Yashwanth Telukuntla
Machine Learning Engineer

(218) 277-2215 | yashwanth.t808@gmail.com | Katy, TX | LinkedIn

PROFESSIONAL SUMMARY

- AI/ML Specialist with 6+ years of experience designing and implementing cutting-edge machine learning solutions, Statistical Modeling, Predictive Modeling, Data Analytics, Data Modeling, Data Architecture, Data Analysis, Data Mining, Text Mining. Expertise in developing and deploying sophisticated AI models to drive data-driven decision-making and innovation.
- Strong foundation in data structures, algorithms, and system architecting, enabling the development of scalable and efficient in large-scale ML systems. Proficient in deploying machine learning models via APIs and collaborating with cross-functional teams to drive data-driven decision-making and innovation.
- Proficient in applying advanced machine learning algorithms for classification, regression, clustering, anomaly detection and dimensionality reduction tasks across diverse business domains.
- Strong background in natural language processing (NLP), including sentiment analysis, text classification, named entity recognition, and large language modeling and Multi-modal using state-of-the-art techniques, text preprocessing, tokenization topic modeling, to extract meaningful insights from unstructured data and support data-driven decision-making.
- Expertise in Deep learning, including convolutional neural networks (CNNs) for image processing and recurrent neural networks (RNNs) and Artificial Neural Networks (ANN) for sequence modeling tasks.
- Skilled in developing and optimizing recommendation systems using collaborative filtering and content-based approaches.
- Proficiency in feature engineering, model training, evaluating, hyperparameter tuning, and cross-validation to ensure optimal AI model performance and generalization.
- Profound knowledge of generative AI techniques, including generative adversarial networks (GANs) and variational autoencoders (VAEs), for creating synthetic data, images, and text.
- Extensive experience with large language models (LLMs) such as GPT 3.5 and GPT 4, BERT, and llama 2 leveraging transfer learning and fine-tuning for various NLP tasks.
- Proficient in working with large-scale language models and neural networks, applying advanced concepts in context learning and few-shot learning.
- Expertise in Retrieval-Augmented Generation (RAG) techniques, combining the power of LLMs with external knowledge bases like Vector Data Bases to enhance accuracy and relevance of AI-generated content.
- Experience with reinforcement learning techniques for decision-making processes in complex, dynamic environments.
- Adept at implementing computer vision algorithms for object detection, image segmentation, and facial recognition tasks.
- Expertise in time series analysis and forecasting using both traditional statistical methods and modern deep learning approaches.
- Skilled in developing and deploying chatbots and conversational AI systems, integrating advanced NLP techniques for enhanced user interactions. Implemented A/B testing frameworks for ML models to optimize performance and user experience.
- Worked with CI/CD pipelines for ML models, ensuring continuous integration, testing, and deployment of AI solutions with a strong background in data versioning and model versioning using tools like DVC (Data Version Control) and MLflow.
- Expertise in monitoring ML models in production, including setting up alerts for model drift and implementing automated retraining pipelines.
- Mentored junior developers on the effective use of design patterns, contributing to improved code quality across the team and conducted code reviews and resolved complex technical challenges.
- Knowledge in ethical AI practices, including bias detection and mitigation in machine learning models

CERTIFICATIONS

- Generative AI Professional (Oracle)
- Machine Learning specialization (Coursera)

EDUCATION

University at Albany, SUNY, Albany, New York, USA.

Master of Science in Computer Science (Concentration in AI)

TECHNICAL SKILLS

Programming Languages: Python, R, Java, C/C++, YAML, JSON, HTML, CSS.

Databases: ChromaDB, Pinecone, Oracle, MySQL, HBase, MongoDB, GraphRAG, Neo4j.

Frameworks: TensorFlow, Keras, PyTorch, Scikit-learn, XGBoost, LSTM, Fast API, Streamlit, Django, Flask, NumPy, pandas, matplotlib, SciPy, Llama Index.

Natural Language Processing (NLP): NLTK, spaCy, Hugging Face Transformers, Gensim.

Version controls and Tools: Git, GitHub, GitLab, Bitbucket.

Data Visualization Tools: Tableau, Power BI, Plotly.

Development Tools & IDEs: PyCharm, Visual Studio Code, JupyterLab, Eclipse, IntelliJ IDEA.

AWS Ecosystem: S3Bucket, Data Lake, AWS Lambda, Bedrock, Sagemaker.

Azure Ecosystem: Azure DataLake, ADF, Databricks, Azure SQL, Azure OpenAI.

GCP Ecosystem: Google AI Platform, Vertex AI, Google Analytics, Google Cloud Big Data.

Monitoring & Logging Tools: Prometheus, Grafana, ELK Stack (Elasticsearch) .

MLOps & DevOps Tools: MLflow, Kubeflow, Airflow, Docker, Kubernetes, Jenkins, Circle CI, GitLab CI/CD.

Hadoop Components / Big Data: HDFS, MapReduce, PIG, Apache Spark, Hadoop, Kafka, Snowflake.

Other Skills: Collaboration, Interpersonal skills, Problem solving, information retrieval.

WORK EXPERIENCE

Client: American National Insurance Company, Galveston, Texas, USA July 2023 - Present

Role: Sr. Machine Learning Engineer

Description: American National Insurance Company (ANICO) is a prominent financial services and insurance company offers a wide range of insurance products and financial services across the United States. Designed and developed scalable Generative AI solutions to enhance user engagement and experience, while ensuring efficient data handling, security, and compliance across various applications.

Responsibilities:

- Designed and developed scalable Generative AI solutions using llama 2 and Langchain frameworks, building intelligent applications to enhance user engagement and experience.
- Implemented Retrieval-Augmented Generation (RAG) techniques by integrating external data sources and embeddings, utilizing Hugging Face Transformers to improve AI model responses and context awareness.
- Created and maintained AI applications using Langchain and LangSmith, integrating embedding techniques for seamless data flow and enhanced semantic understanding in AI models.
- Crafted and optimized prompts for large language models to elicit desired outputs, enhancing the quality and relevance of generated content and experimented with various prompt structures and techniques to guide models towards specific behaviours and responses.
- Fine-tuned transformer-based models using PyTorch and TensorFlow, tailoring llama 2 and Hugging Face models for domain-specific applications to improve accuracy and relevance.
- Integrated Hugging Face Transformers into AI pipelines, leveraging pre-trained models and applying fine-tuning with PyTorch and TensorFlow to enhance natural language understanding capabilities.
- Automated AI workflows and model training processes using AWS services including Lambda, Sagemaker, and Step Functions, ensuring efficient CI/CD pipelines and rapid deployment of Generative AI models.
- Performed ongoing monitoring, automation, and refinement of AI models using LangSmith and AWS CloudWatch, ensuring optimal performance and accuracy across applications.
- Ensured data security and compliance standards were met when handling sensitive data within AI models, adhering to best practices and security protocols across OpenAI, Langchain, and AWS services.

- Collaborated with cross-functional teams, providing technical guidance and mentorship on Generative AI technologies such as OpenAI, Langchain, embeddings, Hugging Face Transformers, PyTorch, and TensorFlow, fostering a learning environment and staying updated with industry trends.

Environment: NLTK, spaCy, Open AI, llama 2, GPT 3.5, GPT 4, API, Azure, AWS, CI/CD, Language Models, Text Generation, Text Summarization, Vector database, embeddings, Sentiment Analysis, Jenkins, Kubernetes, Fast API, Oracle, Power BI, PySpark, Python, Spark Streaming, Agile Methodology.

Client: Methodist Health System, Dallas, Texas, USA **Nov 2021 - June 2023**

Role: Machine Learning

Description: Methodist Health System is a non-profit healthcare provider It operates a network of hospitals, outpatient clinics, and specialized health centres. Involved in transforming and analysing data, developing predictive models, and implementing machine learning techniques to enhance data-driven decision-making.

Responsibilities:

- Collaborated with cross-functional teams to gather requirements, define project scopes, and translate business needs into technical specifications, ensuring that NLP and ML solutions are aligned with organizational goals and deliver measurable value.
- Built and maintained scalable machine learning pipelines using MLflow model registry and auto logging features. This included setting up automated logging of model parameters and metrics and managing model versions to streamline the deployment process.
- Monitored and maintained model performance through continuous evaluation using quality metrics such as precision, recall, F1-score, ROC-AUC, and confusion matrices, enabling proactive adjustments and ensuring sustained accuracy and reliability of deployed models.
- Optimized ML workflows by automating repetitive tasks, integrating version control systems, and streamlining data ingestion processes, thereby enhancing productivity and ensuring consistency across various stages of model development and deployment.
- Designed and deployed scalable machine learning models using Google Vertex AI on Google Cloud Platform (GCP), enhancing prediction accuracy and operational efficiency by leveraging GCP's robust infrastructure and integrated ML tools.
- Implemented end-to-end MLOps pipelines with GCP tools such as Cloud Build, Cloud Source Repositories, and Cloud Deploy, streamlining model training, testing, and deployment processes to ensure seamless integration and continuous delivery of ML models.
- Managed large-scale data ingestion and preprocessing workflows on GCP using Google Cloud Dataflow and Cloud Storage, ensuring high-quality datasets for ML model training by utilizing GCP's scalable data processing capabilities.
- Monitored and logged model performance and infrastructure health using Google Cloud Monitoring and Logging, facilitating proactive issue detection and resolution, and ensuring the reliability and scalability of NLP and ML applications deployed on GCP.

Environment: Docker, Kubernetes, AWS, EC2, S3, Lambda, Python, Jupyter Notebook, Databricks, GIT, Microservices, Named Entity Recognition (NER), Part-of-Speech (POS), Unix/Linux, Airflow, Data Visualization, Data Mining, Grafana, Model Monitoring, Gensim, TensorFlow, Keras.

Client: Deutsche Bank, Bangalore, India **Mar 2020 - Jul 2021**

Role: Machine Learning Engineer

Description: Deutsche Bank is a global investment bank offering financial products and services to corporations, governments, and institutional investors. Contributed to the development of machine learning models and statistical analysis to solve business problems and improve data-driven decision-making processes.

Responsibilities:

- Implemented comprehensive MLOps pipelines that automated the end-to-end lifecycle of machine learning models, including data collection, cleaning, preprocessing, transforming, training, validation, deployment, and monitoring, thereby enhancing efficiency and reducing time-to-market.
- Ensured data security and compliance by implementing robust access controls, encryption methods, and data governance practices, adhering to industry standards and regulatory requirements to safeguard sensitive information and maintain data integrity.
- Collaborated with data engineers and operation team to collect data from internal system to fit the analytical requirements.
- Utilized containerization technologies such as Docker and orchestration tools like Kubernetes to deploy and scale machine learning applications efficiently, ensuring high availability, scalability, and streamlined resource management in production environments.
- Served ML models via RESTful APIs to integrate with front-end applications, enhancing accessibility and user experience.

- Conducted extensive data preprocessing and feature engineering using techniques like normalization, dimensionality reduction, feature selection, and embedding generation, ensuring high-quality inputs for machine learning models and improving overall model performance.
- Collaborated with team members and translated functional requirements to technical requirements for development.

Environment: Python, Pandas, Scikit learn, logistic Regression, LSTM, Random Forest, Decision Trees, Support Vector Machines (SVM), Word2Vec, CNN, Random Forest, TensorFlow, Keras, Seaborn, Text Classification, Preprocessing, Model Versioning.

Client: Schneider Electric, Schneider Electric, India **June 2018 - Feb 2021**

Role: Data Engineer

Description: Schneider Electric is a global leader in digital transformation of energy management and automation. Involved in developing and managing ETL processes, automating data workflows, and creating dashboards to ensure efficient data integration and reporting across various platforms.

Responsibilities:

- Develop ETL processes to extract data from various sources, transform it into a usable format, and load it into data storage solutions.
- Implementing and Managing ETL solutions and automating operational processes and responsible for ETL and data validation using SQL Server Integration Services.
- Worked on building dashboards in Tableau with ODBC connections from different sources like Big Query/ presto SQL engine and developed stored procedures in MS SQL to fetch the data from different servers using FTP and processed these files to update the tables.
- Involved in using SAP and transactions done in SAP - SD Module for handling customers of the client and generating the sales reports.
- Building/Maintaining Docker container clusters managed by Kubernetes Linux, Bash, Git, Docker. Implemented a continuous delivery (CI/CD) pipeline with Docker for custom application images in the cloud using Jenkins
- Hands on experience with building data pipelines in python/PySpark/Hive SQL/Presto and Monitored Data Engines to
- Created ETL Pipeline using Spark and Hive for ingest data from multiple sources.
- Carried out data transformation and cleansing using SQL queries, Python and PySpark.
- Expertise knowledge in Hive SQL, Presto SQL and Spark SQL for ETL jobs and using the right technology to get the job done.

Environment: CI/CD, Cluster, Data Factory, Docker, ETL, Factory, HBase, Hive, Java, Jenkins, Kubernetes, pivot tables, Spark, Python, ETL, Power BI, Tableau, Hive/Hadoop, Snowflakes, Power BI.