VAISHNAVI KUKKALA

AI/ML Engineer

USA | +1 (571) 637-9044 | Email: kukkalavaishnavi07@gmail.com LinkedIn | github.io/portfolio/

SUMMARY

- An **AI/ML Engineer with around 4 years of experience** in building end-to-end ML pipelines using Python, PySpark, and AWS SageMaker, automating data preprocessing, feature engineering, and model deployment.
- Strong with NLP, computer vision, time-series forecasting, and anomaly detection, leveraging BERT, LSTM, XGBoost, and deep learning architectures to deliver impactful solutions.
- Proficient in MLOps practices with MLflow, Airflow, Docker, and Kubernetes, ensuring reproducibility, scalability, and zero-downtime deployments in production environments.
- Proven track record of business impact, including $2.5M+ fraud detection savings, reductions in manual work, and measurable improvements in forecasting accuracy and customer retention.

PROFESSIONAL EXPERIENCE

AI/ML Engineer | TCS (Internship) | USA **Jan 2025 – Current**

- Designed and productionized scalable ML pipelines using Python, PySpark, and SageMaker, automating data preprocessing, feature engineering, hyperparameter tuning (Optuna), and deployment that helps in cutting deployment time by 70% and improved churn prediction accuracy.
- Developed BERT, TCN, and XGBoost models for financial time-series forecasting and fraud detection, processing 2M+ daily transactions in real time with less than 2% false positives, preventing $1.8M in annual losses.
- Participated in implementing NLP pipelines using Hugging Face Transformers, spaCy, OpenCV, and YOLOv5 for sentiment analysis, OCR, and compliance automation that led to a reduction in manual review effort by 60% and ensured nearly 100% classification accuracy.
- Built robust MLOps frameworks with MLflow, Airflow, Docker, and Jenkins, enabling automated retraining, drift monitoring, reproducible deployments, and scalable SageMaker inference endpoints with zero downtime.
- Collaborated with cross-functional teams, translating business requirements into AI-driven solutions, performing A/B testing, and delivering measurable improvements in customer retention, portfolio performance, and operational efficiency.

AI/ML Engineer | HCL Tech | India **June 2020 – July 2023**

- Developed anomaly detection models using ensemble methods on healthcare claims and billing codes, detecting $2.5M+ in fraudulent transactions annually with 96% precision and 91% recall.
- Designed hospital capacity forecasting models with ARIMA, Prophet, and LSTM to predict ER visits, patient inflow, and equipment utilization, improving forecasting accuracy and reducing patient wait times.
- Automated clinical document processing like discharge summaries, prescriptions, and pathology reports using BERT and custom NLP pipelines to cut down manual audit efforts while ensuring HIPAA compliance.
- Built personalized care recommendation engines with collaborative and demographic filtering, increasing patient engagement by 20% and boosting repeat visit rates.
- Engineered HIPAA-compliant ETL pipelines using PySpark and Apache Airflow to integrate multi-source EHR and hospital datasets, improving ingestion speed by 60% and enabling real-time analytics.
- Deployed scalable AI microservices with Flask/FastAPI, Docker, and Azure Kubernetes, achieving nearly 100% uptime and enabling on-demand scaling across 6+ healthcare facilities.
- Assisted while implementing Explainable AI (XAI) frameworks such as SHAP, LIME, and ELI5 to enhance clinical trust, supporting FDA audit approvals, and increasing physician adoption of ML outputs.
- Mentored junior engineers and contributed to AI CoE initiatives, standardizing ML templates, reviewing code, and leading knowledge-sharing sessions on Responsible AI and regulated model deployment.

TECHNICAL SKILLS

Programming & Scripting:	Python, R programming, Scala, MATLAB, Java, C/C#, .NET
Machine Learning:	Regression, Classification, Clustering, Feature Engineering, Model Tuning (Grid/Random/Optuna)
Deep Learning:	TensorFlow, PyTorch, Keras for CNNs, RNNs, LSTMs, Transformers, GANs, and attention-based architectures
NLP & Text Mining:	BERT, GPT-based LLMs, spaCy, NLTK, Hugging Face Transformers, NER, and Summarization
Computer Vision:	OpenCV, YOLO, Detectron2 for image recognition, object detection, segmentation, and OCR applications
Data Engineering & Big Data:	Apache Spark, PySpark, Hadoop, Kafka, Airflow for large-scale ETL, batch/stream processing
Databases & Querying:	SQL, NoSQL (MongoDB, Cassandra), Snowflake, BigQuery
Cloud & MLOps:	AWS (SageMaker, S3, Lambda, EKS), Azure ML, GCP Vertex AI, MLflow, Kubeflow, Docker, Kubernetes
Data Visualization & BI Tools:	Tableau, Power BI, Matplotlib, Seaborn, Plotly
CI/CD & DevOps for AI:	Jenkins, GitHub Actions, Terraform, Ansible for automation and scalable AI/ML deployments
Statistical Analysis:	Time-series forecasting (ARIMA, Prophet, LSTM), hypothesis testing, Bayesian modeling
Agile & Collaboration Tools:	Jira, Confluence, Git, Bitbucket, Slack, MS Teams, backlog refinement, and cross-team collaboration

EDUCATION

Master of Science in Data Science | University of New Haven, West Haven, CT, USA **May 2025**

CERTIFICATION

Microsoft Certified: Azure Fundamentals (AZ-900) – Microsoft **| Oct 2022**
Tata - GenAI Powered Data Analytics Job Simulation by Forage **| Issued Jul 2025**
Python for Data Science by IBM **| Issued May 2023**

PUBLICATIONS

1.) Integrating MobileNetV3 and SqueezeNet for Multi-class Brain Tumor Classification | **Link**
2.) Miniaturized Planar Dual Band Monopole UWB Antenna using Capacitively Loaded Loop Resonator with Notch Characteristics | **Link**

PROJECTS

Plagiarism Detection using Transformers **| Dec 2024**
Object Detection using yolov5s **| Dec 2024**