Lathish Balaji Baskaran

Glassboro, NJ | lathishb510@gmail.com | (856) 462-2845 | Github

SUMMARY

Data Scientist with extensive experience in machine learning, AI, and data engineering. Skilled in building ML models, automating data workflows, and integrating AI solutions into real-world applications. Proficient in tools like TensorFlow, PyTorch, and LangChain, with a focus on data processing, predictive modeling, and delivering insights through Tableau. Passionate about AI, with a focus on advancing skills in predictive modeling and deep learning.

WORK EXPERIENCE

Data Scientist, Rowan University   **May 2024 - Present**

Job Description:

- Designed and developed a natural language-powered chatbot using LangChain to generate answers from structured databases, ensuring accurate data retrieval, reducing errors through ethical guardrails, and delivering real-time insights with intuitive visualizations via a user-friendly interface.
- Integrated APIs across multiple platforms, including the Informatica Business Terms API, using Python with Tableau and TabPy, enabling seamless data exchange, delivering real-time updates, and reducing dashboard latency and manual handling by 70%.
- Built and optimized ETL processes on the Veera platform, improving data pipeline efficiency of data pipelines, which led to faster data processing, improved dashboarding, and more accurate predictive modeling.
- Developed an R script to automate parsing and restructuring of ETL job data from the Veera platform, enabling seamless integration with Informatica and improving data lineage and job management.
- Integrated ThoughtSpot Sage Search within a website and trained ThoughtSpot's generative AI model to enhance data retrieval accuracy. Optimized the user interface for intuitive and efficient end-user interaction, ensuring seamless integration and improved accessibility of insights.
- Designed and executed SQL queries for comprehensive data extraction, reporting, and analysis, ensuring data integrity and usability for diverse institutional needs.
- Developed dashboards and visualizations using Tableau for multiple departments.
- Developed and maintained Cognos reports, ensuring accurate data representation and creating detailed documentation for processes to support scalability and ease of maintenance.
- Automated the extraction, standardization, and analysis of 700+ resumes using Python, leveraging Regex and Fitz for precise data parsing, delivering actionable insights for university research while reducing manual workload and improving accuracy and efficiency.

Graduate Research Assistant, Rowan University   **May 2023 – May 2024**

Job Description:

- Developed and implemented a machine learning model using PCA to identify the most influential variables for classifying research universities in Carnegie rankings, revealing key drivers of institutional performance and areas for growth.
- Enhanced the model architecture by integrating encoding layers with ReLU activation and dropout layers to improve pattern recognition and mitigate overfitting, while leveraging PCA for dimensionality reduction to focus on the most impactful features.
- Performed extensive data cleaning and preprocessing to address inconsistencies, missing values, and outliers in large datasets, enabling the machine learning model to uncover meaningful patterns that provided actionable insights, shaped strategic decisions, and guided future research initiatives.
- Integrated the model into a web application using Flask, making it easily accessible to research teams and automating the generation of data-driven insights for improved institutional research.

Junior Data Analyst, Vinayak Communication   **Jan 2022 - Dec 2022**

Job Description:

- Utilized Python and Excel to analyze large datasets, focusing on sales trends and performance for a wholesale and retail electronics company.
- Developed machine learning models using TensorFlow to predict areas for improvement and identify high-potential regions for sales growth. Specifically, built and trained neural networks to capture complex patterns in the data, such as seasonality and regional sales variations.
- Built data visualizations using Python (Matplotlib, Seaborn) to present complex findings in an easy-to-understand format, aiding strategic decision-making.
- Leveraged historical data to drive actionable insights, enhancing the company's sales strategy and operational efficiency.

SKILLS

Programming Languages: Python, SQL, R, JavaScript

Database Management Systems: Oracle, MySQL, Mongo DB

Big Data Technologies: PySpark, SparkSQL

Data Visualization: Tableau, Cognos, Power BI

Machine Learning libraries: TensorFlow, PyTorch, Keras, Hugging Face, Scikit-learn, Pandas, NumPy

Cloud Technologies: Microsoft Azure (Azure AI Studio, Azure OpenAI, Azure Machine Learning), Amazon Web Services (AWS IAM, S3, EC2, SageMaker), Databricks

Web Development : Docker, Flask, Streamlit

ETL Tools: Veera Construct, SAS, Tableau Prep Builder

PROJECTS

Database Query and Interaction Chatbot

- Designed a natural language-powered chatbot using LangChain to interpret user queries and generate structured database queries, ensuring seamless and intuitive interaction with data sources.
- Converted datasets into embeddings to facilitate efficient semantic search, enabling accurate matching between user queries and relevant structured and unstructured data.
- Enhanced data retrieval accuracy by employing the Retrieval-Augmented Generation (RAG) method, integrating real-time knowledge base information with language model outputs.
- Trained the model to minimize hallucinations by incorporating strict guardrails, adhering to AI ethics principles, and preparing it for responsible deployment of generative AI.
- Automated the extraction of key insights and created real-time visualizations to provide actionable data instantly.
- Developed a user-friendly chatbot interface that offered comprehensive, real-time responses and intuitive user experience.

Generative AI-Enhanced Tableau Dashboard

- Designed a generative AI-powered Tableau dashboard, utilizing a custom tableau extension to provide real-time explanations for terms when hovered over, enhancing data comprehension.
- Built a Tableau extension using JavaScript to detect hover events and trigger AI-generated explanations, seamlessly integrating natural language understanding into the dashboard experience.
- Leveraged a pre-stored dataset to fetch contextual information for each term, using generative AI models to produce detailed and accurate explanations.
- Automated the delivery of concise, real-time term explanations directly within the dashboard, streamlining the learning process for users.

Traffic Violations Analysis

- Used Databricks to manage large-scale data engineering tasks on traffic violations datasets.
- Utilized PySpark to wrangle and preprocess massive datasets, leveraging distributed computing for seamless handling.
- Performed in-depth exploratory data analysis (EDA) with Pandas and NumPy, uncovering key insights that guided our modeling approach.
- Developed and trained a Random Forest model to predict the likelihood of future traffic violations, fine-tuning it to improve accuracy and using feature importance analysis to understand key factors.
- Created visualizations in Tableau to clearly highlight data patterns, model predictions, and underlying causes of traffic violations, making it easier to understand the factors contributing to these incidents.
- Applied parallel processing to efficiently train and evaluate the Random Forest model on the dataset.

EDUCATION

ROWAN UNIVERSITY, Glassboro, NJ   **Jan 2023 - Dec 2024**

Master of Science in Data Science

Related Courses: Deep Learning, Data Mining, Data Warehousing, Visual analytics

CERTIFICATIONS

Azure AI Engineer – Issued by Microsoft