Yuvaraja Reddy Avuthu

+1 716-910-7802 | yuvaraja.avuthu@gmail.com | Linkedin | Portfolio | Github | Hugging Face

SUMMARY

With 3+ years of experience as an AI/ML Engineer specializing in Generative AI, Large Language Models (LLMs), NLP, and machine learning, with expertise in AWS and GCP cloud platforms. Skilled in deploying scalable AI solutions using FastAPI, Docker, and frameworks like LangChain, OpenAI, and Hugging Face, delivering impactful applications in conversational AI, text generation, and real-time analytics.

TECHNICAL SKILLS

- **Generative AI (Gen AI)**: LangChain, GPT-3.5 Turbo, OpenAI, Hugging Face Transformers, Vector Databases (ChromaDB, Pinecone, FAISS), Llama3, Reinforcement Learning (RLHF), Prompt Engineering, Few-Shot Learning, RAG, LoRA, Fine-Tuning
- **NLP**: SpaCy, NLTK, Gensim, Hugging Face, Stanford NLP, TextBlob, Word2Vec, TF-IDF, BERT, Transformers, RNN, TensorFlow, PyTorch
- **Machine Learning & Deep Learning**: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, CNN, RNN, LSTMs, Transformers, ARIMA, Transfer Learning, Regression, Decision Trees, Random Forest, SVM, KNN, K-Means, Hierarchical Clustering, PCA, Naive Bayes, Predictive Modeling, Feature Engineering.
- **Cloud Platforms & DevOps:** AWS (S3, SageMaker, Lambda, Glue, EMR, RedShift, IAM, Kinesis, SNS, Athena, CloudWatch, Bedrock, API Gateway), GCP (Vertex AI, BigQuery), Docker, Kubernetes, Jenkins, Bitbucket, Fast API, RESTful API.
- **Big Data & Databases:** Apache Spark (SQL, Streaming), Kafka, Snowflake, Hadoop, Apache Airflow, ETL/ELT, PostgreSQL, MySQL, MongoDB, Oracle.
- **Programming, Visualization & Tools:** Python, SQL, Git, Pandas, NumPy, Jupyter, Anaconda, PySpark, Tableau, Looker Studio, Power BI.
- **Data Engineering & Orchestration:** Data Extraction, Data Validation, Dimensional Modeling, Data Warehousing, Workflow Orchestration

EXPERIENCE

Data Scientist (AI/ML Engineer) | BCBSA | Aug 2024 – Present

- Conducted statistical analysis and developed predictive models, reducing feature engineering workflows' runtime by 20% and improving machine learning model training efficiency.
- Extracted and analyzed large datasets using PySpark, SQL, and AWS tools, generating actionable insights that led to a 15% increase in product performance and revenue optimization.
- Designed and implemented GPT-like large language model prototypes, achieving 10% efficiency gains over baseline transformer models and enabling scalable deployment for real-world applications.
- Applied transfer learning and fine-tuning techniques to LLMs, improving performance for domain-specific tasks like conversational AI, sentiment analysis, and document summarization.
- Conducted prompt engineering experiments to optimize LLM outputs, saving 15% in content generation costs.
- Implemented natural language processing (NLP) for text classification, sentiment analysis & recommendation systems.
- Developed a cloud-native MLOps platform on AWS for scalable AI/ML deployment and management using Python, TensorFlow ensuring 99.9% availability while handling petabytes of medical imaging data.
- Employed advanced optimization techniques such as quantization and pruning (TensorFlow-Model-Optimization) to reduce ML model size by 60% without sacrificing accuracy.
- Collaborated with business and technical teams to analyze KPIs and recommend process enhancements, improving service delivery metrics by 20%.
- Built LLM-powered RAG applications with BigQuery-based feature engineering, improving model inference performance for NLP tasks such as sentiment analysis and document summarization.
- Established a scalable data ingestion pipeline on AWS (S3, Glue, Athena) for processing and storing terabytes of medical data from diverse sources, enduring reliability and scalability.
- Utilized pre-built machine learning algorithms (scikit-learn, XGBoost) for predictive analytics on medical data, resulting in an 18% improvement in diagnosis accuracy.
- Stayed up to date with the latest AI/ML trends and technologies, integrating emerging techniques like deep learning, reinforcement learning, and transfer learning into projects.

Data Scientist Analyst | Accenture AI | Jan 2021 – Jan 2023

- Led end-to-end NLP pipeline development for aviation insurance underwriting, fine-tuning spaCy's (en-core-web-lg) pre-trained NER model on 10K+ documents (PDF/DOC/images) to extract 8+ entity types with 92% accuracy.
- Applied statistical and machine learning techniques like regression, clustering, classification, and deep learning to drive business outcomes.
- Collaborated with cross-functional computer vision teams to preprocess 5K+ scanned documents using OCR pipelines, enabling high-accuracy text extraction from images and improving downstream NLP model performance by 18% for named entity recognition (NER) and classification tasks.
- Designed and implemented a Decision Tree classifier using Scikit-learn, leveraging NER-extracted entities and web-scraped risk metrics to automate 80% of insurance decision workflows, significantly reducing manual underwriting time and enhancing process accuracy.
- Engineered a scalable web scraping framework using BeautifulSoup and Selenium, automating real-time data extraction from 4+ external risk sources; integrated structured outputs into risk assessment pipelines, improving model feature richness and prediction reliability by 20%.
- Developed and deployed containerized FastAPI microservices with Docker, achieving sub-500ms response times for high-frequency NER and recommendation API calls, enhancing system scalability, fault tolerance, and real-time decision support for underwriting applications.
- Designed cross-sell recommender system using KMeans clustering (customer segments) and KNN (policy similarity), building Excel dashboards with pivot tables to visualize recommendations, increasing upsell conversion by 35% for aviation clients.
- Spearheaded Agile workflows via Jira/Confluence, documenting model versions and maintaining 99% uptime for production systems through Git-based CI/CD pipelines.
- Enhanced predictive analytics with Scikit-learn, building machine learning models using Random Forest, ARIMA, LSTM, and Logistic Regression to forecast financial trends, and risk exposure while ensuring statistical significance via ANOVA.
- Implemented AWS Lambda to automate data transformations and trigger event-driven workflows, enhancing efficiency in financial data processing.
- Coordinated cross-functional teams to resolve critical system issues, achieving 99.9% uptime and reducing 50% incident resolution time.

Data Analyst | Skyline Infrastructure | May 2020 – Dec 2020

- Utilized SQL to query and analyze large datasets, generating actionable insights that supported process optimization and strategic decision-making.
- Assisted in identifying operational inefficiencies and recommending process improvements to enhance performance.
- Assisted in developing resource allocation models by performing data analysis with SQL, contributing to a 15% improvement in project delivery timelines.
- Used tools like Excel, Python, R, and statistical software to perform quantitative analysis and modeling.
- Designed and maintained dashboards to monitor key performance indicators (KPIs) using SQL and data visualization tools, enabling real-time decision-making for operations teams.
- Operated financial market research, collected data and performed analysis by using statistical software
- Performed sensitivity analysis and risk assessment using SQL and statistical techniques to evaluate the impact of variable changes on business outcomes, enhancing decision-making accuracy under uncertainty.

PUBLICATIONS

Performance Comparison of Machine Learning Algorithms in Symbol Detection Using OFDM Jan 2022
Inventive Communication and Computational Technologies

MSE and BER Analysis of Text, Audio and Image Transmission Using ML Based OFDM Jan 2021
IEEE International Conference for Innovation in Technology (INOCON)

EDUCATION

- **Master's in data science** from University at Buffalo, SUNY, NY – USA.
- **Bachelor's in Electrical and Communication Engineering** from Amrita Vishwa Vidyapeetham, - INDIA

CERTIFICATIONS

AWS Cloud | Gen AI | Python | NLP | Deep Learning | Machine Learning