Sravani Ganta

sravaniganta1999@gmail.com | +1 (226)-972-3043 | linkedin | Github

Summary

Experienced professional with 3+ years of expertise in designing and implementing scalable AI/ML pipelines, specializing in Generative AI, Retrieval-Augmented Generation architectures, Azure Cognitive Services, and SQL Server management. Proficient in optimizing workflows, developing innovative solutions, and leveraging advanced technologies to drive automation and enable enterprise digital transformation. Skilled in integrating AI capabilities into cloud environments to enhance operational efficiency and deliver robust, data-driven solutions.

Experience

LTIMindtree, Canada **April 2024 – Present**

Role: Cloud Engineer

- Designed and deployed AI-driven financial applications leveraging Azure OpenAI models and RAG with Azure Cognitive Search, improving document retrieval accuracy by 20% through custom ranking.
- Developed AI-powered healthcare solutions using Azure Cognitive Services for NLP tasks such as text analytics, sentiment analysis, speech-to-text, and real-time translation, reducing manual effort by 40%.
- Built scalable ML pipelines on Azure ML, automating data preprocessing, model training, and deployment.
- Containerized and deployed applications on Azure Web Apps, integrating APIs for seamless interactions.
- Optimized Generative AI workflows in PromptFlow, increasing chatbot accuracy and reducing misunderstandings by 20%.
- Engineered AI solutions using Azure AI Foundry to drive enterprise digital transformation.
- Applied transfer learning and domain-specific data to optimize AI models for enhanced predictions, image recognition, and chatbot functionality.

University of Waterloo **September 2023 – December 2023**

Role: Teaching Assistant, Supervisor: Prof. Abdalla Mohamed Hussein, Course: ECE 358 - Computer Networks

- Provided guidance and assistance to undergraduate students in problem-solving and assignments
- Conducted online office hours and offered support to students enrolled in the course

University of Waterloo **April 2023 – August 2023**

Role: Machine learning Researcher, Supervisor: Prof. En Hui Yang

- Designed and implemented weight initialization strategies to transfer knowledge from larger models to smaller models, utilizing pre-trained ResNet and ShuffleNet architectures coupled with Kaiming normalization on CIFAR-100 and ImageNet datasets.
- Achieved 10% faster convergence and a 0.5% increase in accuracy for child models through optimized strategies.

Accenture Private Ltd, India **October 2020 – August 2022**

Role: Database Administrator

- Installed, configured, and upgraded Microsoft SQL Server in complex on-premises and clustered environments, ensuring high availability, scalability, and optimal system reliability.
- Migrated SQL Server databases across environments with minimal downtime, ensuring data integrity and reliability.
- Designed and implemented High Availability and Disaster Recovery solutions, including Database Mirroring, Log Shipping, Transactional Replication, and SQL Server failover clusters, minimizing data loss and enhancing disaster recovery processes.
- Configured and maintained SQL Server maintenance plans to address fragmentation and disk space issues, ensuring successful execution of backups, index rebuilds, and statistics updates through proactive monitoring.
- Managed SQL Server authentication, roles, and permissions, implementing organizational security policies, and configured Linked Servers, Endpoints, Operators, and Database Email for seamless integration and proactive alerting across environments.
- Provided 24/7 on-call support for critical database environments, ensuring stability and availability by conducting consistency checks and implementing granular backup and recovery strategies.

Skills

- **Languages**: Python, SQL, C, ASP.Net, KQL, HTML
- **Libraries &frameworks:** TensorFlow, PyTorch, Keras, NumPy, Pandas, PySpark, NLTK, NLP
- **Development Tools:** Visual Studio, GitHub, Jupyter Notebook
- **Cloud Platforms:** Azure (ML Studio, AI Foundry, OpenAI, Cognitive Services), AWS
- **Data Processing:** Data Wrangling, Cleaning, Manipulation, Exploration, and Visualization
- **Databases:** Microsoft SQL server, DBA, Database Management systems, RDBMS

Education

University of Waterloo **September 2022 – December 2023**

Master of Engineering in Electrical and Computer Engineering

Jawaharlal Nehru Technological University **June 2016 – August 2020**

Bachelor of Technology in Electrical and Instrumentation Engineering

Research

Weight Initialization Techniques for Enhanced Deep Learning | *Resnet, Shufflenet, Kaiming Normalization, Weight Initialization.*

- Implemented an innovative weight initialization method that included down sampling the larger model weights to align with the smaller model layers, followed by a process of random sampling and Kaiming normalization. Additionally, developed Resnet and Shufflenet models with diverse configurations, involving weight transfers between fully connected and convolutional layers, both with and without Kaiming normalization, evaluated on CIFAR-100 and Imagenet datasets.
- Achieved a notable 10% faster convergence rate and an accuracy improvement of 0.5% when compared to the baseline of smaller models on both datasets when weights are initialized from the larger model convolutional layer to the smaller model convolutional layer with Kaiming Normalization.

Projects

SQL Migration, Pipeline Dynamics, and Power BI Insights | *Microsoft SQL Server, Microsoft Azure, Data Visualization, ETL, Power BI*

- Orchestrated the migration of an on-premises SQL database to Azure, ensuring seamless data integration from various sources including CSV files. This foundational work laid the groundwork for advanced data manipulation and analysis.
- Engineered a robust ETL pipeline leveraging Azure Data Factory, facilitating the efficient transfer of data from on-premises environments to Azure Storage. This included the creation of Azure resources and the configuration of a self-hosted integration runtime for cloud connectivity.
- Implemented Azure Databricks for enhanced data processing, employing Spark SQL for table aggregation. This step was crucial for handling large volumes of data and performing complex computations efficiently.
- Finalized the data processing framework with the integration of Power BI, enabling sophisticated data visualization and insights. This allowed for the dynamic representation of data analytics, aiding in decision-making processes.

Admin Login page for Chatbot Management| *React, Next.js, JavaScript, Tailwind CSS*

- Developed an admin page using Next.js and React to efficiently manage chatbot user information, including user activity, last seen status, payment dates, and renewal countdowns.
- Designed and implemented an intuitive user interface for admin operations.
- Successfully fetched and displayed user data, formatted for readability.
- Calculated and displayed renewal countdown based on payment dates.

Medical Condition Classification and Drug Review Analysis | *NLP, python, sentiment analysis*

- Developed an NLP and ML-based drug recommendation system using trigram TF-IDF vectorization and the Passive Aggressive Classifier, achieving a 98.2% accuracy rate in evaluations with a drug dataset. This involved preprocessing text data by removing HTML, special characters, and stopwords, and applying stemming and lemmatization.
- Implemented a comprehensive text analysis approach, utilizing both bag of words and TFIDF vectorizer techniques, and integrated Naive Bayes and Passive Aggressive Classifier models for sentiment analysis and drug recommendation.

Google Cloud Data Pipelining with Airflow | *Google API, Airflow, AWS, EC2, python*

- Developed and automated a data pipeline to extract, transform, and clean YouTube video reviews using Google API, Python, and Apache Airflow, storing data securely on Amazon S3.
- Designed and implemented ETL processes to extract data from multiple sources, transform it using custom Python scripts, and load it into a centralized data warehouse.
- Deployed the pipeline on AWS EC2 instances, ensuring high availability and scalability to handle large volumes of data.
- Implemented monitoring and alerting mechanisms using AWS CloudWatch and Airflow's built-in features to ensure the reliability and performance of the data pipeline.
- Automated the data pipeline with Apache Airflow and stored the processed data securely on Amazon S3.

Clustering and Classification with Deep Learning | *Machine Learning, Data Visualization, Deep Neural Networks, Clustering*

- Designed, implemented, and compared multiple deep learning architectures for data classification, including a default architecture and a customized architecture utilizing various Deep Neural Network (DNN) variants. This process involved comprehensive training and evaluation to determine the optimal model for the dataset.
- Conducted advanced clustering techniques to analyze and segment datasets, creating intuitive visualizations to communicate patterns and insights effectively. Leveraged tools like Matplotlib, Seaborn, and Plotly to enhance interpretability and aid in data-driven decision-making.
- Implemented diverse machine learning algorithms, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Naïve Bayes, Decision Trees, Random Forests, and Gradient Tree Boosting, to classify datasets with high accuracy. Developed models to address specific data challenges and improve outcomes.

Certifications

- Azure OpenAI Specialist
- Azure Cognitive Services Specialist
- Machine learning LLM
- AZAI Gen AI Specialist
- AZAI AzureML Advanced
- AZ-303
- Microsoft SQL