VAISHNAVI KUKKALA

AI/ML Engineer

USA | +1 (571) 637-9044 | Email: vaishnavi@careerattainment.com | OPT

LinkedIn | github.io/portfolio/

SUMMARY

An AI/ML Engineer with 4+ years of experience designing and scaling ML solutions across computer vision, forecasting, and data analytics domains. Proficient in Python, scikit-learn, and cloud-based ML platforms (AWS, GCP, Azure). Experienced in automating data pipelines, performing statistical modeling, and building visualization dashboards using Power BI and Excel. Adept at leveraging AI assistants (ChatGPT, Copilot, Claude) for research, analytics, and narrative generation. Proven record of translating data insights into executive-level recommendations and delivering measurable business impact in manufacturing and healthcare sectors.

PROFESSIONAL EXPERIENCE

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

- Design end-to-end ML pipelines on AWS SageMaker and Vertex AI using Python & PySpark, cutting model deployment time by 70%.
- Develop advanced LLM pipelines using RAG architecture for technical documentation analysis, creating GenAI apps using LangChain that reduced manual review time by 45% through intelligent query resolution.
- Create a GenAI app using LangChain for maintenance troubleshooting, leveraging RAG architecture to access technical manuals and historical repair data, and cutting diagnostic time by 30%.
- Develop predictive maintenance models using XGBoost and CNN architectures, enhancing with transformer fine-tuning for time-series data from 200+ factory assets, achieving 94% prediction accuracy with integrated LLMOps monitoring.
- Build multimodal AI inspection systems combining computer vision (YOLOv5) and textual data analysis, reducing manual inspection workload by 60% while maintaining 99.8% defect detection accuracy.
- Implement cloud-native ML frameworks with MLflow & Kubernetes, establishing comprehensive LLMOps monitoring for automated retraining and achieving zero-downtime deployments across production AI microservices.
- Performed statistical data correlation and forecasting analysis using Excel, Pandas, and Power BI dashboards to support management decisions in production analytics.
- Used ChatGPT and GitHub Copilot to debug, optimize Python scripts, and accelerate report generation, reducing development time by 35%.
- Leveraged Claude and Gemini for literature synthesis, competitive benchmarking, and model documentation.
- Build real-time anomaly detection systems deployed as FastAPI inference services, processing Kafka streams to identify production issues within seconds and reducing quality incidents by 25%.
- Created executive-level PowerPoint presentations translating ML insights into actionable business recommendations for directors and product teams.

AI/ML Engineer | HCL Tech (Full-time) | India **June 2020 – July 2023**

- Developed anomaly detection models using ensemble methods on medical claims data, identifying $2.5M+ in annual claim anomalies with 96% precision and 91% recall.
- Designed hospital capacity forecasting models with ARIMA, Prophet, and LSTM to predict ER visits and bed utilization, reducing patient wait times by less than 4 minutes.
- Automated clinical document processing for discharge summaries and prescriptions using BERT and custom NLP pipelines, cutting manual audit efforts by 60% while maintaining HIPAA compliance.
- Built personalized suggestion systems with collaborative filtering, increased 20% patient engagement rates & reduced 15% readmission rates.
- Engineered HIPAA-compliant ETL pipelines using PySpark and Apache Airflow to process multi-source EHR data, improving data ingestion speed and enabling real-time clinical analytics.
- Deployed scalable AI microservices with FastAPI, Docker, and Azure Kubernetes, achieving 99.9% uptime across 6+ healthcare facilities while ensuring PHI security.
- Implemented Explainable AI frameworks using SHAP and LIME to enhance clinical trust, supporting FDA audit approvals and increasing physician adoption of ML outputs.
- Developed interactive Tableau dashboards for monitoring KPIs and defect rates, improving visibility across business teams.
- Built demand and resource forecasting models using scikit-learn and statsmodels; optimized scheduling scenarios using PuLP-based linear programming.
- Mentored junior engineers and contributed to AI Centre of Excellence initiatives, standardizing ML templates and leading knowledge-sharing sessions on Responsible AI in healthcare.

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), RAG
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 Summarisation
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), 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, LLM Pipelines
Statistical Analysis:	Time-series forecasting (ARIMA, Prophet, LSTM), hypothesis testing, Bayesian modelling
Forecasting & Optimization:	Scikit-learn, statsmodels, PuLP, OR-Tools, Excel Solver
Agile & Collaboration Tools:	Jira, Confluence, Git, Bitbucket, Slack, MS Teams, backlog refinement, and cross-team collaboration
Business Communication:	PowerPoint storytelling, AI-assisted narrative writing (ChatGPT, Gemini), Data storytelling.

CERTIFICATION

Microsoft Certified: Azure Fundamentals (AZ-900) – Microsoft **| Oct 2022**

Tata - GenAI Powered Data Analytics Job Simulation by Forage **| Issued Jul 2025**

Microsoft Excel Data Analyst Associate (PL-300) | **Issued Jan 2025**
Learning Power BI Desktop | **Issued Jan 2025**
Python for Data Science by IBM | **Issued May 2023**

PUBLICATIONS

- Integrating MobileNetV3 and SqueezeNet for Multi-class Brain Tumour Classification |
  Link
- Miniaturised 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**

EDUCATION

Master of Science in Data Science | University of New Haven, West Haven, CT, USA | GPA: 3.94/4 **May 2025**
Bachelor's in Electronics and Communication Engineering | Vardhaman College of Engineering, Hyderabad, TS, India **June 2021**