Madhuri Paluri

Newark, California
551-344-7396 palurimadhuri22@gmail.com linkedin.com/in/madhuri-paluri-822716198/
github.com/MadhuriPa98

Professional Summary

5+ years of experience in machine learning, data science working with large volumes of unstructured, structured data building scalable applications using Docker, Kubernetes and deploying them on AWS or Azure in Agile methodology.

Education

Stevens Institute of Technology **Aug. 2022 – May 2024**
MS Thesis in Computer Science *Hoboken, New Jersey*

Thesis Title : A Efficient Sub-Sentence classification using student teacher networks

Relevant Coursework

- Data Structures
- Machine learning
- Deep Learning
- Natural Language processing
- Artificial Intelligence
- Computer vision
- Reinforcement learning
- Parallel Programming
- Algorithms

Technical Skills

Programming Languages: Python, Java, C, HTML/CSS, JavaScript, SQL
Developer Tools: VS Code, Eclipse, Google Cloud Platform, Android Studio
Technologies/Frameworks: Linux, Jenkins, GitHub, PyTorch, Tensorflow
Cloud Technologies: DataBricks, Ansible, Terraform, Kinesis, EMR, DynamoDB, Jenkins, sagemaker, CloudWatch, AWS Glue, AWS Lambda
BigData Tools: Spark, Kafka, Airflow, Snowflake, Hive, Sqoop, Terraform, Presto, Reltio, Redshift
DataBases: Oracle, MySQL, PostgreSQL, SQL Server, MongoDB

Experience

Stevens Institute of Technology **August 2023 – Current**
Research Assistant *Hoboken, NewJersey*

- Worked on **Hallucination Boundary detection task** in conversational chats with the help of student teacher networks
- Constructed a specialized dataset of 35,000 user chatbot data which contains hallucination start and end tokens using GPT 3.5 and GPT 4 API
- Experimented with various encoder models like BERT and RoBERTA in case of student and teacher networks and compared accuracy, memory, and Latency
- Developed an AI powered chatbot using LLAMA 2 7B by including retrieval augmentation technique (RAG) which improved conversational capability and boosted information retrieval
- Developed an intent identification model using LLAMa2 by classifying the user utterances and achieved an accuracy of 84%
- Utilized vector databases like RAG using Pinecone to effectively catalog all incidents by creating a repository of historical user resolutions implemented by technical support staff and business users
- Used Langchain framework to integrate with databases like Oracle DB, Postgres SQL, applied prompt engineering techniques
- Experimented several language models like Mistral and Gemini to compare retrieval efficiency and summarization quality

Tata Consultancy Services **July 2019 – July 2022**
Data Scientist *Hyderabad, India*

- Developed an end-to-end ML pipeline starting from data curation, training, validation, testing to deployment using PyTorch for intent identification model using PySpark,fine-tuning self-supervised networks on custom datasets, resulting in a notable 93% accuracy
- Spearheaded topic modeling using LDA and multi label classification to cluster conversational data between users and customer care using transformer models like BERT and T5 using PyTorch framework
- Performed dimensionality reduction techniques like PCA/SVD to reduce BERT embedding size from 1024 to 400 for achieving noise filtering thereby retaining 90% of accuracy
- Leveraged Ml Flow with databricks for workflow creation and artifactory management and reduced $1.5M dollars by incorporating efficient spot instances and parallelizing operations using batch and stream pipelines

- Collaborated with cross functional teams streamlining the integration of the machine learning models using Kubernetes and deployed them on Amazon EC2 using AWS which increased customer- client interaction time by 80%
- Maintained Customized Machine Learning recommendation systems pipeline for new policy engagement hosted through Airflow and Kubeflow on Google Cloud thereby fostering business by +3.5M in year
- Developed Pydantic models in python and deployed them using FastAPI for managing JSON data and integrating with frontend for seamless building of apps
- Maintained feature tables in snowflake thereby continuously integrating with Machine learning models in databricks for building scalable production ready applications.

Accenture **May 2018 – April 2019**

Data Scientist *Banglore, India*

- Enhanced PySpark query optimization by 45% using window functions, Joins, Filters and broadcast Joins and parallelizing of operations
- Developed credit risk model using logistic regression and decision trees reducing false positives by 23% using ensemble models
- Implemented high-performance computing on AWS Kinesis and building real-time analysis with Kafka and Spark Streaming
- Automated weekly report generation and visualization dashboards for communicating the performance indicators with business leaders using Tableau

Projects

JOE CRM – A replacement for Traditional CRM systems | *Groq, Llama3, LECL, BraveAPI, VERCEL* **May 2024**

- Revolutionized traditional data driven CRM systems to action driven using LLAMA3 70B which creates a campaign plan for prospective customers by sending highly customized emails.
- Used Langchain Expression language (LECL) for multi-step conversations thereby enabling sophisticated prompts and deployed it using FASTAPI.
- Used Brave API to get the relevant history of customers and curated prompts for GROQ powered LLAMA3 which helps in analyzing users for customer success managers.
- Leveraged VERCEL for building action centric front end user interface and deployed JOE CRM on AWS.

SubSentence classification using Knowledge distillation | *BERT, RoBERTa, PyTorch* **December 2023 - May 2024**

- Detected the start and end tokens of hallucination,sentiment analysis and text entailment of conversational chat data.
- Retained student accuracy by 95% with reduction in size up to 66% compared to teacher model which is significant in case of edge device deployment

Certifications

- **AWS Certified Data Engineer**, AWS.
- **Microsoft Certified Data Engineer DP-203**, Microsoft.
- **ML Devops Engineer**, Udacity

Awards/Achievements

Best Employee Award **2020 – 2022**

Tata Consultancy services *Hyderabad,India*

- Received Best Employee award while working for TCS for making impact in the team through innovation and research which fostered client business.
- Solved Major Production issues which saved $400K to client.

Achievements **05-11-2022**

Meta Hackathon *California,USA*

- Achieved 4th place in Hackathon out of 55 teams conducted by META where we developed JOE CRM which automates traditional CRM systems using LLAM3 and GROQ .