning models using convolutional neural networks and sophisticated computer vision techniques to analyze complex medical imaging data, achieving 94% validation accuracy and reducing diagnostic processing time by 40%
- Engineered and optimized over 15 critical behavioral and physiological features including gaze deviation, interaction latency, and attention metrics using cosine similarity, significantly improving model precision by 38% for real-time patient monitoring applications
- Developed highly efficient real-time gaze-tracking algorithms with linear transformation and normalization techniques to ensure consistent, high-quality data input across diverse devices, boosting model reliability by 25%
- Built scalable and automated machine learning pipelines on AWS Lambda, Glue, and S3 using PyTorch and TensorFlow, streamlining training and retraining workflows and reducing data engineering overhead by 60%
- Collaborated closely with leading healthcare researchers to translate complex clinical diagnostic parameters into actionable machine learning features, contributing to the development of explainable AI systems fully aligned with FDA regulatory guidelines
- Created Power BI dashboards integrated with SQL databases to continuously monitor key model performance metrics including inference speed, attention drift, and threshold sensitivity, providing actionable insights for product and research teams

Goldman Sachs, Jersey City, NJ **Jan 2024– Dec 2024**

AI & Deep Learning Developer

- Architected cutting-edge deep learning models for financial time-series forecasting, significantly improving asset price prediction accuracy by 22%, enhancing algorithmic trading strategies and portfolio optimization
- Implemented NLP pipelines using BERT transformers to analyze financial news and social media sentiment, enhancing trading signal precision by 18% and improving market responsiveness
- Executed scalable, cloud-native machine learning workflows on Google Cloud Platform BigQuery and Vertex AI, reducing model retraining time and risk evaluation latency by 50% while ensuring high availability
- Applied unsupervised learning and autoencoder techniques for transaction anomaly detection, reducing false positives by 30% and strengthening fraud detection capabilities
- Streamlined end-to-end CI/CD pipelines for machine learning lifecycle management with Docker, Kubernetes, and GitHub Actions, reducing deployment times by 40% and ensuring seamless production rollouts
- Partnered with quantitative analysts, traders, and compliance teams to ensure models adhered to regulatory standards and business KPIs, improving transparency, reliability, and stakeholder trust

Accenture, Noida, India **Jan 2019–Jul 2022**

Machine Learning Engineer

- Devised deep learning models including convolutional and recurrent neural networks for automated document classification and object detection, increasing operational throughput by 40% and accelerating enterprise AI adoption
- Optimized and fine-tuned NLP models using BERT and GPT-3 via Hugging Face Transformers for sentiment analysis and intent detection, improving classification accuracy by 20% and reducing customer escalation times
- Formed reusable PyTorch and TensorFlow pipelines with automated hyperparameter tuning and batch inference, reducing model training time by 60% and ensuring reproducibility and scalability across deployments
- Formulated and deployed real-time AI services using OpenCV and LSTM networks to detect abnormal behaviors in video streams, reducing manual fraud monitoring efforts by 55% and strengthening compliance controls
- Transformed legacy machine learning models to AWS SageMaker using containerized deployments and established API Gateway and Lambda endpoints, reducing deployment time by 50% and enabling seamless model scaling
- Orchestrated model evaluation, logging, and drift detection with MLflow and AWS CloudWatch, enhancing governance, reducing production incidents by 40%, and ensuring continuous model performance
- Crafted and maintained Tableau dashboards integrated with prediction confidence intervals and model impact KPIs, boosting stakeholder decision-making efficiency by 30% and accelerating data-driven strategy adoption across departments