SAI VISHWA GADDAM

♦ Fremont, CA, USA♦+1(510)-361-8334 ♦saivishwa.gaddam@gmail.com ♦ LinkedIn ♦ GitHub

Skills

- **Programming and Development:** Python, R, Flask, GitHub Actions, Docker.
- **Data Science:** Data Science, Statistics, Exploratory Data Analysis (EDA), Data Visualization.
- **Machine Learning**: Machine Learning, Deep Learning, Natural Language Processing (NLP), Generative AI.
- **MLOps and Development**: MLOps, ModelOps, PyTorch, LangChain, WhyLabs, Transformers, MLflow ,Splunk, LLM, LLM Fine-tuning, Prompt Engineering, Datadog, Azure Promptflow
- **Database Management**: SQL, MySQL, MSSQL, PostgreSQL, NoSQL, RDBMS.
- **Data Analysis and Visualization:** ETL, Excel, Tableau, Power BI, DVC, A/B Testing, Data warehouse.
- **Big Data and Cloud Technologies:** AWS, Azure, Hadoop, PySpark, ONNX Runtime**,** Azure DataBricks, GCP.
- **Additional Skills**: Statistical Analysis, Business Insights,vector database, Stable Diffusion, computer vision.

Work Experience

LTIMindtree 08/2024 -
Data Scientist

- Implemented end-to-end workflows in Azure Promptflow, where datasets were provided as input to generate LLM responses and evaluated using AI-assisted metrics such as coherence, groundedness, relevance, and fluency.
- Designed and implemented evaluation frameworks for Large Language Models (LMMs) using Azure AI and Python scripting, Incorporating AI-assisted and traditional metrics like coherence and groundedness.
- Designed and implemented data pipelines and analytics solutions using Azure Databricks, enabling efficient processing and analysis of large-scale datasets for business insights.
- Built a FastAPI-based prompt management interface to track and manage prompts efficiently, enabling performance analysis using both AI-assisted metrics (e.g., coherence, relevance, fluency) and the traditional metrics (e,g. accuracy, F1 Score) at the date and prompt level.
- Collaborated with engineering teams to prototype low-latency LLM pipelines, reducing response time by 30% via parallelization.
- Leveraged Azure Document intelligence to extract specific metadata from PDFs, such as bolded items, numerical data, and section headers, automating structured data extraction with 80% accuracy.
- Automated metadata extraction workflows from PDFs using GPT-4 and related models, streamlining document processing.
- Developed a BERT-based contradiction detection system, utilizing feedback on neural networks to enhance precision in identifying Conflicting statements withing large datasets.

KloudDB
Mlops Engineer 06/2023-05/2024

- Collaborated with cross-functional teams to design and deliver advanced data solutions, aligning analytics strategies with organizational objectives and improving data-driven innovation.
- Deployed ML models with Docker and MLflow, focusing on on-device efficiency (pruning, quantization) for edge deployment.
- Leverage Docker for efficient deployment of machine learning models, enhancing deployment speed and reducing errors during production integration.
- Established a robust model versioning system using MLflow, enabling efficient tracking and rollback of models, reducing operational overhead.

Lizard Monitoring 05/2020 - 07/2022
Data Scientist

- Transformed unsupervised data into supervised data by utilizing NLP techniques to extract keywords from comments associated with alerts. Processed over 14,000 comments using NLTK libraries and created a word cloud to identify key Terms.
- Classified alerts into five distinct types by matching extracted keywords with a predefined list, accurately categorizing over 95% of alerts and enhancing the analysis of alert data.
- Developed and documented an LSTM-based machine learning model using sensor data from refrigeration units to predict refrigeration failures, achieving an accuracy of 86%, which significantly enhanced the reliability of failure predictions.
- Conducted time series analysis to identify defrost cycles, creating comprehensive tables that contained defrost cycle data from all sensors. This allowed for the differentiation between alarms caused by defrost cycles and actual refrigeration failures, improving alarm accuracy.
- Designed a real-time anomaly detection pipeline (PySpark + AWS Lambda) processing 10K+ events/sec with <100ms latency
- Implemented CI/CD pipelines for automating the deployment and monitoring of machine learning models, ensuring

continuous integration, quicker updates, and reduced downtime.

Incrivel Soft Solutions 06/2019 - 05/2020
Machine Learning Engineer Intern

- Architected and implemented a deep learning model using TensorFlow and Keras for time-series forecasting, achieving a 20% reduction in prediction error compared to traditional methods.
- Developed and fine-tuned ensemble learning models, including Gradient Boosting and Random Forest, to improve prediction accuracy and robustness, handling imbalanced datasets using techniques such as SMOTE.
- Designed and implemented Cross-validation and statistical tests including ANOVA, Chi-square test to verify the
- models' significance.

Projects

Agentic AI for Revenue Decline Root Cause Analysis

- Developed an **AI agent** using **LangGraph** and **OpenAI (GPT-4)** to automate root cause analysis of revenue decline through dynamic database interaction.
- Built a **self-correcting pipeline** with conditional edges and nodes (e.g., model_check_query, correct_query) to address syntax/logic errors, improving success rate by 50%.
- Automated end-to-end analysis, reducing manual root cause investigation time by 60% for stakeholders.
- Integrated **LangGraph's StateGraph** to orchestrate multi-step workflows, enabling seamless transitions between table selection, query generation, and result interpretation.

Predictive Analytics for Airbnb Listings Project [Project]

- Built scalable ETL pipelines using Apache Spark to ingest, clean, and preprocess 450,000+ records, addressing missing values and outliers for structured data warehousing in Google BigQuery.
- orchestrated workflows with Apache Airflow to automate data validation, feature engineering, and model training, reducing manual intervention by 40%.
- Designed a Google BigQuery-based data warehouse to centralize cleaned datasets, enabling efficient querying for cross-functional teams (product, analytics).
- Automated model retraining pipelines using Airflow DAGs, scheduling weekly updates to maintain predictive accuracy while aligning with data governance policies.

Chicken-Disease-Classification-Project [Project]

- Orchestrated an end-to-end MLOps pipeline for chicken disease classification, integrating MLflow, AWS, Azure, and DVC. Automated model training and deployment using CI/CD pipelines.
- Implemented DVC for efficient data versioning, ensuring consistency and reproducibility across the pipeline. Utilized AWS and Azure for scalable model training, packaging, and deployment, reducing training time by 50%.
- Implemented model quantization for ResNet-50, reducing inference time by 60% while retaining 95% accuracy.

Education

California State University, East Bay 08/2022 - 05/2024

- **Masters in Data Science**
  - **Relevant Coursework:** Python, R, Probability & Random Variables, Linear and Logistic models, ANOVA, data analytics, data visualization, Deep Learning, statistical Methods, Natural Language Processing.

Teegala Krishna Reddy College, Hyderabad 08/2016 -10/2020

- **Bachelor's in Computer Science**
  - **Relevant Coursework:** Operating systems, Computer Organization, Data Structures & Algorithms, Data Mining, Data Modeling, Python Programming, Cloud Computing, and Web Technologies.

Certifications

1. AWS Cloud Practitioner
2. Azure AZ-900
3. Azure AI Engineer Associate
4. Databricks Generative AI Associate
5. AWS Machine Learning Specialty

Publications

Title: Developing Solutions for Employee Retention Management using Machine Learning
Journal: International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS), January 2023, Volume 04, Issue 01
Summary: Developed a predictive model using the XGBoost algorithm to forecast employee turnover with 92% accuracy. Explored synthetic data generation with GANs to address dataset imbalance.Conducted extensive statistical analysis to identify key factors influencing attrition and proposed actionable strategies to enhance employee retention. This work is detailed in a published paper in IRJMETS.